Elect. Comm. in Probab. **11** (2006), 45–57

ELECTRONIC COMMUNICATIONS in PROBABILITY

MEASURE CONCENTRATION FOR COMPOUND POISSON DISTRIBUTIONS

I. KONTOYIANNIS[1]
Division of Applied Mathematics, Brown University, 182 George Street, Providence, RI 02912, USA.
email: `yiannis@aueb.gr`

M. MADIMAN
Division of Applied Mathematics, Brown University, 182 George Street, Providence, RI 02912, USA.
email: `mokshay.madiman@yale.edu`

Submitted 19 June 2005, *accepted in final form* 4 May 2006

AMS 2000 Subject classification: 60E07, 60E15, 46N30, 39B62
Keywords: Compound Poisson measure, measure concentration, entropy method, logarithmic-Sobolev inequality, polynomial tails, Herbst argument

Abstract

We give a simple development of the concentration properties of compound Poisson measures on the nonnegative integers. A new modification of the Herbst argument is applied to an appropriate modified logarithmic-Sobolev inequality to derive new concentration bounds. When the measure of interest does not have finite exponential moments, these bounds exhibit optimal *polynomial* decay. Simple new proofs are also given for earlier results of Houdré (2002) and Wu (2000).

1 Introduction

Concentration of measure is a well-studied phenomenon, and in the past 30 years or so it has been explored through a wide array of tools and techniques; [14][11][12] offer broad introductions. Results in this area are equally well motivated by theoretical questions (in areas such as geometry, functional analysis and probability), as by numerous applications in different fields including the analysis of algorithms, mathematical physics and empirical processes in statistics.

From the probabilistic point of view, measure concentration describes situations where a random variable is strongly concentrated around a particular value. This is typically quantified by the rate of decay of the probability that the random variable deviates from that value (usually its mean or median) by a certain amount. As a simple concrete example consider a function

[1]SUPPORTED IN PART BY A SLOAN FOUNDATION RESEARCH FELLOWSHIP AND BY NSF GRANT #0073378-CCR.

$f(W)$ of a Poisson(λ) random variable W; if $f:\mathbb{Z}_+\to\mathbb{R}$ is 1-Lipschitz, i.e., $|f(k)-f(k+1)|\le 1$ for all $k\in\mathbb{Z}_+=\{0,1,2,\ldots\}$, then [2],

$$\Pr\big\{f(W)-E[f(W)]>t\big\}\le\exp\Big\{-\frac{t}{4}\log\Big(1+\frac{t}{2\lambda}\Big)\Big\}. \qquad (1)$$

Although the distribution of $f(W)$ may be quite complex, (1) provides a simple, explicit bound on the probability that it deviates from its mean by an amount t. This is a general theme: Under appropriate conditions, it is possible to derive useful, accurate bounds of this type for a large class of random variables with complex and often only partially known distributions. We also note that the consideration of Lipschitz functions is motivated by applications, but it is also related to fundamental concentration properties captured by isoperimetric inequalities [11].

The bound (1) was established in [2] using the so-called "entropy method," pioneered by Ledoux [9][10][11]. The entropy method consists of two steps. First, a (possibly modified) logarithmic-Sobolev inequality is established for the distribution of interest. Recall that, for an arbitrary probability measure μ and any nonnegative function f on the same space, the entropy functional $\mathrm{Ent}_\mu(f)$ is defined by

$$\mathrm{Ent}_\mu(f)=\textstyle\int f\log f\,d\mu-(\int f\,d\mu)\log(\int f\,d\mu),$$

whenever all the above integrals exist. In the case of the Poisson, Bobkov and Ledoux [2] proved the following modified log-Sobolev inequality: Writing P_λ for the Poisson(λ) measure, for any function $f:\mathbb{Z}_+\to\mathbb{R}$ with positive values,

$$\mathrm{Ent}_{\mathrm{P}_\lambda}(f)\le\lambda E_{\mathrm{P}_\lambda}\Big[\frac{1}{f}|Df|^2\Big],$$

where $Df(k)=f(k+1)-f(k)$, $k\ge 0$, is the discrete gradient, and E_μ denotes the expectation operator with respect to a measure μ. In fact, they also established the following sharper bound which we will use below; for any function f on $\mathbb{Z}_+$,

$$\mathrm{Ent}_{\mathrm{P}_\lambda}\big(e^f\big)\le\lambda E_{\mathrm{P}_\lambda}\Big[e^f\big\{|Df|e^{|Df|}-e^{|Df|}+1\big\}\Big]. \qquad (2)$$

The second step in the entropy method is the so-called Herbst argument: Starting from some Lipschitz function f, the idea is to use the modified log-Sobolev inequality to obtain an upper bound on the entropy of $e^{\tau f}$, and from that to deduce a differential inequality for the moment-generating function $F(\tau)=E[e^{\tau f}]$ of f. Then, solving the differential inequality yields an upper bound on $G(\tau)$, and this leads to a concentration bound via Markov's inequality.

Our main goal in this work is to carry out a similar program for an arbitrary *compound Poisson* measure on $\mathbb{Z}_+$. Recall that for any $\lambda>0$ and any probability measure Q on the natural numbers $\mathbb{N}=\{1,2,\ldots\}$, the compound Poisson distribution CP(λ,Q) is the distribution of the random sum $Z\stackrel{\mathcal{D}}{=}\sum_{i=1}^W X_i$, where $W\sim$ Poisson(λ) and the X_i are independent random variables with distribution Q on $\mathbb{N}$, also independent of W; we denote the CP(λ,Q) measure by $\mathrm{CP}_{\lambda,Q}$. The class of compound Poisson distributions is much richer than the one-dimensional Poisson family. In particular, the CP(λ,Q) law inherits its tail behavior from Q: CP(λ,Q) has finite variance iff Q does, it has exponentially decaying tails iff Q does, and so on [13]. It is in part from this versatility of tail behavior that the compound Poisson distribution draws its importance in many applications. Alternatively, CP(λ,Q) is characterized as the infinite divisible law without a Gaussian component and with Lévy measure λQ.

From the above discussion we observe that the Herbst argument is heavily dependent on the use of moment-generating functions, which assumes the existence of exponential moments. Our main technical contribution is a modification of the Herbst argument for the case when the random variables of interest do *not* satisfy such exponential integrability conditions. We derive what appear to be perhaps the first concentration inequalities for a class of infinitely divisible random variables that have finite variance but do not have finite exponential moments. Apart from the derivation of the present results, the modified Herbst argument is applicable in a variety of other cases and may be of independent interest. In particular, this approach can be applied to prove analogous inequalities for compound Poisson vectors, as well as concentration bounds for more general infinitely divisible laws.
Our starting point is the following modified log-Sobolev inequality for the compound Poisson measure $\mathrm{CP}_{\lambda,Q}$.

Theorem 1. [Modified Log-Sobolev Inequality for Compound Poisson Measures] For any $\lambda > 0$, any probability measure Q on $\mathbb{N}$ and any bounded $f : \mathbb{Z}_+ \to \mathbb{R}$,

$$\mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}\left(e^f\right) \le \lambda \sum_{j\ge 1} Q_j E_{\mathrm{CP}_{\lambda,Q}}\Big[e^f\big\{|D^j f|e^{|D^j f|} - e^{|D^j f|} + 1\big\}\Big], \tag{3}$$

where $D^j f(k) = f(k+j) - f(k)$, for $j, k \in \mathbb{Z}_+$.

This can be derived easily from [15, Cor 4.2] of Wu, which was established using elaborate stochastic calculus techniques. In Section 3 we also give an alternative, elementary proof, by tensorizing the Bobkov-Ledoux result (2). Note the elegant similarity between the bounds in (2) and (3).
We then apply the Herbst argument (both the classical form and our modified version) to establish concentration bounds for $\mathrm{CP}(\lambda, Q)$ measures under various assumptions on the tail behavior of Q. These are stated in Section 2 and proved in Section 4. For example, we establish the following polynomial concentration result. Recall that a function $f : \mathbb{Z}_+ \to \mathbb{R}$ is K-Lipschitz if $|f(j+1) - f(j)| \le K$, for all $j \in \mathbb{Z}_+$.

Corollary 2. [Polynomial Concentration] Suppose that Z has $\mathrm{CP}(\lambda, Q)$ distribution where Q has finite moments up to order L,

$$L = \sup\left\{\tau \ge 1 \,:\, \textstyle\sum_{j\ge 1} j^\tau Q_j < \infty\right\} > 1,$$

and write q_r for its integer moments,

$$q_r = \textstyle\sum_{j\ge 1} j^r Q_j, \quad r \in \mathbb{N}.$$

If $f : \mathbb{Z}_+ \to \mathbb{R}$ is K-Lipschitz, then for any positive integer $n < L$ and any $t > 0$ we have,

$$\Pr\big\{|f(Z) - E[f(Z)]| > t\big\} \le A \cdot B^n \cdot t^{-n},$$

where for the constants A, B we can take,

$$\begin{aligned} A &= \exp\Big\{\lambda \sum_{r=1}^{n} \binom{n}{r} q_r\Big\} \\ B &= 2|f(0)| + 2K\lambda q_1 + 1. \end{aligned}$$

Various stronger and more general results are given in Section 2. There, at the price of more complex constants, we get bounds which, for large t, are of (the optimal) order $t^{-L+\delta}$ for any $\delta > 0$. Note that although the bound of Corollary 2 is not useful for small t, it is in general impossible to obtain meaningful results for arbitrary $t > 0$. For example, if f is the identity function and $Z \sim \text{Poisson}(\lambda)$ where λ is of the form $m + 1/2$ for an integer m, then $|Z - E(Z)| \geq 1/2$ with probability 1; a more detailed discussion is given in Section 2.

As noted above, the results of Section 2 appear to be some of the first *non-exponential* concentration bounds that have been derived, with the few recent exceptions discussed next. Of the extensive current literature on concentration, our results are most closely related to the work of Houdré and his co-authors. Using sophisticated technical tools derived from the "covariance representations" developed in [7][8], Houdré [5] obtained concentration bounds for Lipschitz functions of infinitely divisible random vectors with finite *exponential* moments. In [6], truncation and explicit computations were used to extend these results to the class of stable laws on $\mathbb{R}^d$, and the preprint [4] extends them further to a large class of functionals on Poisson space. To our knowledge, the results in [6][4] are the only concentration bounds with power-law decay to date. But when specialized to scalar random variables they only apply to distributions with infinite variance, whereas our results hold for compound Poisson random variables with a finite Lth moment for any $L > 1$. Although the methods of [6][4] as well as the form of the results themselves are very different from those derived here, more detailed comparisons are possible as outlined in Section 2. Finally, the recent paper [3] contains a different extension of the Herbst argument to cases where exponential moments do not exist. The focus there is on *moment* inequalities for functions of independent random variables, primarily motivated by statistical applications.

2 Concentration Bounds

The following result illustrates the potential for using the Herbst argument even in cases where the existence of exponential moments fails or cannot be assumed.

Theorem 3. [Power-law Concentration] Suppose that Z has $\text{CP}(\lambda, Q)$ distribution where Q has finite moments up to order L,

$$L = \sup\left\{\tau \geq 1 \,:\, \textstyle\sum_{j\geq 1} j^\tau Q_j < \infty\right\} > 1,$$

and write $q_1 = \sum_{j\geq 1} j\, Q_j$ for its first moment.

(i) If $f : \mathbb{Z}_+ \to \mathbb{R}$ is K-Lipschitz, then for any $t > 0$ and $\epsilon > 0$ we have,

$$\Pr\left\{|f(Z) - Ef(Z)| > t\right\} \leq \exp\left\{\inf_{0<\alpha<L}\left[I_\epsilon(\alpha) + \alpha \log\left(\frac{2|f(0)| + 2K\lambda q_1 + \epsilon}{t}\right)\right]\right\}, \quad (4)$$

where

$$\begin{aligned} I_\epsilon(\alpha) &= \lambda \sum_{j\geq 1} Q_j \left\{C_{j,\epsilon}^\alpha - 1 - \alpha \log C_{j,\epsilon}\right\} \\ C_{j,\epsilon} &= 1 + \frac{jK}{\epsilon}. \end{aligned}$$

(ii) The upper bound (4) is meaningful (less than 1) iff $t > T := 2|f(0)| + 2K\lambda q_1 + \epsilon$, and then,

$$\Pr\left\{|f(Z) - Ef(Z)| > t\right\} \leq \exp\left\{-\int_0^{\log(t/T)} i_\epsilon^{-1}(s)ds\right\}$$

where $i_\epsilon(\alpha) := I'_\epsilon(\alpha) = \lambda \sum_{j\geq 1} Q_j[C^{\alpha}_{j,\epsilon} - 1]\log C_{j,\epsilon}$.

Remarks.
1. Taking $\alpha = L - \delta$ for any $\delta > 0$ in the exponent of (4), we get a bound on the tails of $f(Z)$ of order $t^{-(L-\delta)}$ for large t. By considering the case where f is the identity function $f(k) = k$, $k \in \mathbb{Z}_+$, we see that this power-law behavior is in fact optimal. In particular, this shows that the tail of the CP(λ, Q) law decays like the tail of Q, giving a quantitative version of a classical result from [13].
2. As will become evident from the proof, Theorem 3 holds for any random variable Z with law μ instead of $\mathrm{CP}_{\lambda,Q}$, as long as μ satisfies the log-Sobolev inequality of Theorem 1 with respect to some probability measure Q on $\mathbb{N}$ and some $\lambda > 0$, and assuming that μ has finite moments up to order L. The bound (4) remains exactly the same, except that the first moment $M_1 = E[Z]$ of μ replaces λq_1.
3. Integrability properties follow immediately from the theorem: For any K-Lipschitz function f, $E_{\mathrm{CP}_{\lambda,Q}}[|f|^\tau] < \infty$ for all $\tau < L$, and the same holds for any law μ as in the previous remark.
4. Since the support of CP(λ, Q) is $\mathbb{Z}_+$, we would expect the range of f to be highly disconnected. Therefore, it is natural to consider the (typical) class of functions $f : \mathbb{Z}_+ \to \mathbb{R}$ whose mean under $\mathrm{CP}_{\lambda,Q}$ is not in the range of f. Indeed, if we assume that $|f(j) - E[f(Z)]| \geq \epsilon > 0$, for all j, then, under the same assumptions as in Theorem 3, the exponent in the bound (4) simplifies to $\inf_{0<\alpha<L}[I_\epsilon(\alpha) + \alpha \log(D/t)\Big]$, where $D := E|f(Z) - E[f(Z)]|$. obvious that any bound can only be useful for $t > D_0 := \min_{k\in\mathbb{Z}_+} |f(k) - E[f(Z)]|$, since the probability that $|f(Z) - E[f(Z)]| \geq D_0$ is equal to 1. Moreover, D and D_0 coincide in many special cases, e.g., when the range of f is a lattice in $\mathbb{R}$ and its mean $E[f(Z)]$ is on the midpoint between two lattice points. In this sense, the restriction $t > D$ is quite natural.
6. The expression $2|f(0)| + 2K\lambda q_1$ in Theorem 3 is simply an upper bound to the constant $D = E|f(Z) - E[f(Z)]|$. In both cases, when $L > 2$ it is possible to obtain potentially sharper results by bounding D above using Jensen's inequality by, $\big[\{K^2\lambda q_2 + \{|f(0)| + K\lambda q_1\}^2\big]^{\frac{1}{2}}$, where q_2 is the second moment of Q. Similar expressions can be derived in the case of higher moments.
7. The most closely related results to our power-law concentration bounds appear to be in the recent preprint [4].[2] The main assumptions in [4] are that the random variable of interest has infinite variance, and also certain growth conditions.[3] Because of the infinite-variance assumption, the majority of the results in this paper (corresponding to $L > 2$) apply to cases that are not covered in [4]. As for the growth conditions, they are convenient to check in several important special classes, e.g., for α-stable laws on $\mathbb{R}$, but they can be unwieldy in the compound Poisson case, especially as they depend on Q in an intricate way. On the other hand, if Q has infinite variance, [4, Cor. 5.3] gives optimal-order bounds, including the case when Q has infinite mean, for which our results do not apply.

Next we show how the Herbst argument can be used to recover precisely a result of [5] in the case when we have exponential moments.

[2] The results in [4] are stated in the much more general setting of functionals on an abstract Poisson space. Using the Wiener-Ito decomposition, any infinitely divisible random variable can be represented as a Poisson stochastic integral, which in turn can be realized as a “nice” functional on Poisson space.

[3] Actually, the relevant bounds in [4] specialized to Lipschitz functions of CP(λ, Q) random variables require that the probability measure Q be *non-atomic*, which excludes all the cases we consider. However, shortly after the first writing of this paper, C. Houdré in a personal communication informed us that this assumption can be removed by an appropriate construction. Although we did not check every detail of this generalization for ourselves, in our comparison we assume that the results in [4] remain valid (and their statements unchanged) even if Q is atomic.

Theorem 4. [Exponential Concentration] [5] Suppose that Z has CP(λ, Q) distribution where Q has finite exponential moments up to order M,

$$M = \sup\left\{\tau \geq 0 \;:\; \textstyle\sum_{j\geq 1} e^{\tau j}\, Q_j < \infty\right\} > 0.$$

If $f : \mathbb{Z}_+ \to \mathbb{R}$ is K-Lipschitz, then for any $t > 0$ we have,

$$\Pr\left\{f(Z) - Ef(Z) > t\right\} \;\leq\; \exp\left\{\inf_{0<\alpha<M/K}[H(\alpha) - \alpha t]\right\} \;=\; \exp\left\{-\int_0^t h^{-1}(s)\, ds\right\}, \tag{5}$$

where $H(\alpha) := \lambda \sum_{j\geq 1} Q_j[e^{\alpha K j} - 1 - \alpha K j]$, and h^{-1} is the inverse of $h(\alpha) := H'(\alpha)$.

Remarks.
8. Theorem 1 of [5] gives concentration bounds for a class of infinitely divisible laws with finite exponential moments, and in the compound Poisson case it reduces precisely to (5), which also applies to any random variable Z whose law satisfies the result of Theorem 1. It is also interesting to note that Theorem 4 can be derived by applying [15, Prop 3.2] to a compound Poisson random variable (constructed via the Wiener-Ito decomposition), and then using Markov's inequality.
9. Theorems 3 and 4 easily generalize to Hölder continuous functions. In the discrete setting of $\mathbb{Z}_+$, f is K-Lipschitz iff it is Hölder continuous for every exponent $\beta \geq 1$ with the same constant K. But if f is Hölder continuous with exponent $\beta < 1$, this more stringent requirement makes it possible to strengthen Theorem 3 and Theorem 4, by respectively redefining, $C_{j,\epsilon} = 1 + \frac{j^\beta K}{\epsilon}$, and

$$H(\alpha) = \lambda \sum_{j\geq 1} Q_j\left[e^{\alpha K j^\beta} - 1 - \alpha K j^\beta\right].$$

10. While all our power-law results dealt with two-sided deviations, the bound in Theorem 4 is one-sided. The reason for this discrepancy is that the last step in all the relevant proofs is an application of Markov's inequality, which leads us to restrict attention to nonnegative random variables. When exponential moments exist, the natural consideration of the exponential of the random variable takes care of this issue, but in the case of regular moments we are forced to take absolute values.

3 Proof of Theorem 1

An alternative representation for the law of a CP(λ, Q) random variable Z is in terms of the series

$$Z \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\infty} j\, Y_j, \qquad Y_j \sim \text{Poisson}(\lambda Q_j), \tag{6}$$

where the Y_j are independent.
For each n, let μ_n denote the joint (product) distribution of $(Y_1, \ldots, Y_n)$. In this instance, the tensorization property of the entropy [1][10][11] can be expressed as

$$\mathrm{Ent}_{\mu_n}(G) \leq \sum_{j=1}^{n} E\Big[\mathrm{Ent}_{\mathrm{P}_{\lambda Q_j}}\big(G_j(Y_1, \ldots, Y_{j-1}, \cdot, Y_{j+1}, \ldots, Y_n)\big)\Big], \tag{7}$$

where $G : \mathbb{Z}_+^n \to \mathbb{R}_+$ is an arbitrary function, and the entropy on the right-hand side is applied to the restriction G_j of G to its jth co-ordinate. Now given an f as in the statement of the theorem, define the functions $G : \mathbb{Z}_+^n \to \mathbb{R}_+$ and $H : \mathbb{Z}_+^n \to \mathbb{R}_+$ by

$$H(y_1, \dots, y_n) = f\Big(\sum_{k=1}^{n} k y_k\Big), \quad y_1^n \in \mathbb{Z}_+^n,$$

and $G = e^H$. Let $\bar{\mu}_n$ denote the distribution of the sum $S_n := \sum_{k=1}^n kY_k$ and write $H_j : \mathbb{Z}_+ \to \mathbb{R}$ for the restriction of H to the variable y_j with the remaining y_i's fixed. Applying (7) to G we obtain,

$$\begin{aligned} \mathrm{Ent}_{\bar{\mu}_n}(e^f) \;=\; \mathrm{Ent}_{\mu_n}(G) \;&\leq\; \sum_{j=1}^{n} E\Big[\mathrm{Ent}_{\mathrm{P}_{\lambda Q_j}}\Big(G_j(Y_1, \dots, Y_{j-1}, \cdot, Y_{j+1}, \dots, Y_n)\Big)\Big] \\ &=\; \sum_{j=1}^{n} E\Big[\mathrm{Ent}_{\mathrm{P}_{\lambda Q_j}}\big(e^{H_j(Y_1,\dots,Y_{j-1},\cdot,Y_{j+1},\dots,Y_n)}\big)\Big]. \end{aligned}$$

Using the Bobkov-Ledoux inequality (2) to bound each term in the above sum, and noting that, trivially, $DH_j(y_1, \dots, y_n) = D^j f(\sum_{k=1}^n ky_k)$,

$$\begin{aligned} \mathrm{Ent}_{\bar{\mu}_n}(e^f) \;&\leq\; \sum_{j=1}^{n} \lambda Q_j E_{\mu_n}\Big[e^H\big\{|DH_j|e^{|DH_j|} - e^{|DH_j|} + 1\big\}\Big] \\ &=\; \lambda \sum_{j=1}^{n} Q_j E_{\bar{\mu}_n}\Big[e^f\big\{|D^j f|e^{|D^j f|} - e^{|D^j f|} + 1\big\}\Big], \\ &\leq\; \lambda \sum_{j=1}^{\infty} Q_j E_{\bar{\mu}_n}\Big[e^f\big\{|D^j f|e^{|D^j f|} - e^{|D^j f|} + 1\big\}\Big], \end{aligned} \tag{8}$$

where the last inequality follows from the fact that $xe^x - e^x + 1 \geq 0$ for $x \geq 0$.
Finally, we want to take the limit as $n \to \infty$ in (8). Since $\bar{\mu}_n \Rightarrow \mathrm{CP}_{\lambda,Q}$ as $n \to \infty$ by (6), and since f is bounded, by bounded convergence

$$\mathrm{Ent}_{\bar{\mu}_n}(e^f) \to \mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}(e^f), \quad n \to \infty. \tag{9}$$

Similarly, changing the order of summation and expectation in the right-hand side of (8) by Fubini, taking $n \to \infty$ by bounded convergence, and interchanging the order again, it converges to

$$\lambda \sum_{j \geq 1} Q_j E_{\mathrm{CP}_{\lambda,Q}}\Big[e^f\big\{|D^j f|e^{|D^j f|} - e^{|D^j f|} + 1\big\}\Big].$$

This together with (9) implies that (8) yields the required result upon taking $n \to \infty$. □

4 Concentration Proofs

For notational convenience we define the function $\eta(x) := xe^x - e^x + 1$, $x \in \mathbb{R}$, and note that it is non-negative; it achieves its minimum at 0; it is strictly convex on $(-1, \infty)$ and strictly

concave on $(-\infty,-1)$; it decreases from 1 to 0 as x increases to zero, and it is increasing to infinity for $x>0$.
The main technical ingredient of the paper is the following proposition, which is based on a modification of the Herbst argument.

Proposition 5. Suppose that Z has $\mathrm{CP}(\lambda,Q)$ distribution where Q has finite moments up to order $L>1$. If $f:\mathbb{Z}_+\to\mathbb{R}$ is bounded and K-Lipschitz, then for $t>0$, $\epsilon>0$ and $\alpha\in(0,L)$, we have,

$$\Pr\{|f(Z)-Ef(Z)|>t\}\le\exp\Big\{I_\epsilon(\alpha)+\alpha E[\log g_\epsilon(Z)]-\alpha\log t\Big\},$$

where $I_\epsilon(\alpha)$ is defined as in Theorem 3 and

$$g_\epsilon(x):=|f(x)-E[f(Z)]|\,\mathbb{I}_{\{\,|f(x)-E[f(Z)]|\ge\epsilon\,\}}+\epsilon\,\mathbb{I}_{\{\,|f(x)-E[f(Z)]|<\epsilon\,\}}.$$

Proof of Proposition 5. Since f is bounded, by its definition g_ϵ is also bounded above by $2\|f\|_\infty+\epsilon$ and below by ϵ. Therefore, the (polynomial) moment-generating function $G(\tau):=E[g_\epsilon(Z)^\tau]$ is well-defined for all $\tau>0$. Moreover, since both g_ϵ and $\log g_\epsilon$ are bounded, dominated convergence justifies the following differentiation under the integral,

$$G'(\tau)=E\left[\frac{\partial}{\partial\tau}e^{\tau\log g_\epsilon(Z)}\right]=E\big[g_\epsilon(Z)^\tau\log g_\epsilon(Z)\big]$$

so we can relate $G(\tau)$ to the entropy of g_ϵ^τ,

$$\mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}(g_\epsilon^\tau)=\tau G'(\tau)-G(\tau)\log G(\tau)=\tau^2G(\tau)\frac{d}{d\tau}\left[\frac{\log G(\tau)}{\tau}\right]. \tag{10}$$

In order to bound this entropy we will apply Theorem 1 to the function $\phi(x):=\tau\log g_\epsilon(x)$. First we observe that g_ϵ can be written as the composition $g_\epsilon=h\circ(f-E[f(Z)])$, where it is easy to verify that the function $h(x):=|x|\mathbb{I}_{\{|x|\ge\epsilon\}}+\epsilon\mathbb{I}_{\{|x|<\epsilon\}}$ is 1-Lipschitz. And since f is K-Lipschitz by assumption, g_ϵ is itself K-Lipschitz. Hence we can bound $D^j\phi$ as

$$D^j\phi(x)=\tau\log\left|\frac{g_\epsilon(x+j)}{g_\epsilon(x)}\right|\le\tau\log\left(1+\left|\frac{D^jg_\epsilon(x)}{g_\epsilon(x)}\right|\right)\le\tau\log\left(1+\frac{jK}{\epsilon}\right)=\tau\log C_{j,\epsilon}.$$

The same argument also yields a corresponding lower bound, so that $|D^j\phi(x)|\le\tau\log C_{j,\epsilon}$. Applying Theorem 1 to ϕ gives,

$$\mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}(g_\epsilon^\tau)=\mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}(e^\phi)\le\lambda\sum_{j\ge1}Q_jE_{\mathrm{CP}_{\lambda,Q}}[e^\phi\eta(|D^j\phi|)]\le\lambda G(\tau)\sum_{j\ge1}Q_j\eta(\tau\log C_{j,\epsilon}),$$

since $\eta(x)$ is increasing for $x\ge0$. Combining this with (10) we obtain the following differential inequality valid for all $\tau>0$:

$$\frac{d}{d\tau}\left[\frac{\log G(\tau)}{\tau}\right]\le\lambda\sum_{j\ge1}Q_j\frac{\eta(\tau\log C_{j,\epsilon})}{\tau^2}.$$

To solve, we integrate with respect to τ on $(0,\alpha]$ to obtain, for any $\alpha < L$,

$$\begin{aligned}\frac{\log G(\alpha)}{\alpha} - E[\log g_\epsilon(Z)] &\le \lambda \sum_{j\ge 1} Q_j \int_0^\alpha \frac{\eta(\tau \log C_{j,\epsilon})}{\tau^2} d\tau \\ &= \lambda \sum_{j\ge 1} Q_j \log C_{j,\epsilon} \int_0^{\alpha \log C_{j,\epsilon}} \frac{\eta(s)}{s^2}\, ds \\ &= \lambda \sum_{j\ge 1} Q_j \log C_{j,\epsilon} \left[\frac{e^s - 1 - s}{s}\right]_0^{\alpha \log C_{j,\epsilon}} \\ &= I_\epsilon(\alpha)/\alpha,\end{aligned}$$

or, equivalently,

$$G(\alpha) \le \exp\bigl\{\alpha E[\log g_\epsilon(Z)] + I_\epsilon(\alpha)\bigr\}, \tag{11}$$

where the exchange of sum and integral is justified by Fubini's theorem since all the quantities involved are nonnegative. To complete the proof we observe that $g_\epsilon \ge |f - E[f(Z)]|$, so that by (11) and an application of Markov's inequality,

$$\begin{aligned}\Pr\bigl\{|f(Z) - E[f(Z)]| > t\bigr\} &\le \Pr\bigl\{g_\epsilon(Z) > t\bigr\} \\ &= \Pr\bigl\{g_\epsilon(Z)^\alpha > t^\alpha\bigr\} \\ &\le t^{-\alpha} \cdot G(\alpha) \\ &\le \exp\Bigl\{I_\epsilon(\alpha) + \alpha E[\log g_\epsilon(Z)] - \alpha \log t\Bigr\}.\end{aligned}$$

□

Remark 11. The starting point of the above derivation – namely the use of Markov's inequality – can be combined with any number of methods that can be subsequently used to control the (polynomial) moment-generating function of $f(Z)$. For example, in the specialized setting of Corollary 2 with n even, as pointed out by the referee, bounds that are sometimes better than those of Corollary 2 can be obtained by simply noting that

$$\Pr\{|f(Z) - Ef(Z)| > t\} \le \frac{E[|f(Z) - Ef(Z)|^n]}{t^n} \le \frac{2^n K^n E|Z|^n}{t^n},$$

and then expressing EZ^n in terms of the moments of Q by differentiating the characteristic function of Z. However, such methods do not seem to be applicable when the best polynomial decay one can expect has an odd degree (consider, for instance, a Q whose mass function Q_j is proportional to $(j^5 \log j)^{-1}$, where one can explicitly evaluate the bound of Corollary 2), or more generally when we seek to obtain the optimal rate of decay t^{-L}, for arbitrary real $L > 1$ as in Theorem 3.

Using Proposition 5 we can prove Theorem 3 and Corollary 2.

Proof of Theorem 3. The first step is to bring the upper bound in Proposition 5 into a more tractable form. Observe that by its definition, $g_\epsilon(x) \le |f(x) - E[f(Z)]| + \epsilon$, so that, by Jensen's inequality, for a function f satisfying the hypotheses of Proposition 5,

$$E[\log g_\epsilon(Z)] \le \log E[g_\epsilon(Z)] \le \log \Bigl[E\bigl\{|f(Z) - E[f(Z)]|\bigr\} + \epsilon\Bigr]. \tag{12}$$

Thus the upper bound in Proposition 5 can be weakened to

$$\Pr\big\{|f(Z)-E[f(Z)]|>t\big\}\le\exp\Big\{I_\epsilon(\alpha)+\alpha\log\big(D+\epsilon\big)-\alpha\log t\Big\}, \tag{13}$$

where $D:=E\big\{|f(Z)-E[f(Z)]|\big\}$. Next we use the Lipschitz property of f to obtain an upper bound for the above exponent which is uniform over all f with $f(0)$ fixed. Since $f(j)\in[f(0)-Kj,f(0)+Kj]$, we have $|f(j)|\le|f(0)|+Kj$, and hence

$$D\le 2E|f(Z)|\le 2|f(0)|+2K\lambda q_1,$$

where we used the fact that the mean of the CP(λ,Q) law is λq_1. Substituting in (13) and taking the infimum over α yields the required result (4), and it only remains to remove the boundedness assumption on f. But since the bound itself only depends on f via $f(0)$ and K, truncating f at level $\pm n$ and passing to the limit $n\to\infty$ proves part (a).
With $T=2|f(0)|+2K\lambda q_1+\epsilon$, in order to evaluate the exponent

$$\inf_{0<\alpha<L}\big[I_\epsilon(\alpha)+\alpha\log\big(T/t\big)\big] \tag{14}$$

in (4), we calculate the first two derivatives of $I_\epsilon(\alpha)$ with respect to α as,

$$I'_\epsilon(\alpha)=\lambda\sum_{j\ge1}Q_j[C^\alpha_{j,\epsilon}-1]\log C_{j,\epsilon}\quad\text{ and }\quad I''_\epsilon(\alpha)=\lambda\sum_{j\ge1}Q_jC^\alpha_{j,\epsilon}(\log C_{j,\epsilon})^2,$$

where the exchange of differentiation and expectation is justified by dominated convergence; observe that, since $C_{j,\epsilon}>1$, both are positive for all $\alpha>0$. In particular, since $I_\epsilon(\alpha)>0$, the exponent (14) can only be negative (equivalently, the bound in (4) can only be less than 1) if the second term in (14) is negative, i.e., if $t>T$. On the other hand, since $I'_\epsilon(0)=0$ and $I''_\epsilon(\alpha)>0$ for all α, we see that $I_\epsilon(\alpha)$ is locally quadratic around $\alpha=0$. This means that, as long as $t>T$, choosing α sufficiently small we can make (14) negative, therefore the bound of the theorem is meaningful precisely when $t>T$.
To obtain the alternative representation, fix any $\epsilon>0$ and set $i_\epsilon(\alpha)=I'_\epsilon(\alpha)$. Since $I''_\epsilon(\alpha)$ is strictly positive, for $t>T$ the expression $I_\epsilon(\alpha)+\alpha\log(T/t)$ is uniquely minimized at $\alpha^*>0$ which solves $i_\epsilon(\alpha)=\log(t/T)>0$. Hence, for all $t>T$, integrating by parts,

$$\begin{aligned}
\min_{0<\alpha<L}\Big[I_\epsilon(\alpha)+\alpha\log(T/t)\Big] &= I_\epsilon(\alpha^*)+\alpha^*\log(T/t)\\
&= \int_0^{\alpha^*}i_\epsilon(s)ds+\alpha^*\log(T/t)\\
&= \int_0^{i_\epsilon(\alpha^*)}x\,di_\epsilon^{-1}(x)+\alpha^*\log(T/t)\\
&= i_\epsilon(\alpha^*)i_\epsilon^{-1}\big(i_\epsilon(\alpha^*)\big)-\int_0^{i_\epsilon(\alpha^*)}i_\epsilon^{-1}(x)dx+\alpha^*\log(T/t)\\
&= -\int_0^{\log(t/T)}i_\epsilon^{-1}(x)dx,
\end{aligned}$$

which proves part (b). $\square$

Proof of Corollary 2. This is an application of Theorem 3 for specific values of α and ϵ: Bounding the infimum by the value at $\alpha = n$ and taking $\epsilon = K$,

$$\Pr\left\{|f(Z) - Ef(Z)| > t\right\} \leq \exp\left\{I_K(n) + n\log\left(\frac{2|f(0)| + 2K\lambda q_1 + K}{t}\right)\right\}. \tag{15}$$

Using the binomial theorem to expand $I_1(n)$,

$$\begin{aligned} I_K(n) \;=\; \lambda\sum_{j\geq 1} Q_j\left\{C_{j,K}^n - 1 - n\log C_{j,K}\right\} \;&=\; \lambda\sum_{j\geq 1} Q_j\left\{(1+j)^n - 1\right\} - \lambda n\sum_{j\geq 1} Q_j\log(1+j) \\ &\leq\; \lambda\sum_{j\geq 1} Q_j\sum_{r=1}^{n}\binom{n}{r}j^r \;\leq\; \lambda\sum_{r=1}^{n}\binom{n}{r}q_r \end{aligned}$$

Substituting this bound into (15) and rearranging yields the result. □

Next we go on to prove the exponential concentration result Theorem 4 using the classical Herbst argument in conjunction with the modified log-Sobolev inequality of Theorem 1.

Proof of Theorem 4. We proceed similarly to the proof of Proposition 5. Assume f is a bounded and K-Lipschitz, and let $F(\tau) = E[\exp\{\tau f(Z)\}]$, $\tau > 0$ be the moment-generating function of $f(Z)$. Dominated convergence justifies the differentiation

$$F'(\tau) = E[f(Z)e^{\tau f(Z)}],$$

so we can relate $F'(\tau)$ to the entropy of $e^{\tau f}$ by

$$\mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}(e^{\tau f}) = \tau F'(\tau) - F(\tau)\log F(\tau) = \tau^2 F(\tau)\frac{d}{d\tau}\left[\frac{\log F(\tau)}{\tau}\right]. \tag{16}$$

Since f is K-Lipschitz, the function $g := \tau f$ is τK-Lipschitz, so that $|D^j g| \leq \tau K j$. Applying Theorem 1 to g,

$$\mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}(e^{\tau f}) = \mathrm{Ent}_{\mathrm{CP}_{\lambda,Q}}(e^{g}) \leq \lambda\sum_{j\geq 1} Q_j E\Big[e^{g(Z)}\eta(|D^j g(Z)|)\Big] \leq \lambda F(\tau)\sum_{j\geq 1} Q_j\,\eta(j\tau K).$$

Combining this with (16) yields

$$\frac{d}{d\tau}\left[\frac{\log F(\tau)}{\tau}\right] \leq \lambda\sum_{j\geq 1} Q_j\left\{\frac{\eta(j\tau K)}{\tau^2}\right\},$$

and integrating with respect to τ from 0 to $\alpha > 0$ we obtain

$$\begin{aligned} \frac{\log F(\alpha)}{\alpha} - E[f(Z)] \;&\leq\; \lambda\int_0^{\alpha}\sum_{j\geq 1} Q_j\left\{\frac{\eta(j\tau K)}{\tau^2}\right\}d\tau \\ &=\; \lambda\sum_{j\geq 1} jKQ_j\int_0^{j\alpha K}\frac{\eta(s)}{s^2}\,ds \\ &=\; \lambda\sum_{j\geq 1} Q_j\left[\frac{e^{jK\alpha} - 1 - jK\alpha}{\alpha}\right], \end{aligned}$$

where the exchange of the sum and integral is justified by Fubini's theorem since the integrand is nonnegative. Therefore, we have the following a bound on the moment-generating function F,

$$F(\alpha) \leq \exp\big\{\alpha E[f(Z)] + H(\alpha)\big\}, \quad \alpha > 0, \tag{17}$$

where $H(\alpha) = \lambda \sum_j Q_j \left[e^{jK\alpha} - 1 - jK\alpha\right]$. An application of Markov's inequality now gives

$$\begin{aligned} \Pr\big\{f(Z) - E[f(Z)] > t\big\} &\leq e^{-\alpha t} E\Big[\exp\big\{\alpha[f(Z) - E[f(Z)])]\big\}\Big] \\ &= e^{-\alpha t} F(\alpha) e^{-\alpha E[f(Z)]} \\ &\leq \exp\big\{H(\alpha) - \alpha t\big\}. \end{aligned}$$

The removal of the boundedness assumption is a routine truncation argument as in the proof of Theorem 3 or in [2][5]. In order to obtain the best bound for the deviation probability, we minimize the exponent over $\alpha \in (0, M/K)$. This yields the first expression in Theorem 4; the second representation follows from a standard argument as in the last part of the proof of Theorem 3 or [5]. □

Acknowledgments

We thank Christian Houdré for sending us a copy of the preprint [4], and we also thank Christian Houdré and Andrew Barron for numerous comments on these and other related results. Finally, we thank the referee for his insightful comments and very careful reading of our first manuscript.

References

[1] S.G. Bobkov. Some extremal properties of the Bernoulli distribution. *Teor. Veroyatnost. i Primenen.*, 41(4):877–884, 1996.

[2] S.G. Bobkov and M. Ledoux. On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. *J. Funct. Anal.*, 156(2):347–365, 1998.

[3] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 2005.

[4] J-C. Breton, C. Houdré, and N. Privault. Dimension free and infinite variance tail estimates on Poisson space. *Preprint, available at* http://arxiv.org/abs/math/0412346, 2004.

[5] C. Houdré. Remarks on deviation inequalities for functions of infinitely divisible random vectors. *Ann. Probab.*, 30(3):1223–1237, 2002.

[6] C. Houdré and P. Marchal. On the concentration of measure phenomenon for stable and related random vectors. *Ann. Probab.*, 32(2):1496–1508, 2004.

[7] C. Houdré, V. Pérez-Abreu, and D. Surgailis. Interpolation, correlation identities, and inequalities for infinitely divisible variables. *J. Fourier Anal. Appl.*, 4(6):651–668, 1998.

[8] C. Houdré and N. Privault. Concentration and deviation inequalities in infinite dimensions via covariance representations. *Bernoulli*, 8(6):697–720, 2002.

[9] M. Ledoux. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics*, volume 1648 of *Lecture Notes in Math.*, pages 165–294. Springer, Berlin, 1996.

[10] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 1:63–87 (electronic), 1997.

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

[12] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Algorithms Combin., 16, Springer, Berlin, 1998.

[13] K-I. Sato. *Lévy processes and infinitely divisible distributions*, volume 68. Cambridge University Press, Cambridge, 1999.

[14] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, No. 81:73–205, 1995.

[15] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields*, 118(3):427–438, 2000.