

Gram Charlier and Edgeworth expansion for sample variance

Eric Benhamou¹

Abstract

In this paper, we derive a valid Edgeworth expansions for the Bessel corrected empirical variance when data are generated by a strongly mixing process whose distribution can be arbitrarily. The constraint of strongly mixing process makes the problem not easy. Indeed, even for a strongly mixing normal process, the distribution is unknown. Here, we do not assume any other assumption than a sufficiently fast decrease of the underlying distribution to make the Edgeworth expansion convergent. This results can obviously apply to strongly mixing normal process and provide an alternative to the work of [6] and [4].

Mathematics Subject Classification: 62E10; 62E15

Keywords: sample variance; Edgeworth expansion

1 Introduction

Let X_1, \dots, X_n be a random sample and define the sample variance statistic

¹ A.I. Square Connect, 35 Boulevard d'Inkermann 92200 Neuilly sur Seine, France.
LAMSADE, Universit Paris Dauphine, Place du Marchal de Lattre de Tassigny, 75016
Paris, France.
E-mail: eric.benhamou@aisquareconnect.com, eric.benhamou@dauphine.eu,

as:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad X_n = (X_1, \dots, X_n)^T \quad (1)$$

where \bar{X}_n is the empirical mean, s_n^2 the Bessel corrected empirical variance also called sample variance, and X_n the vector of the full history of this random sample. We are interested in the distribution of the sample variance under very weak conditions, namely that it admits a valid Edgeworth expansion.

It is insightful to notice that even with the additional constraint of a multi dimensional Gaussian distribution $\mathcal{N}(0, \Sigma)^2$ for the underlying random vector X_n , the distribution of the sample variance is not known. In this particular setting, the sample variance is the squared norm of a multi dimensional Gaussian and can be seen as the linear combination of independent but not Homoscedastic variables. Standard theory states that the sample variance of a collection of independent and identically distributed normal variables follows a chi-squared distribution. But in this particular case, the different variables $X_i - \bar{X}_n$ are not independent and the result can not apply. Though they marginally have the same variance (if we conditioned by \bar{X}_n), they are correlated with each other.

Hence, in general, the distribution of the sample variance in the normal case is not a known distribution. There is however one special case, where it is a χ^2 distribution. Remember that when $X \sim \mathcal{N}(0, \Sigma)$, $X^T B X$ has a χ^2 distribution of degree d the rank of B if and only if $B \Sigma B = B$ (see for instance page 413 remark following theorem 1 of [3]). Notice that in our case, the B matrix is $B = I_n - \frac{1}{n} 1_n^T 1_n$, where I_n is the identity matrix of size n and 1_n is the vector of size n filled with 1 as we can write

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (2)$$

$$= \frac{1}{n-1} X^T (I_n - \frac{1}{n} 1_n^T 1_n) X \quad (3)$$

we obtain that it is a χ^2 distribution if and only if

$$(I_n - \frac{1}{n} 1_n^T 1_n) \Sigma (I_n - \frac{1}{n} 1_n^T 1_n) = I_n - \frac{1}{n} 1_n^T 1_n$$

²with *Sigma* arbitrarily

In other cases, the sample variance is a linear combination of Gamma distribution, and one has to rely on approximations as explained in [6] and [4].

This simple example explains the interest in deriving an approximation of the distribution of the sample variance by means of Gram Charlier and Edgeworth expansion.

2 Gram Charlier and Edgeworth expansion

2.1 Key concepts

GramCharlier expansion³, and Edgeworth expansion⁴, are series that approximate a probability distribution in terms of its cumulants. The series are the same but, they differ in the ordering of their terms. Hence the truncated series are different, as well as the accuracy of truncating the series. The key idea in these two series is to expand the characteristic function in terms of the characteristic function of a known distribution with suitable properties, and to recover the concerned distribution through the inverse Fourier transform.

In our case, a natural candidate to expand around is the normal distribution as the central limit theorem and its different extensions to non independent and non identically distributed variable state that the resulting distribution is a normal distribution (or in the most general case to truncated symmetrical and α stable distributions ⁵.

Let denote by \hat{f} (respectively $\hat{\phi}$ the characteristic function of our distribution whose density function is f (respectively ϕ), and κ_j its cumulants (respectively γ_j). Cumulants definition state

$$\hat{f}(t) = \exp \left[\sum_{j=1}^{\infty} \kappa_j \frac{(it)^j}{j!} \right] \quad \text{and} \quad \hat{\phi}(t) = \exp \left[\sum_{j=1}^{\infty} \gamma_j \frac{(it)^j}{j!} \right], \quad (4)$$

which gives the following formal identity:

$$\hat{f}(t) = \exp \left[\sum_{j=1}^{\infty} (\kappa_j - \gamma_j) \frac{(it)^j}{j!} \right] \hat{\phi}(t). \quad (5)$$

³named in honor of the Danish mathematician, Jrgen Pedersen Gram and the Swedish astronomer, Carl Charlier

⁴named in honor of the Anglo-Irish philosopher, Francis Ysidro Edgeworth

⁵please see the extension of CLT in [2]

Using Fourier transform property that say $(it)^j \hat{\phi}(t)$ is the Fourier transform of $(-1)^j [D^j \phi](-x)$, where D is the differential operator with respect to x , we get the formal expansion:

$$f(x) = \exp \left[\sum_{j=1}^{\infty} (\kappa_j - \gamma_j) \frac{(-D)^j}{j!} \right] \phi(x) \quad (6)$$

If ϕ is chosen as the normal density $\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$ with mean and variance as given by f , that is, mean $\mu = \kappa_1$ and variance $\sigma^2 = \kappa_2$, then the expansion becomes

$$f(x) = \exp \left[\sum_{j=3}^{\infty} \kappa_j \frac{(-D)^j}{j!} \right] \phi(x), \quad (7)$$

since $\gamma_j = 0$ for all $j > 2$, as higher cumulants of the normal distribution are 0. By expanding the exponential and collecting terms according to the order of the derivatives, we arrive at the GramCharlier A series. Such an expansion can be written compactly in terms of Bell polynomials as

$$\exp \left[\sum_{j=3}^{\infty} \kappa_j \frac{(-D)^j}{j!} \right] = \sum_{j=0}^{\infty} B_j(0, 0, \kappa_3, \dots, \kappa_j) \frac{(-D)^j}{j!}. \quad (8)$$

Since the j -th derivative of the Gaussian function ϕ is given in terms of Hermite polynomial as

$$\phi^{(j)}(x) = \frac{(-1)^j}{\sigma^j} He_j \left(\frac{x-\mu}{\sigma} \right) \phi(x), \quad (9)$$

this gives us the final expression of the Gram-Charlier A series as

$$f(x) = \phi(x) \sum_{j=0}^{\infty} \frac{1}{j! \sigma^j} B_j(0, 0, \kappa_3, \dots, \kappa_j) He_j \left(\frac{x-\mu}{\sigma} \right). \quad (10)$$

If we include only the first two correction terms to the normal distribution, we obtain

$$f(x) = \phi(x) \left[1 + \frac{\kappa_3}{3! \sigma^3} He_3 \left(\frac{x-\mu}{\sigma} \right) + \frac{\kappa_4}{4! \sigma^4} He_4 \left(\frac{x-\mu}{\sigma} \right) + R_5 \right], \quad (11)$$

with $He_3(x) = x^3 - 3x$ and $He_4(x) = x^4 - 6x^2 + 3$ and

$$R_5(x) = \sum_{j=5}^{\infty} \frac{1}{j! \sigma^j} B_j(0, 0, \kappa_3, \dots, \kappa_j) He_j \left(\frac{x-\mu}{\sigma} \right).$$

If in the above expression, the cumulant are function of a parameter $\frac{1}{n}$, we can rearrange terms by power of $\frac{1}{n}$ and find the Edgeworth expansion.

2.2 Cumulant for weak conditions

In order to derive our Gram Charlier or Edgeworth expansion, we need to compute in full generality our different cumulants. Using similar techniques as in [1], we can get the various cumulants as follows:

The first two cumulants are easy and given by:

$$\kappa_1 = (\dot{\mu}_2)_1 - (\dot{\mu}_1^2)_1 \quad (12)$$

$$\kappa_2 = \frac{A_{1,0}^2}{n-1} + \frac{A_{0,1}^2}{n} + \frac{A_{1,1}^2}{(n-1)n} + R_2 \quad (13)$$

with the different numerator terms given by:

$$A_{1,0}^2 = -4(\dot{\mu}_1^4)_2 + 8(\dot{\mu}_1^2 \dot{\mu}_2)_2 - (\dot{\mu}_2^2)_2 \quad (14)$$

$$A_{0,1}^2 = (\dot{\mu}_4)_2 - 4(\dot{\mu}_1 \dot{\mu}_3)_2 \quad (15)$$

$$A_{1,1}^2 = 6(\dot{\mu}_1^4)_2 - 12(\dot{\mu}_1^2 \dot{\mu}_2)_2 + 3(\dot{\mu}_2^2)_2 \quad (16)$$

$$R_2 = (\dot{\mu}_2^2)_2 - 2(\dot{\mu}_1^2 \dot{\mu}_2)_2 + (\dot{\mu}_1^4)_2 - (\kappa_1)^2 \quad (17)$$

with the natural symmetric moment estimators whose expressions are provided in appendix section A.1. The term R_2 is the second order rest and is equal to zero if the sample is i.i.d. For general case, this term does not cancel out and should be taken into account. It can be rewritten as

$$R_2 = (\dot{\mu}_2^2)_2 - (\dot{\mu}_2)_1^2 + 2((\dot{\mu}_2)_1(\dot{\mu}_1^2)_1 - (\dot{\mu}_1^2 \dot{\mu}_2)_2) + (\dot{\mu}_1^4)_2 - (\dot{\mu}_1^2)_1^2 \quad (18)$$

The third cumulant is more involved and given by:

$$\begin{aligned} \kappa_3 = & \frac{A_{2,0}^3}{(n-1)^2} + \frac{A_{1,1}^3}{(n-1)n} + \frac{A_{0,2}^3}{n^2} + \frac{A_{2,1}^3}{(n-1)^2 n} + \frac{A_{1,2}^3}{(n-1)^2 n} \\ & + \frac{A_{2,2}^3}{(n-1)^2 n^2} + R_3 \end{aligned} \quad (19)$$

where the different numerator terms are given by:

$$\begin{aligned} A_{2,0}^3 = & -40(\dot{\mu}_1^6)_3 + 120(\dot{\mu}_1^4 \dot{\mu}_2)_3 - 56(\dot{\mu}_1^3 \dot{\mu}_3)_3 - 78(\dot{\mu}_1^2 \dot{\mu}_2^2)_3 + 48(\dot{\mu}_1 \dot{\mu}_2 \dot{\mu}_3)_3 \\ & + 2(\dot{\mu}_2^3)_3 - 6(\dot{\mu}_3^2)_3 \end{aligned} \quad (20)$$

$$A_{1,1}^3 = 18(\dot{\mu}_1^2 \dot{\mu}_4)_3 - 3(\dot{\mu}_2 \dot{\mu}_4)_3 \quad (21)$$

$$A_{0,2}^3 = (\dot{\mu}_6)_3 - 6(\dot{\mu}_1 \dot{\mu}_5)_3 \quad (22)$$

$$\begin{aligned} A_{2,1}^3 = & 136(\dot{\mu}_1^6)_3 - 408(\dot{\mu}_1^4 \dot{\mu}_2)_3 + 160(\dot{\mu}_1^3 \dot{\mu}_3)_3 + 288(\dot{\mu}_1^2 \dot{\mu}_2^2)_3 - 144(\dot{\mu}_1 \dot{\mu}_2 \dot{\mu}_3)_3 \\ & - 24(\dot{\mu}_2^3)_3 + 12(\dot{\mu}_3^2)_3 \end{aligned} \quad (23)$$

$$A_{1,2}^3 = 15(\dot{\mu}_2\dot{\mu}_4)_3 - 30(\dot{\mu}_1^2\dot{\mu}_4)_3 \quad (24)$$

$$A_{2,2}^3 = -120(\dot{\mu}_1^6)_3 + 360(\dot{\mu}_1^4\dot{\mu}_2)_3 - 120(\dot{\mu}_1^3\dot{\mu}_3)_3 - 270(\dot{\mu}_1^2\dot{\mu}_2^2)_3 + 120(\dot{\mu}_1\dot{\mu}_2\dot{\mu}_3)_3 \\ + 30(\dot{\mu}_2^3)_3 - 10(\dot{\mu}_3^2)_3 \quad (25)$$

$$R_3 = -3\dot{\mu}_2'\dot{\mu}_1' + 2(\kappa_1)^3 \quad (26)$$

with the natural symmetric moment estimators given in appendix section A.2. The fourth cumulant is even more involved and given by:

$$\kappa_4 = \frac{A_{3,0}^4}{(n-1)^3} + \frac{A_{2,1}^4}{(n-1)^2n} + \frac{A_{1,2}^4}{(n-1)n^2} + \frac{A_{0,3}^4}{n^3} + \frac{A_{3,1}^4}{(n-1)^3n} + \frac{A_{2,2}^4}{(n-1)^2n^2} \\ + \frac{A_{1,3}^4}{(n-1)n^3} + \frac{A_{3,2}^4}{(n-1)^3n^2} + \frac{A_{2,3}^4}{(n-1)^2n^3} + \frac{A_{3,3}^4}{(n-1)^3n^3} + R_4 \quad (27)$$

with the different numerator terms are given by:

$$A_{3,0}^4 = -672(\dot{\mu}_1^8)_4 + 2688(\dot{\mu}_1^6\dot{\mu}_2)_4 - 1216(\dot{\mu}_1^5\dot{\mu}_3)_4 - 3120(\dot{\mu}_1^4\dot{\mu}_2^2)_4 + 400(\dot{\mu}_1^4\dot{\mu}_4)_4 \\ + 2240(\dot{\mu}_1^3\dot{\mu}_2\dot{\mu}_3)_4 + 960(\dot{\mu}_1^2\dot{\mu}_2^3)_4 - 384(\dot{\mu}_3^2\dot{\mu}_1^2)_4 - 480(\dot{\mu}_1^2\dot{\mu}_2\dot{\mu}_4)_4 \\ - 624(\dot{\mu}_1\dot{\mu}_2^2\dot{\mu}_3)_4 + 144(\dot{\mu}_1\dot{\mu}_3\dot{\mu}_4)_4 - 6(\dot{\mu}_2^4)_4 + 96(\dot{\mu}_2\dot{\mu}_3^2)_4 \\ - 3(\dot{\mu}_4^2)_4 + 12(\dot{\mu}_2^2\dot{\mu}_4)_4 \quad (28)$$

$$A_{2,1}^4 = -128(\dot{\mu}_1^3\dot{\mu}_5)_4 + 96(\dot{\mu}_1\dot{\mu}_2\dot{\mu}_5)_4 - 24(\dot{\mu}_3\dot{\mu}_5)_4 \quad (29)$$

$$A_{1,2}^4 = 32(\dot{\mu}_1^2\dot{\mu}_6)_4 - 4(\dot{\mu}_2\dot{\mu}_6)_4 \quad (30)$$

$$A_{0,3}^4 = (\dot{\mu}_8)_4 - 8(\dot{\mu}_1\dot{\mu}_7)_4 \quad (31)$$

$$A_{3,1}^4 = 3792(\dot{\mu}_1^8)_4 - 15168(\dot{\mu}_1^6\dot{\mu}_2)_4 + 6144(\dot{\mu}_1^5\dot{\mu}_3)_4 + 18144(\dot{\mu}_1^4\dot{\mu}_2^2)_4 - 1920(\dot{\mu}_1^4\dot{\mu}_4)_4 \\ - 11520(\dot{\mu}_1^3\dot{\mu}_2\dot{\mu}_3)_4 - 6336(\dot{\mu}_2^3\dot{\mu}_1^2)_4 + 1680(\dot{\mu}_1^2\dot{\mu}_3^2)_4 + 2520(\dot{\mu}_1^2\dot{\mu}_2\dot{\mu}_4)_4 \\ + 3600(\dot{\mu}_1\dot{\mu}_2^2\dot{\mu}_3)_4 - 624(\dot{\mu}_1\dot{\mu}_3\dot{\mu}_4)_4 + 234(\dot{\mu}_2^4)_4 - 432(\dot{\mu}_2\dot{\mu}_3^2)_4 + 33(\dot{\mu}_4^2)_4 \\ - 252(\dot{\mu}_2^2\dot{\mu}_4)_4 \quad (32)$$

$$A_{2,2}^4 = 400(\dot{\mu}_1^3\dot{\mu}_5)_4 - 336(\dot{\mu}_1\dot{\mu}_2\dot{\mu}_5)_4 + 48(\dot{\mu}_3\dot{\mu}_5)_4 \quad (33)$$

$$A_{1,3}^4 = 28(\dot{\mu}_2\dot{\mu}_6)_4 - 56(\dot{\mu}_1^2\dot{\mu}_6)_4 \quad (34)$$

$$A_{3,2}^4 = -7440(\dot{\mu}_1^8)_4 + 29760(\dot{\mu}_1^6\dot{\mu}_2)_4 - 10880(\dot{\mu}_1^5\dot{\mu}_3)_4 - 36480(\dot{\mu}_1^4\dot{\mu}_2^2)_4 \\ + 3104(\dot{\mu}_1^4\dot{\mu}_4)_4 + 20992(\dot{\mu}_1^3\dot{\mu}_2\dot{\mu}_3)_4 + 13824(\dot{\mu}_1^2\dot{\mu}_2^3)_4 - 2752(\dot{\mu}_1^2\dot{\mu}_3^2)_4 \\ - 4368(\dot{\mu}_1^2\dot{\mu}_2\dot{\mu}_4)_4 - 7248(\dot{\mu}_1\dot{\mu}_2^2\dot{\mu}_3)_4 + 976(\dot{\mu}_1\dot{\mu}_3\dot{\mu}_4)_4 - 738(\dot{\mu}_2^4)_4 \\ + 800(\dot{\mu}_2\dot{\mu}_3^2)_4 - 57(\dot{\mu}_4^2)_4 + 612(\dot{\mu}_2^2\dot{\mu}_4)_4 \quad (35)$$

$$A_{2,3}^4 = -336(\dot{\mu}_1^3\dot{\mu}_5)_4 + 336(\dot{\mu}_1\dot{\mu}_2\dot{\mu}_5)_4 - 56(\dot{\mu}_3\dot{\mu}_5)_4 \quad (36)$$

$$\begin{aligned}
A_{3,3}^4 = & 5040(\hat{\mu}_1^8)_4 - 20160(\hat{\mu}_1^6\hat{\mu}_2)_4 + 6720(\hat{\mu}_1^5\hat{\mu}_3)_4 + 25200(\hat{\mu}_1^4\hat{\mu}_2^2)_4 \\
& - 1680(\hat{\mu}_1^4\hat{\mu}_4)_4 - 13440(\hat{\mu}_1^3\hat{\mu}_2\hat{\mu}_3)_4 - 10080(\hat{\mu}_1^2\hat{\mu}_2^3)_4 + 1680(\hat{\mu}_1^2\hat{\mu}_3^2)_4 \\
& + 2520(\hat{\mu}_1^2\hat{\mu}_2\hat{\mu}_4)_4 + 5040(\hat{\mu}_1\hat{\mu}_2^2\hat{\mu}_3)_4 - 560(\hat{\mu}_1\hat{\mu}_3\hat{\mu}_4)_4 + 630(\hat{\mu}_2^4)_4 \\
& - 560(\hat{\mu}_2\hat{\mu}_3^2)_4 + 35(\hat{\mu}_4^2)_4 - 420(\hat{\mu}_2^2\hat{\mu}_4)_4
\end{aligned} \tag{37}$$

with the natural symmetric moment estimators given in appendix section A.3.

Proof. See appendix section B. □

3 Conclusion

In this paper, we have derived the most general formula for the Gram Charlier and the resulting Edgeworth expansion for the sample variance under very weak conditions. Our formula does not assume that the underlying sample is independent neither identically distributed. This formula can therefore be applied to strong mixing processes like sample of an auto regressive process of order 1 (AR(1)). It extends in particular the work of [5].

A Notations

A.1 Empirical Moments of order 1 and 2 Notation

We adopt the following notations

- moment of order 1:

$$(\hat{\mu}_2)_1 = \frac{\sum_{i=1}^n \mathbb{E}[X_i^2]}{n} \tag{38}$$

$$(\hat{\mu}_1^2)_1 = \frac{\sum_{i \neq j} \mathbb{E}[X_i X_j]}{n(n-1)} \tag{39}$$

- moments of order 2:

$$(\acute{\mu}_4)_2 = \frac{\sum_{i=1}^n \mathbb{E}[X_i^4]}{n} \quad (40)$$

$$(\acute{\mu}_1 \acute{\mu}_3)_2 = \frac{\sum_{i \neq j} \mathbb{E}[X_i^3 X_j]}{n(n-1)} \quad (41)$$

$$(\acute{\mu}_2^2)_2 = \frac{\sum_{i \neq j} \mathbb{E}[X_i^2 X_j^2]}{n(n-1)} \quad (42)$$

$$(\acute{\mu}_1^2 \acute{\mu}_2)_2 = \frac{\sum_{i \neq j \neq k} \mathbb{E}[X_i^2 X_j X_k]}{n(n-1)(n-2)} \quad (43)$$

$$(\acute{\mu}_1^4)_2 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E}[X_i X_j X_k X_l]}{n(n-1)(n-2)(n-3)} \quad (44)$$

A.2 Empirical Moments of order 3 Notation

$$(\acute{\mu}_1^6)_3 = \frac{\sum_{i \neq j \neq k \neq l \neq m \neq n} \mathbb{E}[X_i X_j X_k X_l X_m X_n]}{n(n-1)(n-2)(n-3)(n-4)(n-5)} \quad (45)$$

$$(\acute{\mu}_1^4 \acute{\mu}_2)_3 = \frac{\sum_{i \neq j \neq k \neq l \neq m} \mathbb{E}[X_i^2 X_j X_k X_l X_m]}{n(n-1)(n-2)(n-3)(n-4)} \quad (46)$$

$$(\acute{\mu}_1^3 \acute{\mu}_3)_3 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E}[X_i^3 X_j X_k X_l]}{n(n-1)(n-2)(n-3)} \quad (47)$$

$$(\acute{\mu}_1^2 \acute{\mu}_2^2)_3 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E}[X_i^2 X_j^2 X_k X_l]}{n(n-1)(n-2)(n-3)} \quad (48)$$

$$(\acute{\mu}_1^2 \acute{\mu}_4)_3 = \frac{\sum_{i \neq j \neq k} \mathbb{E}[X_i^4 X_j X_k]}{n(n-1)(n-2)} \quad (49)$$

$$(\acute{\mu}_1 \acute{\mu}_2 \acute{\mu}_3)_3 = \frac{\sum_{i \neq j \neq k} \mathbb{E}[X_i^3 X_j^2 X_k]}{n(n-1)(n-2)} \quad (50)$$

$$(\acute{\mu}_1 \acute{\mu}_5)_3 = \frac{\sum_{i \neq j} \mathbb{E}[X_i^5 X_j]}{n(n-1)} \quad (51)$$

$$(\acute{\mu}_2^3)_3 = \frac{\sum_{i \neq j \neq k} \mathbb{E}[X_i^2 X_j^2 X_k^2]}{n(n-1)(n-2)} \quad (52)$$

$$(\acute{\mu}_2 \acute{\mu}_4)_3 = \frac{\sum_{i \neq j} \mathbb{E}[X_i^4 X_j^2]}{n(n-1)} \quad (53)$$

$$(\acute{\mu}_3^2)_3 = \frac{\sum_{i \neq j} \mathbb{E}[X_i^3 X_j^3]}{n(n-1)} \quad (54)$$

$$(\acute{\mu}_6)_3 = \frac{\sum_i \mathbb{E}[X_i^6]}{n} \quad (55)$$

A.3 Empirical Moments of order 4 Notation

$$(\mu_1^8)_4 = \frac{\sum_{i \neq j \neq k \neq l \neq m \neq n \neq o \neq p} \mathbb{E} [X_i X_j X_k X_l X_m X_n X_o X_p]}{n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)(n-7)} \quad (56)$$

$$(\mu_1^6 \mu_2)_4 = \frac{\sum_{i \neq j \neq k \neq l \neq m \neq n \neq o} \mathbb{E} [X_i^2 X_j X_k X_l X_m X_n X_o]}{n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)} \quad (57)$$

$$(\mu_1^5 \mu_3)_4 = \frac{\sum_{i \neq j \neq k \neq l \neq m \neq n} \mathbb{E} [X_i^3 X_j X_k X_l X_m X_n]}{n(n-1)(n-2)(n-3)(n-4)(n-5)} \quad (58)$$

$$(\mu_1^4 \mu_2^2)_4 = \frac{\sum_{i \neq j \neq k \neq l \neq m \neq n} \mathbb{E} [X_i^2 X_j^2 X_k X_l X_m X_n]}{n(n-1)(n-2)(n-3)(n-4)(n-5)} \quad (59)$$

$$(\mu_1^4 \mu_4)_4 = \frac{\sum_{i \neq j \neq k \neq l \neq m} \mathbb{E} [X_i^4 X_j X_k X_l X_m]}{n(n-1)(n-2)(n-3)(n-4)} \quad (60)$$

$$(\mu_1^3 \mu_5)_4 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E} [X_i^5 X_j X_k X_l]}{n(n-1)(n-2)(n-3)} \quad (61)$$

$$(\mu_1^3 \mu_2 \mu_3)_4 = \frac{\sum_{i \neq j \neq k \neq l \neq m} \mathbb{E} [X_i^3 X_j^2 X_k X_l X_m]}{n(n-1)(n-2)(n-3)(n-4)} \quad (62)$$

$$(\mu_1^2 \mu_2 \mu_4)_4 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E} [X_i^4 X_j^2 X_k X_l]}{n(n-1)(n-2)(n-3)} \quad (63)$$

$$(\mu_1^2 \mu_2^3)_4 = \frac{\sum_{i \neq j \neq k \neq l \neq m} \mathbb{E} [X_i^2 X_j^2 X_k^2 X_l X_m]}{n(n-1)(n-2)(n-3)(n-4)} \quad (64)$$

$$(\mu_1^2 \mu_3^2)_4 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E} [X_i^3 X_j^3 X_k X_l]}{n(n-1)(n-2)(n-3)} \quad (65)$$

$$(\mu_1^2 \mu_6)_4 = \frac{\sum_{i \neq j \neq k} \mathbb{E} [X_i^6 X_j X_k]}{n(n-1)(n-2)} \quad (66)$$

$$(\mu_1 \mu_2^2 \mu_3)_4 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E} [X_i^3 X_j^2 X_k^2 X_l]}{n(n-1)(n-2)(n-3)} \quad (67)$$

$$(\mu_1 \mu_2 \mu_5)_4 = \frac{\sum_{i \neq j \neq k} \mathbb{E} [X_i^5 X_j^2 X_k]}{n(n-1)(n-2)} \quad (68)$$

$$(\mu_1 \mu_3 \mu_4)_4 = \frac{\sum_{i \neq j \neq k} \mathbb{E} [X_i^4 X_j^3 X_k]}{n(n-1)(n-2)} \quad (69)$$

$$(\mu_1 \mu_7)_4 = \frac{\sum_{i \neq j} \mathbb{E} [X_i^7 X_j]}{n(n-1)} \quad (70)$$

$$(\mu_2^2 \mu_4)_4 = \frac{\sum_{i \neq j \neq k} \mathbb{E} [X_i^4 X_j^2 X_k^2]}{n(n-1)(n-2)} \quad (71)$$

$$(\mu_2^4)_4 = \frac{\sum_{i \neq j \neq k \neq l} \mathbb{E} [X_i^2 X_j^2 X_k^2 X_l^2]}{n(n-1)(n-2)(n-3)} \quad (72)$$

$$(\mu_2 \mu_6)_4 = \frac{\sum_{i \neq j} \mathbb{E} [X_i^6 X_j^2]}{n(n-1)} \quad (73)$$

$$(\mu_2 \mu_3^2)_4 = \frac{\sum_{i \neq j \neq k} \mathbb{E} [X_i^3 X_j^3 X_k^2]}{n(n-1)(n-2)} \quad (74)$$

$$(\mu_3 \mu_5)_4 = \frac{\sum_{i \neq j} \mathbb{E} [X_i^5 X_j^3]}{n(n-1)} \quad (75)$$

$$(\mu_4^2)_4 = \frac{\sum_{i \neq j} \mathbb{E} [X_i^4 X_j^4]}{n(n-1)} \quad (76)$$

$$(\mu_8)_4 = \frac{\sum_i \mathbb{E} [X_i^8]}{n} \quad (77)$$

B Cumulant computation

First of all, the first four cumulants, denoted by κ_i for $ii = 1, \dots, 4$ are obtained through standard relationships with respect to moments denoted by μ'_i as follows:

$$\kappa_1 = \mu'_1 \quad (78)$$

$$\kappa_2 = \mu'_2 - \mu_1'^2 \quad (79)$$

$$\kappa_3 = \mu'_3 - 3\mu'_2 \mu'_1 + 2\mu_1'^3 \quad (80)$$

$$\kappa_4 = \mu'_4 - 4\mu'_1 \mu'_3 - 3\mu_2'^2 + 12\mu_1'^2 \mu'_2 - 6\mu_1'^4 \quad (81)$$

Hence we are left with computing the first four moments of the sample variance. The first two moments of the sample variance are easy to compute and given for instance in [1]. For the cumulant of order 3 and 4, we first compute the different moments and then regroup the terms.

B.1 Third Moment for $s_n^2 \text{Sn2}$

Let us do some routine algebraic computation. We have

$$s_n^6 = \frac{1}{n^3(n-1)^3} \left((n-1) \sum_{i=1}^n X_i^2 - \sum_{i \neq j} X_i X_j \right)^3 \quad (82)$$

$$\begin{aligned} &= \frac{1}{n^3(n-1)^3} \left((n-1)^3 \left(\sum_{i=1}^n X_i^2 \right)^3 - 3(n-1)^2 \left(\sum_{i=1}^n X_i^2 \right)^2 \left(\sum_{k \neq l} X_k X_l \right) \right. \\ &\quad \left. + 3(n-1) \left(\sum_{i=1}^n X_i^2 \right) \left(\sum_{k \neq l} X_k X_l \right)^2 + \left(\sum_{k \neq l} X_k X_l \right)^3 \right) \quad (83) \end{aligned}$$

Let us expand. The first expansion $(\sum_{i=1}^n X_i^2)^3$ is easy and immediate:

$$\left(\sum_{i=1}^n X_i^2 \right)^3 = \sum_{i=1}^n X_i^6 + 3 \sum_{i \neq j} X_i^4 X_j^2 + \sum_{i \neq j \neq k} X_i^2 X_j^2 X_k^2 \quad (84)$$

In the expansion of $(\sum_{i=1}^n X_i^2)^2 (\sum_{j \neq k} X_j X_k)$, the possibilities are:

- $\sum_{i \neq j} X_i^5 X_j$
- $\sum_{i \neq j \neq k} X_i^4 X_j X_k$
- $\sum_{i \neq j} X_i^3 X_j^3$
- $\sum_{i \neq j \neq k} X_i^3 X_j^2 X_k$
- $\sum_{i \neq j \neq k \neq l} X_i^2 X_j^2 X_k X_l$

In the expansion of $(\sum_{i=1}^n X_i^2) (\sum_{j \neq k} X_j X_k)^2$, the possibilities are:

- $\sum_{i \neq j} X_i^4 X_j^2$
- $\sum_{i \neq j \neq k} X_i^4 X_j X_k$
- $\sum_{i \neq j \neq k} X_i^3 X_j^2 X_k$
- $\sum_{i \neq j \neq k \neq l} X_i^3 X_j X_k X_l$
- $\sum_{i \neq j \neq k} X_i^2 X_j^2 X_k^2$
- $\sum_{i \neq j \neq k \neq l} X_i^2 X_j^2 X_k X_l$

- $\sum_{i \neq j \neq k \neq l \neq m} X_i^2 X_j X_k X_l X_m$

In the expansion of $(\sum_{i \neq j} X_i X_j)^3$, the possibilities are:

- $\sum_{i \neq j} X_i^3 X_j^3$
- $\sum_{i \neq j \neq k} X_i^3 X_j^2 X_k$
- $\sum_{i \neq j \neq k} X_i^2 X_j^2 X_k^2$
- $\sum_{i \neq j \neq k \neq l} X_i^3 X_j X_k X_l$
- $\sum_{i \neq j \neq k \neq l} X_i^2 X_j^2 X_k X_l$
- $\sum_{i \neq j \neq k \neq l \neq m} X_i^2 X_j X_k X_l X_m$
- $\sum_{i \neq j \neq k \neq l \neq m \neq n} X_i X_j X_k X_l X_m$

which leads to

$$\begin{aligned}
\mathbb{E}[s_n^8] = & -\frac{(n-5)(n-4)(n-3)(n-2)\mu_1^6}{(n-1)^2 n^2} + \frac{3(n-5)(n-4)(n-3)(n-2)\mu_2\mu_1^4}{(n-1)^2 n^2} \\
& + \frac{4(n-3)(n-2)(3n-5)\mu_3\mu_1^3}{(n-1)^2 n^2} - \frac{3(n-3)(n-2)(n^2-6n+15)\mu_2^2\mu_1^2}{(n-1)^2 n^2} \\
& - \frac{3(n-5)(n-2)\mu_4\mu_1^2}{(n-1)n^2} - \frac{12(n-2)(n^2-4n+5)\mu_2\mu_3\mu_1}{(n-1)^2 n^2} - \frac{6\mu_5\mu_1}{n^2} \\
& - \frac{2(3n^2-6n+5)\mu_3^2}{(n-1)^2 n^2} + \frac{3(n^2-2n+5)\mu_2\mu_4}{(n-1)n^2} + \frac{\mu_6}{n^2} \\
& + \frac{(n-2)(n^3-3n^2+9n-15)\mu_2^3}{(n-1)^2 n^2}
\end{aligned} \tag{85}$$

Regrouping all the terms leads to

$$\begin{aligned}
\mathbb{E}[s_n^6] = & M_{0,0}^3 + \frac{M_{1,0}^3}{n-1} + \frac{M_{0,1}^3}{n} + \frac{M_{2,0}^3}{(n-1)^2} + \frac{M_{1,1}^3}{(n-1)n} + \frac{M_{0,2}^3}{n^2} \\
& + \frac{M_{2,1}^3}{(n-1)^2 n} + \frac{M_{1,2}^3}{(n-1)n^2} + \frac{M_{2,2}^3}{(n-1)^2 n^2}
\end{aligned} \tag{86}$$

with

$$M_{2,0}^3 = -60\mu_1^6 + 180\mu_1^4\mu_2 - 68\mu_1^3\mu_3 - 129\mu_1^2\mu_2^2 + 60\mu_2\mu_3\mu_1 + 13\mu_2^3 - 6\mu_3^2 \quad (87)$$

$$M_{0,0}^3 = -\mu_1^6 + 3\mu_1^4\mu_2 - 3\mu_1^2\mu_2^2 + \mu_2^3 \quad (88)$$

$$M_{2,2}^3 = -120\mu_1^6 + 360\mu_1^4\mu_2 - 120\mu_1^3\mu_3 - 270\mu_1^2\mu_2^2 + 120\mu_1\mu_2\mu_3 + 30\mu_2^3 - 10\mu_3^2 \quad (89)$$

$$M_{1,0}^3 = 12\mu_1^6 - 36\mu_1^4\mu_2 + 12\mu_1^3\mu_3 + 27\mu_1^2\mu_2^2 - 12\mu_1\mu_2\mu_3 - 3\mu_2^3 \quad (90)$$

$$M_{2,1}^3 = 154\mu_1^6 - 462\mu_2\mu_1^4 + 172\mu_1^3\mu_3 + 333\mu_1^2\mu_2^2 - 156\mu_1\mu_2\mu_3 - 33\mu_2^3 + 12\mu_3^2 \quad (91)$$

$$M_{1,0}^3 = -3\mu_1^2\mu_4 + 3\mu_2\mu_4 \quad (92)$$

$$M_{1,2}^3 = 15\mu_2\mu_4 - 30\mu_1^2\mu_4 \quad (93)$$

$$M_{1,1}^3 = 21\mu_1^2\mu_4 - 6\mu_2\mu_4 \quad (94)$$

$$M_{0,2}^3 = \mu_6 - 6\mu_1\mu_5 \quad (95)$$

Using previous results in the relationship between cumulant and moment leads to the result. \square

B.2 Fourth Moment for $s_n^2 \text{Sn2}$

Let us do some routine algebraic computation. We have

$$s_n^8 = \frac{1}{n^4(n-1)^4} \left((n-1) \sum_{i=1}^n X_i^2 - \sum_{i \neq j} X_i X_j \right)^4 \quad (96)$$

$$\begin{aligned} &= \frac{1}{n^4(n-1)^4} \left((n-1)^4 \left(\sum_{i=1}^n X_i^2 \right)^4 - 4(n-1)^3 \left(\sum_{i=1}^n X_i^2 \right)^3 \left(\sum_{k \neq l} X_k X_l \right) + 6(n-1)^2 \right. \\ &\quad \left. \left(\sum_{i=1}^n X_i^2 \right)^2 \sum_{k \neq l} X_k X_l \right)^2 + 4(n-1) \left(\sum_{i=1}^n X_i^2 \right) \left(\sum_{k \neq l} X_k X_l \right)^3 + \left(\sum_{k \neq l} X_k X_l \right)^4 \right) \quad (97) \end{aligned}$$

Again, one needs to expand all the terms and look at all the various possibilities to demonstrate the following relationship where we have regrouped

against each of the symmetric empirical moment estimator

$$\begin{aligned}
\mathbb{E}[s_n^8] = & \frac{(n-7)(n-6)(n-5)(n-4)(n-3)(n-2)\mu_1^8}{(n-1)^3n^3} - \frac{4(n-7)(n-6)(n-5)(n-4)(n-3)(n-2)\mu_2\mu_1^6}{(n-1)^3n^3} \\
& - \frac{8(n-5)(n-4)(n-3)(n-2)(3n-7)\mu_3\mu_1^5}{(n-1)^3n^3} + \frac{6(n-5)(n-4)(n-3)(n-2)(n^2-10n+35)\mu_2^2\mu_1^4}{(n-1)^3n^3} \\
& + \frac{2(n-4)(n-3)(n-2)(3n^2-30n+35)\mu_4\mu_1^4}{(n-1)^3n^3} + \frac{16(n-4)(n-3)(n-2)(3n^2-20n+35)\mu_2\mu_3\mu_1^3}{(n-1)^3n^3} \\
& + \frac{8(n-3)(n-2)(3n-7)\mu_5\mu_1^3}{(n-1)^2n^3} - \frac{4(n-4)(n-3)(n-2)(n^3-9n^2+45n-105)\mu_2^3\mu_1^2}{(n-1)^3n^3} \\
& + \frac{8(n-3)(n-2)(9n^2-30n+35)\mu_3^2\mu_1^2}{(n-1)^3n^3} - \frac{12(n-3)(n-2)(n^3-9n^2+35n-35)\mu_2\mu_4\mu_1^2}{(n-1)^3n^3} \\
& - \frac{4(n-7)(n-2)\mu_6\mu_1^2}{(n-1)n^3} - \frac{24(n-3)(n-2)(n^3-7n^2+25n-35)\mu_2^2\mu_3\mu_1}{(n-1)^3n^3} \\
& - \frac{8(n-2)(3n^3-21n^2+45n-35)\mu_3\mu_4\mu_1}{(n-1)^3n^3} - \frac{24(n-2)(n^2-4n+7)\mu_2\mu_5\mu_1}{(n-1)^2n^3} \\
& - \frac{8\mu_7\mu_1}{n^3} + \frac{(n-3)(n-2)(n^4-4n^3+18n^2-60n+105)\mu_2^4}{(n-1)^3n^3} \\
& - \frac{8(n-2)(3n^3-15n^2+35n-35)\mu_2\mu_3^2}{(n-1)^3n^3} + \frac{(3n^4-12n^3+42n^2-60n+35)\mu_4^2}{(n-1)^3n^3} \\
& + \frac{6(n-2)(n^4-4n^3+16n^2-40n+35)\mu_2^2\mu_4}{(n-1)^3n^3} - \frac{8(3n^2-6n+7)\mu_3\mu_5}{(n-1)^2n^3} \\
& + \frac{4(n^2-2n+7)\mu_2\mu_6}{(n-1)n^3} + \frac{\mu_8}{n^3} \tag{98}
\end{aligned}$$

Once this is done, one needs to collect all the terms and can get the final result. \square

References

- [1] Eric Benhamou, A few properties of sample variance, *arxiv*, (September-October, 2018).
- [2] B.V. Gnedenko and A.N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Cambridge, Mass, 1954.
- [3] Robert V. Hogg, Joseph W. McKean and Allen T. Craig, *Introduction to Mathematical Statistics*, Pearson, 4th edition, 1978.

- [4] A. M. Mathai, Storage capacity of a dam with gamma type inputs, *Annals of the Institute of Statistical Mathematics*, **3**, (1982), 591-597.
- [5] Anna Mikusheva, Second Order Expansion of the t-statistics in AR(1) models, *Econometric Theory*, Cambridge University Press, **31**(3), (2015), 426-448, DOI=10.1017/S0266466614000383.
- [6] P. G. Moschopoulos, The distribution of the sum of independent gamma random variables, *Annals of the Institute of Statistical Mathematics*, **37**, (1985), 541-544.