# Sharp Large Deviations for empirical correlation coefficients

Thi Kim Tien Truong, Marguerite Zani

# Sharp Large Deviations for empirical correlation coefficients

T.K.T. Truong [*] , M. Zani [*†]

**Abstract**

In this paper, we study Sharp Large Deviations for empirical Pearson coefficients, i.e. $r_n = \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)/\sqrt{\sum_{i=1}(X_i - \bar{X}_n)^2 \sum_{i=1}(Y_i - \bar{Y}_n)^2}$ or $\tilde{r}_n = \sum_{i=1}^n (X_i - \mathbb{E}(X))(Y_i - \mathbb{E}(Y))/\sqrt{\sum_{i=1}(X_i - \mathbb{E}(X))^2 \sum_{i=1}(Y_i - \mathbb{E}(Y))^2}$ (when the expectations are known). Our framework is for random samples $(X_i, Y_i)$ either Spherical or Gaussian. In each case, we follow the scheme of Bercu et al. We also compute the Bahadur exact slope in the Gaussian case.

*Keywords:* Pearson's Empirical Correlation Coefficient, Sharp Large Deviations, Spherical Distribution, Gaussian distribution.

## 1 Introduction

The Beauvais–Pearson linear correlation coefficient between two real random variables $X$ and $Y$ is defined by

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}},$$

whenever this quantity exists. Such quantities were formally defined more than a century ago by Pearson [22, 23]. The correlation describes the linear relation between two random variables. It is clear from Cauchy–Schwartz inequality that the absolute value of $\rho$ is less than or equal to 1. Moreover, $\rho = \pm 1$ if and only if $X$ and $Y$ are linearly related. When $\rho = 0$ we say that $X$ and $Y$ are uncorrelated, i.e. linearly independent. The empirical counterpart is the following. Let us consider two samples $\mathbf{X} = (X_1, \cdots, X_n)$ and $\mathbf{Y} = (Y_1, \cdots, Y_n)$. The so–called empirical Pearson correlation coefficient is given by

$$r_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2 \sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}\,, \tag{1}$$

where $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ and $\bar{Y}_n = \frac{1}{n}\sum_{k=1}^n Y_k$ are the empirical means of the samples. Whenever $E(\mathbf{X})$ and $E(\mathbf{Y})$ are both known, we consider $\tilde{r}_n$:

$$\tilde{r}_n = \frac{\sum_{i=1}^n (X_i - E(X_i))(Y_i - E(Y_i))}{\sqrt{\sum_{i=1}^n (X_i - E(X_i))^2 \sum_{i=1}^n (Y_i - E(Y_i))^2}}\,. \tag{2}$$

[*]Institut Denis Poisson, Université d'Orléans, Université de Tours, CNRS, Route de Chartres, B.P. 6759, 45067, Orléans cedex 2, France

[†]Corresponding author: marguerite.zani@univ-orleans.fr

The study of the correlation coefficients is detailed in many references (see e.g. [19] or [25]) and it is shown that many competing correlation indexes are special cases of Pearson's correlation coefficient ([24]). The asymptotic behaviour of $(r_n)_n$, $(\tilde{r}_n)_n$ is worth considering. It is clear that when $\mathbf{X}$ and $\mathbf{Y}$ are independent, $r_n$, $\tilde{r}_n \to 0$ when $n \to \infty$. Moreover, whenever $(\mathbf{X}, \mathbf{Y})$ are sampled from a known distribution $(X, Y)$, $r_n$, $\tilde{r}_n \to \rho$ when $n \to \infty$. In this paper, we study Sharp Large Deviations (SLD) associated to these asymptotics.

Large Deviations for empirical correlation coefficients have been studied by Si [26] in the Gaussian case. We extend his results to SLD in the spherical and Gaussian cases. It can be noticed here that for the Gaussian case, we prove SLD on a resctricted domain of $\rho$ since the convexity properties of the functions are only true for $0 \leq |\rho| \leq \rho_0$, where $\rho_0$ is explicitely defined. This point was not noticed in [26] since the large deviations are given through a contraction principle which is actually not valid (the function used is not continuous). We stress the fact that things have to be handled in a different way for the case $|\rho| > \rho_0$, and there is no proof that the rate function should be the same.

We consider here the asymptotic development of

$$P(r_n \geq c) \text{ or, equivalently } P(\tilde{r}_n \geq c)$$

for $0 < c < 1$. We follow the scheme of Bercu et al. [5, 6] and split:

$$P(r_n \geq c) = A_n B_n \,,$$

where

$$A_n = \exp[n(L_n(\lambda_c) - c\lambda_c)] \,, \tag{3}$$

$$B_n = \mathbb{E}_n[\exp[-n\lambda_c(r_n - c)]\mathbb{1}_{r_n \geq c}] \ , \tag{4}$$

$L_n$ is the normalized cumulant generating function (n.c.g.f.) of $r_n$, $L$ its limit as $n \to \infty$ and $\lambda_c$ is the unique $\lambda$ such that $L'(\lambda_c) = c$. We perform the following change of probability

$$\frac{dQ_n}{dP} = e^{\lambda_c n r_n - n L_n(\lambda_c)} \,,$$

and $\mathbb{E}_n$ is the expectation under this new probability $Q_n$. The key point here is to develop the characteristic function $\Phi_n$ of $\frac{\sqrt{n}(r_n - c)}{\sigma_c}$. We use an expansion already computed in the i.i.d. case by Cramér (see [10], Lemma 2, p.72) and Esseen [13].

Such studies have been done in the context of small ball deviations and Gaussian seminorms by Ibragimov [16], Li [17], Sytaya [27], Zolotarev [31], with an anaytic point of vue and different asymptotics. Independently, with large deviations techniques, was done a similar work by Dembo, Meyer-Wolf and Zeitouni [11, 18]. We can also cite works in other contexts: Ben Arous [4] on asymptotic expansion of the heat kernel associated with an hypoelliptic operator (in small time), and Bolthausen [7] on the limiting behaviour of the partition function for random vectors in Banach spaces in a general i.i.d. case.

The paper is organized as follows: in Sections 2 and 3, we present the SLD results in the spherical and Gaussian cases; Section 4 is devoted to the proofs. In Section 5, we briefly extend our results to any order developments as we present an application to Bahadur exact slopes for the test based on $r_n$ in the Gaussian case. Finally, in an Appendix, we give some more details and references on the Laplace method.

## 2 Spherical distribution

In this section, we study empirical correlations coefficients (1) and (2) under the following spherical distribution assumption. We denote by $\mathbf{v}'$ the transpose of vector $\mathbf{v}$.

**Assumption 2.1** *Let* $\mathbf{X} = (X_1, \cdots, X_n)'$ *and* $\mathbf{Y} = (Y_1, \cdots, Y_n)'$, $n \in \mathbb{N}$, $n > 2$, *be two independent random vectors where* $\mathbf{X}$ *has a* $n$*-variate spherical distribution with* $P(\mathbf{X} = (0, \cdots, 0)') = 0$ *and* $\mathbf{Y}$ *has any distribution with* $P(\mathbf{Y} \in \{\mathbf{1}\}) = 0$ *(where* $\mathbf{1} = \{k(1, \cdots, 1)', k \in \mathbb{R}\}$*).*

### 2.1 SLDP for $r_n$

In order to derive SLD for $(r_n)_n$ we compute the n.c.g.f.

$$L_n(\lambda) = \frac{1}{n} \log E(e^{n\lambda r_n}) \,. \tag{5}$$

The asymptotics of $L_n$ are given in the following proposition:

**Proposition 2.2** *For any* $\lambda \in \mathbb{R}$, *we have*

$$E\left(e^{n\lambda r_n}\right) = \frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\Gamma(\frac{n-2}{2})} e^{nh(r_0(\lambda))} \left(\frac{c_0(\lambda)}{\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right)\right) , \tag{6}$$

*where*

- $h(r) = \lambda r + \frac{1}{2}\log(1 - r^2)$,
- $r_0(\lambda)$ *is the unique root in* $]-1, 1[$ *of* $h'(r) = 0$, *i.e.*

$$r_0(\lambda) = \frac{-1 + \sqrt{1 + 4\lambda^2}}{2\lambda} \,, \tag{7}$$

- $g(r) = (1 - r^2)^{-2}$ *and* $c_0(\lambda) = \sqrt{\frac{2\pi}{|h''(r_0(\lambda))|}} g(r_0(\lambda)) \,.$

*Therefore*

$$L_n(\lambda) = L(\lambda) - \frac{1}{n}\left[\frac{1}{2}\log\sqrt{1 + 4\lambda^2} - \frac{3}{2}\log\frac{1 + \sqrt{1 + 4\lambda^2}}{2}\right] + O\left(\frac{1}{n^2}\right) . \tag{8}$$

*where* $L$ *is the limit normalized* log*–Laplace transform of* $r_n$*:*

$$L(\lambda) = h(r_0(\lambda)). \tag{9}$$

The proof of this proposition is postponed to Section 4. Now we have the following SLDP:

**Theorem 2.3** *For any $0 < c < 1$, under Assumption (2.1), we have*

$$P(r_n \geq c) = \frac{e^{-nL^*(c) - \frac{1}{2}\log(1+4\lambda_c^2) + \frac{3}{2}\log\frac{1+\sqrt{1+4\lambda_c^2}}{2}}}{\lambda_c \sigma_c \sqrt{2\pi n}}(1 + o(1)), \tag{10}$$

*where*

- *$\lambda_c$ is the unique solution of $L'(\lambda) = c$, i.e. $\lambda_c = \dfrac{c}{1-c^2}$,*
- *$\sigma_c^2 = L''(\lambda_c) = \dfrac{(1-c^2)^2}{1+c^2}$,*
- *$L^*(y) = -\frac{1}{2}\log(1-y^2)$.*

*Proof:*

To prove the SLD for $(r_n)_n$, we proceed as in Bercu et al. [5, 6]. The following lemma, which proof is given in the Section 4, gives some basic properties of $L$:

**Lemma 2.4** *Let $L(\lambda) = h(r_0(\lambda))$ where $h$ and $r_0$ are defined in Proposition 2.2, we have*

- *$L$ is defined on $\mathbb{R}$ and $L$ is $\mathcal{C}^\infty$ on its domain.*
- *$L$ is a strictly convex function on $\mathbb{R}$, $L$ reaches its minimum at $\lambda = 0$. Moreover for any $\lambda \in \mathbb{R}$, $L'(\lambda) \in ]-1, 1[$.*
- *The Legendre dual of $L$ is defined on $]-1, 1[$ and computed as*

$$L^*(y) = \sup_{\lambda \in \mathbb{R}}\{\lambda y - L(\lambda)\} = -\frac{1}{2}\log(1-y^2). \tag{11}$$

.

Let $0 < c < 1$ and $\lambda_c > 0$ such that $L'(\lambda_c) = c$. Then

$$L^*(c) = c\lambda_c - L(\lambda_c),$$

We denote by $\sigma_c^2 = L''(\lambda_c)$, and define the following change of probability:

$$\frac{dQ_n}{dP} = e^{\lambda_c n r_n - nL_n(\lambda_c)}. \tag{12}$$

The expectation under $Q_n$ is denoted by $E_n$. We write

$$P(r_n \geq c) = A_n B_n, \tag{13}$$

where

$$A_n = \exp[n(L_n(\lambda_c) - c\lambda_c)],$$
$$B_n = E_n(\exp[-n\lambda_c(r_n - c)]\mathbb{1}_{r_n \geq c}).$$

On the one hand, from (8)

$$A_n = \exp[-nL^*(c) - \frac{1}{4}\log(1+4\lambda_c^2) + \frac{3}{2}\log\frac{1+\sqrt{1+4\lambda^2}}{2}]\left(1+O\left(\frac{1}{n}\right)\right).$$

On the other hand, let us denote by

$$U_n = \frac{\sqrt{n}(r_n - c)}{\sigma_c},$$

$$\Phi_n(u) = E_n(e^{iuU_n}) = \exp(-\frac{iu\sqrt{n}}{\sigma_c}c + nL_n(\lambda_c + \frac{iu}{\sigma_c\sqrt{n}}) - nL_n(\lambda_c))\,.$$

We have the following technical results on $\Phi_n$, proved in Section 4.

**Lemma 2.5** *For any $K \in \mathbb{N}^*$, $\eta > 0$, for $n$ large enough and any $u \in \mathbb{R}$,*

$$|\Phi_n(u)| \leq \frac{1}{|\lambda_c + \frac{iu}{\sigma_c\sqrt{n}}|^K}\frac{c_0^K(\lambda)}{c_0(\lambda)}(1+\eta)\,. \tag{14}$$

*where $c_0$ and $c_0^K$ are the first coefficients in Laplace's method (see Theorem 5.6), corresponding respectively to*

$$g(r) = (1-r^2)^{-2}$$

*and*

$$g^K(r) = (2r)^K(1-r^2)^{-K-2}\,.$$

From lemma above, choosing $K \geq 2$, we see that $\Phi_n$ is in $L^2$ and by Parseval formula,

$$B_n = E_n[e^{-\lambda_c\sigma_c\sqrt{n}U_n}\mathbb{1}_{U_n\geq 0}] = \frac{1}{2\pi}\int_{\mathbb{R}}\left(\frac{1}{\lambda_c\sigma_c\sqrt{n}+iu}\right)\Phi_n(u)du = \frac{C_n}{\lambda_c\sigma_c\sqrt{2\pi n}},$$

where

$$C_n = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\left(1+\frac{iu}{\lambda_c\sigma_c\sqrt{n}}\right)^{-1}\Phi_n(u)du.$$

The key point here is to study the asymptotics of $\Phi_n$.

**Lemma 2.6** *We have*

$$\lim_{n\to\infty}\Phi_n(u) = e^{-u^2/2} \text{ and } \lim_{n\to\infty}C_n = 1\,.$$

From lemma above, which proof is postponed to Section 4, we have equation (10).

□

### 2.2 Known expectation

In case $E(\mathbf{X})$ and $E(\mathbf{Y})$ are both known, we consider $\tilde{r}_n$ given in formula (2) which can be written as follows

$$\tilde{r}_n = \frac{(\mathbf{X} - E(\mathbf{X}))' (\mathbf{Y} - E(\mathbf{Y}))}{\|\mathbf{X} - E(\mathbf{X})\| \, \|\mathbf{Y} - E(\mathbf{Y})\|} . \tag{15}$$

We can derive a SLD result similar to the previous one. The following proposition gives the expression of the n.c.g.f. of $\tilde{r}_n$:

**Proposition 2.7** *For any $\lambda \in \mathbb{R}$, we have*

$$E(e^{n\lambda\tilde{r}_n}) = \frac{\Gamma(\frac{n}{2})}{\pi^{1/2}\Gamma(\frac{n-1}{2})} e^{nh(r_0(\lambda))} \left( \frac{\tilde{c}_0(\lambda)}{\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right) \right) , \tag{16}$$

*where*

- $h(r) = \lambda r + \frac{1}{2}\log(1 - r^2)$,
- $r_0(\lambda)$ *is the unique root in* $]-1, 1[$ *of* $h'(r) = 0$, *i.e.*

$$r_0(\lambda) = \frac{-1 + \sqrt{1 + 4\lambda^2}}{2\lambda} ,$$

- $\tilde{g}(r) = (1 - r^2)^{-3/2}$ *and* $\tilde{c}_0(\lambda) = \sqrt{\frac{2\pi}{|h''(r_0(\lambda))|}} \tilde{g}(r_0(\lambda))$ .

*The n.c.g.f. of $\tilde{r}_n$ is*

$$\tilde{L}_n(\lambda) = h(r_0(\lambda)) - \frac{1}{n}\left[ \frac{1}{2}\log\sqrt{1 + 4\lambda^2} - \log\frac{1 + \sqrt{1 + 4\lambda^2}}{2} \right] + O\left(\frac{1}{n^2}\right) . \tag{17}$$

This proposition is proved in Section 4. We have the following SLDP:

**Theorem 2.8** *For any $0 < c < 1$, under Assumption* (2.1), *we have*

$$P(\tilde{r}_n \geq c) = \frac{\exp^{-nL^*(c) - \frac{1}{4}\log(1+4\lambda_c^2) + \log\frac{1+\sqrt{1+4\lambda_c^2}}{2}}}{\lambda_c \sigma_c \sqrt{2\pi n}} (1 + o(1)). \tag{18}$$

*Proof:*

The proof of Theorem 2.8 is exactly similar to the one of Theorem 2.3 and formula (10) is changed to (18) according to the way formula (8) is changed to (17).

□

## 3 Gaussian case

**Assumption 3.1** *Let $(X, Y)$ be a $\mathbb{R}^2$-valued Gaussian random vector where $\sigma_1^2 = Var(X)$, $\sigma_2^2 = Var(Y)$ and $\rho$ is the correlation coefficient: $Cov(X, Y) = \rho\sigma_1\sigma_2$. We consider $(\mathbf{X}, \mathbf{Y}) = \{(X_i, Y_i), i = 1, \cdots n\}$ an i.i.d. sample of $(X, Y)$.*

### 3.1 General case

We deal with the Pearson coefficient given in (1). As presviously mentioned, Large deviations for $(r_n)_n$ are detailed in the paper of Si [26]. It can be noted that the contraction principle used by Si is not valid here. The rate function is correct however, but only on some domain of $\rho$. We can give an expression of the normalized log–Laplace transform $L_n$ given by (5).

**Proposition 3.2** *Let us define*

$$\rho_0 := \frac{\sqrt{3+2\sqrt{3}}}{3}\,.$$

*For any $\lambda \in \mathbb{R}$ and $\rho$ such that $|\rho| \leq \rho_0$, we have the n.c.g.f. of $r_n$:*

$$L_n(\lambda) = \overline{h}(r_0(\lambda)) + \frac{1}{2}\log(1-\rho^2) + \frac{1}{n}\left[\log \overline{g}_\rho(r_0(\lambda)) - \frac{1}{2}\log|\overline{h}''(r_0(\lambda))|\right] + O\left(\frac{1}{n^2}\right), \quad (19)$$

*in which*

- $\overline{h}(r) = \lambda r - \log(1-\rho r) + \frac{1}{2}\log(1-r^2)$,
- $r_0(\lambda)$ *is the unique real root in* $]-1,1[$ *of* $\overline{h}'(r) = 0$,
- $\overline{g}_\rho(r) = (1-\rho^2)^{-1/2}(1-\rho r)^{3/2}(1-r^2)^{-2}$.

The proof of this proposition is postponed to Section 4. We prove the following SLDP:

**Theorem 3.3** *For any $0 \leq \rho < c < 1$ and $|\rho| \leq \rho_0$ (with the notations of Proposition 3.2), we have*

$$P(r_n \geq c) = \frac{e^{-nL^*(c)+\log \overline{g}_\rho(r_0(\lambda_c))-\frac{1}{2}\log|\overline{h}''(r_0(\lambda_c))|}}{\lambda_c \sigma_c \sqrt{2\pi n}}(1+o(1))\,, \quad (20)$$

*where for any $-1 < y < 1$,*

$$L^*(y) = \log\left(\frac{1-\rho y}{\sqrt{(1-\rho^2)}\sqrt{(1-y^2)}}\right). \quad (21)$$

*Proof:*

Following the Proof of Theorem 2.3, we can easy obtain (20). Note that the rate function in Si [26] matches our (21).

□

### 3.2 Known expectations

In case $E(X)$ and $E(Y)$ are both known; and $\rho = 0$, we have the following result

**Proposition 3.4** *The n.c.g.f. of $\tilde{r}_n$ is given for any $\lambda \in \mathbb{R}$ by*

$$L_n(\lambda) = h(u_0(\lambda)) - \frac{1}{4n}\log(1+4\lambda^2) + O\left(\frac{1}{n^2}\right), \quad (22)$$

*where*

- $h(r) = \lambda r + \frac{1}{2}\log(1-r^2)$,
- $u_0(\lambda)$ *is the unique solution of* $h'(\lambda) = 0$ *in* $]-1,1[$.

The proof is postponed to Section 4. The SLDP is therefore:

**Theorem 3.5** *When* $\rho = 0$ *and under Assumption 3.1, for* $0 < c < 1$*, we have*

$$P(\tilde{r}_n \geq c) = \frac{e^{-nL^*(c)-\frac{1}{4}\log(1-4\lambda_c^2)}}{\lambda_c \sigma_c \sqrt{n}}(1+o(1)), \tag{23}$$

*where* $L^*$ *is given in Theorem 2.3.*

# 4 Proofs

## 4.1 Proof of Proposition 2.2

We know from Muirhead (Theorem 5.1.1, [19]) that

$$(n-2)^{1/2}\frac{r_n}{(1-(r_n)^2)^{1/2}}$$

has a Student's $t_{n-2}$-distribution. Hence the density function of $r_n$ is

$$f_n(r) = \frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\Gamma(\frac{n-2}{2})}(1-r^2)^{(n-4)/2} \quad (-1 < r < 1). \tag{24}$$

Applying Theorem 5.6, we get

$$\begin{aligned} E\left(e^{n\lambda r_n}\right) &= \int_{-1}^{1} e^{n\lambda r} f_n(r)dr = \int_{-1}^{1} e^{n\lambda r}\frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\Gamma(\frac{n-2}{2})}(1-r^2)^{(n-4)/2}dr \\ &= \frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\Gamma(\frac{n-2}{2})} e^{nh(r_0(\lambda))}\left(\frac{c_0(\lambda)}{\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right)\right). \end{aligned}$$

where $h$, $r_0$ and $c_0$ are given in Proposition 2.2.

So we have

$$E\left(e^{n\lambda r_n}\right) = \frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\Gamma(\frac{n-2}{2})}\sqrt{\frac{2\pi}{n}} e^{nh(r_0(\lambda))}\frac{g(r_0(\lambda))}{\sqrt{|h''(r_0(\lambda))|}}\left(1+O\left(\frac{1}{n}\right)\right) \tag{25}$$

$$= \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})}\sqrt{\frac{2}{n}}\, e^{nh(r_0(\lambda))}\frac{1}{(1-r_0(\lambda)^2)\sqrt{1+r_0(\lambda)^2}}\left(1+O\left(\frac{1}{n}\right)\right) \tag{26}$$

From the duplication formula (see e.g. Olver [21])

$$2^{2z-1}\Gamma(z)\Gamma(z+\frac{1}{2}) = \sqrt{\pi}\Gamma(2z)\,,$$

as well as the Stirling formula (see [21])

$$\log\Gamma(z) = z\log z - z - \frac{1}{2}\log z + \log\sqrt{2\pi} + O\left(\frac{1}{Re(z)}\right) \text{ , as } Re(z) \to \infty\,,$$

formula (26) above becomes

$$E(e^{n\lambda r}) = e^{nh(r_0(\lambda))}\frac{1}{(1-r_0(\lambda)^2)\sqrt{1+r_0(\lambda)^2}}\left(1+O\left(\frac{1}{n}\right)\right)\,.$$

With the expression of $r_0$, we get formula (8).

### 4.2 Proof of Lemma 2.4

We can explicit the full expression of $L$:

$$L(\lambda) = \frac{-1+\sqrt{1+4\lambda^2}}{2} - \frac{1}{2}\log(\frac{1+\sqrt{1+4\lambda^2}}{4})\,. \tag{27}$$

It is easy to see that $L$ is defined on $\mathbb{R}$, $\mathcal{C}^\infty$ on its domain.

From the definition of $L$ we can deduce

$$L'(\lambda) = r_0(\lambda) + h'(r_0(\lambda)) = r_0(\lambda), \tag{28}$$

and by construction of $r_0$, $L' \in ]-1,1[$. Now we can compute

$$L''(\lambda) = r_0'(\lambda) = \frac{1}{2\lambda^2}\left(1 - \frac{1}{\sqrt{1+4\lambda^2}}\right), \tag{29}$$

and it is easily seen that $L''(\lambda) > 0$ for any $\lambda \in \mathbb{R}^*$ and $L''(0)$ can be defined by continuity as 1. Hence $L$ is strictly convex on $\mathbb{R}$ and has its minimum at $\lambda = 0$. Moreover, if we have

$$L'(\lambda_c) = r_0(\lambda_c) = c,$$

then $0 < c < 1$ implies $\lambda_c > 0$ and we can obtain

$$4\lambda_c(\lambda_c(1-c^2) - c) = 0.$$

This leads us to the expression

$$\lambda_c = \frac{c}{1-c^2}\,.$$

Hence the preceding expression yields

$$\sigma_c^2 = L''(\lambda_c) = \frac{(1-c^2)^2}{1+c^2}\,.$$

### 4.3 Proof of Lemmas 2.5 and 2.6

The proof of Lemma 2.5 is based on iterated integrations by parts. We detail below the steps.

$$\begin{aligned}\Phi_n(u) = E_n(e^{iuU_n}) &= \int_{\mathbb{R}} e^{iu\frac{\sqrt{n}(r-c)}{\sigma_c}} f_n(r) e^{\lambda_c n r - nL_n(\lambda_c)} dr \\ &= \mathbf{\Gamma}_n \, e^{-iu\frac{\sqrt{n}c}{\sigma_c}} e^{-nL_n(\lambda_c)} \int_{-1}^{1} e^{(iu\frac{\sqrt{n}}{\sigma_c}+\lambda_c n)r} (1-r^2)^{n/2-2} dr,\end{aligned}$$

where, for seek of simplicity, we denote by

$$\mathbf{\Gamma}_n = \frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\Gamma(\frac{n-2}{2})}\,. \tag{30}$$

For $K \in \mathbb{N}^*$, performing $K$ integrations by part, since $f_n$ is zero at $-1$ and $1$ when $n$ is large enough, we get:

$$\begin{aligned}\Phi_n(u) = \mathbf{\Gamma}_n e^{-iu\frac{\sqrt{n}c}{\sigma_c}} e^{-nL_n(\lambda_c)} \times \cdots \\ \cdots \times \frac{(\frac{n}{2}-2)(\frac{n}{2}-3)\cdots(\frac{n}{2}-K-1)}{\left(iu\frac{\sqrt{n}}{\sigma_c}+\lambda_c n\right)^K} \int_{-1}^{1} e^{(iu\frac{\sqrt{n}}{\sigma_c}+\lambda_c n)r} (-2r)^K (1-r^2)^{n/2-2-K} dr.\end{aligned}$$

Hence,

$$|\Phi_n(u)| \leq \mathbf{\Gamma}_n e^{-nL_n(\lambda_c)} \frac{(\frac{n}{2}-2)(\frac{n}{2}-3)\cdots(\frac{n}{2}-K-1)}{\left|iu\frac{\sqrt{n}}{\sigma_c}+\lambda_c n\right|^K} \int_{-1}^{1} e^{\lambda_c n r} (2r)^K (1-r^2)^{n/2-2-K} dr.$$

Using Laplace's method once again (see the Appendix), for a given $\eta > 0$ we can find $N$ large enough such that for any $n \geq N$,

$$|\Phi_n(u)| \leq \frac{1}{|\lambda_c + \frac{iu}{\sqrt{n}\sigma_c}|^K} \frac{c_0^K(\lambda)}{c_0(\lambda)} (1+\eta)\,. \tag{31}$$

□

To prove Lemma 2.6, we first split $C_n$ into two terms:

$$C_n = \frac{1}{\sqrt{2\pi}} \int_{|u|\leq n^\alpha} \left(1 + \frac{iu}{\lambda_c \sigma_c \sqrt{n}}\right)^{-1} \Phi_n(u) du + \frac{1}{\sqrt{2\pi}} \int_{|u|>n^\alpha} \left(1 + \frac{iu}{\lambda_c \sigma_c \sqrt{n}}\right)^{-1} \Phi_n(u) du. \tag{32}$$

For the second term in the RHS of (32) we have

$$\begin{aligned}\left|\int_{|u|>n^\alpha} \frac{1}{\left(1+\frac{iu}{\lambda_c\sigma_c\sqrt{n}}\right)}\Phi_n(u)du\right| &\leq \int_{|u|>n^\alpha} \frac{1}{\left|1+\frac{iu}{\lambda_c\sigma_c\sqrt{n}}\right|}\Phi_n(u)du \\ &\leq \int_{|u|>n^\alpha} \frac{1}{|\lambda_c|^K\left|1+\frac{iu}{\lambda_c\sigma_c\sqrt{n}}\right|^{K+1}}du\frac{c_0^K(\lambda_c)}{c_0(\lambda_c)}(1+\eta) \\ &\leq \frac{c_0^K(\lambda_c)}{|\lambda_c|^K c_0(\lambda_c)}(1+\eta)\int_{|u|>n^\alpha}\frac{1}{\left(1+\frac{u^2}{\lambda_c^2\sigma_c^2 n}\right)^{(K+1)/2}}du \\ &\leq \frac{c_0^K(\lambda_c)}{|\lambda_c|^K c_0(\lambda_c)}(1+\eta)(\lambda_c^2\sigma_c^2 n)^{(K+1)/2}2\frac{n^{-\alpha K}}{K}.\end{aligned}$$

In order to have a negligible term, it is enough to have $-K\alpha+\frac{K+1}{2}<0$, i.e. fixing $K=3$, $\alpha=\frac{3}{4}$. Now for the domain $\{|u|\leq n^\alpha\}$, we study more precisely the expression

$$\Phi_n(u)=E_n(e^{iuU_n})=\exp\left[-\frac{iu\sqrt{n}}{\sigma_c}c+nL_n(\lambda_c+\frac{iu}{\sigma_c\sqrt{n}})-nL_n(\lambda_c)\right]. \tag{33}$$

We first remark that $E(e^{n\lambda r_n})$ is analytic in $\lambda$ on $\mathbb{R}$, hence it can be expanded by analytic continuation and $L_n(\lambda+iy)$ for $\lambda,y\in\mathbb{R}$ is well defined. From the analyticity, we can expand in Taylor series the expression (33) above.

$$\begin{aligned}\Phi_n(\lambda_c) &= \exp\{-iu\frac{\sqrt{n}c}{\sigma_c}+n\sum_{k=1}^{\infty}\left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k\frac{L_n^{(k)}(\lambda_c)}{k!}\} \\ &= \exp\{-iu\frac{\sqrt{n}c}{\sigma_c}+n\frac{iu}{\sigma_c\sqrt{n}}L_n'(\lambda_c)+n\sum_{k\geq 2}\left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k\frac{L_n^{(k)}(\lambda_c)}{k!}\}.\end{aligned} \tag{34}$$

We detail now a development of $L_n$ – and its derivatives – which will be useful in the whole paper.

**Technical Lemma 4.1** *For any $\lambda\in\mathbb{R}$, we have*

$$L_n(\lambda)=h(r_0(\lambda))+\frac{1}{n}\log\mathbf{\Gamma}_n-\frac{1}{2n}\log n+\frac{1}{n}R_0(\lambda)+\frac{1}{n}\sum_{p\geq 1}\frac{R_p(\lambda)}{n^p p!}, \tag{35}$$

*where $\mathbf{\Gamma}_n$ is defined in* (30) *and*

$$R_0(\lambda) = \log c_0(\lambda), \tag{36}$$

$$R_p(\lambda) = \sum_{1\leq s\leq p}(-1)^{s-1}(s-1)!B_{p,s}(c_1,c_2,\cdots)c_0^{-s}, \tag{37}$$

*where the coefficients $c_i$ are given by Laplace development (see Appendix) and $B_{p,s}$ are the partial exponential Bell polynomials (see* (75)*).*

*Proof of Technical Lemma 4.1:*
From the Appendix we can develop

$$E(e^{n\lambda r_n}) = \frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\Gamma(\frac{n-2}{2})} \frac{e^{nh(r_0(\lambda)}}{\sqrt{n}} \sum_{p\geq 0} \frac{c_p(\lambda)}{(2p)!n^p}, \tag{38}$$

where

$$c_p(\lambda) = \sqrt{\frac{2\pi}{|h''(r_0(\lambda))|}} \sum_{k=0}^{2p} \left( \begin{array}{c} 2p \\ k \end{array} \right) g^{(2p-k)}(r_0(\lambda)) \cdot \sum_{m=0}^{k} B_{k,m} \left( \frac{h^{(3)}(r_0(\lambda))}{2.3}, \ldots, \frac{h^{(k-m+3)}(r_0(\lambda))}{(k-m+2)(k-m+3)} \right) \frac{(2m+2p-1)!!}{|h''(t_0)|^{m+p}}. \tag{39}$$

From Faà di Bruno formula (see e.g. formula [5c] of Comtet [8]):

$$\log E(e^{n\lambda r_n}) = nh(r_0(\lambda)) + \log \left( \frac{\Gamma(\frac{n-1}{2})}{\sqrt{n}\pi^{1/2}\Gamma(\frac{n-2}{2})} \right) + \log c_0(\lambda) + \sum_{p\geq 1} \frac{R_p(\lambda)}{n^p p!}, \tag{40}$$

where $R_p$ is defined in formula (37) above. Hence the formula (35) is proven. □

From expressions (37) and (39), we see that $R_p$ is a polynomial in $g^{(s)}(r_0(\lambda))$ and $h^{(s)}(r_0(\lambda))$ where the derivatives are taken with respect to $r$. The function $r_0(\lambda)$ is $\mathcal{C}^\infty$ on $\mathbb{R}$. We can therefore express the derivatives of $L_n$ as follows:

$$L_n^{(k)}(\lambda) = L^{(k)}(\lambda) + \frac{R_0^{(k)}(\lambda)}{n} + \frac{1}{n} \sum_{p\geq 1} \frac{R_p^{(k)}(\lambda)}{n^p p!}. \tag{41}$$

Back to formula (34), and from the choice of $\lambda_c$, we have

$$\frac{\partial}{\partial \lambda} h(r_0(\lambda))\Big|_{\lambda=\lambda_c} = L'(\lambda_c) = c$$

and

$$\begin{aligned}
\Phi_n(u) &= \exp\{\frac{iu\sqrt{n}}{\sigma_c}[L_n'(\lambda_c) - c] + n \sum_{k\geq 2} \left( \frac{iu}{\sigma_c\sqrt{n}} \right)^k \frac{L_n^{(k)}(\lambda_c)}{k!}\} \\
&= \exp\{\frac{iu}{\sqrt{n}\sigma_c}[R_0'(\lambda) + \sum_{p\geq 1} \frac{R_p'(\lambda)}{n^p p!}] - \frac{u^2}{2\sigma_c^2} L_n''(\lambda_c) + n \sum_{k\geq 3} \left( \frac{iu}{\sigma_c\sqrt{n}} \right)^k \frac{L_n^{(k)}(\lambda_c)}{k!}\} \\
&= \exp\{-\frac{u^2}{2} + \sum_{k=3}^{2p} \left( \frac{iu}{\sigma_c\sqrt{n}} \right)^k \frac{nL^{(k)}(\lambda_c)}{k!} + \sum_{k=1}^{2p} \left( \frac{iu}{\sigma_c\sqrt{n}} \right)^k \frac{R_0^{(k)}(\lambda_c)}{k!} + \sum_{k\geq 1} \left( \frac{iu}{\sigma_c\sqrt{n}} \right)^k \frac{1}{k!} \sum_{p\geq 1} \frac{R_p^{(k)}(\lambda_c)}{n^p p!}\}.
\end{aligned} \tag{42}$$

For $p$ large enough such that $\{u^k/(\sqrt{n})^{k+2p}\}$ is bounded on $\{|u| \leq n^\alpha\}$, we can have a uniform bound on the rest of the sum in the last term on the RHS above. Hence we can write, for a given $m \in \mathbb{N}$ large enough

$$\Phi_n(u) = \exp\{-\frac{u^2}{2} + \sum_{k=3}^{2m+3} \left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k \frac{nL^{(k)}(\lambda_c)}{k!} + \sum_{k=1}^{2m+1} \left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k \frac{R_0^{(k)}(\lambda_c)}{k!} + \sum_{k=1}^{2m+1}\sum_{p=1}^{s(m)} \left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k \frac{1}{k!}\frac{R_p^{(k)}(\lambda_c)}{n^p p!}\} + O(\frac{1+|u|^{2m+4}}{n^{m+1}})\}. \quad (43)$$

We follow the scheme of Cramer [10] Lemma 2, p.72 (see also Bercu and Rouault [6]), and we get the wanted results.

□

**Remark 4.2** *A thorough study of expressions $L_n^{(k)}$ and $R_p^{(k)}$ are given in [30].*

### 4.4 Proof of Proposition 2.7

By symmetry, the mean $E(\mathbf{X}) = 0$ if it exists. Then, $\tilde{r}_n$ from (15) becomes

$$\tilde{r}_n = \frac{\mathbf{X}'(\mathbf{Y} - E(\mathbf{Y}))}{\|\mathbf{X}\| \, \|\mathbf{Y} - E(\mathbf{Y})\|}. \quad (44)$$

Applying Theorem 1.5.7 from Muirhead [19], with $\alpha = \dfrac{\mathbf{Y} - E(\mathbf{Y})}{\|\mathbf{Y} - E(\mathbf{Y})\|} \in \mathbb{R}^n$, then

$$(n-1)^{1/2}\frac{\tilde{r}_n}{(1-\tilde{r}_n^2)^{1/2}}$$

has a $t_{n-1}$-distribution. Comparing to $r_n$, the degree of the $t$-distribution is one degree less than $\tilde{r}_n$.

Hence the density function of $\tilde{r}_n$ is

$$\frac{\Gamma(\frac{n}{2})}{\pi^{1/2}\Gamma(\frac{n-1}{2})}(1-r^2)^{(n-3)/2}, \quad (-1 < r < 1). \quad (45)$$

Applying Laplace's method we get

$$\begin{aligned} E\left(e^{n\lambda\tilde{r}_n}\right) &= \int_{-1}^{1} e^{n\lambda r}\frac{\Gamma(\frac{n}{2})}{\pi^{1/2}\Gamma(\frac{n-1}{2})}(1-r^2)^{(n-3)/2}dr \\ &= \frac{\Gamma(\frac{n}{2})}{\pi^{1/2}\Gamma(\frac{n-1}{2})}e^{nh(r_0(\lambda))}\left(\frac{\tilde{c}_0(\lambda)}{\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right)\right), \end{aligned}$$

where $h$, $r_0$ and $c_0$ are given in Proposition 2.7. Then

$$\begin{aligned} E\left(e^{n\lambda\tilde{r}_n}\right) &= \frac{\Gamma(\frac{n}{2})}{\pi^{1/2}\Gamma(\frac{n-1}{2})}\sqrt{\frac{2\pi}{n}}e^{nh(r_0(\lambda))}\frac{\tilde{g}(r_0(\lambda))}{\sqrt{|h''(r_0(\lambda))|}}\left(1+O\left(\frac{1}{n}\right)\right) \\ &= e^{nh(r_0(\lambda))}\frac{1}{\sqrt{(1-r_0^2(\lambda))(1+r_0^2(\lambda))}}\left(1+O\left(\frac{1}{n}\right)\right). \end{aligned} \quad (46)$$

And we can obtain formula (17) from the expression of $r_0$.

### 4.5 Proof of Proposition 3.2

From Muirhead, we know that the density function of a $n+1$ sample correlation coefficient $r_{n+1}$ is given by

$$\frac{(n-1)\Gamma(n)}{\Gamma(n+1/2)\sqrt{2\pi}}(1-\rho^2)^{n/2}(1-\rho r)^{-n+1/2}(1-r^2)^{(n-3)/2} \\ {}_2F_1\left(\frac{1}{2},\frac{1}{2};n+\frac{1}{2};\frac{1}{2}(1+\rho r)\right) \qquad (-1<r<1).$$

where ${}_2F_1$ is the hypergeometric function (see [21]). Hence Laplace transform is

$$E\left(e^{(n+1)\lambda r_{n+1}}\right) = \frac{(n-1)\Gamma(n)}{\Gamma(n+1/2)\sqrt{2\pi}}(1-\rho^2)^{n/2} \\ \int_{-1}^{1} e^{(n+1)\lambda r}(1-\rho r)^{-n+1/2}(1-r^2)^{(n-3)/2}{}_2F_1\left(\frac{1}{2},\frac{1}{2};n+\frac{1}{2};\frac{1}{2}(1+\rho r)\right)dr.$$

Looking for a limit as $n\to\infty$, we can use the following result due to Temme [28, 29] (see also [14]): the function ${}_2F_1$ has the following Laplace transform representation

$${}_2F_1(a,b,c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-zt)^a}dt \tag{47}$$

and

$${}_2F_1(a,b,c+\lambda;z) \sim \frac{\Gamma(c+\lambda)}{\Gamma(c+\lambda-b)}\sum_{s=0}^{\infty}\mathrm{f}_s(z)\frac{(b)_s}{\lambda^{b+s}}\,, \tag{48}$$

where the equivalent is for $\lambda\to+\infty$ and

$$\mathrm{f}(t) = \left(\frac{e^t-1}{t}\right)^{b-1} e^{(1-c)t}(1-z+ze^{-t})^{-a}\,,$$

$$\mathrm{f}(t) = \sum_{s=0}^{\infty}\mathrm{f}_s(t)t^s.$$

In our case, we get as $n\to\infty$:

$${}_2F_1\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}+n;\frac{1}{2}(1+\rho r)\right) \sim \frac{\Gamma(\frac{1}{2}+n)}{\Gamma(n)}\left(\frac{1}{\sqrt{n}}+\frac{2+\rho r}{8n^{3/2}}+o\left(\frac{1}{n^{3/2}}\right)\right)\,. \tag{49}$$

Hence we have to deal with the following integral:

$$\int_{-1}^{1} e^{(n+1)\lambda r}(1-\rho r)^{-n+1/2}(1-r^2)^{(n-3)/2}\left(1+\frac{2+\rho r}{8n}+o\left(\frac{1}{n}\right)\right)dr. \tag{50}$$

Neglecting the terms of lower order in $n$ we focus on

$$\int_{-1}^{1} e^{(n+1)\lambda r}(1-\rho r)^{-n+1/2}(1-r^2)^{(n-3)/2}dr = \int_{-1}^{1} e^{n\overline{h}(r)}\overline{g}(r)dr\,, \tag{51}$$

where

$$\overline{h}(r) = \lambda r - \log(1-\rho r) + \frac{1}{2}\log(1-r^2)\,, \tag{52}$$

$$\overline{g}(r) = e^{\lambda r}\sqrt{(1-\rho r)}(1-r^2)^{-3/2}\,.$$

The following lemma details the properties of the function $\bar{h}$:

**Lemma 4.3** *For any $\rho \in ]-1,1[$ and $r \in ]-1,1[$, the function $\bar{h}$ of formula (52) is defined for any $\lambda \in \mathbb{R}$. Moreover the equation $\bar{h}'(r) = 0$ has at least one solution in $]-1,1[$ and $\bar{h}''(r) < 0$ on $]-1,1[$ for any $|\rho| \leq \rho_0$ where $\rho_0 = \dfrac{\sqrt{3+2\sqrt{3}}}{3}$.*

*Proof:*
We compute easily

$$\overline{h}'(r) = \lambda + \frac{\rho}{1-\rho r} - \frac{r}{1-r^2}$$

and see that $H(r) = \bar{h}'(r)(1-r^2) = 0$ has at least one root in $]-1,1[$ (since $H(-1)H(1) < 0$). Hence there exists at least one solution $r_0 \in ]-1,1[$ such that $\bar{h}'(r) = 0$. Next, we compute

$$\bar{h}''(r) = \frac{\rho^2}{(1-\rho r)^2} - \frac{1+r^2}{(1-r^2)^2}$$

and we have

$$\bar{h}''(r) < 0 \text{ for any } r \in ]-1,1[ \iff |\rho| \leq \rho_0 := \frac{\sqrt{3+2\sqrt{3}}}{3}\,. \qquad \square$$

We know from Si [26] that the rate function in this case is

$$I_\rho(s) = \log\left(\frac{1-\rho s}{\sqrt{(1-\rho^2)}\sqrt{(1-s^2)}}\right) \text{ for } -1 < s < 1\,. \tag{53}$$

As previously said, even if this function was obtained by a contraction principle which is not applicable here (the function involved is not continuous, see Dembo and Zeitouni for more details [12]), we claim that the expression of the rate function above is nevertheless correct in the given domain $\{|\rho| \leq \rho_0\}$. We prove it below. We have

$$L(\lambda) = \bar{h}(r_0(\lambda)) + \frac{1}{2}\log(1-\rho^2)\,,$$

where $r_0$ satisfies

$$\bar{h}'(r_0(\lambda)) = 0.$$

Now we compute

$$L'(\lambda) = r_0(\lambda) + r_0'(\lambda)\bar{h}'(r_0(\lambda)) = r_0(\lambda). \tag{54}$$

For every $-1 < c < 1$ and $\lambda_c$ such that $L'(\lambda_c) = c$, we have

$$\begin{aligned} L^*(c) &= c\lambda_c - L(\lambda_c) \\ &= c\lambda_c - \{\lambda_c r_0(\lambda_c) + \frac{1}{2}\log(1 - r_0^2(\lambda_c)) - \log(1 - \rho r_0(\lambda_c)) + \frac{1}{2}\log(1-\rho^2)\} \\ &= -\frac{1}{2}\log(1-c^2) + \log(1-\rho c) - \frac{1}{2}\log(1-\rho^2) = \log\frac{1-\rho c}{\sqrt{1-c^2}\sqrt{1-\rho^2}}. \end{aligned}$$

From the dual properties of Legendre transform, the condition of Laplace's method $\overline{h}''(r) < 0$ is compatible with the condition of convexity of $I_\rho$ in $]-1,1[$. Indeed, for $\rho_0 < |\rho| < 1$, $I_\rho$ is not convex. From that point, under condition $|\rho| \leq \rho_0$, we can get

$$E\left(e^{(n+1)\lambda r_{n+1}}\right) = \frac{n-1}{\sqrt{2n\pi}}(1-\rho^2)^{n/2}\sqrt{\frac{2\pi}{n}}e^{n\overline{h}(r_0(\lambda))}\frac{\overline{g}(r_0(\lambda))}{\sqrt{|\overline{h}''(r_0(\lambda))|}}\left(1+O\left(\frac{1}{n}\right)\right) \tag{55}$$

$$= e^{(n+1)\overline{h}(r_0(\lambda))}\frac{(1-\rho^2)^{n/2}(1-\rho r_0(\lambda))^{3/2}}{(1-r_0^2(\lambda))^2\sqrt{|\overline{h}''(r_0(\lambda))|}}\left(1+O\left(\frac{1}{n}\right)\right). \tag{56}$$

We can adjust the size of sample into $n$ and obtain

$$E\left(e^{n\lambda r_n}\right) = e^{n\overline{h}(r_0(\lambda))}\frac{(1-\rho^2)^{(n-1)/2}(1-\rho r_0(\lambda))^{3/2}}{(1-r_0^2(\lambda))^2\sqrt{|\overline{h}''(r_0(\lambda))|}}\left(1+O\left(\frac{1}{n}\right)\right), \tag{57}$$

which leads us to (19). We give below two graphics, one for $\rho = \rho_0 - 0.1$ and one for $\rho = \rho_0 + 0.1$. We can clearly see the change of convexity.

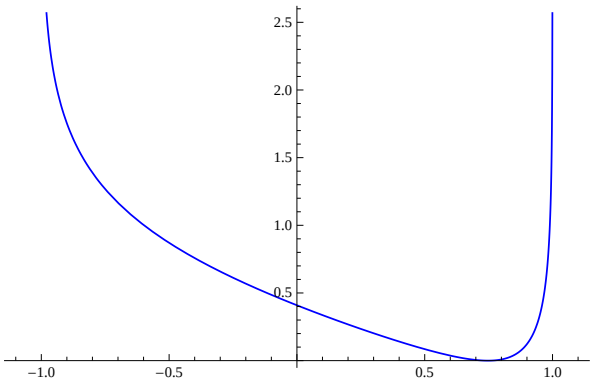


Figure 1: $I_\rho$ for $\rho = \rho_0 - 0.1$

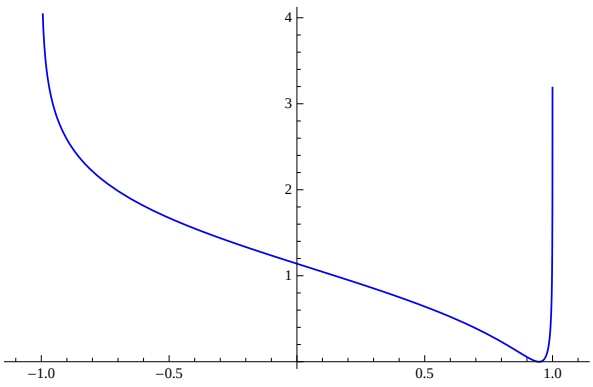


Figure 2: $I_\rho$ for $\rho = \rho_0 + 0.1$

### 4.6 Proof of Proposition 3.4

For the asymptotics of $L_n$ in this case, we follow the steps of Si [26]. Up to considering $X_1 = X - E(X)$ and $Y_1 = Y - E(Y)$, we can boil down to $E(X) = E(Y) = 0$.

If we denote by $\langle , \rangle$ the Euclidean scalar product in $\mathbb{R}^2$, and

$$\tilde{X} = \left( \frac{X_1}{\sqrt{\sum_{i=1}^n X_i^2}}, \cdots, \frac{X_n}{\sqrt{\sum_{i=1}^n X_i^2}} \right) , \quad \tilde{Y} = \left( \frac{Y_1}{\sqrt{\sum_{i=1}^n Y_i^2}}, \cdots, \frac{Y_n}{\sqrt{\sum_{i=1}^n Y_i^2}} \right) ,$$

therefore

$$\tilde{r}_n = \langle \tilde{X}, \tilde{Y} \rangle. \tag{58}$$

Large deviations for $(\tilde{r}_n)_n$ are proved in [26]. We derive here the corresponding sharp principle. Since $\tilde{X}$, $\tilde{Y}$ are independent random variables with uniform distribution $\tilde{\sigma}_n$ on the unit sphere $\mathcal{S}^{n-1}$ of $\mathbb{R}^n$, we can compute

$$E\left(e^{\lambda \tilde{r}_n}\right) = \iint\limits_{\mathcal{S}^{n-1}\times\mathcal{S}^{n-1}} e^{\lambda\langle x,y\rangle} \tilde{\sigma}_n(dx)\tilde{\sigma}_n(dy)dxdy \tag{59}$$

$$= \frac{a_{n-1}}{a_n} \int_{-1}^{1} e^{\lambda u} \left(\sqrt{1-u^2}\right)^{n-1} du\,, \tag{60}$$

where $a_n$ is the area of the unit sphere:

$$a_i = \frac{2\pi^{\frac{i+1}{2}}}{\Gamma(\frac{i+1}{2})}\,.$$

In order to get the SLD, we want to compute the normalized log–Laplace transform: for any $\lambda \in \mathbb{R}$, from Stirling formula (see [21]), we get easily

$$\frac{a_{n-1}}{a_n} = \sqrt{\frac{n}{2\pi}} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Then we can write

$$\int_{-1}^{1} e^{n\lambda u}\left(\sqrt{1-u^2}\right)^{n-1} du = \int_{-1}^{1} e^{nh(u)} g(u) du\,,$$

where $h(u) = \lambda u + \frac{1}{2}\log(1-u^2)$ and $g(u) = \frac{1}{\sqrt{1-u^2}}$. We apply Laplace's method to get:

$$\int_{-1}^{1} e^{nh(u)} du = e^{nh(u_0(\lambda))}\left(\frac{c_0(\lambda)}{\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right)\right), \tag{61}$$

where

$$u_0(\lambda) = \frac{-1+\sqrt{1+4\lambda^2}}{2\lambda}, \quad c_0(\lambda) = \sqrt{\frac{2\pi}{h''(u_0(\lambda))|}} g(u_0(\lambda))\,.$$

This leads to

$$L_n(\lambda) = h(u_0(\lambda)) - \frac{1}{2n}\log(1+4\lambda^2) + O\left(\frac{1}{n^2}\right). \tag{62}$$

# 5 Further results

## 5.1 Any order development

We present in this section a way to extend the results of Sections 2 and 3 to higher orders. Moreover, whenever functions involved are smooth enough, these techniques can be applied and the asymptotics are given in other cases.

**Theorem 5.1** *In the framework of Sections 2 and 3, for any $0 < c < 1$, there exists a sequence $(\delta_{c,k})_k$ such that*

$$P(r_n \geq c) = \frac{e^{-nL^*(c)+R_0(\lambda_c)}}{\lambda_c \sigma_c \sqrt{2\pi n}}\left[1 + \sum_{k=1}^{p} \frac{\delta_{c,k}}{n^k} + O\left(\frac{1}{n^{p+1}}\right)\right]. \tag{63}$$

*Proof:*

For seek of simplicity, we only present here the proof for $(r_n)_n$ in the spherical case. Similarly to the proof of Theorem 2.3, we briefly give the main ideas: From the decomposition $P(r_n \geq c) = A_n B_n$, in which

$$\begin{aligned} A_n &= \exp[n(L_n(\lambda_c) - c\lambda_c)] \\ &= \exp[-nL^*(c) + R_0(\lambda_c) + \sum_{p\geq 1} \frac{R_p(\lambda_c)}{n^p (2p)!}] \\ &= \exp[-nL^*(c) + R_0(\lambda_c))]\left(1 + \sum_{p\geq 1} \frac{\eta_p(\lambda_c)}{n^p (2p)!}\right), \end{aligned}$$

where $(\eta_p)_p$ is a sequence of smooth functions of $\lambda$. Recall that we can develop $L_n$ as in (35) and $L_n^{(k)}$ as in (41) and from the development of $\Phi_n$ in (42),

$$\left(1+\frac{iu}{\lambda_c\sigma_c\sqrt{n}}\right)^{-1}\Phi_n(u)=e^{-\frac{u^2}{2\sigma_c}}\left(1+\sum_{k=1}^{2p+1}\frac{P_{p,k}(u)}{n^{k/2}}+\frac{1+u^{6(p+1)}}{n^{p+1}}\,O(1)\right), \tag{64}$$

where $P_{p,k}$ are polynomials in odd powers of $u$ for $k$ odd, and polynomials in even powers of $u$ for $k$ even. From that points, we can complete the proof of Theorem 5.1.

□

## 5.2 Correlation test and Bahadur exact slope

### 5.2.1 Bahadur slope

Let us recall here some basic facts about Bahadur exact slopes of test statistics. For a reference, see [2] and [20]. Consider a sample $X_1,\cdots,X_n$ having common law $\mu_\theta$ depending on a parameter $\theta\in\Theta$. To test $(H_0):\theta\in\Theta_0$ against the alternative $(H_1):\theta\in\Theta_1=\Theta\backslash\Theta_0$, we use a test statistic $S_n$, large values of $S_n$ rejecting the null hypothesis. The $p$-value of this test is by definition $G_n(S_n)$, where

$$G_n(t)=\sup_{\theta\in\Theta_0}P_\theta(S_n\geq t).$$

The Bahadur exact slope $c(\theta)$ of $S_n$ is then given by the following relation

$$c(\theta)=-2\liminf_{n\to\infty}\frac{1}{n}\log\left(G_n(S_n)\right). \tag{65}$$

Quantitatively, for $\theta\in\Theta_1$, the larger $c(\theta)$ is, the faster $S_n$ rejects $H_0$.

A theorem of Bahadur (Theorem 7.2 in [3]) gives the following characterization of $c(\theta)$: suppose that $\lim_n n^{-1/2}S_n=b(\theta)$ for any $\theta\in\Theta_1$, and that $\lim_n n^{-1}\log\left(G_n(n^{1/2}t)\right)=-I(t)$ under any $\theta\in\Theta_0$. If $I$ is continuous on an interval containing $b(\Theta_1)$, then $c(\theta)$ is given by:

$$c(\theta)=2I(b(\theta)). \tag{66}$$

### 5.2.2 Correlation in the Gaussian case

In the Gaussian case, under Assumption 3.1, we have the following strong law of large numbers:

$$r_n\to\rho=\mathrm{cov}(X,Y). \tag{67}$$

We wish to test $H_0:\ \rho=0$ against the alternative $H_1:\ \rho\neq0$. It is obvious that under $H_1$,

$$\lim_{n\to\infty}r_n=\rho,$$

and this limit is continuous when $\rho\neq0$.

Besides, we have here

$$G_n(t)=\sup_{\rho\in\Theta_0}P_\rho(\sqrt{n}r_n\geq t)$$

and

$$\frac{1}{n}\log G_n(\sqrt{n}t) \to -\frac{1}{2}\log(1-t^2)\,.$$

Therefore the Bahadur slope is

$$c(\rho) = \log(1-\rho^2)\,. \tag{68}$$

We show that this statistic is optimal in a certain sense. In the framework above, to test $\theta \in \Theta_0$ against the alternative $\theta \in \Theta_1$ we define the likelihood ratio:

$$\lambda_n = \frac{\sup_{\theta\in\Theta_0}\prod_{i=1}^n \mu_\theta(x_i)}{\sup_{\theta\in\Theta_1}\prod_{i=1}^n \mu_\theta(x_i)}$$

and the related statistic:

$$\hat{S}_n = \frac{1}{n}\log\lambda_n. \tag{69}$$

Bahadur showed in [1] that $\hat{S}_n$ is optimal in the following sense: for any $\theta \in \Theta_1$,

$$\lim_{n\to\infty}\frac{1}{n}\log G_n(\hat{S}_n) = -J(\theta), \tag{70}$$

where $J$ is the infimum of the Kullback–Leibler information:

$$J(\theta) = \inf\{K(\theta,\theta_0), \theta_0 \in \Theta_0\} \tag{71}$$

and

$$K(\theta,\theta_0) = -\int \log[\frac{\mu_{\theta_0}(x)}{\mu_\theta(x)}]d\mu_\theta\,. \tag{72}$$

**Definition 5.2** *Let $T_n$ be a statistic in the parametric framework defined above, then if $c(\theta)$ is the Bahadur slope of $T_n$, we have*

$$c(\theta) \le 2J(\theta)$$

*and $T_n$ is said to be optimal if the upper bound is reached.*

We have the following result on the statistic $r_n$.

**Proposition 5.3** *The sequence of empirical coefficients $(r_n)_n$ is asymptotically optimal in the Bahadur sense ([1]).*

*Proof:*

We can easily compute the Kullback–Liebler information in this case:

Let $\theta = (\mu,\Sigma)$ corresponds to the distribution of $(X,Y)$ in the case $\theta \in \Theta_1$ and $\theta = (\mu_0,\Sigma_0)$ for $\theta \in \Theta_0$. Since $\rho = 0$ in the case $\theta \in \Theta_0$, the matrix $\Sigma_0$ is diagonal.

$$K(\theta,\theta_0) = -\frac{1}{2}\log|\Sigma| + \frac{1}{2}\log|\Sigma_0| - 1 + \frac{1}{2}\mathrm{tr}\Sigma_0^{-1}[\Sigma - (\mu-\mu_0)^t(\mu-\mu_0)], \tag{73}$$

where $|\Sigma|$ stands for the determinant of $\Sigma$. The infimum in (73) is reached when $\mu_0 = \mu$ and the diagonal terms in $\Sigma_0$ are the ones of $\Sigma$.

Hence,

$$J(\theta) = \inf_{\theta_0\in\Theta_0} K(\theta,\theta_0) = -\frac{1}{2}\log|\Sigma| + \frac{1}{2}\log\sigma_{11} + \frac{1}{2}\log\sigma_{22} = -\frac{1}{2}\log(1-\rho^2).$$

□

## Appendix: Laplace method

We present here some well known results about asymptotics of Laplace transforms. More precisely, we consider integrals of type

$$I(x) = \int_a^b e^{xp(t)} q(t) dt \tag{74}$$

and its asymptotics as $x \to \infty$. Details and references can be found in Olver [21] and Queffelec and Zuily [15]. The explicit computations are also done in [30]. Let us first recall some definitions (for more details, see Comtet [8, 9]).

**Definition 5.4** *Partial exponential Bell polynomials are defined for any positive integers $k \leq n$ by*

$$B_{n,k}(x_1, x_2, \cdots, x_{n-k+1}) = \sum \frac{n!}{c_1! c_2! \cdots c_{n-k+1}!} \left(\frac{x_1}{1!}\right)^{c_1} \left(\frac{x_2}{2!}\right)^{c_2} \cdots \left(\frac{x_{n-k+1}}{(n-k+1)!}\right)^{c_{n-k+1}}, \tag{75}$$

*where the sum is taken over all positive integers $c_1, c_2 \cdots, c_{n-k+1}$ such that*

$$c_1 + c_2 + \cdots + c_{n-k+1} = k,$$
$$c_1 + 2c_2 + \cdots + (n-k+1)c_{n-k+1} = n.$$

**Definition 5.5** *The complete exponential Bell polynomials are defined by*

$$B_0 = 1,$$
$$\forall n \geq 1\,, \quad B_n = \sum_{k=1}^{n} B_{n,k}.$$

*where $B_{n,k}$ are partial exponential Bell polynomials defined above.*

**Theorem 5.6** *Let $(a, b)$ be a non-empty open interval, possibly non bounded and $t_0$ be some point in $(a, b)$. Denote by $V_{t_0}$ a neighborhood of $t_0$ such that $p, q : (a, b) \to \mathbb{R}$ are functions of class $\mathcal{C}^\infty(V_{t_0})$.*

*We suppose that*

*i) $p$ is measurable on $(a, b)$,*

*ii) The maximum of $p$ is reached at $t_0$ (i.e. $p'(t_0) = 0$ and $p''(t_0) < 0$),*

*iii) There exists $x_0$ such that $\int_a^b e^{x_0 p(t)} |q(t)| dt < +\infty$.*

*Then there exist coefficients $c_0(t_0), c_1(t_0), \ldots$ depending on derivatives of $p$ and $q$ at $t_0$, such that for any $N \geq 0$, as $x \to +\infty$ we have*

$$\int_a^b e^{xp(t)} q(t) dt = e^{xp(t_0)} \left( \frac{c_0(t_0)}{\sqrt{x}} + \frac{c_1(t_0)}{2!\, x^{3/2}} + \cdots + \frac{c_N(t_0)}{(2N)!\, x^{N+1/2}} + O\left(\frac{1}{x^{N+3/2}}\right)\right). \tag{76}$$

*Moreover,* $(c_N)_N$ *can be computed as*

$$c_N(t_0) = \sqrt{\frac{2\pi}{|p''(t_0)|}} \sum_{k=0}^{2N} \left( \begin{array}{c} 2N \\ k \end{array} \right) q^{(2N-k)}(t_0) \sum_{m=0}^{k} B_{k,m}\left( \frac{p^{(3)}(t_0)}{2.3}, \ldots, \frac{p^{(k-m+3)}(t_0)}{(k-m+2)(k-m+3)} \right) \frac{(2m+2N-1)!!}{|p''(t_0)|^{m+N}}.$$

*where* $B_{k,m}$ *are the Bell polynomials defined above and* $(2n+1)!! = 1.3.5.\ldots.(2n+1)$.

## Acknowlegments

The authors wish to thank B. Bercu for valuable and fruitful discussions.

## References

[1] R.R. Bahadur. An optimal property of the likelihood ratio statistic. In *Proceedings on the Fifth Berkeley Symp.on Math. Stat. and Prob.*, volume 1, pages 13–26, 1967.

[2] R.R. Bahadur. Some limit theorems in statistics. In *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia, Pa.: SIAM, Society for Industrial and Applied Mathematics, 1971.

[3] R.R. Bahadur. Large deviations of the maximum likelihood estimate in the markov chain case. In *Recent advances in Stat.*, pages 273–286. Pap. in Honor of H. Chernoff, 1983.

[4] G. Ben Arous. Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus. *Ann. Sci. École Norm. Sup. (4)*, 21:307–331, 1988.

[5] B. Bercu, F. Gamboa, and M. Lavielle. Sharp large deviations for Gaussian quadratic forms with applications. *ESAIM PS*, 4:1–24, 2000.

[6] B. Bercu and A. Rouault. Sharp large deviations for the Ornstein–Uhlenbeck process. *Theory Probab. Appl.*, 46(1):1–19, 2002.

[7] E. Bolthausen. Laplace approximations for sums of independent random vectors. *Prob. Theory and Rel. Fields*, 76:167–206, 1987.

[8] L. Comtet. *Analyse Combinatoire, Tome 1.* Presses Universitaires de France, 1970.

[9] L. Comtet. *Analyse Combinatoire, Tome 2.* Presses Universitaires de France, 1970.

[10] H. Cramér. *Random Variables and Probability Distributions.* Cambridge University Press, 1937.

[11] A. Dembo, E. Mayer-Wolf, and O. Zeitouni. Exact behaviour for Gaussian seminorms. *Stat. and Prob. Letters*, 23:275–280, 1995.

[12] A. Dembo and O. Zeitouni. *Large deviations techniques and applications (second edition).* Springer, 1998.

[13] C. Esseen. Fourier analysis of distribution functions. a mathematical study of the laplace–gaussian law. *Acta Mathematica*, 77:1–125, 1945.

[14] C. Ferreira, José L. Lopez, and Ester Pérez Sinusía. The gauss hypergeometric function $f(a, b, c; z)$ for large $c$. *Journal of Comp. and Applied Math.*, 197:568–577, 2006.

[15] H.Quéffelec and C. Zuily. *Analyse pour l'agrégation, 4e ed.* Dunod, 2013.

[16] I.A. Ibragimov. Hitting probability of a gaussian vector with values in a hilbert space in a sphere of small radius. *J.Sov. Math*, 20:2164–2174, 1982.

[17] W.V. Li. Comparison results for the lower tail of gaussian seminorms, comparison results for the lower tail of gaussian seminorms, comparison results for the lower tail of gaussian seminorms. *J. Theor. Probab.*, 5:1–31, 1992.

[18] E. Mayer-Wolf and O. Zeitouni. The probability of small gaussian ellipsoids and associated conditional moments. *Annals of Proba*, 21:14–24, 1993.

[19] Muirhead. Aspects of multivariate statistical theory. In *Wiley Series in Probability and Mathematical Statistics*. Wiley and Sons, 1982.

[20] Y. Nikitin. *Asymptotic efficiency of non parametric tests.* Cambridge University Press, Cambridge, 1995.

[21] Frank W. J. Olver. *Asymptotics and Special Functions.* AK Peters, 1997.

[22] K. Pearson. Notes on regression and inheritance in the case of two parents. *Royal Society Proceedings*, 58(241), 1895.

[23] K. Pearson. Notes on the history of correlation. *Biometrika*, 13:25–45, 1920.

[24] J. L. Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[25] L. Sachs. *Applied Statistics. A handbook of Techniques.* Springer Verlag, Berlin–Heidelberg–New York, 1978.

[26] Shen Si. Large deviation for the empirical correlation coefficient of two Gaussian random variables. *Acta Mathematica Scientia*, 27B(4):821–828, 2007.

[27] G.N. Sytaya. On some asymptotic representations of the gaussian measure in a hilbert space. In *Theory of Stochastic Processes, Publication 2*, pages 93–104. 1974.

[28] M. Temme, Nico. *Special functions. An introduction to Classical Functions of Mathematical Physics.* Wiley and Sons, 1996.

[29] M. Temme, Nico. Large parameter cases of the gauss hypergeometric function. *Journal of Comp. and Applied Math.*, 153:441–462, 2003.

[30] T.K.T. Truong. *Sharp Large Deviations for some Test Statistics*. PhD thesis, Université d'Orléans, 2018.

[31] M.V. Zolotarev. Asymptotic behaviour of the gaussian mesure in $l_2$. *J.Sov. Math*, 35:2300–2334, 1986.