Entropy, Compound Poisson Approximation, Log-Sobolev Inequalities and Measure Concentration

Ioannis Kontoyiannis[1] and Mokshay Madiman
Division of Applied Mathematics, Brown University
182 George St., Providence RI 02912
{yiannis, mokshay}@dam.brown.edu

Abstract — **The problem of approximating the distribution of a sum $S_n = \sum_{i=1}^n Y_i$ of n discrete random variables Y_i by a Poisson or a compound Poisson distribution arises naturally in many classical and current applications, such as statistical genetics, dynamical systems, the recurrence properties of Markov processes and reliability theory. Using information-theoretic ideas and techniques, we derive a family of new bounds for compound Poisson approximation. We take an approach similar to that of Kontoyiannis, Harremoës and Johnson (2003), and we generalize some of their Poisson approximation bounds to the *compound* Poisson case. Partly motivated by these results, we derive a new logarithmic Sobolev inequality for the compound Poisson measure and use it to prove measure-concentration bounds for a large class of discrete distributions.**

I. Introduction

Consider a sum $S_n = \sum_{i=1}^n Y_i$ of n random variables Y_i taking values in $\{0, 1, 2, \ldots\}$. The study of the distribution P_{S_n} of S_n is an important part of classical probability theory, and it arises naturally in many important applications. In particular, it is often the case that P_{S_n} can be well approximated by a Poisson or a compound Poisson distribution; see, e.g., [1][2][3] and the references therein. This is hardly surprising since, for example, in the special case when the Y_i are independent, the only possible limiting distributions of such sums (as $n \to \infty$) are compound Poisson distributions [4].

The two most well-known classical examples are:

Example 1. *The Poisson Law.* We find it convenient to write each Y_i as a product $B_i X_i$, where B_i is a Bernoulli(p_i) random variable and X_i is independent of B_i and has distribution Q_i, say, on $\{1, 2, \ldots\}$. This representation is unique, with $p_i = \Pr\{Y_i \neq 0\}$ and $Q_i(k) = \Pr\{Y_i = k\}/p_i$, for $k \geq 1$.

Suppose that the X_i are all identically equal to 1, and that the B_i are independent and identically distributed (i.i.d.) Bernoulli(λ/n) random variables for some $\lambda > 0$. Then S_n has a Binomial($n, \lambda/n$) distribution, which for large n is well approximated by the Poisson distribution with parameter λ, Po(λ). In fact, P_{S_n} converges to Po(λ) as $n \to \infty$.

Example 2. *The Compound Poisson Law.* In the same notation as above suppose that the X_i are i.i.d. with common distribution Q, and that the B_i are i.i.d. Bernoulli(λ/n) random variables for some $\lambda > 0$. Then S_n can be expressed as

$$S_n = \sum_{i=1}^n B_i X_i = \sum_{i=1}^{S'_n} X_i,$$

where $S'_n = \sum_{i=1}^n B_i$ has a Binomial($n, \lambda/n$) distribution. [Throughout the paper we take the empty sum $\sum_{i=1}^0 (\cdots)$ to be equal to zero.] Since $P_{S'_n}$ converges to Po(λ) as n grows to infinity, it is easy to see that S_n itself will converge in distribution to the compound Poisson sum

$$\sum_{i=1}^Z X_i, \tag{1}$$

where Z is a Po(λ) random variable independent of the X_i.

Definition. Given an arbitrary probability distribution Q on the positive integers and a $\lambda > 0$, the *compound Poisson distribution* CP(λ, Q) is the distribution of the sum (1), where the X_i are i.i.d. with common distribution Q and Z is an independent Po(λ) random variable. The parameter λ is called the *rate* of CP(λ, Q) and Q is the *base distribution*.

The natural interpretation of the compound Poisson distribution comes from thinking of events occuring at random times, and in clusters; cf. [5]. For example, we can imagine major earthquakes in San Francisco as occuring at approximately the event times of a Poisson process with rate λ/year, and with each major earthquake followed by smaller quakes, or aftershocks, the number of which is varies according to the distribution Q. Then a simple model for the total number of quakes in t years would be a compound Poisson distribuition with rate λt and base distribution Q.

Thus motivated, it is natural to expect that the compound Poisson should provide a good approximation for P_{S_n} under more general conditions than those in Example 2. Intuitively, the minimal requirements for such an approximation to hold should be that:

1. None of the Y_i dominate the sum, i.e., their parameters $p_i = \Pr\{Y_i \neq 0\}$ are all appropriately small;
2. The Y_i are only weakly dependent.

The main result of the following section, Theorem 1, provides a quantitative version of these requirements, as well as a precise bound on the accuracy of the resulting compound Poisson approximation. It generalizes the corresponding Poisson approximation bound of [6]. In Section III we derive a sharper bound on the approximation of P_{S_n} by a general compound Poisson distribution, in terms of a new version of the Fisher information for discrete random variables. This bound belongs to the family of logarithmic Sobolev inequalities. Finally, in Section IV we use this inequality to derive concentration of

[1] Contact author: Prof. Ioannis Kontoyiannis. *Tel.* 401-863-3044; *Fax.* 401-863-1355.

This work was supported in part by a Sloan Research Fellowship and by NSF grant #0073378-CCR.

measure bounds for the tails of Lipschitz functions with respect to a general class of measures that includes the compound Poisson.

It is perhaps worth mentioning that, although all the results we prove here have known counterparts in the simple Poisson case, it is perhaps somewhat surprising that their compound Poisson versions take such simple and rather elegant forms, and admit relatively straightforward proofs. Moving from the Poisson to the much richer class of compound Poisson measures typically requires considerable effort, as illustrated by the difficulties that have been encountered in the compound Poisson approximation literature; see for example the relevant comments [7].

II. An Information-Theoretic Compound Poisson Approximation Bound

We will measure the closeness between P_{S_n} and an appropriately chosen compound Poisson distribution $\mathrm{CP}(\lambda, \bar{Q})$ in terms of the relative entropy $D(P_{S_n} \| \mathrm{CP}(\lambda, \bar{Q}))$, defined as usual by

$$D(P\|Q) = \sum_{s\in S} P(s) \log \Big[\frac{P(s)}{Q(s)}\Big],$$

for any pair of probability distributions on the same finite or countably infinite set S. [Throughout the paper, log denotes the natural logarithm to base e.] Although not a proper metric, relative entropy is an important measure of closeness between probability distributions [8][9] and it can be used to obtain total variation bounds via Pinsker's inequality [9],

$$\|P - Q\|^2_{\mathrm{TV}} \le 2D(P\|Q). \qquad (2)$$

Our first result is a general, non-asymptotic bound for compound Poisson approximation in relative entropy.

Theorem 1. Suppose S_n is the sum $\sum_{i=1}^n Y_i$ of n possibly dependent random variables Y_i with values in $\{0, 1, 2, \ldots\}$. For each $i = 1, 2, \ldots, n$, let p_i denote the probability $\Pr\{Y_i \neq 0\}$ and Q_i denote the conditional distribution of Y_i given that it is not equal to zero, $Q_i(k) = \Pr\{Y_i = k\}/p_i$, $k \ge 1$. Then,

$$D(P_{S_n} \|\mathrm{CP}(\lambda, \bar{Q})) \le \sum_{i=1}^n p_i^2 + \Big[\sum_{i=1}^n H(Y_i) - H(Y_1, \ldots, Y_n)\Big],$$

where $\lambda = \sum_{i=1}^n p_i$ and $\bar{Q}$ is the mixture distribution

$$\bar{Q} = \sum_{i=1}^n \frac{p_i}{\lambda} Q_i.$$

The entropy of a discrete random variable X with distribution P on some set S is defined as usual,

$$H(P) = H(X) = -\sum_{s\in S} P(s) \log P(s)$$

where log denotes the natural logarithm.

Remark. The first term in the right-hand-side of this bound measures the individual smallness of the Y_i, as demanded by requirement 1. above, and the second term measures their degree of dependence: It is the difference between the sum of their individual entropies and their joint entropy, which of course is zero iff the Y_i are independent.

Despite the great generality of its statement, the proof of Theorem 1 is entirely elementary and relies only a couple of basic information-theoretic properties of relative entropy and some simple calculus.

Proof of Theorem 1. Let $Z_1, Z_2, \ldots, Z_n$ be independent compound Poisson random variables with each $Z_i \sim \mathrm{CP}(p_i, Q_i)$, and write $T_n = \sum_{i=1}^n Z_i$ for their sum. By the basic infinite divisibility property of the compound Poisson law, the distribution of T_n is $\mathrm{CP}(\lambda, \bar{Q})$. By the data processing inequality for relative entropy we have,

$$\begin{aligned} D(P_{S_n} \|\mathrm{CP}(\lambda, \bar{Q})) &= D(P_{S_n} \| P_{T_n}) \\ &\le D(P_{Y_1,\ldots,Y_n} \| P_{Z_1,\ldots,Z_n}), \end{aligned}$$

where $P_{Y_1,\ldots,Y_n}$ denotes the joint distribution of the Y_i and similarly for $P_{Z_1,\ldots,Z_n}$. Applying the chain rule for relative entropy gives,

$$\begin{aligned} &D(P_{S_n} \|\mathrm{CP}(\lambda, \bar{Q})) \\ &\quad \le \sum_{i=1}^n D(P_{Y_i} \| P_{Z_i}) + \sum_{i=1}^n H(Y_i) - H(Y_1, \ldots, Y_n), \end{aligned}$$

where P_{Y_i} denotes the marginal distribution of Y_i and similarly for P_{Z_i}. Combining this with the simple estimate of the following lemma completes the proof. □

Lemma. For each $i = 1, 2, \ldots, n$,

$$D(P_{Y_i} \| P_{Z_i}) \le p_i^2.$$

Proof of Lemma. The first term in the relative entropy

$$D(P_{Y_i} \| P_{Z_i}) = \sum_{j=0}^{\infty} P_{Y_i}(j) \log \frac{P_{Y_i}(j)}{P_{Z_i}(j)}$$

corresponding to $j = 0$ is

$$(1 - p_i) \log \frac{1 - p_i}{e^{-p_i}} = (1 - p_i) \log(1 - p_i) + p_i(1 - p_i) \le 0,$$

where the inequality follows from the fact that $\log(1-p) \le -p$ for $p < 1$. To estimate the jth term, let Q_i^{*k} denote the k-fold convolution of Q_i with itself, namely, the law of the sum of k i.i.d. random variables with common distribution Q_i. Then,

$$P_{Z_i}(j) = \sum_{k=1}^{\infty} e^{-p_i} \frac{p_i^k}{k!} Q_i^{*k}(j) \ge e^{-p_i} p_i Q_i(j),$$

where the inequality follows from taking only the first term in the series. Therefore, the jth term of $D(P_{Y_i} \| P_{Z_i})$ is

$$P_{Y_i}(j) \log \frac{P_{Y_i}(j)}{P_{Z_i}(j)} \le [p_i Q_i(j)] \log \frac{p_i Q_i(j)}{e^{-p_i} p_i Q_i(j)} = Q_i(j) p_i^2.$$

Summing over j yields the result. □

Next we would like to examine how accurate an approximation the bound in Theorem 1 provides. Coming back to the basic setting in Example 2, let $S_n = \sum_{i=1}^n Y_i = \sum_{i=1}^n B_i X_i$, where the X_i are i.i.d. with common distribution Q and the B_i are i.i.d. Bernoulli(λ/n) for some $\lambda > 0$. In the notation of Theorem 1 we have $Q_i = \bar{Q} = Q$ and $p_i = \lambda/n$, for all $i = 1, 2, \ldots, n$. Therefore,

$$D(P_{S_n} \|\mathrm{CP}(\lambda, Q)) \le \sum_{i=1}^n (\lambda/n)^2 = \frac{\lambda^2}{n},$$

and by Pinsker's inequality (2),

$$\|P_{S_n} - \mathrm{CP}(\lambda, Q)\|_{\mathrm{TV}} \leq \sqrt{2}\frac{\lambda}{\sqrt{n}}. \tag{3}$$

This offers an explicit error term for the approximation and it is certainly sufficient to prove the convergence (both in total variation and in the stronger sense of relative entropy) of P_{S_n} to $\mathrm{CP}(\lambda, Q)$. But compared to the well-known bound of Le Cam [10],

$$\|P_{S_n} - \mathrm{CP}(\lambda, Q)\|_{\mathrm{TV}} \leq \lambda^2/n,$$

we see that (3) gives a suboptimal rate of convergence.

More generally, applying Theorem 1 to a variety of different examples, including ones with non-trivial dependence in the Y_i, we typically find that the bounds we obtain are strong enough to prove convergence, but their rate is often suboptimal. In order to get sharper rates, at least in the case when the Y_i are independent, we take a hint from [6] and turn to the notion of Fisher information.

III. Fisher Information and a Logarithmic Sobolev Inequality

As discussed in [6], there is no universal way to define the Fisher information of an arbitrary discrete random variable. For our purposes, the following expression turns out to provide a natural candidate.

Definition. Given any $\lambda > 0$ and an arbitrary base distribution Q, we define the *CP-Fisher information* of a random variable X with distribution P on $\{0, 1, 2, \ldots\}$ as,

$$J_{\lambda,Q}(X) = \lambda \sum_{j=1}^{\infty} Q(j) E\Big[\Big(\frac{P(X+j)}{P(X)}\frac{C_{\lambda,Q}(X)}{C_{\lambda,Q}(X+j)} - 1\Big)^2\Big]$$

where $C_{\lambda,Q}$ denote the $\mathrm{CP}(\lambda, Q)$ distribution.

For any random variable X, it is obvious from the definition that $J_{\lambda,Q}(X) = 0$ iff $X \sim \mathrm{CP}(\lambda, Q)$. Our next result says that, moreover, the smaller the value of the CP-Fisher information $J_{\lambda,Q}(X)$, the closer X is to having a $\mathrm{CP}(\lambda, Q)$ distribution.

Theorem 2. Let $X \sim P$ be a nonnegative integer-valued random variable. For any $\lambda > 0$ and any base distribution Q,

$$D(P\|\mathrm{CP}(\lambda, Q)) \leq J_{\lambda,Q}(X),$$

as long as either X has full support (i.e., $P(k) > 0$ for all k) or finite support (i.e., there exists an $M > 0$ such that $P(k) = 0$ for all $k \geq M$).

This inequality can be thought of as a natural logarithmic Sobolev inequality for the compound Poisson measure. Indeed, its proof uses an earlier logarithmic Sobolev inequality by Bobkov and Ledoux [11] for the Poisson measure.

Proof Outline. A simple exercise using characteristic functions shows that an alternative representation for $\mathrm{CP}(\lambda, Q)$ to that given by (1) in its definition above, is as a series

$$\sum_{j=1}^{\infty} jZ_j, \tag{4}$$

where the Z_j are independent Poisson random variables with each $Z_j \sim \mathrm{Po}(\lambda Q(j))$. The starting point for our proof is the logarithmic Sobolev inequality for a Poisson distribution proved by Bobkov and Ledoux in [11], stating that for any P on $\{0, 1, \ldots\}$ and any $\lambda > 0$,

$$D(P\|\mathrm{Po}(\lambda)) \leq \lambda E\Big[\Big(\frac{(X+1)P(X+1)}{\lambda P(X)} - 1\Big)^2\Big].$$

Also see [6] for details.

Step 1. We "tensorize" the Bobkov-Ledoux result to obtain a corresponding inequality for any finite product of (not necessarily identical) Poisson distributions.

Step 2. Since Step 1 holds for the product distribution, it also holds for any deterministic transformation of that product, so we can apply it to the distribution of the finite sum $T_n = \sum_{j=1}^n jZ_j$, where the Z_j are as above. This yields a logarithmic Sobolev inequality for the distribution of T_n.

Step 3. Taking the limit as $n \to \infty$ and using the representation in (4) we obtain the required result. □

An easy corollary is the following equivalent (although seemingly more general) version of the logarithmic Sobolev inequality for $\mathrm{CP}(\lambda, Q)$. First we need to extend the definition of relative entropy. For any random variable X with distribution P on $\{0, 1, 2, \ldots\}$ and any function f on $\{0, 1, 2, \ldots\}$, we define,

$$D(f\|P) = E[f(X)\log f(X)] - E[f(X)]\log E[f(X)],$$

whenever all the expectations make sense.

Corollary 1. For any $\lambda > 0$, any base distribution Q and any function f on $\{0, 1, 2, \ldots\}$ with $f(k) > 0$ for all k, we have,

$$D(f\|\mathrm{CP}(\lambda, Q)) \leq \lambda \sum_{j=1}^{\infty} Q(j) E\Big[f(Z)\Big(\frac{f(Z+j)}{f(Z)} - 1\Big)^2\Big],$$

where $Z \sim \mathrm{CP}(\lambda, Q)$.

From the point of view of compound Poisson approximation, the obvious next step would be to examine the properties of the CP-Fisher information $J_{\lambda,Q}(S_n)$ of *sums* S_n of independent random variables, similar to the development in [6] in the Poisson case. Here, however, we pursue a different direction. In the next section we show how the result of Theorem 2 can be used to obtain concentration properties of Lipschitz functions on $\{0, 1, \ldots\}$.

IV. Measure Concentration

Logarithmic Sobolev inequalities like the one obtained in Theorem 2 are well-known to be intimately connected to concentration properties of Lipschitz functions. In unpublished work in the 1970s, Herbst used Gross' [12] celebrated logarithmic Sobolev inequality for the Gaussian measure to derive concentration properties of Lipschitz functions on $\mathbb{R}^n$. Herbst's proof has been generalized and adapted by many authors since then; see, e.g., [13][14]. Our development follows closely along the lines of the corresponding discussion of Herbst's argument in [11] and [15].

In the following result we give a precise description of the tails of any discrete-Lipschitz function, with respect to the compound Poisson measure or any other measure which satisfies a similar logarithmic Sobolev inequality.

Theorem 3. Let Q be a given base distribution with finite support $\{1, 2, \dots, m\}$. Suppose that for some probability measure μ on $\{0, 1, 2, \dots\}$ there is a fixed, finite constant $C > 0$ such that,

$$D(f\|\mu) \leq C \sum_{j=1}^{m} Q(j) E\Big[f(Z)\Big(\frac{f(Z+j)}{f(Z)} - 1\Big)^2\Big], \quad Z \sim \mu,$$

for every function f on $\{0, 1, \dots\}$ with strictly positive values. Then, for any function g that satisfies the Lipschitz condition

$$\sup_{x \in \{0,1,2,\dots\}} |Dg(x)| \leq 1\,, \tag{5}$$

we have $E_\mu(|g|) < \infty$ and the following tail estimates hold:

(a) For all $t > 0$,

$$\mu\{g \geq E_\mu(g) + t\} \leq \exp\Big\{ - \frac{t}{4m} \log\Big(1 + \frac{t}{m^2 C}\Big)\Big\}$$

(b) For $0 < t \leq 2VC/m$,

$$\mu\{g \geq E_\mu(g) + t\} \leq \exp\Big\{ - \frac{t^2}{8VC}\Big\}.$$

where $V = \sum_{j=1}^{m} j^2 Q(j)$ is the second moment of Q.

Remarks.

(i) The proof of the theorem is somewhat technical, but the main gist of the argument is as follows. First we apply the logarithmic Sobolev inequality (assumed in the theorem) to the function $f = e^{\alpha g}$, where g is assumed to satisfy an appropriate Lipschitz condition. Expanding, we get a differential inequality for the function $L(\alpha) = E_\mu[e^{\alpha g}]$, from which we can deduce an exponential upper bound on $E_\mu[e^{\alpha g}]$. The rest follows by a simple application of Chebychev's inequality.

(ii) Although all the constants in the statement of the theorem are explicit and take a rather simple form, there is no reason to expect that they are optimal.

(iii) From the two bounds in Theorem 3 we see that the tails of any Lipschitz function with respect to a compound Poisson measure with finitely supported base distribution (or any other measure satisfying the analogous logarithmic Sobolev inequality) are Gaussian near the mean and Poisson-like away from the mean. In particular, the following integrability result is an immediate consequence of the bound in (b) above.

Corollary 2. Let Q be a given base distribution with finite support $\{1, 2, \dots, m\}$, and suppose that the probability measure μ satisfies the logarithmic Sobolev inequality in the statement of Theorem 3. Then for any function g that satisfies the Lipschitz condition (5) we have,

$$E_\mu[e^{\theta|g| \log^+ |g|}] < \infty,$$

for all $\theta > 0$ small enough, where $\log^+ x = \max\{\log x, 0\}$.

The assumption that the base distribution Q has finite support is quite restrictive, and was made primarily for technical convenience. It can be replaced by much weaker conditions on the form of Q. In that case, not only the assumptions, but also the form of the bound is somewhat different.

In this context, it is useful to recall the following simple fact about the compound Poisson distribution: For a wide range of distributions Q, the tail of the compound Poisson distribution $\mathrm{CP}(\lambda, Q)$ decays at roughly the same rate as the tail Q. This holds as long as Q has infinite support and a polynomial or exponential tail, but not when it has a super-exponential tail; see, e.g., [16] for details.

Thus it would seem reasonable to expect that this result about the tail of $\mathrm{CP}(\lambda, Q)$ can be extended to a result about the tails of Lipschitz functions with respect to $\mathrm{CP}(\lambda, Q)$. Indeed, we are able to establish the following result in this regard.

Theorem 4. Let Q be a given base distribution, and assume there exists a positive δ such that

$$2\delta < \sup\Big\{\alpha \in \mathbb{R} : \sum_j [Q(j) e^{\alpha j}] < \infty\Big\}.$$

If for some probability measure μ on $\{0, 1, 2, \dots\}$ there is a fixed, finite constant $C > 0$ such that

$$D(f\|\mu) \leq C \sum_{j=1}^{\infty} Q(j) E\Big[f(Z)\Big(\frac{f(Z+j)}{f(Z)} - 1\Big)^2\Big], \quad Z \sim \mu,$$

for every function f on $\{0, 1, \dots\}$ with strictly positive values, Then for any function $g > 0$ that satisfies the Lipschitz condition (5), we have $E_\mu(g) < \infty$ and the following tail estimate holds for all $t > 0$:

$$\mu(\{g \geq E_\mu g + t\}) \leq C' \exp\{-t\delta\}$$

where

$$C' = \exp\Big\{\frac{C\delta}{2} \sum_{j=1}^{\infty} j Q_j (e^{2j\delta} - 1)\Big\}.$$

On the other hand, when the tail of Q has superexponential decay, as is trivially the case when Q has finite support, the tail of the tail of $\mathrm{CP}(\lambda, Q)$ is Poisson-like (see, for e.g., [16]). It is this result that Theorem 3 generalizes from the tail of the compound Poisson distribution to the tails of Lipschitz functions under a certain class of measures satisfying a logarithmic Sobolev inequality.

References

[1] A. Barbour, L. Holst, and S. Janson, *Poisson Approximation.* New York: The Clarendon Press Oxford University Press, 1992.

[2] A. Barbour, O. Chryssaphinou, and E. Vaggelatou, "Applications of compound poisson approximation," In *Statistical Distribution Theory and Inference* (N. Balakrishnan, M. V. Koutras and C. Charalambides, eds.), CRC Press, Boca Raton, FL., 2000.

[3] A. Barbour and O. Chryssaphinou, "Compound Poisson approximation: a user's guide," *Ann. Appl. Probab.*, vol. 11, no. 3, pp. 964–1002, 2001.

[4] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. I.* New York, N.Y.: John Wiley & Sons Inc., 1950.

[5] D. Aldous, *Probability approximations via the Poisson clumping heuristic.* New York: Springer-Verlag, 1989.

[6] I. Kontoyiannis, P. Harremoës, and O. Johnson, "Entropy and the law of small numbers," *Preprint*, November 2003, [Available online at www.dam.brown.edu/people/yiannis].

[7] B. Roos, "Kerstan's method for compound Poisson approximation," *Ann. Probab.*, vol. 31, no. 4, pp. 1754–1771, 2003.

[8] T. Cover and J. Thomas, *Elements of Information Theory.* New York: J. Wiley, 1991.

[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press, 1981.

[10] L. Le Cam, "On the distribution of sums of independent random variables," in *Proc. Internat. Res. Sem., Statist. Lab., Univ. California, Berkeley, Calif.* New York: Springer-Verlag, 1965, pp. 179–202.

[11] S. Bobkov and M. Ledoux, "On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures," *J. Funct. Anal.*, vol. 156, no. 2, pp. 347–365, 1998.

[12] L. Gross, "Logarithmic Sobolev inequalities," *Amer. J. Math.*, vol. 97, no. 4, pp. 1061–1083, 1975.

[13] S. Aida and D. Stroock, "Moment estimates derived from Poincaré and logarithmic Sobolev inequalities," *Math. Res. Lett.*, vol. 1, no. 1, 1994.

[14] L. Gross and O. Rothaus, "Herbst inequalities for supercontractive semigroups," *J. Math. Kyoto Univ.*, vol. 38, no. 2, pp. 295–318, 1998.

[15] M. Ledoux, *The concentration of measure phenomenon.* Providence, RI: American Mathematical Society, 2001.

[16] K. -I. Sato, *Levy processes and infinitely divisible distributions.* Cambridge, UK: Cambridge University Press, 1999.