[2] N. Martin and J. England, *Mathematical Theory of Entropy*, Cambridge University Press, Great Britain, Cambridge, 1985.
[3] U. Krengel, *Ergodic Theorems*, De Gruyter, Berlin, New York, 1985.
[4] Z. I. Bezhaeva and V. I. Oseledetz, *On a variance of the sum of functions of a stationary Markov process*, Theory Probab. Appl., 41 (1996), pp. 633–639 (in Russian).
[5] H. Horowitz, *L_∞-limit theorems for Markov processes*, Israel J. Math., 7 (1969), pp. 60–62.
[6] H. Horowitz, *Strong ergodic theorems for Markov processes*, Proc. Amer. Math. Soc., 23 (1969), pp. 328–334.
[7] H. Horowitz, *Pointwise convergence of iterates of a Harris-recurrent operator*, Israel J. Math., 33 (1979), pp. 177–180.
[8] S. R. Fogel, *Harris operators*, Israel J. Math., 33 (1979), pp. 281–309.
[9] S. R. Fogel and N. A. Goussoub, *Ornstein–Metivier–Brunel theorem revisited*, Ann. Inst. H. Poincaré, 15 (1979), pp. 293–301.
[10] A. Brunel and D. Revus, *Un critèr probabiliste de compacité des groupes*, Ann. Probab., 2 (1974), pp. 745–746.
[11] R. A. Olshen, *The coincidence of measure algebras under an exchangeable probability*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, 18 (1971), pp. 153–166.
[12] H. O. Georgii, *On canonical Gibbs states, symmetric and tail events*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, 33 (1976), pp. 331–341.
[13] E. Hewitt and L. J. Savage, *Symmetric measure on Cartesian products*, Trans. Amer. Math. Soc., 80 (1965), pp. 470–501.
[14] D. Blackwell and D. Freedman, *The tails σ-field of Markov chain and a theorem of Orey*, Ann. Math. Statist., 35 (1964), pp. 1291–1295.
[15] L. A. Grigorenko, *On the σ-algebra of symmetric events for a countable Markov chain*, Theory Probab. Appl., 24 (1979), pp. 199–204.
[16] L. A. Grigorenko, *Symmetric events for some stationary random sequences*, Uspekhi matem. nauk, 39 (1984), pp. 127–128 (in Russian).
[17] B. M. Gurevich, *The local limit theorem for Markov chains and regularity conditions*, Theory Probab. Appl., 13 (1968), pp. 182–188.
[18] A. N. Kolmogorov, *The local limit theorem for classical Markov chains*, Izvestiya AN SSSR, ser. matem., 13 (1949), pp. 281–300 (in Russian).
[19] M. I. Gordin and B. A. Lifshits, *The central limit theorem for the stationary Markov processes*, Dokl. AN SSSR, 239 (1978), pp. 766–767 (in Russian).

SOME EXTREMAL PROPERTIES OF THE BERNOULLI DISTRIBUTION*

S. G. BOBKOV†

(*Translated by A. V. Bulinskii*)

Abstract. The location of n-dimensional Bernoulli distribution is examined within the class of all probability distributions in $\mathbf{R}^n$ with finite first moment being an ordered set with the Choquet ordering.

Key words. Bernoulli distribution, comparison of measures after Choquet, boundedness of stochastic processes linearly generated by independent variables

PII. S0040585X97975769

Denote by M_n the class of all probability distributions μ in $\mathbf{R}^n$ having finite first moment: $\int_{\mathbf{R}^n} \|x\| \, d\mu(x) < \infty$. Let M^n denote a subclass of M_n consisting of product-measures, i.e.,

*Received by the editors June 6, 1994. The research was supported by the International Science Foundation Grant NXZ000.

http://www.siam.org/journals/tvp/41-4/97576.html

†The Syktyvkar State University, Faculty of Mathematics, 167001, Syktyvkar, Oktjabrskij pr. 55.

the probability measures admitting representation as a product $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ of measures on the real line with finite first moment.

For a pair of measures μ, $\nu \in M_n$, we shall write $\mu \prec \nu$ whenever for any convex function $f\colon \mathbf{R}^n \to \mathbf{R}$

$$\int_{\mathbf{R}^n} f\,d\mu \leqq \int_{\mathbf{R}^n} f\,d\nu. \tag{1}$$

The order relation $\prec$ was introduced by Choquet [4] (see also [2, p. 279]; [3, p. 29]) for positive measures on convex compact subsets of locally convex spaces. We employ the relation $\prec$ without assumption on support. Note that the integrals in (1) exist (though, possibly taking value $+\infty$), since any convex function majorizes some affine function but all the linear functionals are summable with respect to the measures in M_n.

The main issue of study in this note is the following: what is the location of n-dimensional Bernoulli distribution P^n within M_n and M^n viewed as partially ordered sets, that is what kind of measures are comparable with P^n? By definition P^n assigns the mass 2^{-n} to every point of the discrete cube $\{-1,1\}^n$, i.e., it is the nth degree of measure P given by $P(\{-1\}) = P(\{1\}) = \frac{1}{2}$. Set $\mathfrak{A}^n = \{\mu \in M^n\colon\ \mu \prec P^n\}$, $\mathfrak{B}^n = \{\mu \in M^n\colon\ P^n \prec \mu\}$ so that

$$\mathfrak{A}^n \prec P^n \prec \mathfrak{B}^n.$$

Observe at once that two measures have equal means (barycenters), provided that they are comparable in M_n. Indeed, application of (1) to linear functions f and $-f$ yields equality in (1) for all the linear f. Therefore, the measures comparable with P^n have zero means. For a measure μ on $\mathbf{R}$, we denote by $m(\mu)$ (any of) its medians; note in this connection that $\int_{\mathbf{R}} |x - m(\mu)|\, d\mu(x)$ is independent of the median choice.

THEOREM 1. *The following relations hold*:

$$\mathfrak{A}^1 = \Big\{\mu \in M^1\colon\ \int_{\mathbf{R}} x\, d\mu(x) = 0,\ \mu\Big([-1,1]\Big) = 1\Big\},$$
$$\mathfrak{B}^1 = \Big\{\mu \in M^1\colon\ \int_{\mathbf{R}} x\, d\mu(x) = 0,\ \int_{\mathbf{R}} \Big|x - m(\mu)\Big|\, d\mu(x) \geqq 1\Big\}.$$

THEOREM 2. *One has* $\mathfrak{A}^n = (\mathfrak{A}^1)^n$, $\mathfrak{B}^n = (\mathfrak{B}^1)^n$ (*in the sense of Cartesian products*).

We now state Theorems 1 and 2 in terms of random variables. Let $\xi = (\xi_1, \ldots, \xi_n)$, $\eta = (\eta_1, \ldots, \eta_n)$ be random vectors composed of random variables (r.v.'s) having finite first moment. In the same way as for measures, denote by $m(\lambda)$ a median of r.v. λ.

COROLLARY 1. *Given that for all* $k = 1, \ldots, n$,

$$\mathbf{E}\xi_k = 0, \quad |\xi_k| \leqq 1 \ a.s., \tag{2}$$

$$\mathbf{E}\eta_k = 0, \qquad \mathbf{E}\Big|\eta_k - m(\eta_k)\Big| \geqq 1, \tag{3}$$

one has for any convex function $f\colon \mathbf{R}^n \to \mathbf{R}$

$$\mathbf{E}\, f(\xi) \leqq \mathbf{E}\, f(\eta). \tag{4}$$

If $\zeta = (\zeta_1, \ldots, \zeta_n)$, ζ_k are independent and $\mathbf{P}\{\zeta_k = \pm 1\} = \frac{1}{2}$ (r.v.'s with the law of distribution P), then conditions (2) and (3) are satisfied simultaneously, so by virtue of (4) for all the convex f

$$\mathbf{E}\, f(\xi) \leqq \mathbf{E}\, f(\zeta) \leqq \mathbf{E}\, f(\eta). \tag{5}$$

The converse claim is also valid: if the first (respectively, the second) inequality in (5) holds for all convex f, then (2) (respectively, (3)) is true.

We present another corollary.

COROLLARY 2. *If independent r.v.'s η_n $(n \geqq 1)$ meet* (3), *then for any convex $f\colon \mathbf{R} \to \mathbf{R}$ one has*

$$\liminf_{n\to\infty} \mathbf{E} f\left(\frac{S_n}{\sqrt{n}}\right) \geqq \mathbf{E}\, f(\lambda), \tag{6}$$

where $S_n = \eta_1 + \cdots + \eta_n$ and r.v. λ has a standard normal distribution.

Within the class $\mathfrak{A}^n$ the distribution P^n has yet another extremal property.

THEOREM 3. *Given the validity of* (2), *one has for all convex smooth functions $f\colon \mathbf{R}^n \to \mathbf{R}$*

$$\mathbf{D}\, f(\xi) \leqq 2\mathbf{E}\Big|\nabla f(\xi)\Big|^2. \tag{7}$$

Here ∇f is the gradient of f, $|\nabla f|^2 = \sum_{k=1}^n (\partial f/\partial x_k)^2$.

The constant value 2 appearing in (7) can be sharpened if and only if none of the ξ_k has Bernoulli distribution, in which case (7) is fulfilled with the constant $1 + \max_{1\leqq k\leqq n} \mathbf{D}\xi_k$ in front of $\mathbf{E}|\nabla f(\xi)|^2$. However, when $\xi_k = \zeta_k$ for some k, (7) turns into equality on $f(x) = \max(x_k + 1, 0)$ (in asymptotic sense, i.e., on some sequence of convex smooth f_m converging to f).

Let $(\xi_n)_{n\geqq 1}$ be a sequence of independent r.v.'s, $\mathbf{E}\xi_n = 0$, having a common nondegenerate distribution with compact support in $\mathbf{R}$. Let $(\eta_n)_{n\geqq 1}$ be another sequence of r.v.'s with the same properties (but having a different common distribution). Consider two stochastic processes:

$$x(t) = \sum_{n=1}^{\infty} a_n(t)\,\xi_n, \qquad y(t) = \sum_{n=1}^{\infty} a_n(t)\,\eta_n, \quad t \in T, \tag{8}$$

linearly generated by r.v.'s ξ_n and η_n, respectively. Here the coefficients a_n are arbitrary functions given on a parametric (abstract) set T, moreover, the series in (8) are assumed to be convergent a.s. for all $t \in T$, i.e., $\sum_{n\geqq 1} a_n(t)^2 < +\infty$.

COROLLARY 3. *A stochastic process $x(t)$, $t \in T$, is sample bounded, i.e.,* $\sup_{t\in T} x(t) < +\infty$ *a.s., if and only if it is a process $y(t)$, $t \in T$.*

If it were true that all the r.v.'s ξ_n have $\mathcal{N}(0,1)$-distribution, then (8) would provide the general form of decomposition of a centered Gaussian stochastic process $x(t)$. In that case, the final solution to the problem of establishing necessary and sufficient conditions of boundedness for $x(t)$ realizations was obtained by M. Talagrand (this problem, with some simplifications in proofs, was elucidated in detail by M. Ledoux [1]). In the case of Bernoullian r.v.'s ξ_n (the corresponding stochastic process is called Bernoullian by M. Ledoux, see [1, p. 73]) no necessary and sufficient conditions of boundedness are known for $x(t)$. Corollary 3 indicates that in the case of bounded r.v.'s ξ_n the problem of establishing conditions for boundedness of $x(t)$ is reduced to studying Bernoullian generating variables.

Proof of Theorem 2 follows easily by induction, if we use the following statement.

LEMMA 1. *Let μ_1, ν_1 and μ_2, ν_2 be probability measures in $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. Then if $\mu_1 \prec \mu_2$ and $\nu_1 \prec \nu_2$, one has $\mu_1 \otimes \nu_1 \prec \mu_2 \otimes \nu_2$.*

As usual the product of measures μ and ν is denoted by $\mu \otimes \nu$.

Proof of Lemma 1. We first remark that definition (1) can be extended to all the functions $f\colon \mathbf{R}^n \longrightarrow (-\infty, +\infty]$ meeting the inequality

$$f\Big(tx + (1-t)\,y\Big) \leqq t f(x) + (1-t)\, f(y), \quad x, y \in \mathbf{R}^n,\ t \in [0, 1]$$

(the value $+\infty$ is admitted). Let $f\colon \mathbf{R}^{n+m} \to \mathbf{R}$ be convex. Then the following function will be "wide sense" convex:

$$g(y) = \int f(x, y)\, d\mu_1(x), \qquad h(y) = \int f(x, y)\, d\mu_2(x),$$

moreover, $g(y) \leqq h(y)$ for all $y \in \mathbf{R}^m$. Applying (1) to g, using the Fubini theorem, we have

$$\int_{\mathbf{R}^{n+m}} f\, d\mu_1 \otimes \nu_1 = \int_{\mathbf{R}^m} g\, d\nu_1 \leqq \int_{\mathbf{R}^m} g\, d\nu_2 \leqq \int_{\mathbf{R}^m} h\, d\nu_2 = \int_{\mathbf{R}^{n+m}} f\, d\mu_2 \otimes \nu_2.$$

The lemma is proved.

Proof of Theorem 1. We shall only consider the measures having zero mean, since, as was already noted, $\mu \prec \nu$ entails $\int x\, d\mu(x) = \int x\, d\nu(x)$.

Consider first the former assertion (the description of class $\mathfrak{A}^1$). When $\mu \prec P$, $\operatorname{supp}(\mu) \subset [-1, 1]$; otherwise, for instance if $\mu((1, +\infty)) > 0$, the value $\int f\, d\mu$ can be arbitrarily large, whereas $f(-1)$ and $f(1)$, hence also $\int f\, dP$, remain constant. Conversely, let $\operatorname{supp}(\mu) \subset [-1, 1]$. Then for all $x \in [-1, 1]$ and any convex f,

$$f(x) \leqq \frac{1-x}{2} f(-1) + \frac{1+x}{2} f(1).$$

Consequently, $\int f\, d\mu \leqq [f(-1) + f(1)]/2 = \int f\, dP$. Thus, $\mu \prec P$.

Now we show that $P \prec \mu \Longleftrightarrow \int_{\mathbf{R}} |x - m(\mu)|\, d\mu(x) \geqq 1$. Any convex function $f\colon \mathbf{R} \to \mathbf{R}$ satisfies the inequalities

$$f(x) \geqq f(-1) + f'(-1-0)\,(x+1) = f_{-1}(x),$$
$$f(x) \geqq f(1) + f'(1+0)\,(x-1) = f_1(x)$$

for arbitrary $x \in \mathbf{R}$. Hence $f(x) \geqq f_0(x) = \max\{f_{-1}(x), f_1(x)\}$ and therefore $\int f\, d\mu \geqq \int f_0 d\mu$. But $f(x) = f_0(x)$ for $x = \pm 1$, so $\int f\, dP = \int f_0 dP$. Consequently, $\int f\, d\mu \geqq \int f\, dP$ is valid for all convex f if and only if the same holds for functions of the f_0-kind, i.e., for the functions of the following form

$$f(x) = \begin{cases} \alpha(x - x_0) + c, & x \leqq x_0, \\ \beta(x - x_0) + c, & x \geqq x_0, \end{cases}$$

where $\alpha \leqq \beta$, $c \in \mathbf{R}$, $|x_0| \leqq 1$. Given $\alpha = \beta$, f is an affine function, hence $\int f\, d\mu = f(0) = \int f\, dP$. If $\alpha < \beta$,

$$\begin{aligned} \int f\, d\mu &= c + \alpha \int_{-\infty}^{x_0} (x - x_0)\, d\mu(x) + \beta \int_{x_0}^{\infty} (x - x_0)\, d\mu(x) \\ &= c - \beta x_0 - (\beta - \alpha) \int_{-\infty}^{x_0} (x - x_0)\, d\mu(x), \\ \int f\, dP &= c - \beta x_0 - (\beta - \alpha) \int_{-\infty}^{x_0} (x - x_0)\, dP(x), \end{aligned}$$

because μ and P have zero means. Thus the sign of the value

$$\int f\, d\mu - \int f\, dP = -(\beta - \alpha) \int_{-\infty}^{x_0} (x - x_0)\, d(\mu - P)\,(x)$$

depends neither on c nor on α and β, so one can set $c = 0$, $\alpha = -1$, $\beta = 1$ yielding $f(x) = |x - x_0|$. Consequently,

$$P \prec \mu \Longleftrightarrow \int |x - x_0|\, d\mu(x) \geqq \int |x - x_0|\, dP(x) \quad \text{for all } x_0 \in [-1,\, 1].$$

But for these x_0 we have $\int |x - x_0|\, dP(x) = 1$ and the following characterization results:

$$P \prec \mu \Longleftrightarrow \big(\forall\, x_0 \in [-1,\, 1]\big) \qquad \int_{\mathbf{R}} |x - x_0|\, d\mu(x) \geqq 1 \tag{9}$$

(μ is assumed to have finite first moment, and, moreover, $\int_{\mathbf{R}} x\, d\mu(x) = 0$). Observe that for all $x_0 \in \mathbf{R}$

$$\int_{\mathbf{R}} |x - x_0|\, d\mu(x) \geqq \left| \int_{\mathbf{R}} (x - x_0)\, d\mu(x) \right| = |x_0|.$$

Therefore, given the validity of the inequality in the right-hand side of (9) for all $x_0 \in [-1, 1]$, the same is true for all $x_0 \in \mathbf{R}$. Consequently,

$$P \prec \mu \Longleftrightarrow \inf_{a \in \mathbf{R}} \int_{\mathbf{R}} |x - a|\, d\mu(x) = \int_{\mathbf{R}} \Big| x - m(\mu) \Big|\, d\mu(x) \geqq 1.$$

Theorem 1 is thus proved.

Remarks. The role played by the independence property in the context of Corollary 1 is a question of interest. In particular, when $\xi = (\xi_1, \ldots, \xi_n)$, $\mathbf{E}\xi_i = 0$, $|\xi_i| \leqq 1$ a.s. and, moreover, the independence of ξ_i is not assumed, it is interesting to know when the distribution μ of a random vector ξ is majorized by the multidimensional Bernoulli distribution P^n. The following simple condition can be proposed.

THEOREM 4. *If, for any* $1 \leqq i_1 < \cdots < i_k \leqq n$,

$$\mathbf{E}\, \xi_{i_1} \cdots \xi_{i_k} = 0, \tag{10}$$

then $\mu \prec P^n$.

Proof of Theorem 4. Set $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \{-1, 1\}^n$, $x \in [-1, 1]^n$ for $p_\varepsilon(x) = 2^{-n} \prod_{i=1}^n (1 + \varepsilon_i x_i)$. Obviously,

$$p_\varepsilon(x) \geqq 0, \qquad \sum_{\varepsilon \in \{-1,1\}^n} p_\varepsilon(x) = 1,$$

since the sum above can be written as

$$\mathbf{E}(1 + \zeta_1 x_1) \cdots (1 + \zeta_n x_n) = \mathbf{E}(1 + \zeta_1 x_1) \cdots \mathbf{E}(1 + \zeta_n x_n) = 1,$$

where $\zeta_1, \ldots, \zeta_n$ are independent Bernoullian variables. In a similar manner one verifies the identity

$$\sum_\varepsilon p_\varepsilon(x)\, \varepsilon = x.$$

Therefore, for any convex f: $\mathbf{R}^n \to \mathbf{R}$

$$f(x) \leqq \sum_\varepsilon p_\varepsilon(x)\, f(\varepsilon),$$

and, having integrated in measure μ this inequality, we obtain

$$\int_{\mathbf{R}} f\, d\mu \leqq \sum_\varepsilon \mathbf{E}\, p_\varepsilon(\xi_1, \ldots, \xi_n)\, f(\varepsilon) = 2^{-n} \sum_\varepsilon f(\varepsilon) = \int_{\mathbf{R}^n} f\, dP^n,$$

since

$$\begin{aligned} \mathbf{E}\, p_\varepsilon(\xi_1, \ldots, \xi_n) &= 2^{-n} \mathbf{E}(1 + \varepsilon_1 \zeta_1) \cdots (1 + \varepsilon_n \zeta_n) \\ &= 2^{-n} \left(1 + \sum_{1 \leqq i_1 < \cdots < i_k \leqq n} \varepsilon_{i_1} \cdots \varepsilon_{i_k} \mathbf{E}\xi_{i_1} \cdots \xi_{i_k} \right) = 2^{-n}. \end{aligned}$$

Consequently, $\mu \prec P^n$.

Example 1. A sequence of r.v.'s $\xi_i(\omega) = \cos(2^i \omega)$, $0 \leqq i \leqq n$, given on a probability space $((-\pi, \pi), d\omega/(2\pi))$ meets (10), so its distribution is majorized by P^{n+1}.

Example 2. For $\varepsilon_1, \ldots, \varepsilon_n = \pm 1$, let random vectors $(\xi_1, \ldots, \xi_n)$ and $(\varepsilon_1 \xi_1, \ldots, \varepsilon_n \xi_n)$ have the same distribution. By Theorem 3, whenever $|\xi_i| \leqq 1$ a.s., one has $\mu \prec P^n$. It is noteworthy that this claim can be converted. Whenever $|\xi_i| \geqq 1$ a.s., $P^n \prec \mu$. This fact is readily established by induction, upon representing μ as a mixture $\int_{\mathbf{R}} \mu_t d\pi(t)$ so that the conditional measure μ_t be concentrated on the hyperplane $x_n = t$, and employing the same reasoning as in the proof of Lemma 1.

Proof of Theorem 3. Assume first $n = 1$ and let f be a convex smooth function on $\mathbf{R}$. Since the derivative f' is nondecreasing for any $x, y \in \mathbf{R}$, one has

$$\Big|f(x) - f(y)\Big| \leqq \max\Big\{\Big|f'(x)\Big|, \Big|f'(y)\Big|\Big\}\, |x-y|,$$

so

$$\Big|f(x) - f(y)\Big|^2 \leqq \Big(\Big|f'(x)\Big|^2, \Big|f'(y)\Big|^2\Big)\,(x-y)^2.$$

Hence, for the variance of f with respect to μ, we get

$$\begin{aligned}
\mathbf{D}f &= \frac{1}{2}\int_{\mathbf{R}}\int_{\mathbf{R}} \Big|f(x) - f(y)\Big|^2 d\mu(x)\, d\mu(y) \\
&\leqq \frac{1}{2}\int_{\mathbf{R}}\int_{\mathbf{R}} \Big(\Big|f'(x)\Big|^2 + \Big|f'(y)\Big|^2\Big)\,(x-y)^2 d\mu(x)\, d\mu(y) \\
&= \frac{1}{2}\int_{\mathbf{R}}\int_{\mathbf{R}} \Big|f'(x)\Big|^2 (x^2+y^2) d\mu(x)\, d\mu(y) + \frac{1}{2}\int_{\mathbf{R}}\int_{\mathbf{R}} \Big|f'(y)\Big|^2 (x^2+y^2)\, d\mu(x)\, d\mu(y) \\
&\quad - \int_{\mathbf{R}}\int_{\mathbf{R}} \Big(f'(x)^2 + f'(y)^2\Big)\, xy\, d\mu(x)\, d\mu(y) = \int_{\mathbf{R}} \Big|f'(x)\Big|^2 (x^2+\sigma^2)\, d\mu(x),
\end{aligned}$$

where $\sigma^2 = \int_{\mathbf{R}} x^2 d\mu(x)$ (writing out the last inequality we have taken into account the hypothesis on zero mean of μ). Note that the smoothness condition can be dropped by rendering to $|f'(x)|$ the value $\max\{|f'(x-0)|, |f'(x+0)|\}$.

We demonstrate by induction that, given

$$\mu = \mu_1 \otimes \cdots \otimes \mu_n, \quad \int_{\mathbf{R}} x\, d\mu_i(x) = 0, \quad \int_{\mathbf{R}} x^2\, d\mu_i(x) = \sigma_i^2,$$

for any convex smooth $f\colon \mathbf{R}^n \to \mathbf{R}$ one has

$$\mathbf{D}\, f \leqq \sum_{i=1}^n \left|\frac{\partial f(x)}{\partial x_i}\right|^2 (x_i^2 + \sigma_i^2)\, d\mu(x). \tag{11}$$

Realizing the inductive passage, consider a convex smooth function $f\colon \mathbf{R}^{n+1} \to \mathbf{R}$. Fix $x_{n+1} \in \mathbf{R}$ and write (11) for the function $x \in \mathbf{R}^n \longrightarrow f(x, x_{n+1})$:

$$\begin{aligned}
\int_{\mathbf{R}^n} f(x,\, x_{n+1})^2 d\mu_1 \otimes \cdots \otimes \mu_n(x) &\leqq \Big[\int_{\mathbf{R}^n} f(x,\, x_{n+1})\, d\mu_1 \otimes \cdots \otimes \mu_n(x)\Big]^2 \\
&+ \sum_{i=1}^n \int_{\mathbf{R}^n} \left|\frac{\partial f(x, x_{n+1})}{\partial x_i}\right|^2 (x_i^2 + \sigma_i^2)\, d\mu_1 \otimes \cdots \otimes \mu_n(x).
\end{aligned} \tag{12}$$

Set $g(x_{n+1}) = \int_{\mathbf{R}^n} f(x,\, x_{n+1})\, d\mu_1 \otimes \cdots \otimes \mu_n(x)$. This function is convex, and one can apply the inequality established for $n = 1$:

$$\begin{aligned}
\int_{\mathbf{R}} g(x_{n+1})^2\, d\mu_{n+1}(x_{n+1}) &\leqq \Big[\int_{\mathbf{R}} g(x_{n+1})\, d\mu_{n+1}(x_{n+1})\Big]^2 \\
&\quad + \int_{\mathbf{R}} \Big|g'(x_{n+1})\Big|^2 (x_{n+1}^2 + \sigma_{n+1}^2)\, d\mu_{n+1}(x_{n+1}).
\end{aligned}$$

But $\int_{\mathbf{R}} g\, d\mu_{n+1} = \int_{\mathbf{R}^{n+1}} f\, d\mu$ by the Fubini theorem and

$$g'(x_{n+1}) = \int_{\mathbf{R}^n} \frac{\partial f(x, x_{n+1})}{\partial x_{n+1}}\, d\mu_1 \otimes \cdots \otimes \mu_n(x),$$

whereas by the Schwarz inequality

$$g'(x_{n+1})^2 \leqq \int_{\mathbf{R}^n} \left| \frac{\partial f(x, x_{n+1})}{\partial x_{n+1}} \right|^2 \mu_1 \otimes \cdots \otimes \mu_n(x).$$

Thus we have

$$\int_{\mathbf{R}} g(x_{n+1})^2 \, d\mu_{n+1}(x_{n+1}) \leqq \left(\int_{\mathbf{R}^{n+1}} f \, d\mu \right)^2 + \int_{\mathbf{R}^{n+1}} \left| \frac{\partial f(x, x_{n+1})}{\partial x_{n+1}} \right|^2 (x_{n+1}^2 + \sigma_{n+1}^2) \, d\mu(x, \, x_{n+1}). \tag{13}$$

Integrating (12) in x_{n+1} variable and taking into account (13) we deduce (11) for $n+1$. Whenever $\operatorname{supp}(\mu_i) \subset [-1, 1]$, (11) implies the inequality

$$\mathbf{D} \, f \leqq \max_{1 \leqq i \leqq n} (1 + \sigma_i^2) \, \mathbf{E} |\nabla f|^2.$$

Theorem 3 is proved.

Proof of Corollary 2. By Theorem 1,

$$\mathbf{E} \, f\left(\frac{S_n}{\sqrt{n}} \right) \geqq \mathbf{E} \, f\left(\frac{\zeta_1 + \cdots + \zeta_n}{\sqrt{n}} \right) = \int_{-\infty}^{+\infty} f(x) \, dF_n(x),$$

where F_n is the distribution function of $(\zeta_1 + \cdots + \zeta_n)/\sqrt{n}$. Without loss of generality, set $f(0) = 0$. Since the distribution associated with F_n is symmetric with respect to zero, we have

$$\int_{-\infty}^{+\infty} f \, dF_n = \int_0^{+\infty} \Big(f(x) + f(-x) \Big) \, dF_n(x) = \int_0^{+\infty} \Big(1 - F_n(x) \Big) \, d\Big(f(x) + f(-x) \Big).$$

The function $G(x) = f(x) + f(-x)$ is nondecreasing on $[0, +\infty)$, and one can therefore apply the Fatou lemma:

$$\begin{aligned} \liminf_{n \to \infty} \int_{-\infty}^{+\infty} f \, dF_n &\geqq \int_0^{+\infty} \liminf_{n \to \infty} \Big(1 - F_n(x) \Big) \, dG(x) \\ &= \int_0^{+\infty} \Big(1 - \Phi(x) \Big) \, dG(x) = \int_0^{+\infty} G(x) \, d\Phi(x), \end{aligned}$$

where Φ is the distribution function of a standard normal r.v. λ. It remains to observe that

$$\int_0^{+\infty} G \, d\Phi < +\infty$$

if and only if

$$\int_{\mathbf{R}} f \, d\Phi < +\infty.$$

In any case

$$\int_0^{+\infty} G \, d\Phi = \int_{\mathbf{R}} f \, d\Phi.$$

Proof of Corollary 3. Note first that the studied suprema of stochastic processes are understood as the structural suprema. The space L^0 of all random variables with the usual order relation is a conditionally complete lattice; moreover, the supremum of any bounded subset $K \subset L^0$ is that of some countable subset in K. Thus one can consider a countable parametric set T in Corollary 3. We show first that

$$\sup_t x(t) < +\infty \text{ a.s. } \Longleftrightarrow \mathbf{E} \sup_t x(t) < +\infty,$$

the same being true for $y(t)$.

When $\mathbf{E}\sup_t x(t) < +\infty$ one obviously has $\sup_t x(t) < +\infty$ a.s. Conversely, without loss of generality assume $|\xi_n| \leqq 1$ a.s., $n \geqq 1$. Since $W = \sup_t x(t) < +\infty$ a.s., it is obvious that the variances

$$\mathbf{D}\, x(t) = \sum_{n=1}^{\infty} \Big|a_n(t)\Big|^2 \mathbf{D}\,\xi_1$$

are bounded by some constant σ^2. Let

$$f_n(x) = \sup_t \sum_{k=1}^{n} a_k(t)\, x_k, \quad x = (x_1, \dots, x_n) \in \mathbf{R}^n.$$

This function is convex and Lipschitzian, moreover, $\|f_n\|_{\mathrm{Lip}} \leqq \sigma$. By Theorem 2 we have $\mathbf{D} f_n \leqq 2\sigma^2$. Assume first T to be finite. Because a.s. $f_n(\xi_1, \dots, \xi_n) \longrightarrow W$ as $n \to \infty$, we conclude (for example, due to the Fatou lemma) that $\mathbf{D}W \leqq 2\sigma^2$. If $T = \{t_n\colon\ n \geqq 1\}$ is infinite, considering r.v.'s $W_n = \sup_{1 \leqq k \leqq n} x(t_k)$ and using the fact that $W_n \to W$ as $n \to \infty$, $\mathbf{D}W_n \leqq 2\sigma^2$, we establish that $\mathbf{D}W \leqq 2\sigma^2$ in the general case. In particular, $\mathbf{E}W < +\infty$ (to avoid the definition of variance by means of $\mathbf{E}W$ one should employ the identity $\mathbf{D}W = \frac{1}{2}\,\mathbf{E}|W - W'|$, where W' is an independent copy of W).

By Theorem 1, $\mathbf{E}\sup_t x(t) < +\infty \Longleftrightarrow \mathbf{E}\sup_t y(t) < +\infty$. Indeed, for any convex function $f\colon \mathbf{R}^n \to \mathbf{R}$, inequality (4) gives

$$\mathbf{E}\, f\left(\frac{\xi}{c}\right) \leqq \mathbf{E}\, f\left(\frac{\eta}{d}\right), \tag{14}$$

where $\xi = (\xi_1, \dots, \xi_n)$, $\eta = (\eta_1, \dots, \eta_n)$, $c = \operatorname{ess\,sup}|\xi_1|$, $d = \mathbf{E}|\eta_1 - m(\eta_1)|$. Applying (14) to f_n, and using homogeneity we arrive at the inequality

$$\mathbf{E}\, f_n(\xi) \leqq \frac{c}{d}\, \mathbf{E} f_n(\eta).$$

Since this inequality is independent of dimension, one can extend it to the infinite-dimensional case.

In conclusion the author wishes to express his gratitude to the anonymous referee for careful reading and useful comments.

REFERENCES

[1] M. Ledoux, *Isoperimetry and Gaussian Analysis*, Ecole d'été de Probabilités de Saint-Flour, 1994.

[2] P.-A. Meyer, *Probability and Potentials*, Blaisdell, Waltham, MA, 1966.

[3] R. Phelps, *Lectures on Choquet's Theorem*, D. Van Nostrand, Princeton, Toronto, London, 1966.

[4] G. Choquet, *Le théorème de représentation intégrale dans les ensembles convexes compacts*, Ann. Inst. Fourier Grenoble, 10 (1960), pp. 333–344.