Bernoulli **7**(3), 2001, 439–451

On Gaussian and Bernoulli covariance representations

SERGEY G. BOBKOV[1], FRIEDRICH GÖTZE[2] and CHRISTIAN HOUDRÉ[3]

[1]*School of Mathematics, University of Minnesota, Minneapolis MN 55455, USA.*
E-mail: bobkov@math.umn.edu
[2]*Department of Mathematics, Bielefeld University, 33501 Bielefeld, Germany.*
E-mail: goetze@mathematik.uni-bielefeld.de
[3]*School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.*
E-mail: houdre@math.gatech.edu

We discuss several applications, to large deviations for smooth functions of Gaussian random vectors, of a covariance representation in Gauss space. The existence of this type of representation characterizes Gaussian measures. New representations for Bernoulli measures are also derived, recovering some known inequalities.

Keywords: Bernoulli sums; covariance identities; Gaussian measures; large deviations

1. Introduction

Let γ_n be the canonical Gaussian measure on $\mathbb{R}^n$ with density $\mathrm{d}\gamma_n(x)/\mathrm{d}x = (2\pi)^{-n/2}\exp(-|x|^2/2)$, and let X and Y be two independent random vectors in $\mathbb{R}^n$ with distribution γ_n. Then, for any smooth functions f and g on $\mathbb{R}^n$,

$$\mathrm{cov}(f(X),\, g(X)) = \int_0^1 \mathrm{E}\langle \nabla f(X),\, \nabla g(\alpha X + \sqrt{1-\alpha^2}\,Y)\rangle\,\mathrm{d}\alpha, \tag{1.1}$$

where $\mathrm{cov}(f,\, g) = \mathrm{E}fg - \mathrm{E}f\,\mathrm{E}g$, where ∇ stands for the usual gradient, and where $\langle\cdot,\,\cdot\rangle$ is the Euclidean inner product in $\mathbb{R}^n$. In an infinite-dimensional setting (for functionals of the Wiener process) the above identity is present in Houdré and Pérez-Abreu (1995), and it was then explicitly connected to the Ornstein–Uhlenbeck semigroup by Ledoux (1995); a related approach is the work of Herbst and Pitt (1991). Indeed, if P_t, $t \geqslant 0$, is the Ornstein–Uhlenbeck semigroup associated with γ_n and acting on the space $L^1(\gamma_n)$ via

$$(P_t g)(x) = \int_{\mathbb{R}^n} g(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\,y)\,\mathrm{d}\gamma_n(y), \qquad x \in \mathbb{R}^n,$$

then (1.1) becomes

$$\mathrm{cov}(f,\, g) = \int_0^{+\infty} \mathrm{E}\langle \nabla f,\, \nabla P_t g\rangle\,\mathrm{d}t. \tag{1.2}$$

The covariance representation (1.1) which somehow hides the presence of the Ornstein–

1350–7265

Uhlenbeck operator was further studied in Houdré *et al*. (1998). There, it is treated as a particular case of a more general representation for the expectation of functions of infinitely divisible random vectors – see Houdré *et al*. (1998) and the references therein. It turns out that (1.1) is a powerful tool in the study of a number of problems on Gaussian measures, such as correlation inequalities and comparison results (Houdré 1998). (It is also useful in the corresponding studies of infinitely divisible measures; Houdré 2000).

In relation to (1.2), properties of the Ornstein–Uhlenbeck and other Markovian semigroups have been deeply investigated in the isoperimetric context by Bakry (1994), Bakry and Ledoux (1996) and Ledoux (1992; 1994; 1995; 1996; 1998; 1999). In particular (Ledoux 1998), the semigroup technique is a tool for recovering the Gaussian isoperimetric inequality due to Sudakov and Tsirel'son (1978) and Borell (1975). This isoperimetric inequality implies the following inequality for the deviations of a Lipschitz function f on $\mathbb{R}^n$ from its medians $m(f)$ with respect to γ_n:

$$\gamma_n\{|f - m(f)| \geqslant h\} \leqslant 2(1 - \Phi(h)), \qquad h > 0. \tag{1.3}$$

Here $\Phi(h) = \gamma_1((-\infty, h])$ is the distribution function of γ_1.

In this paper, we present (among other results) another related application based on (1.1). We prove that, for every Lipschitz function f on $\mathbb{R}^n$, that is, such that $\|f\|_{\mathrm{Lip}} \leqslant 1$,

$$\gamma_n\{|f - \mathrm{E}f| \geqslant h\} \leqslant \mathrm{E}|f - \mathrm{E}f| \frac{\mathrm{e}^{-h^2/2}}{h}, \qquad h > 0. \tag{1.4}$$

The proof of this estimate is given in Section 2, where we also discuss how it improves some known deviation inequalities. However, we would like to stress here that it is unlikely that it is possible to derive (1.4) from (1.3) or from the Gaussian isoperimetric inequality in its full strength. The latter allows one to reduce (1.4) to dimension 1, but even in this case we are faced with analytic difficulties.

Another easy application of (1.1) worth mentioning is the following exponential moment inequality: for any smooth function f on $\mathbb{R}^n$ such that $\mathrm{E}f = 0$,

$$\mathrm{E}\mathrm{e}^f \leqslant \mathrm{E}\mathrm{e}^{|\nabla f|^2}. \tag{1.5}$$

With a worse constant $(\pi^2/8)$ in the exponent, inequality (1.5) is due to Pisier (1986), while as stated above it was proved in Bobkov and Götze (1999). Actually, the result obtained in Bobkov and Götze (1999) shows that under a logarithmic Sobolev inequality, a result similar to (1.5) always holds.

If μ_α denotes the Gaussian measure on $\mathbb{R}^n \times \mathbb{R}^n$ which is the distribution of the random vector $(X, \alpha X + \sqrt{1 - \alpha^2} Y)$, and if π_n denotes the probability measure $\int_0^1 \mu_\alpha \, \mathrm{d}\alpha$, then (1.1) becomes

$$\mathrm{cov}(f(X), g(X)) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \langle \nabla f(x), \nabla g(y) \rangle \, \mathrm{d}\pi_n(x, y). \tag{1.6}$$

It is the form (1.6) that we apply to obtain (1.4), only using the fact that the marginals of π_n are γ_n. One may also wonder whether or not it is possible to find a representation similar to (1.6) for random vectors X in $\mathbb{R}^n$ with other, non-Gaussian distributions. As it turns out, for

$n \geqslant 2$ this is not the case: the existence of a measure π_n in (1.6) implies that X is Gaussian. We prove this characterization at the end of Section 2.

There exists, however, one remarkable exception: if we let ∇ denote a discrete gradient for functions on the discrete cube $\{-1, 1\}^n$, it is possible to find a discrete analogue of (1.6) for the canonical Bernoulli measure μ_n on $\{-1, 1\}^n$ (which assigns the mass 2^{-n} to each point). Our main result in this situation asserts that for all f and g on $\{-1, 1\}^n$,

$$\operatorname{cov}(f, g) = \int_{\{-1,1\}^n} \int_{\{-1,1\}^n} \langle \nabla f(x), \nabla g(y) \rangle \, \mathrm{d}\nu_n(x, y), \tag{1.7}$$

where the mixing probability measure ν_n on $\{-1, 1\}^n \times \{-1, 1\}^n$ has density

$$\frac{\mathrm{d}\nu_n(x, y)}{\mathrm{d}\mu_n(x)\,\mathrm{d}\mu_n(y)} = \int_0^1 \prod_{i=1}^n (1 + \alpha x_i y_i)\,\mathrm{d}\alpha.$$

We prove this representation in Section 3. It also has a semigroup version: introducing a family of operators

$$(Q_t g)(x) = \int_{\{-1,1\}^n} \prod_{i=1}^n (1 + \mathrm{e}^{-t} x_i y_i) g(y)\,\mathrm{d}\mu_n(y), \qquad x \in \{-1, 1\}^n,\ t \geqslant 0,$$

where g is an arbitrary function on $\{-1, 1\}^n$, then (1.7) becomes

$$\operatorname{cov}(f, g) = \int_0^\infty \mathrm{E}\langle \nabla f, \nabla Q_t g \rangle \,\mathrm{d}t. \tag{1.8}$$

The semigroup Q_t enjoy properties similar to the Ornstein–Uhlenbeck semigroup P_t, and we refer to Bakry (1994) for the precise relationship between P_t and Q_t.

The main point of the representation (1.7) is that it can be regarded as an independent approach, via the central limit theorem, to large deviations in Gauss space such as the inequalities (1.4) and (1.5). However, in contrast to the Gaussian case, the range of applications of (1.7) to concrete functions on the discrete cube seems to be much more restricted, due to the fact that the discrete gradients are not local. An easy direct application of (1.7)–(1.8) to functions f on $\{0, 1\}^n$, when we can work freely with the discrete gradient and the discrete perimeter, involves only the classical functions $f(x) = (x_1 + \ldots + x_n)/\sqrt{n}$. In this case, it can be shown that, for a universal constant K,

$$\mu_n\{f \geqslant h\} \leqslant K \frac{\mathrm{e}^{-h^2/2}}{h}, \qquad h > 0. \tag{1.9}$$

As in (1.4), the main point in (1.9) is the factor $1/h$ on the right-hand side. Without this factor, the inequality is trivial and holds for arbitrary linear functions $f(x) = a_1 x_1 + \ldots + a_n x_n$ with $a_1^2 + \ldots + a_n^2 = 1$ (this is the so-called sub-Gaussian inequality for Bernoulli sums), and even more generally, for all f on $\{-1, 1\}^n$ with $|\nabla f| \leqslant 1$, $\mathrm{E} f = 0$ (as a consequence of the Gross logarithmic Sobolev inequality on the discrete cube). However, we do not known if the 'true' Gaussian tails (with the factor $1/h$) can be reached for such a general family of functions. In the class of linear functions, the problem of obtaining (1.9), known in the literature as Eaton's conjecture, was affirmatively solved by

Pinelis (1994) by comparing suitable moments of f with corresponding moments of a normally distributed random variable. We conclude by giving in Section 4 a simple inductive proof of this result.

2. Deviation inequalities in Gauss space

Since on $\mathbb{R}^n \times \mathbb{R}^n$ the probability measure π_n in (1.6) has marginals γ_n, as an immediate consequence we obtain:

Theorem 2.1. *For all locally Lipschitz functions f and g on $\mathbb{R}^n$ such that $\|f\|_{\mathrm{Lip}} \leqslant 1$,*

$$\operatorname{cov}(f, g) \leqslant \mathrm{E}|\nabla g|, \tag{2.1}$$

where the covariance and the expectation are taken with respect to γ_n.

More precisely, as soon as $\mathrm{E}|\nabla g| < +\infty$, we have $\mathrm{E}|f|\,|g| < +\infty$ so that the covariance $\operatorname{cov}(f, g)$ is well defined and satisfies the inequality (2.1).

To better understand how this inequality is sharp, let us observe that for the linear functions $f(x) = \langle a, x\rangle$ with $|a| = 1$ and for the indicator functions g of half-spaces H of the form $H = \{x \in \mathbb{R}^n : \langle a, x\rangle \geqslant c\}$ there is (in an asymptotic sense) equality in (2.1).

Let us apply (2.1) to $g = \mathrm{e}^{tf}$, $t \geqslant 0$, assuming for the time being that $\mathrm{E}f = 0$:

$$\mathrm{E}f\,\mathrm{e}^{tf} = \operatorname{cov}(f, \mathrm{e}^{tf}) \leqslant t\,\mathrm{E}|\nabla f|\mathrm{e}^{tf} \leqslant t\,\mathrm{E}\mathrm{e}^{tf}.$$

For the function $u(t) = \log \mathrm{E}\mathrm{e}^{tf}$, we have $\mathrm{E}f\,\mathrm{e}^{tf} = u'(t)\mathrm{e}^{u(t)}$, hence, $u'(t) \leqslant t$. Since $u(0) = 0$, we conclude that $u(t) \leqslant t^2/2$, that is, $\mathrm{E}\mathrm{e}^{tf} \leqslant \mathrm{e}^{t^2/2}$. Thus, for all f on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \leqslant 1$,

$$\mathrm{E}\mathrm{e}^{tf} \leqslant \mathrm{e}^{t\mathrm{E}f + t^2/2}. \tag{2.2}$$

By Chebyshev's inequality, it follows that, for all $h \geqslant 0$,

$$\gamma_n\{|f - \mathrm{E}f| \geqslant h\} \leqslant 2\,\mathrm{e}^{-h^2/2}. \tag{2.3}$$

Using stochastic integrals, inequality (2.2) was proved by G. Pisier and B. Maurey (Pisier 1986, pp. 180–181). This type of idea goes back to Ibragimov *et al*. (1976). They expressed the functional f in the form $f = \mathrm{E}f + W(\tau)$, where W is the standard Wiener process and τ is a stopping time with $0 \leqslant \tau \leqslant 1$, so that $f - \mathrm{E}f \leqslant \sup_{0 \leqslant t \leqslant 1} W(t)$. Since the last supremum is distributed as $|W(1)|$, they deduced from this the following:

$$\gamma_n\{f - \mathrm{E}f \geqslant h\} \leqslant 2(1 - \Phi(h)).$$

Applying the above to $-f$, we obtain an analogue of (1.3) which is of course stronger than (2.3) for large h:

$$\gamma_n\{|f - \mathrm{E}f| \geqslant h\} \leqslant 4(1 - \Phi(h)). \tag{2.4}$$

Different proofs of (2.2) based on semigroup techniques and on Gross's logarithmic inequality were given by Ledoux (1992; 1994). A stronger infimum-convolution inequality was found by Maurey (1991); see also Bobkov and Götze (1999) for more general measures. We now make precise the argument used in the derivation of (2.2) from (2.1) and show, in particular, how the constant 4 in (2.4) can be improved.

Theorem 2.2. *For every Lipschitz function f on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \leqslant 1$, the function*

$$T_f(h) = \mathrm{e}^{h^2/2}\mathrm{E}(f - \mathrm{E}f)\mathbf{1}_{\{f - \mathrm{E}f \geqslant h\}}$$

is non-increasing in $h \geqslant 0$. In particular, for all $h > 0$,

$$\gamma_n\{f - \mathrm{E}f \geqslant h\} \leqslant \mathrm{E}(f - \mathrm{E}f)\frac{\mathrm{e}^{-h^2/2}}{h}, \tag{2.5}$$

$$\gamma_n\{|f - \mathrm{E}f| \geqslant h\} \leqslant \mathrm{E}|f - \mathrm{E}f|\frac{\mathrm{e}^{-h^2/2}}{h}. \tag{2.6}$$

Note that T_f is a constant function when f is linear. To connect (2.6) with (2.4), let us recall Pisier's inequality $\mathrm{E}|f - \mathrm{E}f| \leqslant (2/\sqrt{2\pi})\mathrm{E}|\nabla f|$ (cf. Pisier 1986) from which it follows that $\mathrm{E}|f - \mathrm{E}f| \leqslant 2/\sqrt{2\pi}$ when $\|f\|_{\mathrm{Lip}} \leqslant 1$. Hence, inequality (2.6) implies that

$$\gamma_n\{|f - \mathrm{E}f| \geqslant h\} \leqslant \frac{2}{\sqrt{2\pi}}\frac{\mathrm{e}^{-h^2/2}}{h}.$$

Recalling the asymptotic relation

$$1 - \Phi(h) = \frac{\mathrm{e}^{-h^2/2}}{h\sqrt{2\pi}}(1 + o(1)), \qquad \text{as } h \to +\infty,$$

we also obtain

$$\gamma_n\{|f - \mathrm{E}f| \geqslant h\} \leqslant 2(1 - \Phi(h))(1 + o(1)), \tag{2.7}$$

as $h \to +\infty$. Thus, for h large, (2.7) is a little better than (2.4). As in (1.3), here the constant 2 is optimal, as the example of linear functions shows. It can be made even smaller in some special situations when we have additional information about the value of $\mathrm{E}|f - \mathrm{E}f|$ (e.g. for $f(x) = \max_{i \leqslant n} x_i$).

Similar inequalities can be written for one-sided deviations, but we cannot then apply Pisier's inequality. Instead, in order to obtain a sharp constant, one can apply a related result of Pinelis (1996) showing that, for every convex function Ψ on $\mathbb{R}^n$, the value of $\mathrm{E}\Psi(f - \mathrm{E}f)$ is maximized within the class of all Lipschitz functions f on $\mathbb{R}^n$ for linear functions of Euclidean norm 1. In particular, taking $\Psi(x) = x^+$, we obtain that $\mathrm{E}(f - \mathrm{E}f)^+ \leqslant 1/\sqrt{2\pi}$. Thus, by (2.6), for all $h > 0$,

$$\gamma_n\{f - \mathrm{E}f \geqslant h\} \leqslant \frac{1}{\sqrt{2\pi}}\frac{\mathrm{e}^{-h^2/2}}{h} = (1 - \Phi(h))(1 + o(1)).$$

Actually, as shown in Bobkov (2000), for every $h > 0$, we have

$$\sup_{\|f\|_{\mathrm{Lip}} \leqslant 1} \gamma_n\{f - \mathrm{E}f \geqslant h\} = 1 - \Phi(\alpha(h)),$$

where $\alpha = \alpha(h)$ is the unique solution of the equation $\alpha\Phi(\alpha) + \varphi(\alpha) = h$.

Proof of Theorem 2.2. One may assume that $\mathrm{E}f = 0$ and that, moreover, as a random variable on $(\mathbb{R}^n, \gamma_n)$, f has a continuous positive density p on the whole real line. Otherwise one can apply the statement of the theorem to functions $f_\delta(x_1, \ldots, x_n, x_{n+1}) = (1-\delta)f(x_1, \ldots, x_n) + \delta x_{n+1}$, $0 < \delta < 1$, and then let δ tend to 0. As in obtaining (2.2), applying (2.1) to $g = U(f)$, where U is a non-decreasing (piecewise) differentiable function on $\mathbb{R}$, gives

$$\mathrm{E} f U(f) \leqslant \mathrm{E} U'(f). \tag{2.8}$$

Let F be the distribution function of f. Given $h > 0$ and $\varepsilon > 0$, applying (2.8) to the function $U(x) = \min((x-h)^+, \varepsilon)$ leads to

$$\int_h^{h+\varepsilon} x(x-h)\,\mathrm{d}F(x) + \varepsilon \int_{h+\varepsilon}^{\infty} x\,\mathrm{d}F(x) \leqslant F(h+\varepsilon) - F(h).$$

Dividing by ε and letting ε tend to 0, we obtain, for all $h > 0$,

$$\int_h^{\infty} x\,\mathrm{d}F(x) \leqslant p(h).$$

Thus, the function $V(h) = \int_h^\infty x\,\mathrm{d}F(x) = \int_h^\infty xp(x)\,\mathrm{d}x$ satisfies the differential inequality $V(h) \leqslant -V'(h)/h$, that is,

$$(\log V(h))' \leqslant (-h^2/2)'.$$

This is equivalent to saying that $\log V(h) + h^2/2$ is non-increasing, and therefore so is the function $T_f(h) = V(h)\exp(h^2/2)$. Combining (2.5) with the same inequality applied to $-f$, we get (2.6). Theorem 2.2 is proved. □

We now present the proof of the exponential inequality (1.5).

Theorem 2.3. *For any smooth function f on $\mathbb{R}^n$ such that $\mathrm{E}f = 0$,*

$$\mathrm{E}\mathrm{e}^f \leqslant \mathrm{E}\mathrm{e}^{|\nabla f|^2}.$$

Proof. Define the constant β by

$$\mathrm{E}_{\pi_n} \mathrm{e}^{\langle \nabla f(X), \nabla f(Y)\rangle} = \mathrm{e}^\beta,$$

where π_n is the measure from (1.6). Defining β is equivalent to saying that, for all $g = g(x, y) \geqslant 0$,

$$\mathrm{E}_{\pi_n}(\langle \nabla f(X), \nabla f(Y)\rangle - \beta)g \leqslant \mathrm{Ent}_{\pi_n} g = \sup_{\{h:\mathrm{E}_{\pi_n}\mathrm{e}^h \leqslant 1\}} \mathrm{E}_{\pi_n} gh.$$

Taking $g = \mathrm{e}^{f(y)}$ gives

$$\begin{aligned} \mathrm{E} f \mathrm{e}^f - \beta \mathrm{E}\mathrm{e}^f &= \mathrm{cov}(f, \mathrm{e}^f) - \beta \mathrm{E}\mathrm{e}^f \\ &= \mathrm{E}_{\pi_n} \langle \nabla f(X), \nabla f(Y) \rangle \mathrm{e}^{f(Y)} - \beta \mathrm{E}\mathrm{e}^f \\ &\leqslant \mathrm{Ent}\, \mathrm{e}^f = \mathrm{E} f \mathrm{e}^f - \mathrm{E}\mathrm{e}^f \log \mathrm{E}\mathrm{e}^f. \end{aligned}$$

Hence, $\beta \geqslant \log \mathrm{E}\mathrm{e}^f$, that is, $\mathrm{e}^\beta \geqslant \mathrm{E}\mathrm{e}^f$, and thus,

$$\mathrm{E}\mathrm{e}^f \leqslant \mathrm{E}_{\pi_n} \mathrm{e}^{\langle \nabla f(X), \nabla f(Y) \rangle} \leqslant \mathrm{E}_{\pi_n} \mathrm{e}^{|\nabla f(X)|^2 + |\nabla f(Y)|^2/2}.$$

The Cauchy–Schwarz inequality finishes the proof. □

To finish this section, we turn to a characterization of Gaussian measures via covariance representations of the form (1.6).

Theorem 2.4. *Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbb{R}^n$, $n \geqslant 2$, and assume there exists a finite measure π on $\mathbb{R}^n \times \mathbb{R}^n$ such that, for all bounded smooth functions f and g on $\mathbb{R}^n$,*

$$\mathrm{cov}(f(X), g(X)) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \langle \nabla f(x), \nabla g(y) \rangle \, \mathrm{d}\pi(x, y). \tag{2.9}$$

Then, for some $a \in \mathbb{R}^n$ and $\sigma \geqslant 0$, X has characteristic function

$$\psi(v) \equiv \mathrm{E}\mathrm{e}^{\mathrm{i}\langle v, X \rangle} = \mathrm{e}^{\mathrm{i}\langle a, v \rangle + \sigma^2 |v|^2/2}, \qquad v \in \mathbb{R}^n.$$

Proof. For the functions $f(x) = \mathrm{e}^{\mathrm{i}\langle v_1, x \rangle}$, $g(x) = \mathrm{e}^{\mathrm{i}\langle v_2, x \rangle}$, the identity (2.9) has the form

$$\psi(v_1 + v_2) - \psi(v_1)\psi(v_2) = -\langle v_1, v_2 \rangle \Psi(v_1, v_2),$$

where Ψ is Fourier transform of the measure π. In particular,

$$\langle v_1, v_2 \rangle = 0 \Rightarrow \psi(v_1 + v_2) = \psi(v_1)\psi(v_2).$$

This extends easily to n vectors $v_1, \ldots, v_n \in \mathbb{R}^n$:

$$\langle v_i, v_j \rangle = 0 \text{ for all } i \neq j \Rightarrow \psi(v_1 + \ldots + v_n) = \psi(v_1) \ldots \psi(v_n).$$

Hence, if $(e_i)_{1 \leqslant i \leqslant n}$ is an orthonormal basis of $\mathbb{R}^n$, for all $t_1, \ldots, t_n \in \mathbb{R}$,

$$\psi(t_1 e_1 + \ldots + t_n e_n) = \psi(t_1 e_1) \ldots \psi(t_n e_n).$$

But this means that the random variables $\langle X, e_1 \rangle, \ldots, \langle X, e_n \rangle$ are independent. In particular, $X_1, \ldots, X_n$ are independent. Now by the assumption that $n \geqslant 2$, we can take two linear forms of these independent random variables, say, $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$, and since Y_1 and Y_2 are independent, we conclude by the Darmois–Skitovitch theorem (cf. Kagan *et al*. 1973) that X_1 and X_2 are Gaussian. In the same way, all the X_i are Gaussian. The rest of the proof is now clear. □

3. Covariance representations on the discrete cube

For functions f on $\{-1, 1\}^n$, we define a discrete gradient $\nabla f(x)$, $x \in \{-1, 1\}^n$, as the vector $(\nabla_1 f(x), \ldots, \nabla_n f(x))$ with coordinates

$$\nabla_k f(x) = \tfrac{1}{2}\{f(x_1, \ldots, x_{k-1}, 1, x_{k+1}, \ldots, x_n) - f(x_1, \ldots, x_{k-1}, -1, x_{k+1}, \ldots, x_n)\}.$$

That is, $\nabla_k f(x)$ represents a sign-invariant normalized increment of f along the kth coordinate. In particular, it does not depend on x_k. There is another important difference operator $\Gamma f(x) = (\Gamma_1 f(x), \ldots, \Gamma_n f(x))$ defined by

$$\Gamma_k f(x) = x_k \nabla_k f(x) = \tfrac{1}{2}\{f(x) - f(s_k(x))\},$$

where

$$[s_k(x)]_i = \begin{cases} x_i, & i \neq k, \\ -x_k, & i = k, \end{cases}$$

is the neighbour of x along the kth coordinate.

Recall that the measure ν_n on $\{-1, 1\}^n \times \{-1, 1\}^n$ is defined by its density

$$\frac{\mathrm{d}\nu_n(x, y)}{\mathrm{d}\mu_n(x)\,\mathrm{d}\mu_n(y)} = \int_0^1 \prod_{i=1}^n (1 + \alpha x_i y_i)\,\mathrm{d}\alpha,$$

with respect to the uniform probability measure μ_n on $\{-1, 1\}^n$. Clearly, ν_n is a probability measure with marginals μ_n. In the rest of this paper, expectations, variances and covariances of functions on $\{-1, 1\}^n$ will always be understood with respect to μ_n. We turn now to the basic covariance representation:

Theorem 3.1. *For all functions f and g on $\{-1, 1\}^n$,*

$$\operatorname{cov}(f, g) = \int_{\{-1,1\}^n} \int_{\{-1,1\}^n} \langle \nabla f(x), \nabla g(y) \rangle\, \mathrm{d}\nu_n(x, y). \tag{3.1}$$

As in the Gaussian case, (3.1) can be rewritten in terms of the semigroup Q_t, where $Q_t(u_1, \ldots, u_n) = (Q_t u_1, \ldots, Q_t u_n)$, and becomes

$$\begin{aligned}\operatorname{cov}(f, g) &= \int_0^\infty \mathrm{E}\langle \Gamma f, Q_t(\Gamma g)\rangle\,\mathrm{d}t \\ &= \int_0^\infty \mathrm{E}\langle \Gamma f, \Gamma Q_t(g)\rangle\,\mathrm{d}t = \int_0^\infty \mathrm{E}\langle \nabla f, \nabla Q_t(g)\rangle\,\mathrm{d}t. \end{aligned}\tag{3.2}$$

For large values of n, the covariance functional can also be developed in terms of the difference operators of higher orders (see also Houdré and Pérez-Abreu 1995; Ledoux 1995; Houdré *et al.* 1998; Houdré 1998). We illustrate this general procedure with the example of the second-order difference operators $\nabla_{ij} f(x) = \nabla_i(\nabla_j f(x))$, where we assume that $n \geqslant 2$, and $1 \leqslant i, j \leqslant n$. Note that $\nabla_{ij} f = \nabla_{ji} f$, that this function does not depend on (x_i, x_j) and that, moreover, $\nabla_{ii} f = 0$. When $n = 2$,

$$\nabla_{1,2} f(x) = \tfrac{1}{4}\{f(1, 1) - f(1, -1) - f(-1, 1) + f(-1, 1)\},$$

and similar expressions can be written in the general case. Introduce the matrix $\nabla^2 f(x) = (\nabla_{ij} f(x))_{i,j=1}^n$ and the trace $\mathrm{tr}(\nabla^2 f(x)\nabla^2 g(y)) = \sum_{i,j}\nabla_{ij} f(x)\nabla_{ij} g(y)$. With this notation, we have:

Theorem 3.2. *For all functions f and g on $\{-1, 1\}^n$, $n \geqslant 2$,*

$$\mathrm{cov}(f, g) = \langle \mathrm{E}\nabla f, \mathrm{E}\nabla g\rangle + \int_{\{-1,1\}^n}\int_{\{-1,1\}^n} \mathrm{tr}(\nabla^2 f(x)\nabla^2 g(y))\,\mathrm{d}\lambda_n(x, y), \tag{3.3}$$

where

$$\frac{\mathrm{d}\lambda_n(x, y)}{\mathrm{d}\mu_n(x)\,\mathrm{d}\mu_n(y)} = \int_0^1 \prod_{i=1}^n (1 + \alpha x_i y_i)(1 - \alpha)\,\mathrm{d}\alpha.$$

Note that $\lambda_n(\{-1, 1\}^n \times \{-1, 1\}^n) = \frac{1}{2}$, and that $2\lambda_n$ has marginals μ_n.

Both sides of the identities (3.1)–(3.3) are bilinear forms in f, g, so to prove them it suffices to verify their validity for some generating system of functions, for example for Walsh's functions. Here, we are in the typical situation where it is more difficult to find an identity than to prove it. So, the proof of Theorem 3.2 is omitted, and we prove Theorem 3.1.

The Walsh functions $\varepsilon_\pi : \{-1, 1\}^n \to \{-1, 1\}$ are defined for an arbitrary set $\pi \subset \{1, \ldots, n\}$ by

$$\varepsilon_\pi(x) = \begin{cases} \prod_{i\in\pi} x_i, & \pi \neq \varnothing, \\ 1, & \pi = \varnothing. \end{cases}$$

For simplicity of notation, we set $\varepsilon_k = \varepsilon_{\{k\}}$, $k = 1, \ldots, n$, and thus $\varepsilon_\pi = \prod_{i\in\pi}\varepsilon_i$ for $\pi \neq \varnothing$. Also, we denote by $\mathbf{1}_\pi$ the indicator function of π, while $\mathbf{1}_{\pi_1=\pi_2}$ is understood as 1 if $\pi_1 = \pi_2$, and 0 otherwise. We recall below some elementary properties of the Walsh functions:

1. $\varepsilon_{\pi_1}\varepsilon_{\pi_2} = \varepsilon_{\pi_1\Delta\pi_2}$, where Δ is the symmetric difference of the corresponding sets;
2. $\mathrm{E}\varepsilon_\pi = 0$, if $\pi \neq \varnothing$; $\mathrm{E}\varepsilon_\varnothing = 1$;
3. $\mathrm{E}\varepsilon_{\pi_1}\varepsilon_{\pi_2} = \mathbf{1}_{(\pi_1=\pi_2)}$;
4. $\Gamma_k\varepsilon_\pi = \varepsilon_\pi\mathbf{1}_\pi(k)$, for all $k = 1, \ldots, n$;
5. $\nabla_k\varepsilon_\pi = \varepsilon_{\pi\setminus\{k\}}\mathbf{1}_\pi(k)$, for all $k = 1, \ldots, n$;
6. $(\nabla^2\varepsilon_\pi)_{ij} \equiv \nabla_{ij}\varepsilon_\pi = \varepsilon_{\pi\setminus\{i,j\}}\mathbf{1}_\pi(i)\mathbf{1}_\pi(j)$, for all $i \neq j$, $1 \leqslant i, j \leqslant n$ $(n \geqslant 2)$. In particular, $\nabla_{ij}\varepsilon_\pi = 0$, if $|\pi| = \mathrm{card}\,\pi \leqslant 1$.

With the above properties, the proof of Theorem 3.1 is easy.

Proof of Theorem 3.1. Let $f = \varepsilon_{\pi_1}$, $g = \varepsilon_{\pi_2}$. We may also assume that $\pi_1, \pi_2 \neq \varnothing$. Then $\mathrm{cov}(f, g) = \mathbf{1}_{\pi_1=\pi_2}$. On the other hand,

$$\nabla_k f(x) = \nabla_k\varepsilon_{\pi_1}(x) = \varepsilon_{\pi_1\setminus\{k\}}(x)\mathbf{1}_{\pi_1}(k),$$

$$\nabla_k g(y) = \nabla_k\varepsilon_{\pi_2}(y) = \varepsilon_{\pi_2\setminus\{k\}}(y)\mathbf{1}_{\pi_2}(k),$$

so

$$\langle \nabla f(x), \nabla g(y) \rangle = \sum_{k=1}^{n} (\varepsilon_{\pi_1 \setminus \{k\}}(x) \mathbf{1}_{\pi_1(k)})(\varepsilon_{\pi_2 \setminus \{k\}}(g) \mathbf{1}_{\pi_2}(k))$$

$$= \sum_{k \in \pi_1 \cap \pi_2} \varepsilon_{\pi_1 \setminus \{k\}}(x) \varepsilon_{\pi_2 \setminus \{k\}}(y).$$

In addition, $\prod_{i=1}^{n}(1 + \alpha x_i y_i) = \sum_{\pi} \alpha^{|\pi|} \varepsilon_{\pi}(x) \varepsilon_{\pi}(y)$, hence,

$$\prod_{i=1}^{n}(1 + \alpha x_i y_i) \langle \nabla f(x), \nabla g(y) \rangle = \sum_{\pi} \sum_{k \in \pi_1 \cap \pi_2} \alpha^{|\pi|} \varepsilon_{\pi}(x) \varepsilon_{\pi_1 \setminus \{k\}}(x) \varepsilon_{\pi}(y) \varepsilon_{\pi_2 \setminus \{k\}}(y).$$

Thus,

$$\int_{\{-1,1\}^n} \int_{\{-1,1\}^n} \prod_{i=1}^{n}(1 + \alpha x_i y_i) \langle \nabla f(x), \nabla g(y) \rangle \, \mathrm{d}\mu_n(x) \, \mathrm{d}\mu_n(y)$$

$$= \sum_{\pi} \sum_{k \in \pi_1 \cap \pi_2} \alpha^{|\pi|} \mathrm{E} \varepsilon_{\pi} \varepsilon_{\pi_1 \setminus \{k\}} \mathrm{E} \varepsilon_{\pi} \varepsilon_{\pi_2 \setminus \{k\}}$$

$$= \sum_{\pi} \sum_{k \in \pi_1 \cap \pi_2} \alpha^{|\pi|} \mathbf{1}_{(\pi = \pi_1 \setminus \{k\})} \mathbf{1}_{(\pi = \pi_2 \setminus \{k\})} = 0, \qquad \text{if } \pi_1 \neq \pi_2.$$

If $\pi_2 = \pi_1$, then one takes $\pi = \pi_1 \setminus \{k\}$ above and, after interchanging the sums, this is equal to $\sum_{k \in \pi_1} \alpha^{|\pi_1|-1} = |\pi_1| \alpha^{|\pi_1|-1}$. Finally, $\int_0^1 |\pi_1| \alpha^{|\pi_1|-1} \, \mathrm{d}\alpha = 1$, and the result follows. □

4. Large deviations for Bernoulli sums

Since on $\{-1, 1\}^n \times \{-1, 1\}^n$ the probability measure ν_n in (3.1) has marginals μ_n, as an immediate consequence of Theorem 3.1 and in complete similarity to the Gaussian case we obtain:

Theorem 4.1. *For all functions f, g on $\{-1, 1\}^n$ such that $|\nabla f(x)| \leq 1$, for all $x \in \{-1, 1\}^n$,*

$$\mathrm{cov}(f, g) \leq \mathrm{E}|\nabla g|, \tag{4.1}$$

where the covariance and the expectation are taken with respect to μ_n.

This is the discrete analogue of the Gaussian inequality (2.1), and at the same time it is more general: if we apply (4.1) in dimension nk to functions of the form $f((x_1 + \ldots + x_k)/\sqrt{k})$, $g((x_1 + \ldots + x_k)/\sqrt{k})$, where $x_1, \ldots, x_k \in \{-1, 1\}^n$ and where f and g are smooth functions on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \leq 1$, in the limit we obtain (2.1), by the central limit theorem in $\mathbb{R}^n$. As a result, we obtain Theorem 2.1 and its consequences for the Gauss space on the basis of Theorem 4.1.

Now let us see how to apply (4.1) to the problem of large deviations on the discrete cube itself. Consider the Bernoulli sums $S_n = \varepsilon_1 + \ldots + \varepsilon_n$. Applying (4.1) to $f = S_n/\sqrt{n}$, $g = \mathbf{1}_{\{S_n/\sqrt{n} \geqslant h\}}$, we obtain

$$h\mu_n\left\{\frac{S_n}{\sqrt{n}} \geqslant h\right\} \leqslant \mathrm{E}\frac{S_n}{\sqrt{n}}\mathbf{1}_{\{S_n/\sqrt{n} \geqslant h\}} \leqslant \mu_n^+\left\{\frac{S_n}{\sqrt{n}} \geqslant h\right\}.$$

Here, we use the notation

$$\begin{aligned}\mu_n^+(A) &= \mathrm{E}|\nabla \mathbf{1}_A| = \mathrm{E}|\Gamma \mathbf{1}_A| \\ &= \frac{1}{2}\mathrm{E}\,\mathrm{card}^{1/2}\{i \leqslant n : (x \in A,\ s_i(x) \notin A) \text{ or } (x \notin A,\ s_i(x) \in A)\},\end{aligned}$$

which defines the discrete perimeter of a set $A \subset \{-1, 1\}^n$ ($\mathbf{1}_A$ is the indicator function of A). Thus, in order to obtain the Gaussian tails, that is, the inequality of the form

$$\mu_n\left\{\frac{S_n}{\sqrt{n}} \geqslant h\right\} \leqslant K\,\frac{\mathrm{e}^{-h^2/2}}{h}, \qquad h \geqslant 1, \tag{4.2}$$

where K is a universal constant, we need just an estimate

$$\mu_n^+\left\{\frac{S_n}{\sqrt{n}} \geqslant h\right\} \leqslant K\mathrm{e}^{-h^2/2}. \tag{4.3}$$

The latter is easily attained by applying Stirling's formula. One could try to prove inequality (4.2) for arbitrary linear functions f, using Theorem 4.1 and the same argument. However, we do not know how to prove an appropriate generalization of the anti-isoperimetric inequality (4.3). Nevertheless, we would like to give here an alternative simple proof of Pinelis's result:

Theorem 4.2 (Pinelis 1994). *There exists a universal constant K such that, for all linear functions $f(x) = a_1x_1 + \ldots + a_nx_n$ with $a_1^2 + \ldots + a_n^2 = 1$, we have*

$$\mu_n\{f \geqslant h\} \leqslant K(1 - \Phi(h)), \qquad h > 0. \tag{4.4}$$

Proof. For $0 < h \leqslant \sqrt{3}$, since $\mu_n\{f \geqslant h\} \leqslant \frac{1}{2}$, inequality (4.4) holds with $K = 1/(2(1 - \Phi(\sqrt{3})))$. Thus, we may assume that $h \geqslant \sqrt{2}$. For these values, we prove (4.4) by induction over n. The case $n = 1$ is clear. Let $n > 1$ and assume (4.4) holds for $n - 1$. If $\sqrt{2} \leqslant h \leqslant \sqrt{3}$, there is nothing to prove, as explained above. Let $h \geqslant \sqrt{3}$. Assume that all the $a_i \geqslant 0$ and set $a = a_n$, $b = \sqrt{1 - a^2}$, so that $f(x) = bg(x_1, \ldots, x_{n-1}) + ax_n$, where g is a linear functional of Euclidean norm 1. Thus,

$$\mu_n\{f \geqslant h\} = \tfrac{1}{2}\left[\mu_{n-1}\left\{g \geqslant \frac{h-a}{b}\right\} + \mu_{n-1}\left\{g \geqslant \frac{h+a}{b}\right\}\right].$$

For $h \geqslant \sqrt{3}$, $(h - a)/b \geqslant \sqrt{2}$, so that we may apply the induction hypothesis; therefore, in order to perform the induction step, we need simply to verify that, for all $h \geqslant \sqrt{3}$ and for all a, $b \geqslant 0$, with $a^2 + b^2 = 1$,

$$\tfrac{1}{2}\left\{\Phi\left(\frac{h-a}{b}\right)+\Phi\left(\frac{h+a}{b}\right)\right\}\geqslant\Phi(h).$$

It suffices to see that the function $u(a)=\Phi((h-a)/b)+\Phi((h+a)/b)$ is non-decreasing in $0\leqslant a<1$. Since

$$b'=-\frac{a}{b},\qquad\left(\frac{h-a}{b}\right)'=-\frac{1-ah}{b^3},\qquad\left(\frac{h+a}{b}\right)'=-\frac{1+ah}{b^3},$$

the inequality $u'(a)\geqslant 0$ is equivalent to

$$(1+ah)\varphi\left(\frac{h+a}{b}\right)\geqslant(1-ah)\varphi\left(\frac{h-a}{b}\right),\tag{4.5}$$

which is trivial for $a\geqslant 1/h$. For $0\leqslant a<1/h$, taking the logarithm, (4.5) becomes

$$v(a)=\log(1+ah)-\log(1-ah)-\frac{2ah}{b^2}\geqslant 0.$$

After another differentiation, we see that $v'(a)\geqslant 0$ if and only if

$$\frac{1}{1-a^2h^2}\geqslant\frac{1+a^2}{b^4},$$

that is, if and only if $a^2(1+h^2)+h^2-3\geqslant 0$. Thus, v is increasing if $h\geqslant\sqrt{3}$, and since $v(0)=0$, we obtain $v(a)\geqslant 0$, for all $a\in[0,1]$. Theorem 4.2 follows. □

Acknowledgements

The research of the first author was supported in part by the Alexander von Humboldt Foundation and by Russian Foundation for Fundamental Research grant 99-01-00247. The research of the third author was supported in part by National Science Foundation grant DMS-9803239. This author enjoyed the hospitality of the Centro de Invesigación en Matemáticas, Guanajuato, Mexico, where part of this research was done.

References

Bakry, D. (1994) L'hypercontractivité et son utilisation en théorie des semi-groupes. In P. Bernard (ed.), *Lectures on Probability Theory: École d'Été de Probabilités de Saint-Flour xxii – 1992*, Lecture Notes in Math. 1581. Berlin: Springer-Verlag.

Bakry, D. and Ledoux, M. (1996) Lévy–Gromov isoperimetric inequality for an infinite-dimensional diffusion generator. *Invent. Math.*, **123**, 259–281.

Bobkov, S.G. (2000) Localization proof of the Bakry–Ledoux isoperimetric inequality, and some applications. Preprint (in Russian).

Bobkov, S.G. and Götze, F. (1999) Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, **163**, 1–28.

Borell, C. (1975) The Brunn–Minkowski inequality in Gauss space. *Invent. Math.*, **30**, 207–211.

Herbst, I. and Pitt, L.D. (1991) Diffusion equation techniques in stochastic monotonicity and positive correlations. *Probab. Theory Related Fields*, **87**, 275–312.

Houdré, C. (1998) Comparison and deviation from a representation formula. In I. Karatzas, B. Rajput and M. Taqqu (eds), *Stochastic Processes and Related Topics: A Volume in Memory of Stamatis Cambanis*, 1943–1995, pp. 207–218. Boston: Birkhäuser.

Houdré, C. (2000) Remarks on deviation inequalities for functions of infinitely divisible random vectors. Preprint.

Houdré, C. and Pérez-Abreu, V. (1995) Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *Ann. Probab.*, **23**, 400–419.

Houdré, C., Pérez-Abreu, V. and Surgailis, D. (1998) Interpolation, correlation identities and inequalities for infinitely divisible processes. *J. Fourier Anal. Appl.*, **4**, 651–668.

Ibragimov, I.A., Sudakov, V.N. and Tsirel'son, B.S. (1976) Norms of Gaussian sample functions. In G. Maruyama and J.V. Prokhorov (eds), *Proceedings of the Third Japan–USSR Symposium on Probability Theory*, Lecture Notes in Math. 550, pp. 20–41. Berlin: Springer-Verlag.

Kagan, A.M., Linnik, Yu.V. and Rao, C.R. (1973) *Characterization Problems in Mathematical Statistics*. New York: Wiley.

Ledoux, M. (1992) A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold. *Geom. Funct. Anal.*, **2**, 444–465.

Ledoux, M. (1994) Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space. *Bull. Sci. Math.*, **118**, 485–510.

Ledoux, M. (1995) L'algèbre de Lie des gradients itérés d'un générateur Markovien – développements de moyennes et entropies. *Ann. Sci. École. Norm. Sup.*, **28**, 435–460.

Ledoux, M. (1996) Isoperimetry and Gaussian analysis. In P. Bernard (ed.), *Lectures on Probability Theory and Statistics: École d'Été de Probabilités de Saint-Flour XXIV – 1994*, Lecture Notes in Math. 1648, pp. 165–294. Berlin: Springer-Verlag.

Ledoux M. (1998) A short proof of the Gaussian isoperimetric inequality. In *High Dimensional Probability* (Oberwolfach, 1996), pp. 229–232. Basel: Birkhäuser.

Ledoux M. (1999) Concentration of measure and logarithmic Sobolev inequalities. In J. Azéma, M. Émery, M. Ledoux and M. Yor (eds), *Séminaire de Probabilités XXXIII*, Lecture Notes in Math. 1709, pp. 120–216. Berlin: Springer-Verlag.

Maurey B. (1991) Some deviation inequalities. *Geom. Funct. Anal.*, **1**, 188–197.

Pinelis, I. (1994) Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.*, **22**, 357–368.

Pinelis, I. (1996) Optimal tail comparison based on comparison of moments. In *High Dimensional Probability* (Oberwolfach, 1996), pp. 297–314. Basel: Birkhäuser.

Pisier, G. (1986) Probabilistic methods in the geometry of Banach spaces. In G. Letta and M. Pratelli (eds), *Probability and Analysis*, Lecture Notes in Math. 1206, pp. 167–241. Berlin: Springer-Verlag.

Sudakov, V.N., Tsirel'son, B.S. (1978) Extremal properties of half-spaces for spherically invariant measures. *J. Soviet Math.*, **9**, 9–18. Translated from: *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.*, **41** (1974), 14–24 (Russian).

Received September 1999 and revised September 2000