

A Criterion for the Compound Poisson Distribution to be Maximum Entropy

Oliver Johnson

Department of Mathematics

University of Bristol

University Walk

Bristol, BS8 1TW, UK.

Email: O.Johnson@bristol.ac.uk

Ioannis Kontoyiannis

Department of Informatics

Athens University of Economics & Business

Patission 76

Athens 10434, Greece.

Email: yiannis@aueb.gr

Mokshay Madiman

Department of Statistics

Yale University

24 Hillhouse Avenue

New Haven, CT 06511, USA.

Email: mokshay.madiman@yale.edu

Abstract—The Poisson distribution is known to have maximal entropy among all distributions (on the nonnegative integers) within a natural class. Interestingly, straightforward attempts to generalize this result to general compound Poisson distributions fail because the analogous result is not true in general. However, we show that the compound Poisson does indeed have a natural maximum entropy characterization when the distributions under consideration are log-concave. This complements the recent development by the same authors of an information-theoretic foundation for compound Poisson approximation inequalities and limit theorems.

I. INTRODUCTION

A particularly appealing way to state the classical central limit theorem is to say that, if X_1, X_2, \dots are independent and identically distributed, continuous random variables with zero mean and unit variance, then the entropy of their normalized partial sums $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ increases with n to the entropy of the standard normal distribution, which is maximal among all random variables with zero mean and unit variance. More precisely, if f_n denotes the density of S_n and ϕ the standard normal density, then, as $n \rightarrow \infty$,

$$h(f_n) \uparrow h(\phi) = \sup\{h(f) : \text{densities } f \text{ with mean 0 and variance 1}\}, \quad (1)$$

where $h(f) = -\int f \log f$ denotes the differential entropy and \log denotes the natural logarithm. Precise conditions under which (1) holds are given in [1][2][3]; also see [4][5][6] and the references therein, where numerous related results are stated, along with their history.

Part of the appeal of this formalization of the central limit theorem comes from its analogy to the second law of thermodynamics: The “state” (meaning the distribution) of the random variables S_n evolves monotonically, until the *maximum entropy* state, the standard normal distribution, is reached. Moreover, the introduction of information-theoretic ideas and techniques in connection with the entropy has motivated numerous related results (and their proofs), generalizing and strengthening the central limit theorem in different directions; see the references mentioned above for details.

The classical Poisson convergence limit theorems, of which the binomial-to-Poisson is the prototypical example, have

also been examined under a similar light. An analogous program has been recently carried out in this case [7] [8] [9][10][11][12]. The starting point is the identification of the Poisson distribution as that which has maximal entropy within a natural class of probability measures. Perhaps the simplest way to state and prove this is along the following lines; first we make some simple definitions:

Definition 1.1: For any parameter vector $\mathbf{p} = (p_1, p_2, \dots, p_n)$ with each $p_i \in [0, 1]$, the sum of independent Bernoulli random variables $B_i \sim \text{Bern}(p_i)$,

$$S_n = \sum_{i=1}^n B_i,$$

is called a *Bernoulli sum*, and its probability mass function is denoted by $b_{\mathbf{p}}(x) := \Pr\{S_n = x\}$, for $x = 0, 1, \dots$. Further, for each $\lambda > 0$, we define the following sets of parameter vectors:

$$\mathcal{P}_n(\lambda) = \{\mathbf{p} \in [0, 1]^n : p_1 + p_2 + \dots + p_n = \lambda\}$$

and

$$\mathcal{P}_{\infty}(\lambda) = \bigcup_{n \geq 1} \mathcal{P}_n(\lambda).$$

Shepp and Olkin [7] (see also Mateev [8]) showed that, for fixed $n \geq 1$, the Bernoulli sum $b_{\mathbf{p}}$ which has maximal entropy among all Bernoulli sums with mean λ , is $\text{Bin}(n, \lambda/n)$, the binomial with parameters n and λ/n ,

$$H(\text{Bin}(n, \lambda/n)) = \max\{H(b_{\mathbf{p}}) : \mathbf{p} \in \mathcal{P}_n(\lambda)\}, \quad (2)$$

where $H(P) = -\sum_x P(x) \log P(x)$ denotes the discrete entropy function. Noting that the binomial $\text{Bin}(n, \lambda/n)$ converges to the Poisson distribution $\text{Po}(\lambda)$ as $n \rightarrow \infty$, and that the classes of Bernoulli sums in (2) are nested, $\{b_{\mathbf{p}} : \mathbf{p} \in \mathcal{P}_n(\lambda)\} \subset \{b_{\mathbf{p}} : \mathbf{p} \in \mathcal{P}_{n+1}(\lambda)\}$, Harremoës [10] noticed that a simple limiting argument gives the following maximum entropy property for the Poisson distribution:

$$H(\text{Po}(\lambda)) = \sup\{H(b_{\mathbf{p}}) : \mathbf{p} \in \mathcal{P}_{\infty}(\lambda)\}. \quad (3)$$

A key property in generalizing and understanding this maximum entropy property further is that of ultra log-concavity;

cf. [13]. The distribution P of a random variable X is *ultra log-concave* if $P(x)/\Pi_\lambda(x)$ is log-concave, that is, if,

$$xP(x)^2 \geq (x+1)P(x+1)P(x-1), \quad \text{for all } x \geq 1. \quad (4)$$

Note that the Poisson distribution as well as all Bernoulli sums are ultra log-concave.

Johnson [12] recently proved the following maximum entropy property for the Poisson distribution, generalizing (3):

$$\begin{aligned} H(\text{Po}(\lambda)) \\ = \max \left\{ H(P) : \text{ultra log-concave } P \text{ with mean } \lambda \right\}. \end{aligned} \quad (5)$$

Partly motivated by the desire to provide an information-theoretic foundation for *compound Poisson limit theorems* and the more general problem of *compound Poisson approximation*, as a first step we consider the problem of generalizing the maximum entropy properties (3) and (5) to the case of *compound Poisson* distributions on \mathbb{Z}_+ . We begin with some definitions:

Definition 1.2: Let P be an arbitrary distribution on $\mathbb{Z}_+ = \{0, 1, \dots\}$, and Q a distribution on $\mathbb{N} = \{1, 2, \dots\}$. The *Q-compound distribution* $C_Q P$ is the distribution of the random sum,

$$\sum_{j=1}^Y X_j, \quad (6)$$

where Y has distribution P and the random variables $\{X_j\}$ are independent and identically distributed (i.i.d.) with common distribution Q and independent of Y . The distribution Q is called a *compounding distribution*, and the map $P \mapsto C_Q P$ is the *Q-compounding operation*. The Q -compound distribution $C_Q P$ can be explicitly written as the mixture,

$$C_Q P(x) = \sum_{y=0}^{\infty} P(y) Q^{*y}(x), \quad x \geq 0, \quad (7)$$

where $Q^{*j}(x)$ is the j th convolution power of Q and Q^{*0} is the point mass at $x = 0$.

Above and throughout the paper, the empty sum $\sum_{j=1}^0(\dots)$ is taken to be zero; all random variables considered are supported on $\mathbb{Z}_+ = \{0, 1, \dots\}$; and all compounding distributions Q are supported on $\mathbb{N} = \{1, 2, \dots\}$.

Example 1.3: Let Q be an arbitrary distribution on \mathbb{N} .

- 1) For any $0 \leq p \leq 1$, the *compound Bernoulli distribution* $\text{CBern}(p, Q)$ is the distribution of the product BX , where $B \sim \text{Bern}(p)$ and $X \sim Q$ are independent. It has probability mass function $C_Q P$, where P is the $\text{Bern}(p)$ mass function, so that, $C_Q P(0) = 1 - p$ and $C_Q P(x) = pQ(x)$ for $x \geq 1$.
- 2) A *compound Bernoulli sum* is a sum of independent compound Bernoulli random variables, all with respect to the same compounding distribution Q : Let X_1, X_2, \dots, X_n be i.i.d. with common distribution Q and B_1, B_2, \dots, B_n be independent $\text{Bern}(p_i)$. We call,

$$\sum_{i=1}^n B_i X_i \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\sum_{i=1}^n B_i} X_j,$$

a *compound Bernoulli sum*; in view of (6), its distribution is $C_Q b_{\mathbf{p}}$, where $\mathbf{p} = (p_1, p_2, \dots, p_n)$.

- 3) In the special case of a compound Bernoulli sum with all its parameters $p_i = p$ for a fixed $p \in [0, 1]$, we say that it has a *compound binomial distribution*, denoted by $\text{CBin}(n, p, Q)$.
- 4) Let $\Pi_\lambda(x) = e^{-\lambda} \lambda^x / x!$, $x \geq 0$, denote the $\text{Po}(\lambda)$ mass function. Then, for any $\lambda > 0$, the *compound Poisson distribution* $\text{CPo}(\lambda, Q)$ is the distribution with mass function $C_Q \Pi_\lambda$,

$$C_Q \Pi_\lambda(x) = \sum_{j=0}^{\infty} \Pi_\lambda(j) Q^{*j}(x) = \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} Q^{*j}(x), \quad (8)$$

for all $x \geq 0$.

In view of the Shepp-Olkin maximum entropy property (2) for the binomial distribution, a first natural conjecture might be that the compound binomial has maximum entropy among all compound Bernoulli sums $C_Q b_{\mathbf{p}}$ with a fixed mean; that is,

$$H(\text{CBin}(n, \lambda/n, Q)) = \max \left\{ H(C_Q b_{\mathbf{p}}) : \mathbf{p} \in \mathcal{P}_n(\lambda) \right\}. \quad (9)$$

But, perhaps somewhat surprisingly, as Chi [14] has noted, (9) fails in general. For example, taking Q to be the uniform distribution on $\{1, 2\}$, $\mathbf{p} = (0.00125, 0.00875)$ and $\lambda = p_1 + p_2 = 0.01$, direct computation shows that,

$$H(\text{CBin}(2, \lambda/2, Q)) < 0.090798 < 0.090804 < H(C_Q b_{\mathbf{p}}).$$

As the Shepp-Olkin result (2) was only seen as an intermediate step in proving the maximum entropy property of the Poisson distribution (3), we may still hope that the corresponding result remains true for compound Poisson measures, namely that,

$$H(\text{CPo}(\lambda, Q)) = \sup \left\{ H(C_Q b_{\mathbf{p}}) : \mathbf{p} \in \mathcal{P}_\infty(\lambda) \right\}. \quad (10)$$

Again, (10) fails in general. For example, taking the same Q, λ and \mathbf{p} as above, yields,

$$H(\text{CPo}(\lambda, Q)) < 0.090765 < 0.090804 < H(C_Q b_{\mathbf{p}}).$$

The main purpose of the present work is to show that, despite these negative results, it is possible to provide natural, broad sufficient conditions, under which the compound binomial and compound Poisson distributions can be shown to have maximal entropy in an appropriate class of measures.

Our main result (proved in Section II) states that, as long as Q and the compound Poisson measure $\text{CPo}(\lambda, Q)$ are log-concave, the same maximum entropy statement as in (5) remains valid in the compound Poisson case:

Theorem 1.4: If the distribution Q on \mathbb{N} and the compound Poisson distribution $\text{CPo}(\lambda, Q)$ are both log-concave, then,

$$\begin{aligned} H(\text{CPo}(\lambda, Q)) = \max \left\{ H(C_Q P) : \right. \\ \left. \text{ultra log-concave } P \text{ with mean } \lambda \right\}. \end{aligned}$$

The notion of log-concavity is central in the development of the ideas in this work. [In a different setting, log-concavity also appears as a natural condition for a different maximum entropy problem considered by Cover and Zhang [15].] Recall that the distribution P of a random variable X on \mathbb{Z}_+ is *log-concave* if its support is a (possibly infinite) interval of successive integers in \mathbb{Z}_+ , and,

$$P(x)^2 \geq P(x+1)P(x-1), \quad \text{for all } x \geq 1. \quad (11)$$

We also recall that most of the commonly used distributions appearing in applications (e.g., the Poisson, binomial, geometric, negative binomial, hypergeometric logarithmic series, or Polya-Eggenberger distribution) are log-concave.

Note that the condition that a probability mass function P be ultra log-concave given in Equation (4) is more restrictive than Equation (11), and equates to requiring that the ratio P/Π_λ form a log-concave sequence for some λ .

In [21] we discuss conditions under which the compound Poisson and compound Bernoulli distributions are log-concave.

II. MAXIMUM ENTROPY PROPERTY OF THE COMPOUND POISSON DISTRIBUTION

Here we show that, if Q and the compound Poisson distribution $\text{CPo}(\lambda, Q) = C_Q \Pi_\lambda$ are both log-concave, then $\text{CPo}(\lambda, Q)$ has maximum entropy among all distributions of the form $C_Q P$, when P has mean λ and is ultra log-concave. Our approach is an extension of the ‘semigroup’ arguments of [12].

We begin by recording some basic properties of log-concave and ultra log-concave distributions:

- (i) If P is ultra log-concave, then from the definitions it is immediate that P is log-concave.
- (ii) If Q is log-concave, then it has finite moments of all orders; see [16, Theorem 7].
- (iii) If X is a random variable with ultra log-concave distribution P , then (by (i) and (ii)) it has finite moments of all orders. Moreover, considering the covariance between the decreasing function $P(x+1)(x+1)/P(x)$ and the increasing function $x(x-1)\cdots(x-n)$, shows that the falling factorial moments of P satisfy,

$$E[(X)_n] := E[X(X-1)\cdots(X-n+1)] \leq (E(X))^n;$$

see [12] and [17] for details.

- (iv) The Poisson distribution and all Bernoulli sums are ultra log-concave.

Recall the following definition from [12]:

Definition 2.1: Given $\alpha \in [0, 1]$ and a random variable $X \sim P$ on \mathbb{Z}_+ with mean $\lambda \geq 0$, let $U_\alpha P$ denote the distribution of the random variable,

$$\sum_{i=1}^X B_i + Z_{\lambda(1-\alpha)},$$

where the B_i are i.i.d. $\text{Bern}(\alpha)$, $Z_{\lambda(1-\alpha)}$ has distribution $\text{Po}(\lambda(1-\alpha))$, and all random variables are independent of each other and of X .

Note that, if $X \sim P$ has mean λ , then $U_\alpha P$ has the same mean. Also, recall the following useful relation that was established in Proposition 3.6 of [12]: For all $y \geq 0$,

$$\frac{\partial}{\partial \alpha} U_\alpha P(y) = \frac{1}{\alpha} \left(\lambda(U_\alpha P(y) - U_\alpha P(y-1)) - ((y+1)U_\alpha P(y+1) - yU_\alpha P(y)) \right) \quad (12)$$

Next we define another transformation of probability distributions P on \mathbb{Z}_+ :

Definition 2.2: Given $\alpha \in [0, 1]$, a distribution P on \mathbb{Z}_+ and a compounding distribution Q on \mathbb{N} , let $U_\alpha^Q P$ denote the distribution $C_Q U_\alpha P$:

$$U_\alpha^Q P(x) := C_Q U_\alpha P(x) = \sum_{y=0}^{\infty} U_\alpha P(y) Q^{*y}(x), \quad x \geq 0.$$

An important observation that will be at the heart of the proof of Theorem 1.4 below is that, for $\alpha = 0$, $U_0^Q P$ is simply the compound Poisson measure $\text{CP}(\lambda, Q)$, while for $\alpha = 1$, $U_1^Q P = C_Q P$. The idea is that U_α^Q interpolates between these two distributions, and that we deduce the result by considering monotonicity properties with respect to α .

In [12], the characterization of the Poisson as a maximum entropy distribution was proved through the decrease of its score function. In an analogous way, following [18], we define the score function of a Q -compound random variable as follows.

Definition 2.3: Given a distribution P on \mathbb{Z}_+ with mean λ , the corresponding Q -compound distribution $C_Q P$ has score function defined by:

$$\begin{aligned} r_{1, C_Q P}(x) &= \frac{\sum_{y=0}^{\infty} (y+1)P(y+1)Q^{*y}(x)}{\lambda \sum_{y=0}^{\infty} P(y)Q^{*y}(x)} - 1 \\ &= \frac{\sum_{y=0}^{\infty} (y+1)P(y+1)Q^{*y}(x)}{\lambda C_Q P(x)} - 1. \end{aligned} \quad (13)$$

Notice that the mean of $r_{1, C_Q P}$ with respect to $C_Q P$ is zero, and that if $P \sim \text{Po}(\lambda)$ then $r_{1, C_Q P}(x) \equiv 0$. Further, when Q is the point mass at 1 this score function reduces to the ‘scaled score function’ introduced in [11]. But, unlike the scaled score function and the alternative score function $r_{2, C_Q P}$ given in [18], this score function is not only a function of the compound distribution $C_Q P$, but also explicitly depends on P . A projection identity and other properties of $r_{1, C_Q P}$ are proved in [18], allowing compound Poisson approximation results to be deduced.

Next we show that, if Q is log-concave and P is ultra log-concave, then the score function $r_{1, C_Q P}(x)$ is decreasing in x .

Lemma 2.4: If P is ultra log-concave and the compounding distribution Q is log-concave, then the score function $r_{1, C_Q P}(x)$ of $C_Q P$ is decreasing in x .

Proof: First we recall Theorem 2.1 of Keilson and Sumita [19], which implies that, if Q is log-concave, then for any $m \geq n$, and for any x :

$$Q^{*m}(x+1)Q^{*n}(x) - Q^{*m}(x)Q^{*n}(x+1) \geq 0. \quad (14)$$

[This can be proved by considering Q^{*m} as the convolution of Q^{*n} and $Q^{*(m-n)}$, and writing

$$\begin{aligned} & Q^{*m}(x+1)Q^{*n}(x) - Q^{*m}(x)Q^{*n}(x+1) \\ &= \sum_l Q^{*(m-n)}(l) \left(Q^{*n}(x+1-l)Q^{*n}(x) - \right. \\ & \quad \left. Q^{*n}(x-l)Q^{*n}(x+1) \right). \end{aligned}$$

Since Q is log-concave, then so is Q^{*n} , cf. [20], so the ratio $Q^{*n}(x+1)/Q^{*n}(x)$ is decreasing in x , and (14) follows.]

By definition, $r_{1,C_Q P}(x) \geq r_{1,C_Q P}(x+1)$ if and only if,

$$\begin{aligned} 0 &\leq \left(\sum_y (y+1)P(y+1)Q^{*y}(x) \right) \left(\sum_z P(z)Q^{*z}(x+1) \right) \\ &\quad - \left(\sum_y (y+1)P(y+1)Q^{*y}(x+1) \right) \left(\sum_z P(z)Q^{*z}(x) \right) \\ &= \sum_{y,z} (y+1)P(y+1)P(z) \times \\ &\quad \times [Q^{*y}(x)Q^{*z}(x+1) - Q^{*y}(x+1)Q^{*z}(x)]. \end{aligned}$$

Noting that for $y = z$ the term in square brackets in the double sum becomes zero, and swapping the values of y and z in the range $y > z$, the double sum in (15) becomes,

$$\begin{aligned} & \sum_{y < z} [(y+1)P(y+1)P(z) - (z+1)P(z+1)P(y)] \times \\ & \quad \times [Q^{*y}(x)Q^{*z}(x+1) - Q^{*y}(x+1)Q^{*z}(x)]. \end{aligned}$$

By the ultra log-concavity of P , the first square bracket is positive for $y \leq z$, and by equation (14) the second square bracket is also positive for $y \leq z$. ■

We remark that, under the same assumptions, and using a very similar argument, an analogous result holds for the score function $r_{2,C_Q P}$ recently introduced in [18].

Combining Lemma 2.4 with equation (12) we deduce the following result, which is the main technical step in the proof of Theorem 1.4 below.

Proposition 2.5: Let P be an ultra log-concave distribution on \mathbb{Z}_+ with mean $\lambda > 0$, and assume that Q and $\text{CPo}(\lambda, Q)$ are both log-concave. Let W_α be a random variable with distribution $U_\alpha^Q P$, and define, for all $\alpha \in [0, 1]$, the function,

$$E(\alpha) := E[-\log C_Q \Pi_\lambda(W_\alpha)].$$

Then $E(\alpha)$ is continuous for all $\alpha \in [0, 1]$, it is differentiable for $\alpha \in (0, 1)$, and, moreover, $E'(\alpha) \leq 0$ for $\alpha \in (0, 1)$. In particular, $E(0) \geq E(1)$.

Proof: We simply sketch the proof and refer to the full paper for details. Recall that,

$$\begin{aligned} U_\alpha^Q P(x) &= C_Q U_\alpha P(x) \\ &= \sum_{y=0}^{\infty} U_\alpha P(y) Q^{*y}(x) \\ &= \sum_{y=0}^x U_\alpha P(y) Q^{*y}(x), \end{aligned}$$

where the last sum is restricted to the range $0 \leq y \leq x$, because Q is supported on \mathbb{N} .

By some technical arguments, we can prove that $E(\alpha)$ is differentiable for all $\alpha \in (0, 1)$ and, in fact, that we can differentiate the series

$$\begin{aligned} E(\alpha) &:= E[-\log C_Q \Pi_\lambda(W_\alpha)] \\ &= - \sum_{x=0}^{\infty} U_\alpha^Q P(x) \log C_Q \Pi_\lambda(x), \end{aligned} \quad (15)$$

term-by-term, to obtain,

$$E'(\alpha) = - \sum_{x=0}^{\infty} \frac{\partial}{\partial \alpha} U_\alpha^Q P(x) \log C_Q \Pi_\lambda(x) \quad (16)$$

$$\begin{aligned} &= \frac{\lambda}{\alpha} \sum_{x=0}^{\infty} \log C_Q \Pi_\lambda(x) \left(U_\alpha^Q P(x) r_{1,U_\alpha^Q P}(x) \right. \\ &\quad \left. - \sum_{v=0}^x Q(v) U_\alpha^Q P(x-v) r_{1,U_\alpha^Q P}(x-v) \right) \end{aligned} \quad (17)$$

$$\begin{aligned} &= \frac{\lambda}{\alpha} \sum_{x=0}^{\infty} U_\alpha^Q P(x) r_{1,U_\alpha^Q P}(x) \\ &\quad \left(\log C_Q \Pi_\lambda(x) - \sum_{v=0}^{\infty} Q(v) \log C_Q \Pi_\lambda(x+v) \right), \end{aligned} \quad (18)$$

where the rearrangement leading to the third equality follows by interchanging the order of (second) double summation and replacing x by $x+v$.

Now we note that, exactly as in [12], the last series above is the covariance between the (zero-mean) function $r_{1,U_\alpha^Q P}(x)$ and the function $(\log C_Q \Pi_\lambda(x) - \sum_v Q(v) \log C_Q \Pi_\lambda(x+v))$, under the measure $U_\alpha^Q P$. Since P is ultra log-concave, so is $U_\alpha P$ [12], hence the score function $r_{1,U_\alpha^Q P}(x)$ is decreasing in x , by Lemma 2.4. Also, the log-concavity of $C_Q \Pi_\lambda$ implies that the second function is increasing, and Chebyshev's rearrangement lemma implies that the covariance is less than or equal to zero, proving that $E'(\alpha) \leq 0$, as claimed.

Finally, the fact that $E(0) \geq E(1)$ is an immediate consequence of the continuity of $E(\alpha)$ on $[0, 1]$ and the fact that $E'(\alpha) \leq 0$ for all $\alpha \in (0, 1)$. ■

Notice that, for the above proof to work, it is not necessary that $C_Q \Pi_\lambda$ be log-concave; the weaker property that $(\log C_Q \Pi_\lambda(x) - \sum_v Q(v) \log C_Q \Pi_\lambda(x+v))$ be increasing is enough, although this is not as simple to verify.

Proof: [of Theorem 1.4] As in Proposition 2.5, let $W_\alpha \sim U_\alpha^Q P = C_Q U_\alpha P$, and let $D(P||Q)$ denote the relative entropy between P and Q ,

$$D(P||Q) := \sum_{x \geq 0} P(x) \log \frac{P(x)}{Q(x)}.$$

Then, noting that $W_0 \sim C_Q \Pi_\lambda$ and $W_1 \sim C_Q P$, we have,

$$\begin{aligned} H(C_Q P) &\leq H(C_Q P) + D(C_Q P||C_Q \Pi_\lambda) \\ &= -E[\log C_Q \Pi_\lambda(W_1)] \\ &\leq -E[\log C_Q \Pi_\lambda(W_0)] \\ &= H(C_Q \Pi_\lambda), \end{aligned}$$

where the first inequality is simply the nonnegativity of relative entropy, and the second inequality is exactly the statement that $E(1) \leq E(0)$, proved in Proposition 2.5. ■

III. DISCUSSION

We have given log-concavity conditions under which compound Poisson distributions have a natural maximum entropy property. Of course, to translate this general result into concrete useful statements, one would like simple conditions on Q under which both $C_Q \Pi_\lambda$ is log-concave. A conjecture regarding such a condition, as well as several cases in which the conjecture is verified, is described in the full paper [21]. The conjecture states that $\text{CPo}(\lambda, Q)$ is log-concave when Q is log-concave and $\lambda Q(1)^2 \geq 2Q(2)$.

In particular, the results there imply the following explicit maximum entropy statements.

- Example 3.1:* 1) Suppose Q is supported on $\{1, 2\}$, with probabilities $Q(1) = q$, $Q(2) = 1 - q$, and consider the class of all Bernoulli sums b_p with mean $p_1 + p_2 + \dots + p_n = \lambda$. The compound Poisson maximum entropy property holds in this case, as long as λ is large enough. More precisely, the distribution $\text{CPo}(\lambda, Q)$ has maximal entropy among all compound Bernoulli sums $C_Q b_p$ with $p_1 + p_2 + \dots + p_n = \lambda \geq \frac{2(1-q)}{q^2}$.
- 2) Suppose Q is geometric with parameter $\alpha \in (0, 1)$, i.e., $Q(x) = \alpha(1 - \alpha)^{x-1}$ for all $x \geq 1$, and again consider the class of a Bernoulli sums b_p with mean λ . Then the compound Poisson distribution $\text{CPo}(\lambda, Q)$ has maximal entropy among all compound Bernoulli sums $C_Q b_p$ with $p_1 + p_2 + \dots + p_n = \lambda \geq \frac{2(1-\alpha)}{\alpha}$.

ACKNOWLEDGMENT

The authors would like to thank Zhiyi Chi for the counterexample mentioned in the introduction.

REFERENCES

- [1] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," *J. Amer. Math. Soc.*, vol. 17, no. 4, pp. 975–982 (electronic), 2004.
- [2] A. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianity of the sum of independent random variables: a simple proof," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4295–4297, 2006.
- [3] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2317–2329, 2007.
- [4] Y. Linnik, "An information-theoretic proof of the Central Limit Theorem with the Lindeberg Condition," *Theory Probab. Appl.*, vol. 4, pp. 288–299, 1959.
- [5] A. R. Barron, "Entropy and the Central Limit Theorem," *Ann. Probab.*, vol. 14, no. 1, pp. 336–342, 1986.
- [6] O. T. Johnson, *Information theory and the Central Limit Theorem*. London: Imperial College Press, 2004.
- [7] L. A. Shepp and I. Olkin, "Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution," in *Contributions to probability*. New York: Academic Press, 1981, pp. 201–206.
- [8] P. Mateev, "The entropy of the multinomial distribution," *Teor. Veroyatnost. i Primenen.*, vol. 23, no. 1, pp. 196–198, 1978.
- [9] I. Johnstone and B. MacGibbon, "Une mesure d'information caractérisant la loi de Poisson," in *Séminaire de Probabilités, XXI*. Berlin: Springer, 1987, pp. 563–573.
- [10] P. Harremoës, "Binomial and Poisson distributions as maximum entropy distributions," *IEEE Trans. Information Theory*, vol. 47, no. 5, pp. 2039–2041, 2001.
- [11] I. Kontoyiannis, P. Harremoës, and O. T. Johnson, "Entropy and the law of small numbers," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 466–472, 2005.
- [12] O. T. Johnson, "Log-concavity and the maximum entropy property of the Poisson distribution," *Stoch. Proc. Appl.*, vol. 117, no. 6, pp. 791–802, 2007.
- [13] R. Pemantle, "Towards a theory of negative dependence," *J. Math. Phys.*, vol. 41, no. 3, pp. 1371–1390, 2000.
- [14] Z. Chi, 2006, personal communication.
- [15] T. M. Cover and Z. Zhang, "On the maximum entropy of the sum of two dependent random variables," *IEEE Trans. Information Theory*, vol. 40, no. 4, pp. 1244–1246, 1994.
- [16] J. Keilson and H. Gerber, "Some results for discrete unimodality," *Journal of the American Statistical Association*, vol. 66, no. 334, pp. 386–389, 1971.
- [17] P. Harremoës, O. T. Johnson, and I. Kontoyiannis, "Thinning and the Law of Small Numbers," in *Proceedings of ISIT 2007, 24th - 29th June 2007, Nice*, 2007, pp. 1491–1495.
- [18] A. Barbour, O. T. Johnson, I. Kontoyiannis, and M. Madiman, "Compound Poisson approximation via local information quantities," 2009, in preparation.
- [19] J. Keilson and U. Sumita, "Uniform stochastic ordering and related inequalities," *Canad. J. Statist.*, vol. 10, no. 3, pp. 181–198, 1982.
- [20] S. Karlin, *Total positivity. Vol. I*. Stanford, Calif: Stanford University Press, 1968.
- [21] O. T. Johnson, I. Kontoyiannis, and M. Madiman, "On the entropy and log-concavity of compound Poisson measures," 2008, submitted; see <http://arxiv.org/abs/arXiv:0805.4112>.