Bernoulli **2**(3), 1996, 249–255

Variance of Lipschitz functions and an isoperimetric problem for a class of product measures

SERGEI G. BOBKOV[1] and CHRISTIAN HOUDRÉ[2]*

[1]*Department of Mathematics, Syktyvkar University, 167001 Syktyvkar, Russia*
[2]*Center for Applied Probability, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30322, USA*

The maximal variance of Lipschitz functions (with respect to the ℓ^1-distance) of independent random vectors is found. This is then used to solve the isoperimetric problem, uniformly in the class of product probability measures with given variance.

Keywords: isoperimetry; Lipschitz function; variance inequality

1. Statements

Let $\xi = (\xi_1, \ldots, \xi_n)$ be a vector of independent random variables with finite variance $\sigma_i^2 = \operatorname{var} \xi_i$, $1 \leq i \leq n$. Denote by $\mathscr{F}_1$ the class of all functions on $\mathbb{R}^n$ which are Lipschitz with respect to the ℓ^1-distance

$$d_1(x,y) = \|x-y\|_1 = \sum_{k=1}^{n} |x_k - y_k|, \qquad x, y \in \mathbb{R}^n.$$

By definition, $f \in \mathscr{F}_1$, if for all $x, y \in \mathbb{R}^n$, $|f(x) - f(y)| \leq d_1(x,y)$. Let $S_n = \xi_1 + \ldots + \xi_n$.

Theorem 1. *In the class $\mathscr{F}_1$, the maximal value of* $\operatorname{var} f(\xi)$ *is attained at the function $f(x) = x_1 + \ldots + x_n$. In other words, for any $f \in \mathscr{F}_1$,*

$$\operatorname{var} f(\xi) \leq \operatorname{var} S_n = \sum_{i=1}^{n} \sigma_i^2. \tag{1.1}$$

Fernique (1981, Theorem 3.2) proved an inequality similar to (1.1) for $f \in \mathscr{F}_1$ convex. However, in that case ξ is only assumed to be symmetrically distributed, i.e. for all $\epsilon_i = \pm 1$, the random vectors $(\epsilon_i \xi_1, \ldots, \epsilon_n \xi_n)$ have the same distribution (of course, this assumption holds if the ξ_i are i.i.d. with a symmetric one-dimensional distribution). In contrast to Fernique's difficult proof, Theorem 1 can easily be obtained by induction.

Theorem 1 also has the following consequence: Denote by $M^n(\sigma)$ the family of all the

*To whom correspondence should be addressed.

product measures $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ on $\mathbb{R}^n$ with given variance $\mathrm{var}(\mu) = \sigma^2$, where

$$\mathrm{var}(\mu) = \sum_{i=1}^{n} \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} t \mathrm{d}\mu_i(t) \right|^2 \mathrm{d}\mu_i(x).$$

Hence, with the above notation, $\mathrm{var}(\mu) = \mathrm{var}\, S_n$. Now, given a set $A \subset \mathbb{R}^n$ and $h > 0$, denote by

$$A^h = A + hB_1 = \{x \in \mathbb{R}^n : d_1(a, x) < h, \quad \text{for some } a \in A\}$$

the open h-neighbourhood of A (B_1 is the open ℓ^1-unit ball in $\mathbb{R}^n$). From Theorem 1 we obtain a solution to the isoperimetric problem with respect to the ℓ^1-distance uniformly in the class $M^n(\sigma)$ controlled by the parameter σ.

Theorem 2. *For any $h > 0, \sigma > 0$ and $p \in (0, 1)$,*

$$\inf_{\mu \in M^n(\sigma)} \inf_{\mu(A) \geq p} \mu(A^h) = \begin{cases} p, & \text{if} \quad h \leq \dfrac{\sigma}{\sqrt{p(1-p)}}, \\ 1 - \dfrac{p\sigma^2}{ph^2 - \sigma^2}, & \text{if} \quad h \geq \dfrac{\sigma}{\sqrt{p(1-p)}}. \end{cases} \tag{1.2}$$

The first infimum in (1.2) is taken over all the $\mu \in M^n(\sigma)$, the second is taken over all the Borel sets A of μ-measure greater or equal to p. In particular, from Theorem 2, we have:

Corollary 3. *Given $\sigma > 0$ and $p \in (0, 1)$, one can guarantee that $\mu(A^h) > p$ regardless of the dimension $n \geq 1$, regardless of the measure $\mu \in M^n(\sigma)$, and regardless of the set $A \subset \mathbb{R}^n$ of μ-measure p, if and only if*

$$h > h(p, \sigma) \equiv \frac{\sigma}{\sqrt{p(1-p)}}.$$

Otherwise, it is possible to have $\mu(A^h) = p$.

Equality in (1.2) is easy to obtain when $n = 1$. Indeed, denote by δ_x the unit mass at the point $x \in \mathbb{R}$. If $h \leq h(p, \sigma)$, take

$$\mu = p\delta_0 + (1 - p)\delta_{h(p,\sigma)}, \qquad A = \{0\}.$$

Then, $\mathrm{var}(\mu) = \sigma^2, A^h = (-h, h)$, so $\mu(A^h) = p = \mu(A)$. If $h \geq h(p, \sigma)$, take

$$\mu = p\delta_0 + q\delta_x + r\delta_h, \qquad A = \{0\},$$

with $r = p\sigma^2/(ph^2 - \sigma^2), q = 1 - p - r, x = rh/(p + q)$. Then it is again easy to verify that $\mathrm{var}(\mu) = \sigma^2$, and that $\mu(A^h) = 1 - p\sigma^2/(ph^2 - \sigma^2)$.

Since equality in (1.2) is attained when $n = 1$, (1.2) will not change if the h-neighbourhood is defined with respect to the ℓ^2-distance, or, more generally, with respect to the ℓ^α-distance in $\mathbb{R}^n, 1 \leq \alpha \leq +\infty$. Indeed, the ℓ^α-unit ball B_α is larger than B_1, hence, $A + hB_1 \subset A + hB_\alpha$, and therefore $\mu(A + hB_1) \leq \mu(A + hB_\alpha)$. Hence, the same inequality holds when one takes the second infimum in (1.2). But all the balls B_α coincide when $n = 1$ (in which case equality in (1.2) is attained).

For individual measures μ (for example, for those having finite exponential moments) there exist estimates for $1-\mu(A^h)$ which decrease exponentially when $h \to +\infty$ (see Talagrand 1994). For example, given $\sigma_i > 0, 1 \leq i \leq n$, let $\mu = \mu_1 \otimes \ldots \otimes \mu_n \in M^n(\sigma)$, where $\mu_i = (\delta_{\sigma_i} + \delta_{-\sigma_i})/2, \sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Then, as shown in Talagrand (1994, Proposition 2.1.1., Theorem 2.4.1.) (see also Ledoux 1994, p. 24, for an extension to non-identical marginals), if $h > 0, \mu(A) = p$, then

$$\mu(A + hB_1) \geq 1 - \frac{1}{p}\exp(-h^2/4\sigma^2). \tag{1.3}$$

When all the $\sigma_i = 1$, the extremal sets minimizing $\mu(A^h)$, while $\mu(A) = p$ is fixed, are known, having been obtained by Harper (1966). If one minimizes $\mu(A^h)$ over all convex sets A, the situation changes considerably, and we are then dealing with a much more powerful concentration principle discovered by Talagrand (1988; 1994). In particular, when all the $\sigma_i = 1$, one has

$$\mu(A + hB_2) \geq 1 - \frac{1}{p}\exp(-h^2/8).$$

In our case, since one is looking for a uniformly minimal value of $\mu(A + hB_1)$, it does not matter whether one considers convex sets or arbitrary sets, since the extremal $A = \{0\}$ is convex.

To complete this section, we give an inequality which is actually equivalent to the second part of (1.2). For non-empty sets $A, B \subset \mathbb{R}^n$, let $d_1(A,B) = \inf\{d_1(a,b) : a \in A, b \in B\}$.

Corollary 4. *For any $\mu \in M^n(\sigma)$, and any non-empty Borel sets $A, B \subset \mathbb{R}^n$,*

$$d_1(A,B) \leq \sigma\sqrt{\frac{1}{\mu(A)} + \frac{1}{\mu(B)}}. \tag{1.4}$$

Let $\mu(A) > 0, \mu(B) > 0$ be such that $\mu(A) + \mu(B) \leq 1$. Then, choosing $B = \{h\}$ with h equal to the right-hand side of (1.4), it is easily seen that equality in (1.4) is attained at the same measure μ and for the same set $A = \{0\}$ as the second inequality in (1.2).

2. Proofs

A statement slightly more general than Theorem 1 will actually be proved. Assume we have n measurable spaces (X_k, Σ_k) and n measurable functions $h_k = h_k(x_k, y_k)$ defined on $X_k \times X_k, 1 \leq k \leq n$, and which vanish on the diagonal $x_k = y_k$. let ξ_k be independent random variables with values in $X_k, 1 \leq k \leq n$, such that

$$2\sigma_k^2 = \mathrm{E}h_k^2(\xi_k, \eta_k) < +\infty,$$

where η_k is an independent copy of ξ_k. Put $\xi = (\xi_1, \ldots, \xi_n)$.

Lemma 5. *Let f be a measurable function defined on $X_1 \times \ldots \times X_n$ such that*

$$|f(x) - f(y)| \leq \sum_{k=1}^{n} h_k(x_k, y_k), \tag{2.1}$$

for $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in X_1 \times \ldots \times X_n$. Then

$$\mathrm{var} f(\xi) \leq \sum_{k=1}^{n} \sigma_k^2. \tag{2.2}$$

Proof. This lemma is proved by induction on the dimension n. For $n = 1$, and since 2var $f(\xi) = \int\int (f(\xi) - f(\eta))^2 \mathrm{d}\mu(\xi)\mathrm{d}\mu(\eta)$, (2.2) is immediate. Assume now that (2.2) is true for n. Denote by μ_{n+1} the distribution of ξ_{n+1}, and by P_n the distribution of the random vector $(\xi_1, \ldots, \xi_n)$, thus $P_{n+1} = P_n \otimes \mu_{n+1}$ is the distribution of $(\xi_1, \ldots, \xi_{n+1})$. Let $f : X_1 \times \ldots \times X_{n+1} \to \mathbb{R}$ satisfy (2.1). Now, fix x_{n+1}. Since the function $g(x_1, \ldots, x_n) = f(x_1, \ldots, x_n, x_{n+1})$ satisfies (2.1), making use of the induction hypotheses and writing (2.2) for g, we obtain:

$$\int g^2 \mathrm{d}P_n \leq \left(\int g \mathrm{d}P_n\right)^2 + \sum_{k=1}^{n} \sigma_k^2. \tag{2.3}$$

The function $m(x_{n+1}) = \int g\mathrm{d}P_n$ is well defined, measurable and as a function of one variable,

$$|m(x_{n+1}) - m(y_{n+1})| \leq h_{n+1}(x_{n+1}, y_{n+1}).$$

Thus m satisfies (2.1), hence

$$\int m^2 \mathrm{d}\mu_{n+1} \leq \left(\int m \mathrm{d}\mu_{n+1}\right)^2 + \sigma_{n+1}^2. \tag{2.4}$$

Integrating (2.3) over X_{n+1} (with respect to μ_{n+1}), and taking into account (2.4), gives (2.2) for f. Lemma 5 and thus Theorem 1 are proved. □

Proof of Theorem 2. Let $A \subset \mathbb{R}^n$ be such that $\mu(A) \leq p$. Since the function $f(x) = \inf_{a \in A} d_1(a, x)$ belongs to $\mathscr{F}_1$, we have, by Theorem 1, that $\mathrm{var} f \leq \sigma^2$. In addition, $f \geq 0$ and $\mu(f = 0) \geq p$. Note also that $A^h = \{x \in \mathbb{R}^n : f(x) < h\}$. To get (1.2), it just remains to appeal to the following result:

Lemma 6. *For any $h > 0, \sigma > 0$ and $p \in (0, 1)$,*

$$\sup P(\xi \geq h) = \begin{cases} 1 - p, & \text{if} \quad h \leq \dfrac{\sigma}{\sqrt{p(1-p)}}, \\[2ex] \dfrac{p\sigma^2}{ph^2 - \sigma^2}, & \text{if} \quad h \geq \dfrac{\sigma}{\sqrt{p(1-p)}}, \end{cases} \tag{2.5}$$

where the supremum is taken over all non-negative random variables ξ on a probability space $(\Omega, \mathscr{B}, P)$ such that $P(\xi = 0) \geq p$ and $\mathrm{var}\, \xi \leq \sigma^2$.

Proof. Denote by $\mathscr{L}(\xi)$ the distribution of ξ. The cases of equality in (2.5) were, in fact, already settled in Section 1. If

$$h \leq h(p, \sigma), \qquad \mathscr{L}(\xi) = p\delta_0 + (1-p)\delta_{h(p,\sigma)},$$

then $\operatorname{var}(\xi) = \sigma^2, P(\xi = 0) = p, P(\xi \geq h) = 1 - p$. If

$$h > h(p, \sigma), \quad \mathscr{L}(\xi) = p\delta_0 + q\delta_x + r\delta_h,$$

where $r = p\sigma^2/(ph^2 - \sigma^2), q = 1 - p - r, x = rh/(p+q)$, then, as easily verified, we have $\operatorname{var}(\xi) = \sigma^2, P(\xi = 0) = p$, and $P(\xi \geq h) = r = p\sigma^2(ph^2 - \sigma^2)$. So, one need only show that whenever $h \geq h(p, \sigma)$,

$$P(\xi \geq h) \leq \frac{p\sigma^2}{ph^2 - \sigma^2}. \tag{2.6}$$

To prove this, we first show, following a suggestion by M. Talagrand, that in (2.6) it suffices to consider only those ξ whose distribution is of the type $\mathscr{L}(\xi) = p_0\delta_0 + p_1\delta_x + p_2\delta_h$, for some $0 \leq x < h$. Then, keeping p_2 constant, we maximize the functional $J = p_2$ over all $p_0 \geq p$ and $x \in [0, h)$ such that $\operatorname{var}(\xi) \leq \sigma^2$.

Note first that in (2.6), ξ can be replaced by $\eta = \min(\xi, h)$ since $P(\eta \geq h) = P(\xi \geq h)$, while $\operatorname{var}(\eta) \leq \operatorname{var}(\xi) \leq \sigma^2$ (η is a Lipschitz function of $\xi : \eta = f(\xi)$, where $f(t) = \min(t, h)$). Then, let ξ take values in $[0, h]$, have distribution ν, and assume that $\nu(0, h) > 0$. Then, ν can (uniquely) be written as

$$\nu = p_0\delta_0 + p_1\lambda + p_2\delta_h,$$

where the distribution λ is concentrated in $(0, h)$. Let ξ_1 be a random variable whose distribution is λ, and let $x = \mathrm{E}(\xi_1)$. Then

$$\begin{aligned}\operatorname{var}(\xi) &= p_1\mathrm{E}(\xi_1^2) + p_2h^2 - (p_1x + p_2h)^2 \\ &= p_1\operatorname{var}(\xi_1) + p_1(1 - p_1)x^2 - 2p_1p_2hx - p_2h^2.\end{aligned}$$

Therefore, given the mean value x, $\operatorname{var}(\xi)$ is minimal if and only if $\lambda = \delta_x$, when var $(\xi_1) = 0$. Thus, ξ can be replaced in (2.6) by a random variable η which takes three values, 0, x and h.

So let us assume that $\mathscr{L}(\xi) = p_0\delta_0 + p_1\delta_x + p_2\delta_h$, $0 \leq x < h$, $p_0 \geq p$, $p_1, p_2 \geq 0$, $p_0 + p_1 + p_2 = 1$. Again, $J = p_2$ is constant. By simple algebra, and for fixed p_0, p_1, p_2, the minimal value of

$$\operatorname{var}(\xi) = (p_1x^2 + p_2h^2) - (p_1x + p_2h)^2,$$

as a function of x in $(0, h)$, is attained at

$$x = \frac{p_2}{p_0 + p_2}h.$$

For this value of x, we find

$$\begin{aligned}
\operatorname{var}(\xi) &= \left(p_1 \frac{p_2^2}{(p_0+p_2)^2} h^2 + p_2 h^2\right) - \left(p_1 \frac{p_2}{p_0+p_2} h + p_2 h\right)^2 \\
&= \left[\frac{p_2}{(p_0+p_2)^2}(p_1 p_2 + (p_0+p_2)^2) - \frac{p_2^2}{(p_0+p_2)^2}(p_1 + (p_0+p_2))^2\right] h^2 \\
&= \frac{p_2}{(p_0+p_2)^2}[p_1 p_2 + (p_0+p_2)^2 - p_2] h^2 \\
&= \frac{p_0 p_2}{p_0+p_2} h^2.
\end{aligned}$$

Now, we have to maximize $J = p_2$ under the condition

$$\operatorname{var}(\xi) = \frac{p_0 p_2}{p_0+p_2} h^2 \leq \sigma^2. \tag{2.7}$$

From (2.7), when p_0 decreases, $\operatorname{var}(\xi)$ also decreases, while $J = p_2 = 1 - p_0 - p_1$ increases (p_1 is fixed). Hence, to conclude, it is enough to consider only the case $p_0 = p$. The possible maximal value $p_2 = 1 - p$ satisfies (2.7) if and only if $p(1-p) \leq \sigma^2/h^2$, that is, if and only if $h \leq h(p,\sigma)$. Otherwise, if $h > h(p,\sigma)$, or even if $h = h(p,\sigma)$, the maximal value of $J = p_2$ is, according to (2.7), the value which satisfies

$$\frac{p p_2}{p+p_2} h^2 = \sigma^2.$$

The only solution to this equation is given by

$$p_2 = \frac{p\sigma^2}{ph^2 - \sigma^2}.$$

Lemma 6 follows. □

Proof of Corollary 4. Let $p = \mu(A), q = \mu(B)$. If $p = 0$ or $q = 0$, there is nothing to prove. If $p + q > 1$, then $A \cap B \neq \emptyset$, so $d_1(A,B) = 0$. Thus, we need consider only the case $p + q \leq 1$. Let $p, q > 0, p + q \leq 1$, and assume $A \cap B = \emptyset$. Note that

$$h \equiv \sigma\sqrt{\frac{1}{p} + \frac{1}{q}} \geq h(p,\sigma).$$

Therefore, by (1.2),

$$1 - \mu(A^h) \leq \frac{p\sigma^2}{ph^2 - \sigma^2} = q,$$

and again by (1.2), for all $h_1 > h$, $1 - \mu(A^{h_1}) < q$. Hence, $B \cap (A^{h_1} \setminus A) \neq \emptyset$, and therefore, $d_1(A,B) \leq h_1$. Letting $h_1 \to h$ completes the proof. □

Acknowledgements

Many thanks are due to M. Talagrand who provided an argument significantly simplifying the original proof of Lemma 6. The research of S.G. Bobkov was supported partly by ISF grants NZX000 and NZX300. The research of C. Houdré was supported partly by an NSF Mathematical Sciences Postdoctoral Fellowship.

References

Fernique, X. (1981) Régularité de fonctions aléatoires non gaussiennes. *Lecture Notes in Math.* 976, pp. 2–74. New York: Springer-Verlag.

Harper, H. (1966) Optimal numbering and isoperimetric problems on graphs. *J. Combin. Theory*, **1,** 385–393.

Ledoux, M. (1994) Isoperimetry and Gaussian analysis. *École d'été de Probabilités de Saint-Flour.*

Talagrand, M. (1988) An isoperimetric theorem on the cube and the Khinchine–Kahane inequalities. *Proc. Amer. Math. Soc.*, **104,** 905–909.

Talagrand, M. (1994) Concentration of measure and isoperimetric inequalities in product spaces. Preprint.

Received July 1995 and revised January 1996