Elect. Comm. in Probab. **13** (2008), 266–271

ELECTRONIC COMMUNICATIONS in PROBABILITY

A NOTE ON THE DISTRIBUTIONS OF THE MAXIMUM OF LINEAR BERNOULLI PROCESSES

SERGEY G. BOBKOV
School of Mathematics, University of Minnesota
127 Vincent Hall, 206 Church St. S.E.
Minneapolis, MN 55455 USA
email: bobkov@math.umn.edu

Submitted May 4, 2007, *accepted in final form* May 5, 2008

AMS 2000 Subject classification: 60J65
Keywords: Linear Bernoulli processes, distribution of the maximum

Abstract

We give a characterization of the family of all probability measures on the extended line $(-\infty, +\infty]$, which may be obtained as the distribution of the maximum of some linear Bernoulli process.

On a probability space $(\Omega, \mathbf{P})$ consider a linear process

$$X(t) = a_0(t) + \sum_{n=1}^{\infty} a_n(t)\xi_n, \quad t \in T, \tag{1}$$

generated by independent, identically distributed random variables ξ_n with $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 = 1$. The coefficients $a_n(t)$ are assumed to be arbitrary functions on the parameter set T, satisfying $\sum_{n=1}^{\infty} a_n(t)^2 < +\infty$ for any $t \in T$, so that the series (1) is convergent a.s. Define

$$M = \sup_t X(t) \tag{2}$$

in the usual way as the essential supremum in the space of all random variables with values in the extended real line (identifying random variables that coincide almost surely; cf. Remark 4 below).

We consider the question on the characterization of the family $\mathcal{F}(L)$ of all possible distribution functions $F(x) = \mathbf{P}\{M \leq x\}$ of M, assuming that the common law L of ξ_n is given. In general, M may take the value $+\infty$ with positive probability, so its distribution is supported on $(-\infty, +\infty]$. Introduce also the collection $\mathcal{F}_0(L)$ of all possible distribution functions of M in (2), such that in the series (1), for all $t \in T$,

$$a_n(t) = 0, \quad \text{for all sufficiently large } n. \tag{3}$$

When ξ_n are standard normal, i.e., $L = N(0,1)$, we deal in (1) with an arbitrary Gaussian random process. As is well-known, for the distribution function F of M, $x_0 = \inf\{x \in \mathbf{R} : F(x) > 0\}$ may be finite, and then it is sometimes called a take-off point of the maximum of the Gaussian process. Moreover, F may have an atom at it. But anyway F is absolutely continuous and strictly increasing on $(x_0, +\infty)$, which follows from the log-concavity of Gaussian measures (cf. also [C], [HJ-S-D]).
A complete characterization of all possible distributions F in the Gaussian case may be derived from the Brunn-Minkowski-type inequality for the standard Gaussian measure γ_n on $\mathbf{R}^n$ due to A. Ehrhard [E]. It states that, for all convex (and in fact, for all Borell measurable, cf. [Bo2]) sets A and B in $\mathbf{R}^n$ of positive measure and for all $\lambda \in (0,1)$,

$$\Phi^{-1}\left(\gamma_n(\lambda A + (1-\lambda)B)\right) \;\geq\; \lambda\, \Phi^{-1}\left(\gamma_n(A)\right) + (1-\lambda)\, \Phi^{-1}\left(\gamma_n(B)\right),$$

where Φ^{-1} denotes the inverse to the standard normal distribution function on the line. This inequality immediately implies that, if F is non-degenerate, the function $U = \Phi^{-1}(F)$ must be concave on $\mathbf{R}$ in the generalized sense as a function with values in $[-\infty, +\infty)$. But the converse is true, as well.
Indeed, suppose $U = \Phi^{-1}(F)$ is concave on $\mathbf{R}$, and for simplicity let F be non-degenerate and do not assign a positive mass to the point $+\infty$. Then F is strictly increasing on $(x_0, +\infty)$, so is its inverse $F^{-1} : (F(x_0), 1) \to (x_0, +\infty)$. Moreover, the inverse function $U^{-1} = F^{-1}(\Phi)$ is convex and strictly increasing on $(U(x_0), +\infty)$. Put $M(x) = U^{-1}(x)$ for $x > U(x_0)$, and if x_0 is finite, $M(x) = x_0$ on $(-\infty, U(x_0)]$. Then M is convex and finite on the whole real line, and therefore admits a representation

$$M(x) = \sup_{t \in T}\, [a_0(t) + a_1(t)x], \quad x \in \mathbf{R},$$

for some coefficients $a_0(t)$, $a_1(t)$. By the construction, M has the distribution function F under the measure γ_1, as was required.
Thus, a given non-degenerate distribution function F belongs to $\mathcal{F}(N(0,1))$, if and only if the function $\Phi^{-1}(F)$ is concave. A similar characterization holds true, when ξ_n's have a shifted one-sided exponential distribution with mean zero. Then, F represents the distribution function of M for some coefficients $a_n(t)$, if and only if the function $\log F$ is concave. This follows from the log-concavity of the multidimensional exponential distribution (which is a particular case of Prékopa's theorem [P]; cf. also [Bo1] for a general theory of log-concave measures).
In both above examples, for the "if" part it sufficies to consider simple linear processes $X(t) = a_0(t) + a_1(t)\xi_1$. Hence, $\mathcal{F}_0(L) = \mathcal{F}(L)$. The situation is completely different, when ξ_n have a symmetric Bernoulli distribution L, i.e., taking the values ± 1 with probability $\frac{1}{2}$. This may be seen from:

Theorem 1. *Any distribution function F, such that $F(x) = 0$, for some $x \in \mathbf{R}$, may be obtained as the distribution function of the supremum M of some linear Bernoulli process X in* (1) *with coefficients, satisfying the property* (3).

In turn, the condition (3) ensures that all random variables $X(t)$ in (1) are bounded from below, so is the random variable M in (2). Therefore, the distribution F of M must be one-sided. Thus, we have a full description of the family $\mathcal{F}_0(L)$ in the Bernoulli case. Removing the condition (3), we obtain a larger family $\mathcal{F}(L)$; however, it is not clear at all how to characterize it.

One should also mention that in the homogeneous case $a_0(t) = 0$, much is known about various properties of M in terms of L, but the characterization problem is more delicate, and it seems no description or even conjecture are known in all above cases.
For the proof of Theorem 1 one may assume that $\Omega = \{-1,1\}^\infty$ is the infinite dimensional discrete cube, equipped with the product Bernoulli measure $\mathbf{P}$. An important property of Ω, which will play the crucial role, is that it represents the collection of all extreme points in the cube $K = [-1,1]^\infty$. More precisely, we apply the following statement.

Lemma 2. *Any lower semi-continuous function $f : \{-1,1\}^\infty \to (-\infty,+\infty]$ is representable as*

$$f(x) \ = \ \sup_{t \in T} \left[a_0(t) + \sum_{n=1}^\infty a_n(t) x_n \right], \qquad x = (x_1, x_2, \dots), \tag{4}$$

for some family of the coefficient functions $a_n(t)$, defined on a countable set T and satisfying the property (3).

Note any function of the form (4) is lower semi-continuous.

Proof. First, more generally, let K be a non-empty, compact convex set in a locally convex space E, and denote by Ω the collection of all extreme points of K. A function $f : \Omega \to (-\infty,+\infty]$ is representable as

$$f(x) = \sup_t f_t(x), \qquad x \in \Omega, \tag{5}$$

for some family $(f_t)_{t \in T}$ of continuous, affine functions on E, if and only if

a) f is lower semi-continuous on Ω;
b) f is bounded from below.

This characterization follows from a theorem, usually attributed to Hervé [H]; see E. M. Alfsen [A], Proposition 1.4.1, and historical remarks. Namely, a point x is an extreme point of K, if and only if $\bar{g}(x) = g(x)$, for any lower semi-continuous function g on K, where $\bar{g}$ denotes the lower envelope of g (i.e., the maximal convex, lower semi-continuous function on K, majorized by g).
Clearly, the equality (5) defines a function with properties $a) - b)$. For the opposite direction one may use an argument, contained in the proof of Corollary 1.4.2 of [A]. If f is bounded and lower semi-continuous on Ω, put $g(x) = \liminf_{y \to x} f(y)$ for $x \in \mathrm{clos}(\Omega)$ and $g = \sup_\Omega f$ on $K \setminus \mathrm{clos}(\Omega)$. Then g is lower semi-continuous on K and $g = f$ on Ω. By Hervé's theorem, $\bar{g}(x) = g(x) = f(x)$, for all $x \in \Omega$. Since $\bar{g}$ is also convex on K, one may apply to it the classical theorem on the existence of the representation

$$\bar{g}(x) = \sup_t f_t(x), \quad x \in K,$$

for some family $(f_t)_{t \in T}$ of continuous, affine functions on E (cf. e.g. [A], Proposition 1.1.2, or [M], Chapter 11). Thus, restricting this representation to Ω, we arrive at (5). Finally, if f is unbounded from above, write $f = \sup_n \min\{f,n\}$ and apply (5) to the sequence $\min\{f,n\}$.
In case of the infinite dimensional discrete cube, the right-hand side of (5) may further be specified. Indeed, any continuous, affine function g on $E = \mathbf{R}^\infty$ has the form $g(x_1, x_2, \dots) = a_0 + \sum_{n=1}^\infty a_n x_n$ with finitely many non-zero coefficients. Therefore, (5) is reduced to the

relation (4) with some coefficient functions $a_n = a_n(t)$, that are defined on non-empty, perhaps, uncountable set T and satisfy the property (3).
The latter implies that the sets $T_N = \{t \in T : a_n(t) = 0, \text{ for all } n > N\}$ are non-empty for all $N \geq N_0$ with a sufficiently large N_0. Define

$$f_N(x) \;=\; \sup_{t \in T_N} \left[a_0(t) + \sum_{n=1}^{\infty} a_n(t) x_n \right] \;=\; \sup_{t \in T_N} \left[a_0(t) + \sum_{n=1}^{N} a_n(t) x_n \right], \tag{6}$$

so that $f = \sup_{N \geq N_0} f_N$. Since for each point $v = (x_1, \dots, x_N)$ in the finite dimensional discrete cube $\{-1,1\}^N$, the second supremum in (6) is asymptotically attained for some sequence of indices in T_N, one may choose a countable subset $T_N(v)$ of T_N, such that

$$\sup_{t \in T_N} \left[a_0(t) + \sum_{n=1}^{N} a_n(t) x_n \right] \;=\; \sup_{t \in T_N(v)} \left[a_0(t) + \sum_{n=1}^{N} a_n(t) x_n \right].$$

Therefore, the set $T'_N = \cup_{v \in \{-1,1\}^N} T_N(v)$ is also countable, is contained in T_N, and by (6),

$$f_N(x) \;=\; \sup_{t \in T'_N} \left[a_0(t) + \sum_{n=1}^{\infty} a_n(t) x_n \right], \qquad \text{for all } x \in \{-1,1\}^\infty.$$

As a result, the supremum in (4) may be restricted to the countable set $\cup_N T'_N$.
Finally, let us note Ω is compact, so the property *b*) is automatically satisfied, when *a*) holds. This yields Lemma 2.

Proof of Theorem 1. According to Lemma 2, we need to show that distributions of lower semi-continuous functions f on $\{-1,1\}^\infty$ under the Bernoulli measure $\mathbf{P}$ fill the family of all one-sided distributions on $(-\infty, +\infty]$. In fact, it is enough to consider the functions of the special form $f(x) = \varphi(Q(x))$, where

$$Q(x) = \sum_{n=1}^{\infty} \frac{x_n + 1}{2^{n+1}}, \qquad x = (x_1, x_2, \dots) \in \{-1,1\}^\infty,$$

and where $\varphi : [0,1] \to (-\infty, +\infty]$ is an arbitrary non-decreasing, left (or, equivalently, lower semi-) continuous function. It is allowed that for some point $p \in [0,1]$, φ jumps to the value $+\infty$, and then we require that $\lim_{s \to p} \varphi(s) = +\infty$, as part of the lower semi-continuity assumption.
The map Q is continuous and pushes forward $\mathbf{P}$ to the normalized Lebesgue measure λ on the unit interval $[0,1]$. Hence, f is lower semi-continuous, and its distribution under $\mathbf{P}$ coincides with the distribution of φ under λ.
It remains to see that, for any one-sided probability measure μ on $(-\infty, +\infty]$, there is an admissible φ with the distribution μ under λ. Let us recall the standard argument (cf. e.g. [Bi], Theorem 14.1). Introduce the distribution function $F(u) = \mu((-\infty, u])$, $-\infty < u \leq +\infty$, and define its "inverse"

$$\varphi(s) = \min\{u : F(u) \geq s\}, \quad 0 < s \leq 1.$$

Also put $\varphi(0) = \lim_{s \to 0} \varphi(s)$. Clearly, φ is non-decreasing. Given a sequence $s_n \uparrow s$, $0 < s_n < s \leq 1$, take minimal values u_n, u, such that $F(u_n) \geq s_n$, $F(u) \geq s$. We have $u_n \uparrow u'$, for some

$u' \le u$. Since $F(u') \ge s_n$, for all n, we get $F(u') \ge s$ and hence $u' \ge u$. This shows that φ is left continuous. Finally, given $s \in (0,1]$ and $\alpha > \varphi(0)$, by the definition, $\varphi(s) \le \alpha \Leftrightarrow F(u) \ge s$, for some $u \le \alpha$. Hence,

$$\{s \in (0,1] : \varphi(s) \le \alpha\} = \{s \in (0,1] : F(u) \ge s, \text{ for some } u \le \alpha\} = (0, F(\alpha)].$$

Thus, φ has the distribution function F under λ. The proof is now complete.

Remark 3. The statement of Theorem 1 remains to hold in case of arbitrary independent random variables ξ_n, taking two values, say, a_n and b_n with probabilities p_n and q_n, satisfying

$$\prod_{n=1}^{\infty} \max\{p_n, q_n\} = 0.$$

In this case, the joint distribution $\mathbf{P}$ of ξ_n's represents a product probability measure on $\prod_{n=1}^{\infty}\{a_n, b_n\}$ without atoms. Let $a_n = -1$ and $b_n = 1$ (without loss of generality). Then, the map Q in the proof of Theorem 1 pushes $\mathbf{P}$ forward to a non-atomic probability measure λ on $[0,1]$, and a similar argument works.

Remark 4. The set $S = S(\Omega, \mathbf{P})$ of all random variables with values in the extended line $(-\infty, +\infty]$ represents a lattice with ordering $X \le Y$ a.s. Given an arbitrary non-empty collection $\{X(t)\}_{t \in T}$ in S, there is a unique element M in S, called the essential (or structural) supremum of the family $\{X(t)\}_{t \in T}$, with the properties that

a) $X(t) \le M$ (a.s), for all $t \in T$;
b) If for all $t \in T$ we have $X(t) \le M'$ (a.s.), $M' \in S$, then $M \le M'$ (a.s.)

It is a well-known general fact that M can be represented as a pointwise supremum $M = \sup_n X(t_n)$ a.s., for some sequence t_n in T (cf. e.g. [K-A]). In particular, the supremum in (2) may always be taken over all t's from a countable subset of T.

Acknowledgement. I am grateful to a referee for valuable comments and references.

References

[A] E. M. Alfsen (1971), Compact convex sets and boundary integrals. Springer-Verlag, New York, Heidelberg, Berlin.

[Bi] P. Billingsley (1979), Probability and measure. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-Chichester-Brisbane, xiv+515 pp.

[Bo1] C. Borell (1974), Convex measures on locally convex spaces. *Ark. Mat.*, 12, 239–252.

[Bo2] C. Borell (2003), The Ehrhard inequality. *C. R. Math. Acad. Sci. Paris*, 337, No. 10, 663–666.

[C] B. S. Cirel'son, Density of the distribution of the maximum of a Gaussian process. *Teor. Verojatnost. i Primenen.*, 20 (1975), No. 4, 865–873 (Russian).

[E] A. Ehrhard (1983), Symetrisation dans l'espace de Gauss. *Math. Scand.*, 53, No. 2, 281–301.

[H] M. Hervé (1961), Sur les représentations intégrales á l'aide des points extrémaux dans un ensemble compact convexe métrisable. (French) *C. R. Acad. Sci. Paris*, 253, 366–368.

[HJ-S-D] J. Hoffmann-Jorgensen, L. A. Shepp, and R. M. Dudley (1979), On the lower tail of Gaussian seminorms. *Ann. Probab.*, 7, No. 2, 319–342.

[K-A] L. V. Kantorovich, and G. P. Akilov (1982), Functional analysis. Translated from the Russian by Howard L. Silcock. Second edition. Pergamon Press, Oxford-Elmsford, N.Y., xiv+589 pp.

[M] P. A. Meyer (1966), Probability and potentials. Blaisdell Publ. Co., Toronto.

[P] A. Prékopa (1971), Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math. (Szeged)*, 32, 301–316.