ANNALES DE L'I. H. P., SECTION B

CRISTINA ANDRIANI

PAOLO BALDI

Sharp estimates of deviations of the sample mean in many dimensions

Annales de l'I. H. P., section B, tome 33, nº 3 (1997), p. 371-385

<http://www.numdam.org/item?id=AIHPB_1997__33_3_371_0>

Ann. Inst. Henri Poincaré,

Vol. 33, n° 3, 1997, p. 371-385. *Probabilités et Statistiques*

Sharp estimates of deviations of the sample mean in many dimensions

by

Cristina ANDRIANI and Paolo BALDI *

Dipartimento di Matematica, Università di Roma-Tor Vergata, Via della Ricerca Scientifica, 00133 Roma.
e-mail: baldi@oscar.proba.jussieu.fr

ABSTRACT. – We give the asymptotics as $n \to \infty$ of the probability for the empirical mean of a sequence of i.i.d. random vectors to be in an open domain whose closure doesn't contain the true value of the mean. Our result generalizes those of Bahadur and Rao and holds under suitable assumptions on the boundary of the domain (which needs not however to be convex) and on the laws of the random vectors (a bounded density is needed).

RÉSUMÉ. – On donne le comportement asymptotique pour $n \to \infty$ de la probabilité que la moyenne empirique d'une suite de variables aléatoires multi-dimensionnelles prenne ses valeurs dans un ouvert ne contenant pas la moyenne dans sa fermeture. Ce résultat généralise celui de Bahadur et Rao et est valable sous certaines hypothèses sur la frontière de l'ouvert (lequel ne doit pas nécessairement être convexe) et sur les lois communes des v.a. (on doit supposer un peu moins que l'existence d'une densité bornée).

1. INTRODUCTION

Cramér's theorem on large deviations for the empirical mean states that if $\{X_n\}_n$ is a sequence of i.i.d. d-dimensional r.v. and we note

* Partially supported by research funds of the MURST "Processi Stocastici"

$\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$, then

$$\text{(1.1)} \qquad \begin{aligned} \overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathrm{P}\{\bar{X}_n \in D\} &\le \inf_{x\in\bar{D}} I(x) \\ \underline{\lim}_{n\to\infty} \frac{1}{n} \log \mathrm{P}\{\bar{X}_n \in D\} &\ge \inf_{x\in\overset{\circ}{D}} I(x) \end{aligned}$$

for every Borel set $D \subset \mathbb{R}^d$, where the functional I is the Fenchel-Legendre transform of the logarithm of the Laplace transform of the law μ of X_1. Of course (1.1) gives the asymptotics of $\log \mathrm{P}\{\bar{X}_n \in D\}$ if the Borel set D is such that $\inf_{x\in\overset{\circ}{D}} I(x) = \inf_{x\in\bar{D}} I(x)$.

By "sharp asymptotics" we mean the asymptotics of $\mathrm{P}\{\bar{X}_n \in D\}$ (as opposed to the asymptotics of its logarithm). The first result in this direction was obtained by Bahadur and Rao [2] for the dimension $d = 1$. If $D = [q, +\infty[$, $q > \mathrm{E}(X_1)$ and q lies in the admissible domain of I (see §2 for the explanation of this), then they prove that

$$\mathrm{P}\{\bar{X}_n \in D\} \sim \frac{\sqrt{I''(q)}}{\sqrt{2\pi n}\, I'(q)} \mathrm{e}^{-nI(q)}$$

The aim of this paper is to give the asymptotics of $\mathrm{P}\{\bar{X}_n \in D\}$ for r.v.'s taking values in $\mathbb{R}^d$. This is a much more complicated situation mainly because the boundary of a Borel set D in $\mathbb{R}^d$ can be more complicated. For this reason we have chosen to deal with the simplest situation: we assume that the Laplace transform of μ is finite in a neighborhood of the origin, and that μ has a bounded density (an assumption which can be however weakened, see Remark 3.2). Moreover we suppose that there exists a unique point $x^* \in \partial A$ at which the $\inf_{x\in\bar{D}} I(x)$ is attained and that x^* is a regular constrained point of minimum for I on ∂D (again see §2 for a precise definition) moreover belonging to the admissible domain.

It is fair to acknowledge that we follow here the path of Borovkov and Rogozin [7]. Actually the content of §3 comes from [7] and is here only in order to make the paper self-contained. Borovkov and Rogozin ([7], Theorem 2) give also an asymptotic estimate for $\mathrm{P}\{\bar{X}_n \in D\}$ reducing it to the asymptotics of certain integrals whose behaviour is however not explicitely studied. In §4 we give an explicit asymptotics (Theorem 4.4) in terms of the Cramér transform I, its derivatives and the shape of ∂D at x^*. Moreover we are able to give a clear geometric meaning of the constant appearing in the asymptotics. By the way we also correct a mistake in the Theorem 2 of Borovkov and Rogozin, as it is easy to detect if one considers the case of $N(0, I)$-distributed r.v.'s, a situation in which the computations can be made directly (see [3], [5] or [12] or even the following §4, for instance).

In §2 we recall some properties of the functional I (which is sometimes referred to as the Cramér transform of μ) and other preliminary statements. §3 contains an asymptotic result for the density of $\bar{X}_n$ from which in §4 we derive the asymptotics.

Some results of sharp asymptotics exist already in the litterature; we wish to recall those of R.Azencott [1], in a infinite dimensional setting, concerning the case of small random perturbations of dynamical systems and, more recently, of M. Iltis [10], in a context similar to the one of this paper, but where the geometric meaning of the asymptotics is not apparent.

It is also useful to point out that we don't need assumptions of the existence of dominating points, in the sense of Ney [11].

2. CRAMER TRANSFORM

Let μ be a probability measure on $\mathbb{R}^d$. We shall denote by v its (complex) Laplace transform

$$v(z) = \int_{\mathbb{R}^d} e^{\langle z,x\rangle}\, d\mu(x) = \mathrm{E}(e^{\langle z,X_1\rangle}) \qquad z \in \mathbb{C}^d$$

so that its Fourier transform is

$$\hat{\mu}(t) = v(it) \qquad t \in \mathbb{R}^d$$

It is well-known that for $d = 1$ the set of all $z \in \mathbb{C}^d$ such that the integral in the definition of v converges, is a strip, that is of the form $J \times i\mathbb{R}$, where J is an interval. In general, $d \geq 1$, v is defined on a set of the form $J \times i\mathbb{R}^d$, where J is a convex subset of $\mathbb{R}^d$. We shall refer to this set as the strip of convergence. We shall write $\mathcal{D}_v = \{\lambda \in \mathbb{R}^d, v(\lambda) < +\infty\}$ and also $\Lambda(\lambda) = \log v(\lambda), \lambda \in \mathbb{R}^d$. By I we denote the Cramér transform of μ:

$$I(x) = \sup_{\lambda\in\mathbb{R}^d} (\langle\lambda, x\rangle - \Lambda(\lambda))$$

As a supremum of linear-affine functions, I is clearly convex and lower semi-continuous. Moreover, since $\lambda \to \langle\lambda, x\rangle - \Lambda(\lambda)$ is concave, one can try to compute the supremum by looking for the critical point, e.g. by solving in λ the equation

$$x = \Lambda'(\lambda) \tag{2.1}$$

If this equation has a solution $\lambda := \lambda(x)$, then clearly

$$I(x) = \langle \lambda(x), x \rangle - \Lambda(\lambda(x)) \tag{2.2}$$

(otherwise the supremum is attained at infinity). If equation (2.1) has a solution $\lambda \in \overset{\circ}{\mathcal{D}}_\Lambda = \overset{\circ}{\mathcal{D}}_v$ we shall say that x lies in the admissible domain of I. In other words the admissible domain Ω is the image of the interior of $\mathcal{D}_\Lambda$ by Λ'. If the support of μ is not contained in a proper hyperplane of $\mathbb{R}^d$, then the covariance matrix of μ is positive definite and thus $\Lambda''(0)$ is invertible. Indeed this argument implies that $\Lambda''(\lambda)$ is always invertible for λ in the domain of v, since $\Lambda''(\lambda)$ coincides (see next paragraph) with the covariance matrix of another probability still having its support not contained in a proper hyperspace. This implies that Λ is strictly convex on $\overset{\circ}{\mathcal{D}}_\Lambda$, so that the solution of (2.1) is unique for x in the admissible comain. Moreover the implicit function theorem gives immediately that the admissible domain is an open set and that $x \to \lambda(x)$ is $\mathcal{C}^\infty$ on Ω.

This also implies that the Cramér transform I is $\mathcal{C}^\infty$ when restricted to the admissible domain. By computing the derivatives in (2.2) one has the relations

$$\begin{aligned} I'(x) &= \lambda(x) \\ I''(x) &= \Lambda''(\lambda(x))^{-1} \end{aligned} \tag{2.3}$$

For more on questions of convex analysis the reader can refer to the books of Rockafellar [13] and Ellis [8].

3. THE ASYMPTOTICS FOR THE DENSITY

Assume that x belongs to the admissible domain. Lat us denote $\mu(dy+x)$ the probability measure on $\mathbb{R}^d$ defined either by $A \to \mu(A+x)$ or

$$\int f(y)\,\mu(dy+x) = \int f(y-x)\,\mu(dy)$$

An important role is going to be played by the following measure

$$\nu_{\lambda,x}(dy) = \frac{e^{\langle \lambda, y+x \rangle}}{v(\lambda)}\mu(dy+x)$$

A straightforward computation gives that its Laplace transform is

$$z \to \frac{e^{-\langle z,x \rangle}\, v(z+\lambda)}{v(\lambda)} \tag{3.1}$$

so that $\nu_{\lambda,x}$ is a probability measure. If x belongs to the admissible domain, then let us set $\mu_x = \nu_{\lambda(x),x}$; (3.1) gives then for the Fourier transform $\hat{\mu}_x$ of μ_x

$$\hat{\mu}_x(t) = \mathrm{e}^{-\langle it,x\rangle - \Lambda(\lambda(x))}\, v(\lambda(x) + it) = \mathrm{e}^{I(x)}\, \mathrm{e}^{-\langle \lambda(x)+it,x\rangle}\, v(\lambda(x) + it)$$

Computing the derivatives one gets easily that μ_x is centered and that its covariance matrix coincides with the Hessian of Λ computed at $\lambda = \lambda(x)$.

As we already remarked if the support of μ is not contained in a proper hyperplane of $\mathbb{R}^d$, then the covariance matrix of μ is (strictly) positive definite. Thus $\Lambda''(0)$ is positive definite. But if the support of μ is not contained in a proper hyperplane, the same is true for μ_x, so that also $\Lambda''(\lambda(x))$ is necessarily positive definite. By (2.3) I'' is strictly positive definite in the admissible domain.

THEOREM 3.1. – *Let $\{X_n\}_n$ be a sequence of $\mathbb{R}^d$-valued r.v.'s; assume that their common law μ has a bounded density with respect to the Lebesgue measure and that its Laplace transform v is finite in a neighborhood of the origin. Then $\bar{X}_n$ has a density g_n for which*

$$g_n(x) \sim \left(\frac{n}{2\pi}\right)^{d/2} \left(\det I''(x)\right)^{1/2} \mathrm{e}^{-nI(x)} \tag{3.2}$$

holds as $n \to \infty$ for every x belonging to the admissible domain Ω. Moreover the above expansion holds uniformly for x in any compact subset of $\mathbb{R}^d$ which is contained in Ω.

Proof. – Let us denote by f_n the density of $X_1 + \ldots + X_n$. Then

$$g_n(x) = n^d f_n(nx) \tag{3.3}$$

Let $\lambda \in \mathbb{R}^d$ be a point in the interior of $\mathcal{D}_\Lambda$: then the Fourier transform of $x \to \mathrm{e}^{\langle \lambda,x\rangle} f_n(x)$ is $t \to v(\lambda + it)^n$. Let us prove that it is integrable for large n. Indeed it suffices to show that $x \to \mathrm{e}^{\langle \lambda,x\rangle} f_1(x)$ is in L^p for some $p > 1$, since this will imply that $t \to v(\lambda + it)$ is in L^q for some $q > 2$, by the Hausdorff-Young inequality and that $t \to v(\lambda + it)^n$ is in L^1 for large n. But

$$|\,\mathrm{e}^{\langle \lambda,x\rangle} f_1(x)|^p = |\,\mathrm{e}^{\langle p\lambda,x\rangle} f_1(x)^p| \le \mathrm{e}^{\langle p\lambda,x\rangle} f_1(x)\|f_1\|_\infty^{p-1}$$

and the left hand term is integrable as soon as p is small enough so that $p\lambda$ still lies in $\mathcal{D}_\Lambda$. Thus for large n the Fourier inversion theorem gives

$$\mathrm{e}^{\langle \lambda,x\rangle} f_n(x) = \frac{1}{(2\pi)^k} \int \mathrm{e}^{-\langle i,t\rangle x}\, v^n(\lambda + it)dt$$

which means

$$f_n(x) = \frac{1}{(2\pi)^k} \int e^{-\langle it-\lambda, x\rangle}\, v^n(\lambda + it)dt$$

and by (3.3)

$$g_n(x) = \left(\frac{n}{2\pi}\right)^d \int \left(e^{-\langle it-\lambda, x\rangle}\, v(\lambda + it)\right)^n dt$$

Since this relationship holds for every λ and x lies in the admissible domain, then it holds also for $\lambda = \lambda(x)$, so that

$$\begin{aligned} g_n(x) &= \left(\frac{n}{2\pi}\right)^d \int \left(e^{-\langle it-\lambda(x), x\rangle}\, v(\lambda(x) + it)\right)^n dt \\ &= \left(\frac{n}{2\pi}\right)^d e^{-nI(x)} \int \left(e^{I(x)}\, e^{-\langle it-\lambda(x), x\rangle}\, v(\lambda(x) + it)\right)^n dt \\ &= \left(\frac{n}{2\pi}\right)^d e^{-nI(x)} \int \hat{\mu}_x(t)^n dt \end{aligned}$$

The last integral can be split

$$\int \hat{\mu}_x(t)^n dt = \int_U \hat{\mu}_x(t)^n dt + \int_{U^C} \hat{\mu}_x(t)^n dt$$

U being a neighborhood of the origin. Since μ_x is absolutely continuous, then $|\hat{\mu}_x(t)| < 1$ and by the Riemann-Lebesgue lemma $\lim_{t\to\infty} |\hat{\mu}_x(t)| = 0$. Thus, for every neighborhood U of the origin there exists a constant $k < 1$ such that

$$\sup_{t\in U^C} |\hat{\mu}_x(t)| < k \tag{3.4}$$

so that

$$\left|\int_{U^C} \hat{\mu}_x(t)^n dt\right| \leq \int_{U^C} |\hat{\mu}_x(t)|^n dt < k^{n-q} \int_{U^C} |\hat{\mu}_x(t)|^q dt$$

where q is such that $\hat{\mu}_x \in L^q$. Thus the contribution of the integral over U^C goes to 0 exponentially fast. As for the other term, by a change of variable

$$\int_U \hat{\mu}_x(t)^n dt = n^{-d/2} \int_{\sqrt{n}U} \hat{\mu}_x\left(\frac{t}{\sqrt{n}}\right)^n dt$$

Since the probability measure $\hat{\mu}_x$ is centered and has covariance matrix $\Lambda''(\lambda(x)) = I''(x)^{-1} := \Gamma_x$, by the Central Limit Theorem

$$\hat{\mu}_x\left(\frac{t}{\sqrt{n}}\right)^n \overset{n\to\infty}{\to} \exp\left(-\frac{1}{2}\langle\Gamma_x t, t\rangle\right) \tag{3.5}$$

In order to apply Lebesgue theorem we remark that, if U is small enough then

$$|\mu_x(t)| \leq 1 - \frac{1}{2}\langle(\Gamma_x - \varepsilon I)t, t\rangle \tag{3.6}$$

for $\varepsilon > 0$, which can be chosen small enough so that $\Gamma_x^\varepsilon = \Gamma_x - \varepsilon I$ is still positive definite. Thus

$$|\hat{\mu}_x\left(\frac{t}{\sqrt{n}}\right)| \leq 1 - \frac{1}{2n}\langle\Gamma_x^\varepsilon t, t\rangle$$

and since $1 - y \leq e^{-y}$

$$|\hat{\mu}_x\left(\frac{t}{\sqrt{n}}\right)|^n \leq \exp\left(-\frac{1}{2}\langle\Gamma_x^\varepsilon t, t\rangle\right)$$

Lebesgue theorem now gives

$$\int_{\sqrt{n}U} \hat{\mu}_x\left(\frac{t}{\sqrt{n}}\right)^n dt \overset{n\to\infty}{\to} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle\Gamma_x t, t\rangle\right) dt = (2\pi)^{d/2} \det \Gamma_x^{-1/2}$$

Since the contribution of the integral over U^C goes to 0 exponentially fast, we have finally

$$\int \hat{\mu}_x(t)^n dt \sim \left(\frac{2\pi}{n}\right)^{d/2} \det \Gamma_x^{-1/2} = \left(\frac{2\pi}{n}\right)^{d/2} \det I''(x)^{1/2}$$

We finish by indicating the arguments which lead to the uniformity of the asymptotics (3.2). First it is easy to prove that the Laplace transform v is uniformly continuous in any strip of the form $K + i\mathbb{R}^d$, where K is any compact such that $K \subset\subset J$ (recall that $J + i\mathbb{R}^d$ is the domain of v). This implies that a constant $k < 1$ can be chosen so that (3.4) holds in a neighborhood of x. The analiticity properties of v moreover imply that the remainder in the development to the second order of $\hat{\mu}_x$ at the origin can be controlled with continuity, so that the convergence in (3.5) and the majorization in (3.6) are uniform in a small neighborhood of x.

REMARK 3.2. – *A closer look to the proof of Theorem 3.1 shows that the assumption of existence of a bounded density for μ can be weakened. Indeed any assumption ensuring that, for some n, μ^{*n} has a density to which the inversion Fourier theorem can be applied is sufficient; for instance that μ^{*n} has a bounded density for some n.*

4. THE ASYMPTOTICS

In this section we state our main result, giving the asymptotics for $\mathrm{P}\{\bar{X}_n \in D\}$. We begin by stating the assumptions.

HYPOTHESIS (A). – μ has a bounded density with respect to the Lebesgue measure and its Laplace transform is finite in a neighborhood of the origin.

DEFINITION 4.1. – *Let $\Psi : \mathbb{R}^d \to \mathbb{R}$ be a smooth function and let $x^* \in \partial D$ be a local constrained minimum of Ψ on ∂D; x^* is said to be non degenerate or regular if for some local system of coordinates $G: \mathbb{R}^{d-1} \supset U \to \partial D$ the Hessian of $\Psi \circ G$ is positive definite at $G^{-1}(x^*)$.*

Let M be a hypersurface of $\mathbb{R}^d$, let z be a point on M and M_z the tangent space to M at z. Let $n(x)$ be a C^∞ unit normal field to M around z. If $X = \sum_{i=1}^d a_i \frac{\partial}{\partial x_i}$ is a vector in M_z consider the transformation

$$X \to Xn \tag{4.1}$$

where Xn is the vector whose j-th component is $\sum_{i=1}^d a_i \frac{\partial n_j}{\partial x_i}(z)$. Then Xn is still a vector in M_z so that (4.1) defines a transformation of M_z into itself which is called the Weingarten map (see Hicks [1965], p.21 e.g.). This map is closely related with the curvature of M at z (ibidem p.24).

The following examples show how to compute the Weingarten map in two typical situations. In doing this one must take care because the choice of the C^∞ normal unit field is not unique, so that the Weingarten map is defined up to a constant -1. In the following examples the C^∞ normal unit field is chosen so that at x_0 it points in the direction of the positive d-axis.

EXAMPLE 4.2. – Let us assume that M is locally the graph of a function g. That is $x_0 \in M$ is such that in a neighborhood of x_0 M consists of the points $(x_1, \ldots, x_{d-1}, g(x_1, \ldots, x_{d-1}))$, $g : \mathbb{R}^{d-1} \to \mathbb{R}$ being a smooth function such that $g'(x_0) = 0$. Then in the system of coordinates $x \to (x_1, \ldots, x_{d-1})$ the Weingarten map of M at x_0 is given by $-g''(x_0)$ (see, e.g., Baldi [3], §5).

EXAMPLE 4.3. – Let us assume that $M = \{x; F(x) = 0\}$ where $F : \mathbb{R}^d \to \mathbb{R}$ is a smooth function such that $F'(x_0) \neq 0$ and even that $F'(x_0)$ points along the d-th coordinate. By the implicit functions theorem there exists a smooth function $g : \mathbb{R}^{d-1} \to \mathbb{R}$ such that $F(x_1, \ldots, x_{d-1}, g(x_1, \ldots, x_{d-1})) = 0$ and $g'(x_0) = 0$. Computing the Hessian of g in terms of the derivatives of F one gets

$$-\frac{\partial^2 g}{\partial x_i \partial x_j} = \frac{F_d^2 F_{ij} + F_d F_j F_{id} + F_d F_i F_{jd} - F_i F_j F_{dd}}{F_d^3} = \frac{F_{ij}}{F_d}$$

for $i, j = 1, \ldots, d-1$. In particular the Weingarten map of a sphere of radius R equals the identity matrix divided by R. Also, for $d = 2$, the Weingarten map reduces to multiplication by the inverse of the radius of the osculating circle (multiplication by 0 if the latter equals ∞).

HYPOTHESIS (B). – The infimum of I over $\bar{D}$ is attained at a unique point $x^* \in \partial D$, which moreover is a regular constrained minimum and belongs to the admissible domain.

THEOREM 4.4. – *Under Hypotheses (A) and (B) the following expansion holds*

$$\mathrm{P}\{\bar{X}_n \in D\} \sim \frac{\det(L_1^{-1}(L_1 - L_2))^{-1/2}}{\sqrt{2\pi n}\ \langle I''(x^*)^{-1} I'(x^*), I'(x^*)\rangle^{1/2}}\ \mathrm{e}^{-nI(x^*)}$$

where L_1 and L_2 are the Weingarten maps at x^* of the hypersurfaces $\{y; I(y) = I(x^*)\}$ and ∂D respectively.

Before the proof of Theorem 4.4 it is worth to point out that in the asymptotics the term before the exponential is always of order $n^{-1/2}$, regardless of the dimension d. Also the asymptotics of Theorem 4.4 depend on the law μ only through the value of its Cramér transform I and of its first two derivatives at x^*.

It is not difficult to realize that if x^* is a regular point, then $L_1 - L_2$ is positive definite (and viceversa!) thus ensuring that $\det(L_1^{-1}(L_1 - L_2)) > 0$.

The quantities appearing in Theorem 4.4 have an intuitive meaning which is easy to understand; $\det(L_1^{-1}(L_1 - L_2))^{-1/2}$ is a measure of the contact between ∂D and the level set $\{y; I(y) = I(x^*)\}$ at their point of tangency x^*. The more these two hypersurfaces are "close" the more the symmetric matrix $L_1 - L_2$ will be small thus giving a large asymptotics. Conversely the quantity $\langle I''(x^*)^{-1} I'(x^*), I'(x^*)\rangle^{1/2}$ measures how fast the action functional I increases as one moves from x^* to the interior of D. It should be recalled that the gradient $I'(x^*)$ is orthogonal to ∂D and points to the interior of D.

The quantity $\det(L_1^{-1}(L_1 - L_2))$ appeared already in a completely different problem of sharp asymptotics (Baldi [3]). It should also be noticed that it doesn't depend on the choice of the normal field.

Also remark that the assumption of existence of a bounded density for μ is needed only in order to apply Theorem 3.1. Thus Hypothesis (A) can be weakened according to Remark (3.2).

Proof of Theorem 4.4. – For every neighborhood V of x^* we can split

$$\mathrm{P}\{\bar{X}_n \in D\} = \mathrm{P}\{\bar{X}_n \in D \cap V\} + \mathrm{P}\{\bar{X}_n \in D \cap V^C\} \tag{4.2}$$

Since I is l.s.c. and $I(x) \to +\infty$ as $|x| \to +\infty$ (consequence of Hypothesis (A)) there exists $\eta > 0$ such that

$$\inf_{y \in D \cap V^C} I(y) > I(x^*) + \eta \tag{4.3}$$

and by Cramér's theorem

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \mathrm{P}\{\bar{X}_n \in D \cap V^C\} \leq -I(x^*) + \eta \tag{4.4}$$

As for the estimation of the other term in (4.2) one has

$$\begin{aligned} &\mathrm{P}\{\bar{X}_n \in D \cap V\} \\ &= \int_{D \cap V} g_n(x)\, dx \sim \left(\frac{n}{2\pi}\right)^{d/2} (\det I''(x^*))^{1/2} \int_{D \cap V} \mathrm{e}^{-nI(x)}\, dx \end{aligned}$$

Let us denote by G a local system of coordinates of ∂D around x^* and assume for simplicity that $G(0) = x^*$. Let us denote by A the Hessian of $I \circ G$ at 0 (it is a positive definite matrix because of Hypothesis (B)). Then classical expansion formulas (see e.g. Bleistein and Handelsman [6], (8.3.63) p. 140) state that

$$\int_{D \cap V} \mathrm{e}^{-nI(x)}\, dx \sim \frac{(2\pi)^{(d-1)/2}}{n^{(d+1)/2}} \frac{|\det(DG(0)^* DG(0))|^{1/2}}{|\det A|^{1/2} |I'(x^*)|} \mathrm{e}^{-nI(x^*)} \tag{4.5}$$

This already proves that the quantity $\mathrm{P}\{\bar{X}_n \in D \cap V^C\}$ in (4.2) is negligible because of (4.4). In a first step we shall assume that $I''(x^*)$ is the identity matrix. We only need to prove that

$$\frac{|\det(DG(0)^* DG(0))|^{1/2}}{|\det A|^{1/2} |\operatorname{grad} I(x^*)|} = \frac{\det(L_1^{-1}(L_1 - L_2))^{-1/2}}{\sqrt{2\pi n}\, \langle I''(x^*)^{-1} I'(x^*), I'(x^*)\rangle^{1/2}} \tag{4.6}$$

By a orthogonal change of coordinates we can assume that locally, around x^*, ∂D is the graph of a smooth function g defined on an open set $U \subset \mathbb{R}^{d-1}$. We can assume moreover that $0 \in U$, that $x^* = (0, \ldots, 0, |x^*|)$ and that at x^* the normal to ∂D is parallel to the positive d-axis (it suffices to choose an isometry which changes x^* to $(0, \ldots, 0, |x^*|)$ and the tangent space to ∂D at x^* to the hyperplane $x_d = |x^*|$). This means in particular that the first derivatives of g at 0 vanish. We can thus consider the local system of coordinates G defined by

$$G(x_1, \ldots, x_d) = (x_1, \ldots, x_{d-1}, g(x_1, \ldots, x_{d-1}))$$

Since all the first derivatives of g vanish at 0 one has

$$DG(0) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ & \ddots & & \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & & 0 \end{pmatrix}$$

so that

$$DG(0)^* DG(0) = I_{d-1} \tag{4.7}$$

where I_{d-1} denotes the $(d-1) \times (d-1)$ identity matrix. The vanishing of the first derivatives of g at 0 also implies, by a straightforward computation, that for $i, j = 1, \dots, d-1$

$$\frac{\partial^2 I \circ G}{\partial x_i \partial x_j}(0) = \frac{\partial^2 I}{\partial x_i \partial x_j}(G(0)) + \frac{\partial I}{\partial x_d}(G(0)) \frac{\partial^2 g}{\partial x_i \partial x_j}(0)$$

Since we assume that $I''(x^*)$ is the identity matrix

$$A = I_{d-1} + \frac{\partial I}{\partial x_d}(G(0)) g''(0)$$

But in this situation we know that $L_2 = -g''(0)$ (Example 4.2) whereas $L_1 = -I_{d-1}(\frac{\partial I}{\partial x_d}(G(0))^{-1}$ (Example 4.3). Thus

$$L_1^{-1}(L_1 - L_2) = A \tag{4.8}$$

Since of course

$$\langle I''(x^*)^{-1} I'(x^*), I'(x^*) \rangle^{1/2} = |I'(x^*)| \tag{4.9}$$

(4.7), (4.8) and (4.9) give (4.5) under the assumption that $I''(x^*)$ is the identity matrix. Let us now remove this assumption. We know that the Hessian $I''(x^*)$ is positive definite. Let C be the (symmetric) square root of $I''(x^*)$, and let us define the r.v.'s $Z_i = CX_i$. One has thus

$$\mathrm{P}\{\bar{X}_n \in D\} = \mathrm{P}\{\bar{Z}_n \in \tilde{D}\}$$

where $\tilde{D} = CD$. It is immediate to check that, denoting with $\tilde{\Lambda}$ and $\tilde{I}$ the log-Laplace and the Cramér transforms respectively of the r.v.'s Z_i, one has the relations

$$\tilde{\Lambda}(\lambda) = \Lambda(C\lambda)$$
$$\tilde{I}(z) = I(C^{-1}z)$$

Moreover the infimum of $\tilde{I}$ over $\tilde{D}$ is attained at $z^* = Cx^*$ and

$$\begin{aligned}\tilde{I}(z^*) &= I(C^{-1}z^*) = I(x^*)\\ |\tilde{I}'(z^*)| &= |C^{-1}I'(x^*)| = \langle I''(x^*)^{-1}I'(x^*), I'(x^*)\rangle^{1/2}\\ \tilde{I}''(z^*) &= C^{-1}I''(C^{-1}z^*)C^{-1} = C^{-1}I''(x^*)C^{-1} = I\end{aligned}$$

Since $\tilde{I}''(z^*)$ is the identity matrix, we can apply Step 2 and obtain the asymptotics

$$\begin{aligned}\mathrm{P}\{\bar{X}_n \in D\} = \mathrm{P}\{\bar{Z}_n \in \tilde{D}\} &\sim \frac{\det(\tilde{L}_1^{-1}(\tilde{L}_1 - \tilde{L}_2))^{-1/2}}{\sqrt{2\pi n}\,|\tilde{I}'(z^*)|}\,\mathrm{e}^{-n\tilde{I}(z^*)}\\ &= \frac{\det(\tilde{L}_1^{-1}(\tilde{L}_1 - \tilde{L}_2))^{-1/2}}{\sqrt{2\pi n}\,\langle I''(x^*)^{-1}I'(x^*), I'(x^*)\rangle^{1/2}}\,\mathrm{e}^{-nI(x^*)}\end{aligned}$$

where $\tilde{L}_1$ and $\tilde{L}_2$ are the Weingarten maps of the hypersurfaces $\partial\tilde{D}$ and $\{y, \tilde{I}(y) = \tilde{I}(x^*)\}$ respectively. Thus we only have to prove that the quantity $\det(L_1^{-1}(L_1 - L_2))$ is invariant by linear transformations. This is proved in the next statement which concludes the proof of Theorem 4.4.

PROPOSITION 4.5. – *Let M_1, M_2 be smooth hypersurfaces of $\mathbb{R}^d$ which are tangent at x_0. Let C be an inversible linear transformation of $\mathbb{R}^d$ and let $\tilde{M}_1, \tilde{M}_2$ be the images through C of M_1 and M_2 respectively. If $L_1, L_2, \tilde{L}_1, \tilde{L}_2$ are the Weingarten maps of M_1 and M_2 at x_0 and of $\tilde{M}_1$ and $\tilde{M}_2$ at $y_0 = Cx_0$ respectively, then the quantities $\det(L_1^{-1}(L_1 - L_2))$ and $\det(\tilde{L}_1^{-1}(\tilde{L}_1 - \tilde{L}_2))$ are equal.*

Proof. – We shall give as granted that the statement is true if C is orthogonal. Otherwise let $F : \mathbb{R}^d \to \mathbb{R}$ be a smooth function such that $M = \{x; F(x) = 0\}$ and $F'(x_0) \neq 0$. By an orthogonal transformation we can always assume that $F'(x_0)$ points along the d-th coordinate. Then we know from Example 4.3 that L_1 is given by H_F, the principal minor of order $d-1$ of the Hessian of F, divided by the modulus of $F'(x_0)$. If $n(x)$ is a vector which is normal to M_1 at x, then $C^{*-1}n$ is normal to $\tilde{M}$ at $y = Cx$. Let O be an orthogonal matrix such that $OC^{*-1}n(x_0)$ points along the d-th coordinate: since $OC^{*-1} = (OC)^{*-1}$ and for the orthogonal transformations the statement is true we can assume that $C^{*-1}n$ still points along the d-th coordinate. Since $\tilde{M} = \{y; F \circ C^{-1}(y) = 0\}$ and the gradient of $F \circ C^{-1}$ points along the d-th coordinate, we need only to compute the principal minor of order $d-1$ of the Hessian of $F \circ C^{-1}$ at y, that we shall denote by $H_{\tilde{F}}$. Now it is known that

$$\operatorname{Hess} F \circ C^{-1}(y) = C^{*-1}\operatorname{Hess} F(x)C^{-1}$$

However, by the assumptions we made, C^{*-1} has the form

$$\begin{pmatrix} & & 0 \\ & R^* & \vdots \\ & & 0 \\ & * & * \end{pmatrix}$$

so that

$$H_{\tilde{F}} = R^* H_F R$$
$$\tilde{L}_1 = \frac{|F'(x_0)|}{|(F \circ C^{-1})'(y_0)|} R^* L_1 R$$

Since the same computation holds for L_2, we have finally

$$\tilde{L}_1^{-1}(\tilde{L}_1 - \tilde{L}_2) = R^{-1}\Big(L_1^{-1}(L_1 - L_2)\Big)R$$

so that the determinants of $\tilde{L}_1^{-1}(\tilde{L}_1 - \tilde{L}_2)$ and $L_1^{-1}(L_1 - L_2)$ are equal.

Remark. – It is possible to prove (4.6) directly, without splitting the proof into two parts according to the assumption that $I''(x^*)$ is the identity matrix or not as we did. The way we have chosen points out a useful invariance property of the form $\det(L_1^{-1}(L_1 - L_2))$.

EXAMPLE 4.6. – Let us assume that the distribution of X_1 is the product of two double exponentials, that is it has density

$$f(x) = \frac{1}{4}\, e^{-|x_1|-|x_2|}$$

Then straightforward, if not amusing, computations give the Laplace transform

$$v(\lambda) = \frac{1}{1-\lambda_1^2}\frac{1}{1-\lambda_2^2}$$
$$\Lambda(\lambda) = -\log(1-\lambda_1^2) - \log(1-\lambda_2^2)$$

The system

$$x = \Lambda'(\lambda)$$

has solution

$$\lambda_1(x) = \frac{1}{x_1}\left(\sqrt{1+x_1^2} - 1\right) \qquad \lambda_2(x) = \frac{1}{x_2}\left(\sqrt{1+x_2^2} - 1\right)$$

(the admisible domain coincides with $\mathbb{R}^d$) so that

$$I(x) = \sqrt{1+x_1^2} - 1 + \sqrt{1+x_2^2} - 1$$
$$+ \log\left(\frac{2}{x_1^2}(\sqrt{1+x_1^2}-1)\right) + \log\left(\frac{2}{x_2^2}(\sqrt{1+x_2^2}-1)\right)$$

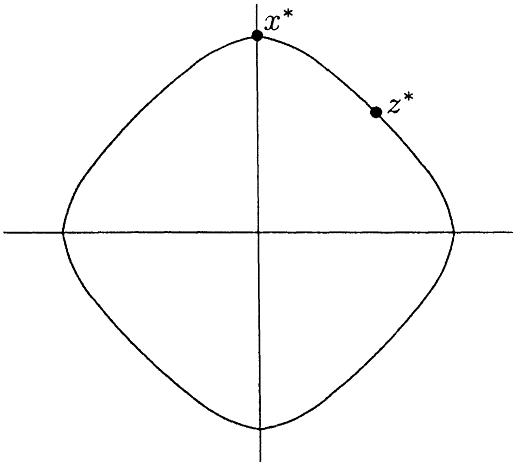

Figure 1. – The aspect of a typical level set of the action functional for the double exponential. At $x^* = (0,1)$ one has $L_1 = 1.20$, whereas at z^*, the point on the same level set which lies on the diagonal, $L_1 = 0.83$ (the osculating circle having obviously a larger radius).

If D is the ball of radius 1 centered at $(0,2)$, then using Lagrange multipliers one finds easily that the minimum of I on ∂D is attained at $x^* = (0,1)$, as the symmetries of the situation suggest. We already know that the Weingarten map of ∂D at x^* is 1, whereas for the level set $\{x; I(x) = I(x^*)\}$ at x^* is

$$-\frac{\frac{\partial^2 I(x^*)}{\partial x_1^2}}{\frac{\partial I(x^*)}{\partial x_2}} = -\frac{\frac{1}{2}}{\sqrt{2}-1} = -\frac{1}{2(\sqrt{2}-1)}$$

So that $\det(L_1^{-1}(L_1 - L_2))^{-1/2} = (2\sqrt{2}-1)^{-1/2}$. Another straightforward computation gives

$$I''(x^*) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1-\frac{\sqrt{2}}{2} \end{pmatrix}$$