Log-concavity, ultra-log-concavity, and a maximum entropy property of discrete compound Poisson measures

Oliver Johnson[a], Ioannis Kontoyiannis[b], Mokshay Madiman[c,*]

[a]*Department of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK.*

[b]*Department of Informatics, Athens University of Economics & Business, Patission 76, Athens 10434, Greece.*

[c]*Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT 06511, USA.*

Abstract

Sufficient conditions are developed, under which the compound Poisson distribution has maximal entropy within a natural class of probability measures on the nonnegative integers. Recently, one of the authors [O. Johnson, *Stoch. Proc. Appl.*, 2007] used a semigroup approach to show that the Poisson has maximal entropy among all ultra-log-concave distributions with fixed mean. We show via a non-trivial extension of this semigroup approach that the natural analog of the Poisson maximum entropy property remains valid if the compound Poisson distributions under consideration are log-concave, but that it fails in general. A parallel maximum entropy result is established for the family of compound binomial measures. Sufficient conditions for compound distributions to be log-concave are discussed and applications to combinatorics are examined; new bounds are derived on the entropy of the cardinality of a random independent set in a claw-free graph, and a connection is drawn to Mason's conjecture for matroids. The present results are primarily motivated by the desire to provide an information-theoretic foundation for compound Poisson approximation and associated limit theorems, analogous to the corresponding developments for the central limit theorem and for Poisson approximation. Our results also demonstrate new links between some probabilistic methods and the combinatorial notions of log-concavity and ultra-log-concavity, and they add to the growing body of work exploring the applications of maximum entropy characterizations to problems in discrete mathematics.

Key words: Log-concavity, compound Poisson distribution, maximum entropy, ultra-log-concavity, Markov semigroup
2000 MSC: 94A17, 60E07, 60E15

*Corresponding author.
Email addresses:* `O.Johnson@bristol.ac.uk` (Oliver Johnson), `yiannis@aueb.gr` (Ioannis Kontoyiannis), `mokshay.madiman@yale.edu` (Mokshay Madiman)

1. Introduction

The primary motivation for this work is the development of an information-theoretic approach to discrete limit laws, specifically those corresponding to compound Poisson limits. Recall that the classical central limit theorem can be phrased as follows: If $X_1, X_2, \ldots$ 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 $n \to \infty$,

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

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 by Barron [5] and Artstein et al. [2]; also see [25, 39, 51] and the references therein.

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 above for details.

Recently, some discrete limit laws have been examined in a similar vein, but, as the discrete entropy $H(P) = -\sum_x P(x) \log P(x)$ for probability mass functions P on a countable set naturally replaces the differential entropy $h(f)$, many of the relevant analytical tools become unavailable. For Poisson convergence theorems, of which the binomial-to-Poisson is the prototypical example, an analogous program has been carried out in [21, 26, 37, 48]. Like with the central limit theorem, there are two aspects to this theory – the Poisson distribution is first identified as that which has maximal entropy within a natural class of probability measures, and then convergence of appropriate sums of random variables to the Poisson is established in the sense of relative entropy (or better still, approximation bounds are obtained that quantify the rate of this convergence).

One of the main goals of this work is to establish a starting point for developing an information-theoretic framework for the much more general class of *compound Poisson* limit theorems.[1] To that end, our first main result, given in Section 2, provides a maximum entropy characterization of compound Poisson laws, generalizing Johnson's characterization [26] of the Poisson distribution. It states that if one looks at the class of all ultra-log-concave distributions on $\mathbb{Z}_+$ with a fixed mean, and then compounds each distribution in this class using a given probability measure on $\mathbb{N}$, then the compound Poisson has maximal entropy in this class, provided it is log-concave.

---

[1] Recall that the compound Poisson distributions are the only infinitely divisible distributions on $\mathbb{Z}_+$, and also they are (discrete) stable laws [50]. In the way of motivation, we may also recall the remark of Gnedenko and Korolev [18, pp. 211-215] that "there should be mathematical … probabilistic models of the universal principle of non-decrease of uncertainty," and their proposal that we should "find conditions under which certain limit laws appearing in limit theorems of probability theory possess extremal entropy properties. Immediate candidates to be subjected to such analysis are, of course, stable laws."

Having established conditions under which a compound Poisson measure has maximum entropy, in subsequent work [4] we consider the problem of establishing compound Poisson limit theorems as well as finite-n approximation bounds, in relative entropy and total variation. The tools developed in the present work, and in particular the definition and analysis of a new score function in Section 3, play a crucial role in these compound Poisson approximation results.

In a different direction, in Section 6 we demonstrate how the present results provide new links between classical probabilistic methods and the combinatorial notions of log-concavity and ultra-log-concavity. Log-concave sequences are well-studied objects in combinatorics, see, e.g., the surveys by Brenti [7] and Stanley [49]. Additional motivation in recent years has come from the search for a theory of negative dependence. Specifically, as argued by Pemantle [44], a theory of negative dependence has long been desired in probability and statistical physics, in analogy with the theory of positive dependence exemplified by the Fortuin-Kasteleyn-Ginibre (FKG) inequality [17] (earlier versions were developed by Harris [23] and, in combinatorics, by Kleitman [34]). But the development of such a theory is believed to be difficult and delicate. One of the major open questions in this area, called the Big Conjecture in Wagner's recent work [52], asserts that, if a probability measure on the Boolean hypercube satisfies certain negative correlation conditions, then the sequence $\{p_k\}$ of probabilities of picking a set of size k, is ultra-log-concave. This is closely related to Mason's conjecture for independent sets in matroids, which asserts that the sequence $\{I_k\}$, where I_k is the number of independent sets of size k in a matroid on a finite ground set, is ultra-log-concave. In Section 6 we describe some simple consequences of our results in the context of matroid theory, and we also discuss an application to bounding the entropy of the size of a random independent set in a claw-free graph.

Before stating our main results in detail, we briefly mention how this line of work connects with the growing body of work exploring applications of maximum entropy characterizations to discrete mathematics. The simplest maximum entropy result states that, among all probability distributions on a finite set S, the uniform has maximal entropy, $\log |S|$. While mathematically trivial, this result, combined with appropriate structural assumptions and various entropy inequalities, has been employed as a powerful tool and has seen varied applications in combinatorics. Examples include Radhakrishnan's entropy-based proof [46] of Bregman's theorem on the permanent of a 0-1 matrix, Kahn's proof [29] of the result of Kleitman and Markowski [35] on Dedekind's problem concerning the number of antichains in the Boolean lattice, the study by Brightwell and Tetali [8] of the number of linear extensions of the Boolean lattice (partially confirming a conjecture of Sha and Kleitman [47]), and the resolution of several conjectures of Imre Ruzsa in additive combinatorics by Madiman, Marcus and Tetali [40]. However, so far, a limitation of this line of work has been that it can only handle problems on finite sets. As modern combinatorics explores more and more properties of countable structures – such as infinite graphs or posets – it is natural that analogues of useful tools such as maximum entropy characterizations in countable settings should develop in parallel. It is particularly natural to develop these in connection with the Poisson and compound Poisson laws, which arise naturally in probabilistic combinatorics; see, e.g., Penrose's work [45] on geometric random graphs.

Section 2 contains our two main results: The maximum entropy characterization of log-concave compound Poisson distributions, and an analogous result for compound binomials. Sections 3 and 4, respectively, provide their proofs. Section 5 discusses conditions for log-concavity, and gives some additional results. Section 6 discusses applications to classical combinatorics, including graph theory and matroid theory. Section 7 contains some concluding remarks, a brief description of potential generalization and extensions, and a discussion of recent, subsequent work by Y. Yu [54], which was motivated

3

by preliminary versions of some of the present results.

2. Maximum Entropy Results

First we review the maximum entropy property of the Poisson distribution.

Definition 2.1. *For any parameter vector* $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ *with each* $p_i \in [0,1]$*, the sum of independent Bernoulli random variables* $B_i \sim \mathrm{Bern}\,(p_i)$*,*

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

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

$$\mathcal{P}_n(\lambda) = \big\{\mathbf{p} \in [0,1]^n \,:\, p_1 + p_2 + \cdots + p_n = \lambda\big\} \quad and \quad \mathcal{P}_\infty(\lambda) = \bigcup_{n \geq 1} \mathcal{P}_n(\lambda).$$

Shepp and Olkin [48] 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 $\mathrm{Bin}(n, \lambda/n)$, the binomial with parameters n and λ/n,

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

where $H(P) = -\sum_x P(x) \log P(x)$ denotes the discrete entropy function. Noting that the binomial $\mathrm{Bin}(n, \lambda/n)$ converges to the Poisson distribution $\mathrm{Po}(\lambda)$ as $n \to \infty$, 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 [21] noticed that a simple limiting argument gives the following maximum entropy property for the Poisson distribution:

$$H(\mathrm{Po}(\lambda)) = \sup\Big\{H(b_{\mathbf{p}}) \,:\, \mathbf{p} \in \mathcal{P}_\infty(\lambda)\Big\}. \tag{3}$$

A key property in generalizing and understanding this maximum entropy property further is that of ultra log-concavity; cf. [44]. The distribution P of a random variable X is *ultra log-concave* if $P(x)/\Pi_\lambda(x)$ is log-concave, that is, if,

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

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

Johnson [26] recently proved the following maximum entropy property for the Poisson distribution, generalizing (3):

$$H(\mathrm{Po}(\lambda)) = \max\Big\{H(P) \,:\, \text{ultra log-concave } P \text{ with mean } \lambda\Big\}. \tag{5}$$

As discussed in the Introduction, we wish to generalize the maximum entropy properties (2) and (3) to the case of *compound Poisson* distributions on $\mathbb{Z}_+$. We begin with some definitions:

4

Definition 2.2. *Let P be an arbitrary distribution on $\mathbb{Z}_+ = \{0, 1, \ldots\}$, and Q a distribution on $\mathbb{N} = \{1, 2, \ldots\}$. The Q-compound distribution $C_Q P$ is the distribution of the random sum,*

$$\sum_{j=1}^{Y} X_j, \tag{6}$$

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,*

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

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

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

Example 2.3. *Let Q be an arbitrary distribution on $\mathbb{N}$.*

1. *For any $0 \leq p \leq 1$, the* compound Bernoulli distribution $\mathrm{CBern}\,(p, Q)$ *is the distribution of the product BX, where $B \sim \mathrm{Bern}(p)$ and $X \sim Q$ are independent. It has probability mass function $C_Q P$, where P is the $\mathrm{Bern}\,(p)$ 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 $X_1, X_2, \ldots, X_n$ be i.i.d. with common distribution Q and $B_1, B_2, \ldots, B_n$ be independent $\mathrm{Bern}(p_i)$. We call,*

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

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

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 $\mathrm{CBin}(n, p, Q)$.*

4. *Let $\Pi_\lambda(x) = e^{-\lambda} \lambda^x / x!$, $x \geq 0$, denote the $\mathrm{Po}(\lambda)$ mass function. Then, for any $\lambda > 0$, the* compound Poisson distribution $\mathrm{CPo}(\lambda, Q)$ *is the distribution with mass function $C_Q \Pi_\lambda$:*

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

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,

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

But, perhaps somewhat surprisingly, as Chi [11] 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,

$$H(\mathrm{CBin}(2, \lambda/2, Q)) < 0.090798 < 0.090804 < H(C_Q b_{\mathbf{p}}). \tag{10}$$

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,

$$H(\mathrm{CPo}(\lambda, Q)) = \sup \Big\{ H(C_Q b_{\mathbf{p}}) \ : \ \mathbf{p} \in \mathcal{P}_\infty(\lambda) \Big\}. \tag{11}$$

Again, (11) fails in general. For example, taking the same Q, λ and $\mathbf{p}$ as above, yields,

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

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 first result (a more general version of which is proved in Section 3) states that, as long as Q and the compound Poisson measure $\mathrm{CPo}(\lambda, Q)$ are log-concave, the same maximum entropy statement as in (5) remains valid in the compound Poisson case:

Theorem 2.4. *If the distribution Q on $\mathbb{N}$ and the compound Poisson distribution $\mathrm{CPo}(\lambda, Q)$ are both log-concave, then,*

$$H(\mathrm{CPo}(\lambda, Q)) = \max \Big\{ H(C_Q P) \ : \ \text{ultra log-concave } P \text{ with mean } \lambda \Big\}.$$

The notion of log-concavity is central in the development of the ideas in this work. 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,

$$P(x)^2 \geq P(x+1)P(x-1), \quad \text{for all } x \geq 1. \tag{12}$$

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 ultra-log-concavity of P, defined as in equation (4), is more restrictive than log-concavity, and it is equivalent to the requirement that ratio P/Π_λ is a log-concave sequence for some (hence all) $\lambda > 0$.

Our second result states that (9) *does* hold, under certain conditions on Q and $\mathrm{CBin}(n, \lambda, Q)$:

Theorem 2.5. *If the distribution Q on $\mathbb{N}$ and the compound binomial distribution $\mathrm{CBin}(n, \lambda/n, Q)$ are both log-concave, then,*

$$H(\mathrm{CBin}(n, \lambda/n, Q)) = \max \Big\{ H(C_Q b_{\mathbf{p}}) \ : \ \mathbf{p} \in \mathcal{P}_n(\lambda) \Big\},$$

6

as long as the tail of Q satisfies either one of the following properties: (a) *Q has finite support; or* (b) *Q has tails heavy enough so that, for some $\rho, \beta > 0$ and $N_0 \geq 1$, we have, $Q(x) \geq \rho^{x^{\beta}}$, for all $x \geq N_0$.*

The proof of Theorem 2.5 is given in Section 4. As can be seen there, conditions (a) and (b) are introduced purely for technical reasons, and can probably be significantly relaxed; see Section 7 for a further discussion.

It remains an open question to give *necessary and sufficient* conditions on λ and Q for the compound Poisson and compound binomial distributions to have maximal entropy within an appropriately defined class. As a first step, one may ask for natural conditions that imply that a compound binomial or compound Poisson distribution is log-concave. We discuss several such conditions in Section 5.

In particular, the discussion in Section 5 implies the following explicit maximum entropy statements.

Example 2.6. 1. *Let Q be an arbitrary log-concave distribution on $\mathbb{N}$. Then Lemma 5.1 combined with Theorem 2.5 implies that the maximum entropy property of the compound binomial distribution in equation (9) holds, for all λ large enough. That is, the compound binomial $\mathrm{CBin}(n, \lambda/n, Q)$ has maximal entropy among all compound Bernoulli sums $C_Q b_{\mathbf{p}}$ with $p_1 + p_2 + \cdots + p_n = \lambda$, as long as $\lambda \geq \frac{nQ(2)}{Q(1)^2 + Q(2)}$.*

2. *Let Q be an arbitrary log-concave distribution on $\mathbb{N}$. Then Theorem 5.5 combined with Theorem 2.4 implies that the maximum entropy property of the compound Poisson $CPo(\lambda, Q)$ holds if and only if $\lambda \geq \frac{2Q(2)}{Q(1)^2}$.*

As mentioned in the introduction, the above results can be used in order to gain better understanding of ultra-log-concave sequences in combinatorics. Specifically, as discussed in more detail in Section 6, they can be used to estimate how "spread out" these sequences in terms of the entropy.

3. Maximum Entropy Property of the Compound Poisson Distribution

Here we show that, if Q and the compound Poisson distribution $CPo(\lambda, Q) = C_Q \Pi_\lambda$ are both log-concave, then $CPo(\lambda, Q)$ 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 [26].

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 [32, 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 [26] and [22] for details.

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

Recall the following definition from [26]:

Definition 3.1. *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. Bern (α), $Z_{\lambda(1-\alpha)}$ has distribution $\mathrm{Po}(\lambda(1-\alpha))$, 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 [26]: For all $y \geq 0$,

$$\frac{\partial}{\partial \alpha} U_\alpha P(y) = \frac{1}{\alpha} \left(\lambda(U_\alpha P(y) - U_\alpha P(y-1) - ((y+1)U_\alpha P(y+1) - y U_\alpha P(y))) \right). \tag{13}$$

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

Definition 3.2. *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 $C_Q U_\alpha P$:*

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

Work of Chafaï [10] suggests that the semigroup of Definition 3.1 may be viewed as the action of the $M/M/\infty$ queue. Similarly the semigroup of Definition 3.2 corresponds to the marginal distributions of a continuous-time hidden Markov process, where the underlying Markov process is the $M/M/\infty$ queue and the output at each time is obtained by a compounding operation.

Definition 3.3. *For a distribution P on $\mathbb{Z}_+$ with mean ν, its* size-biased *distribution $P^\#$ on $\mathbb{Z}_+$ is defined by*

$$P^\#(y) = \frac{(y+1)P(y+1)}{\nu}.$$

An important observation that will be at the heart of the proof of Theorem 2.4 below is that, for $\alpha = 0$, $U_0^Q P$ is simply the compound Poisson measure $\mathrm{CP}(\lambda, Q)$, while for $\alpha = 1$, $U_1^Q P = C_Q P$. The following lemma gives a rough bound on the third moment of $U_\alpha^Q P$:

Lemma 3.4. *Suppose P is an ultra log-concave distribution with mean $\lambda > 0$ on $\mathbb{Z}_+$, and let Q be a log-concave compounding distribution on $\mathbb{N}$. For each $\alpha \in [0,1]$, let W_α, V_α be random variables with distributions $U_\alpha^Q P = C_Q U_\alpha P$ and $C_Q (U_\alpha P)^\#$, respectively. Then the third moments $E(W_\alpha^3)$ and $E(V_\alpha^3)$ are both bounded above by,*

$$\lambda q_3 + 3\lambda^2 q_1 q_2 + \lambda^3 q_1^3,$$

where q_1, q_2, q_3 denote the first, second and third moments of Q, respectively.

Proof. Recall that, as stated in properties (ii) and (iii) in the beginning of Section 3, Q has finite moments of all orders, and that the nth falling factorial moment of any ultra log-concave random variable Y with distribution R on $\mathbb{Z}_+$ is bounded above by $(E(Y))^n$. Now for an arbitrary ultra log-concave distribution R, define random variables $Y \sim R$ and $Z \sim C_Q R$. If r_1, r_2, r_3 denote the first three moments of $Y \sim R$, then,

$$
\begin{aligned}
E(Z^3) &= q_3 r_1 + 3 q_1 q_2 E[(Y)_2] + q_1^3 E[(Y)_3] \\
&\leq q_3 r_1 + 3 q_1 q_2 r_1^2 + q_1^3 r_1^3.
\end{aligned}
\tag{14}
$$

Since the map U_α preserves ultra log-concavity [26], if P is ultra log-concave then so is $R = U_\alpha P$, so that (14) gives the required bound for the third moment of W_α, upon noting that the mean of the distribution $U_\alpha P$ is equal to λ.

Similarly, size-biasing preserves ultra log-concavity; that is, if R is ultra log-concave, then so is $R^\#$, since $R^\#(x+1)(x+1)/R^\#(x) = (R(x+2)(x+2)(x+1))/(R(x+1)(x+1)) = R(x+2)(x+2)/R(x+1)$ is also decreasing. Hence, $R' = (U_\alpha P)^\#$ is ultra log-concave, and (14) applies in this case as well. In particular, noting that the mean of $Y' \sim R' = (U_\alpha P)^\# = R^\#$ can be bounded in terms of the mean of $Y \sim R$ as,

$$
E(Y') = \sum_x x \frac{(x+1)U_\alpha P(x+1)}{\lambda} = \frac{E[(Y)_2]}{E(Y)} \leq \frac{\lambda^2}{\lambda} = \lambda,
$$

the bound (14) yields the required bound for the third moment of V_α. $\qquad\square$

In [26], the characterization of the Poisson as a maximum entropy distribution was proved through the decrease of its score function. In an analogous way, we define the score function of a Q-compound random variable as follows, cf. [4],

Definition 3.5. *Given a distribution P on $\mathbb{Z}_+$ with mean λ, the corresponding Q-compound distribution $C_Q P$ has score function defined by,*

$$
r_{1,C_Q P}(x) = \frac{C_Q(P^\#)(x)}{C_Q P(x)} - 1, \quad x \geq 0.
$$

More explicitly, one can write

$$
r_{1,C_Q P}(x) = \frac{\sum_{y=0}^\infty (y+1)P(y+1)Q^{*y}(x)}{\lambda \sum_{y=0}^\infty P(y)Q^{*y}(x)} - 1.
\tag{15}
$$

Notice that the mean of of $r_{1,C_Q P}$ with respect to $C_Q P$ is zero, and that if $P \sim \mathrm{Po}(\lambda)$ 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 [37]. But, unlike the scaled score function and an alternative score function given in [4], 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 [4].

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 3.6. *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 [33], which implies that, if Q is log-concave, then for any $m \geq n$, and for any x:

$$Q^{*m}(x+1)Q^{*n}(x) - Q^{*m}(x)Q^{*n}(x+1) \geq 0. \tag{16}$$

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

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

Since Q is log-concave, then so is Q^{*n}, cf. [30], so the ratio $Q^{*n}(x+1)/Q^{*n}(x)$ is decreasing in x, and (16) follows.]

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

$$\begin{aligned}
0 \quad\leq\quad & \left(\sum_y (y+1)P(y+1)Q^{*y}(x)\right)\left(\sum_z P(z)Q^{*z}(x+1)\right) \\
& - \left(\sum_y (y+1)P(y+1)Q^{*y}(x+1)\right)\left(\sum_z P(z)Q^{*z}(x)\right) \\
=\quad & \sum_{y,z}(y+1)P(y+1)P(z)\left[Q^{*y}(x)Q^{*z}(x+1) - Q^{*y}(x+1)Q^{*z}(x)\right]. \tag{17}
\end{aligned}$$

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 (17) becomes,

$$\sum_{y<z}\left[(y+1)P(y+1)P(z) - (z+1)P(z+1)P(y)\right]\left[Q^{*y}(x)Q^{*z}(x+1) - Q^{*y}(x+1)Q^{*z}(x)\right].$$

By the ultra log-concavity of P, the first square bracket is positive for $y \leq z$, and by equation (16) the second square bracket is also positive for $y \leq z$. $\qquad\square$

We remark that, under the same assumptions, and using a very similar argument, an analogous result holds for some alternative score functions recently introduced in [4] and in related work.

Combining Lemmas 3.6 and 3.4 with equation (13) we deduce the following result, which is the main technical step in the proof of Theorem 2.4 below.

Proposition 3.7. *Let P be an ultra log-concave distribution on $\mathbb{Z}_+$ with mean $\lambda > 0$, and assume that Q and $\mathrm{CPo}(\lambda, Q)$ 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, $E(0) \geq E(1)$.

10

Proof. Recall that,

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

where the last sum is restricted to the range $0 \leq y \leq x$, because Q is supported on $\mathbb{N}$. Therefore, since $U_\alpha P(x)$ is continuous in α [26], so is $U_\alpha^Q P(x)$, and to show that $E(\alpha)$ is continuous it suffices to show that the series,

$$E(\alpha) := E[-\log C_Q \Pi_\lambda (W_\alpha)] = -\sum_{x=0}^{\infty} U_\alpha^Q P(x) \log C_Q \Pi_\lambda(x), \tag{18}$$

converges uniformly. To that end, first observe that log-concavity of $C_Q \Pi_\lambda$ implies that $Q(1)$ is nonzero. [Otherwise, if $i > 1$ be the smallest integer i such that $Q(i) \neq 0$, then $C_Q \Pi_\lambda(i+1) = 0$, but $C_Q \Pi_\lambda(i)$ and $C_Q \Pi_\lambda(2i)$ are both strictly positive, contradicting the log-concavity of $C_Q \Pi_\lambda$.] Since $Q(1)$ is nonzero, we can bound the compound Poisson probabilities as,

$$1 \geq C_Q \Pi_\lambda(x) = \sum_y [e^{-\lambda} \lambda^y / y!] Q^{*y}(x) \geq e^{-\lambda} [\lambda^x / x!] Q(1)^x, \quad \text{for all } x \geq 1,$$

so that the summands in (18) can be bounded,

$$0 \leq -\log C_Q \Pi_\lambda(x) \leq \lambda + \log x! - x \log(\lambda Q(1)) \leq C x^2, \quad x \geq 1, \tag{19}$$

for a constant $C > 0$ that depends only on λ and $Q(1)$. Therefore, for any $N \geq 1$, the tail of the series (18) can be bounded,

$$0 \leq -\sum_{x=N}^{\infty} U_\alpha^Q P(x) \log C_Q \Pi_\lambda(x) \leq C E[W_\alpha^2 \mathbb{I}_{\{W_\alpha \geq N\}}] \leq \frac{C}{N} E[W_\alpha^3],$$

and, in view of Lemma 3.4, it converges uniformly.

Therefore, $E(\alpha)$ is continuous in α, and, in particular, convergent for all $\alpha \in [0, 1]$. To prove that it is differentiable at each $\alpha \in (0, 1)$ we need to establish that: (i) the summands in (18) are continuously differentiable in α for each x; and (ii) the series of derivatives converges uniformly.

Since, as noted above, $U_\alpha^Q P(x)$ is defined by a finite sum, we can differentiate with respect to α under the sum, to obtain,

$$\frac{\partial}{\partial \alpha} U_\alpha^Q P(x) = \frac{\partial}{\partial \alpha} C_Q U_\alpha P(x) = \sum_{y=0}^{x} \frac{\partial}{\partial \alpha} U_\alpha P(y) Q^{*y}(x). \tag{20}$$

And since $U_\alpha P$ is continuously differentiable in $\alpha \in (0, 1)$ for each x (cf. [26, Proposition 3.6] or equation (13) above), so are the summands in (18), establishing (i); in fact, they are infinitely differentiable, which can be seen by repeated applications of (13). To show that the series of derivatives converges uniformly, let α be restricted in an arbitrary open interval $(\epsilon, 1)$ for some $\epsilon > 0$. The relation (13) combined with

11

(20) yields, for any x,

$$\frac{\partial}{\partial \alpha} U_\alpha^Q P(x)$$

$$= \sum_{y=0}^{x} \left(\lambda (U_\alpha P(y) - U_\alpha P(y-1)) - ((y+1)U_\alpha P(y+1) - y U_\alpha P(y)) \right) Q^{*y}(x)$$

$$= -\frac{1}{\alpha} \sum_{y=0}^{x} ((y+1)U_\alpha P(y+1) - \lambda U_\alpha P(y)) \left(Q^{*y}(x) - Q^{*y+1}(x) \right)$$

$$= -\frac{1}{\alpha} \sum_{y=0}^{x} ((y+1)U_\alpha P(y+1) - \lambda U_\alpha P(y)) Q^{*y}(x)$$

$$\quad + \sum_{v=0}^{x} Q(v) \frac{1}{\alpha} \sum_{y=0}^{x} ((y+1)U_\alpha P(y+1) - \lambda U_\alpha P(y)) Q^{*y}(x-v)$$

$$= -\frac{\lambda}{\alpha} U_\alpha^Q P(x) \left(\frac{\sum_{y=0}^{x}(y+1)U_\alpha P(y+1)Q^{*y}(x)}{\lambda U_\alpha^Q P(x)} - 1 \right)$$

$$\quad + \frac{\lambda}{\alpha} \sum_{v=0}^{x} Q(v) U_\alpha^Q P(x-v) \left(\frac{\sum_{y=0}^{x}(y+1)U_\alpha P(y+1)Q^{*y}(x-v)}{\lambda U_\alpha^Q P(x-v)} - 1 \right)$$

$$= -\frac{\lambda}{\alpha} \left(U_\alpha^Q P(x) r_{1,U_\alpha^Q P}(x) - \sum_{v=0}^{x} Q(v) U_\alpha^Q P(x-v) r_{1,U_\alpha^Q P}(x-v) \right). \tag{21}$$

Also, for any x, by definition,

$$|U_\alpha^Q P(x) r_{1,U_\alpha^Q P}(x)| \leq C_Q (U_\alpha P)^{\#}(x) + U_\alpha^Q P(x),$$

where, for any distribution P, we write $P^{\#}(y) = P(y+1)(y+1)/\lambda$ for its size-biased version. Hence for any $N \geq 1$, equations (21) and (19) yield the bound,

$$\left| \sum_{x=N}^{\infty} \frac{\partial}{\partial \alpha} U_\alpha^Q P(x) \log C_Q \Pi_\lambda(x) \right|$$

$$\leq \sum_{x=N}^{\infty} \frac{C \lambda x^2}{\alpha} \left\{ C_Q (U_\alpha P)^{\#}(x) + U_\alpha^Q P(x) + \sum_{v=0}^{x} Q(v)[C_Q(U_\alpha P)^{\#}(x-v) + U_\alpha^Q P(x-v)] \right\}$$

$$= \frac{2C}{\alpha} E \left[\left(V_\alpha^2 + W_\alpha^2 + X^2 + X V_\alpha + X W_\alpha \right) \mathbb{I}_{\{V_\alpha \geq N, \, W_\alpha \geq N, \, X \geq N\}} \right]$$

$$\leq \frac{C'}{\alpha} \left\{ E[V_\alpha^2 \mathbb{I}_{\{V_\alpha \geq N\}}] + E[W_\alpha^2 \mathbb{I}_{\{W_\alpha \geq N\}}] + E[X^2 \mathbb{I}_{\{X \geq N\}}] \right\}$$

$$\leq \frac{C'}{N \alpha} \left\{ E[V_\alpha^3] + E[W_\alpha^3] + E[X^3] \right\},$$

where $C, C' > 0$ are appropriate finite constants, and the random variables $V_\alpha \sim C_Q(U_\alpha P)^{\#}$, $W_\alpha \sim U_\alpha^Q P$ and $X \sim Q$ are independent. Lemma 3.4 implies that this bound converges to zero uniformly

12

in $\alpha \in (\epsilon, 1)$, as $N \to \infty$. Since $\epsilon > 0$ was arbitrary, this establishes that $E(\alpha)$ is differentiable for all $\alpha \in (0, 1)$ and, in fact, that we can differentiate the series (18) term-by-term, to obtain,

$$E'(\alpha) = -\sum_{x=0}^{\infty} \frac{\partial}{\partial \alpha} U_\alpha^Q P(x) \log C_Q \Pi_\lambda(x) \tag{22}$$

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

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

where the second equality follows from using (21) above, and 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 [26], 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 $U_\alpha P$ [26], hence the score function $r_{1,U_\alpha^Q P}(x)$ is decreasing in x, by Lemma 3.6. 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 $\alpha \in (0, 1)$. $\square$

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.

We can now state and prove a slightly more general form of Theorem 2.4. Recall that the relative entropy between distributions P and Q on $\mathbb{Z}_+$, denoted by $D(P\|Q)$, is defined by

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

Theorem 3.8. *Let P be an ultra-log-concave distribution on $\mathbb{Z}_+$ with mean λ. If the distribution Q on $\mathbb{N}$ and the compound Poisson distribution $C_Q \Pi_\lambda$ are both log-concave, then*

$$D(C_Q P \| C_Q \Pi_\lambda) \leq H(C_Q \Pi_\lambda) - H(C_Q P).$$

Proof. As in Proposition 3.7, let $W_\alpha \sim U_\alpha^Q P = C_Q U_\alpha P$. Noting that $W_0 \sim C_Q \Pi_\lambda$ and $W_1 \sim C_Q P$, we have

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

where the inequality is exactly the statement that $E(1) \leq E(0)$, proved in Proposition 3.7. $\square$

Since $0 \leq D(C_Q P \| C_Q \Pi_\lambda)$, Theorem 2.4 immediately follows.

13

4. Maximum Entropy Property of the Compound Binomial Distribution

Here we prove the maximum entropy result for compound binomial random variables, Theorem 2.5. The proof, to some extent, parallels some of the arguments in [21][42][48], which rely on differentiating the compound-sum probabilities $b_{\mathbf{p}}(x)$ for a given parameter vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ (recall Definition 2.1 in the Introduction), with respect to an individual p_i. Using the representation,

$$C_Q b_{\mathbf{p}}(y) = \sum_{x=0}^{n} b_{\mathbf{p}}(x) Q^{*x}(y), \quad y \geq 0, \tag{23}$$

differentiating $C_Q b_{\mathbf{p}}(x)$ reduces to differentiating $b_{\mathbf{p}}(x)$, and leads to an expression equivalent to that derived earlier in (21) for the derivative of $C_Q U_\alpha P$ with respect to α.

Lemma 4.1. *Given a parameter vector* $\mathbf{p} = (p_1, p_2, \ldots, p_n)$, *with* $n \geq 2$ *and each* $0 \leq p_i \leq 1$, *let,*

$$\mathbf{p_t} = \left(\frac{p_1 + p_2}{2} + t, \frac{p_1 + p_2}{2} - t, p_3, \ldots, p_n \right),$$

for $t \in [-(p_1 + p_2)/2, (p_1 + p_2)/2]$. *Then,*

$$\frac{\partial}{\partial t} C_Q b_{\mathbf{p_t}}(x) = (-2t) \sum_{y=0}^{n} b_{\widetilde{\mathbf{p}}}(y) \left(Q^{*(y+2)}(x) - 2Q^{*(y+1)}(x) + Q^{*y}(x) \right), \tag{24}$$

where $\widetilde{\mathbf{p}} = (p_3, \ldots, p_n)$.

Proof. Note that the sum of the entries of $\mathbf{p}_t$ is constant as t varies, and that $\mathbf{p_t} = \mathbf{p}$ for $t = (p_1 - p_2)/2$, while $\mathbf{p_t} = ((p_1 + p_2)/2, (p_1 + p_2)/2, p_3, \ldots, p_n)$ for $t = 0$. Writing $k = p_1 + p_2$, $b_{\mathbf{p_t}}$ can be expressed,

$$
\begin{aligned}
b_{\mathbf{p_t}}(y) &= \left(\frac{k^2}{4} - t^2 \right) b_{\widetilde{\mathbf{p}}}(y - 2) + \left(k \left(1 - \frac{k}{2} \right) + 2t^2 \right) b_{\widetilde{\mathbf{p}}}(y - 1) \\
&\quad + \left(\left(1 - \frac{k}{2} \right)^2 - t^2 \right) b_{\widetilde{\mathbf{p}}}(y),
\end{aligned}
$$

and its derivative with respect to t is,

$$\frac{\partial}{\partial t} b_{\mathbf{p_t}}(y) = -2t \left(b_{\widetilde{\mathbf{p}}}(y - 2) - 2 b_{\widetilde{\mathbf{p}}}(y - 1) + b_{\widetilde{\mathbf{p}}}(y) \right).$$

The expression (23) for $C_Q b_{\mathbf{p_t}}$ shows that it is a finite linear combination of compound-sum probabilities $b_{\mathbf{p_t}}(x)$, so we can differentiate inside the sum to obtain,

$$
\begin{aligned}
\frac{\partial}{\partial t} C_Q b_{\mathbf{p_t}}(x) &= \sum_{y=0}^{n} \frac{\partial}{\partial t} b_{\mathbf{p_t}}(y) Q^{*y}(x) \\
&= -2t \sum_{y=0}^{n} \left(b_{\widetilde{\mathbf{p}}}(y - 2) - 2 b_{\widetilde{\mathbf{p}}}(y - 1) + b_{\widetilde{\mathbf{p}}}(y) \right) Q^{*y}(x) \\
&= -2t \sum_{y=0}^{n-2} b_{\widetilde{\mathbf{p}}}(y) \left(Q^{*(y+2)}(x) - 2Q^{*(y+1)}(x) + Q^{*y}(x) \right),
\end{aligned}
$$

since $b_{\widetilde{\mathbf{p}}}(y) = 0$ for $y \leq -1$ and $y \geq n - 1$. $\qquad \square$

14

Next we state and prove the equivalent of Proposition 3.7 above. Note that the distribution of a compound Bernoulli sum is invariant under permutations of the Bernoulli parameters p_i. Therefore, the assumption $p_1 \geq p_2$ is made below without loss of generality.

Proposition 4.2. *Suppose that the distribution Q on $\mathbb{N}$ and the compound binomial distribution* $\mathrm{CBin}(n, \lambda/n, Q)$ *are both log-concave; let* $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ *be a given parameter vector with* $n \geq 2$, $p_1 + p_2 + \ldots + p_n = \lambda > 0$, *and* $p_1 \geq p_2$; *let* W_t *be a random variable with distribution* $C_Q b_{\mathbf{p_t}}$; *and define, for all* $t \in [0, (p_1 - p_2)/2]$, *the function,*

$$E(t) := E[-\log C_Q b_{\overline{\mathbf{p}}}(W_t)],$$

where $\overline{\mathbf{p}}$ *denotes the parameter vector with all entries equal to* λ/n. *If* Q *satisfies either of the conditions:* (a) Q *finite support; or* (b) Q *has tails heavy enough so that, for some* $\rho, \beta > 0$ *and* $N_0 \geq 1$, *we have,* $Q(x) \geq \rho^{x^\beta}$, *for all* $x \geq N_0$, *then* $E(t)$ *is continuous for all* $t \in [0, (p_1 - p_2)/2]$, *it is differentiable for* $t \in (0, (p_1 - p_2)/2)$, *and, moreover,* $E'(t) \leq 0$ *for* $t \in (0, (p_1 - p_2)/2)$. *In particular,* $E(0) \geq E((p_1 - p_2)/2)$.

Proof. The compound distribution $C_Q b_{\mathbf{p_t}}$ is defined by the finite sum,

$$C_Q b_{\mathbf{p_t}}(x) = \sum_{y=0}^{n} b_{\mathbf{p_t}}(y) Q^{*y}(x),$$

and is, therefore, continuous in t. First, assume that Q has finite support. Then so does $C_Q b_{\mathbf{p}}$ for any parameter vector $\mathbf{p}$, and the continuity and differentiability of $E(t)$ are trivial. In particular, the series defining $E(t)$ is a finite sum, so we can differentiate term-by-term, to obtain,

$$
\begin{aligned}
E'(t) &= -\sum_{x=0}^{\infty} \frac{\partial}{\partial t} C_Q b_{\mathbf{p_t}}(x) \log C_Q b_{\overline{\mathbf{p}}}(x) \\
&= 2t \sum_{x=0}^{\infty} \sum_{y=0}^{n-2} b_{\widetilde{\mathbf{p}}}(y) \left(Q^{*(y+2)}(x) - 2Q^{*(y+1)}(x) + Q^{*y}(x) \right) \log C_Q b_{\overline{\mathbf{p}}}(x) \qquad (25) \\
&= 2t \sum_{y=0}^{n-2} \sum_{z=0}^{\infty} b_{\widetilde{\mathbf{p}}}(y) Q^{*y}(z) \sum_{v,w} Q(v) Q(w) \Bigg[\log C_Q b_{\overline{\mathbf{p}}}(z + v + w) - \log C_Q b_{\overline{\mathbf{p}}}(z + v) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \log C_Q b_{\overline{\mathbf{p}}}(z + w) + \log C_Q b_{\overline{\mathbf{p}}}(z) \Bigg], \quad (26)
\end{aligned}
$$

where (25) follows by Lemma 4.1. By assumption, the distribution $C_Q b_{\overline{\mathbf{p}}} = \mathrm{CBin}(n, \lambda/n, Q)$ is log-concave, which implies that, for all z, v, w such that $z + v + w$ is in the support of $\mathrm{CBin}(n, \lambda/n, Q)$,

$$\frac{C_Q b_{\overline{\mathbf{p}}}(z)}{C_Q b_{\overline{\mathbf{p}}}(z + v)} \leq \frac{C_Q b_{\overline{\mathbf{p}}}(z + w)}{C_Q b_{\overline{\mathbf{p}}}(z + v + w)}.$$

Hence the term in square brackets in equation (26) is negative, and the result follows.

Now, suppose condition (b) holds on the tails of Q. First we note that the moments of W_t are all uniformly bounded in t: Indeed, for any $\gamma > 0$,

$$E[W_t^\gamma] = \sum_{x=0}^{\infty} C_Q b_{\mathbf{p_t}}(x) x^\gamma = \sum_{x=0}^{\infty} \sum_{y=0}^{n} b_{\mathbf{p_t}}(y) Q^{*y}(x) x^\gamma \leq \sum_{y=0}^{n} \sum_{x=0}^{\infty} Q^{*y}(x) x^\gamma \leq C_n q_\gamma, \qquad (27)$$

15

where C_n is a constant depending only on n, and q_γ is the γth moment of Q, which is of course finite; recall property (ii) in the beginning of Section 3.

For the continuity of $E(t)$, it suffices to show that the series,

$$E(t) := E[-\log C_Q b_{\overline{\mathbf{P}}}(W_t)] = -\sum_{x=0}^{\infty} C_Q b_{\mathbf{P_t}}(x) \log C_Q b_{\overline{\mathbf{P}}}(x), \tag{28}$$

converges uniformly. The tail assumption on Q implies that, for all $x \geq N_0$,

$$1 \geq C_Q b_{\overline{\mathbf{P}}}(x) = \sum_{y=0}^{n} b_{\overline{\mathbf{P}}}(y) Q^{*y}(x) \geq \lambda(1-\lambda/n)^{n-1} Q(x) \geq \lambda(1-\lambda/n)^{n-1} \rho^{x^\beta},$$

so that,

$$0 \leq -\log C_Q b_{\overline{\mathbf{P}}}(x) \leq Cx^\beta, \tag{29}$$

for an appropriate constant $C > 0$. Then, for $N \geq N_0$, the tail of the series (28) can be bounded,

$$0 \leq -\sum_{x=N}^{\infty} C_Q b_{\mathbf{P_t}}(x) \log C_Q b_{\overline{\mathbf{P}}}(x) \leq CE[W_t^\beta \mathbb{I}_{\{W_t \geq N\}}] \leq \frac{C}{N} E[W_t^{\beta+1}] \leq \frac{C}{N} C_n q_{\beta+1},$$

where the last inequality follows from (27). This obviously converges to zero, uniformly in t, therefore $E(t)$ is continuous.

For the differentiability of $E(t)$, note that the summands in (18) are continuously differentiable (by Lemma 4.1), and that the series of derivatives converges uniformly in t; to see that, for $N \geq N_0$ we apply Lemma 4.1 together with the bound (29) to get,

$$\left| \sum_{x=N}^{\infty} \frac{\partial}{\partial t} C_Q b_{\mathbf{P_t}}(x) \log C_Q b_{\overline{\mathbf{P}}}(x) \right|$$

$$\leq 2t \sum_{x=N}^{\infty} \sum_{y=0}^{n} b_{\widetilde{\mathbf{P}}}(y) \left(Q^{*(y+2)}(x) + 2Q^{*(y+1)}(x) + Q^{*y}(x) \right) Cx^\beta$$

$$\leq 2Ct \sum_{y=0}^{n} \sum_{x=N}^{\infty} \left(Q^{*(y+2)}(x) + 2Q^{*(y+1)}(x) + Q^{*y}(x) \right) x^\beta,$$

which is again easily seen to converge to zero uniformly in t as $N \to \infty$, since Q has finite moments of all orders. This establishes the differentiability of $E(t)$ and justifies the term-by-term differentiation of the series (18); the rest of the proof that $E'(t) \leq 0$ is the same as in case (a). $\square$

Note that, as with Proposition 3.7, the above proof only requires that the compound binomial distribution $\mathrm{CBin}(n, \lambda/n, Q) = C_Q b_{\overline{\mathbf{P}}}$ satisfies a property weaker than log-concavity, namely that the function, $\log C_Q b_{\overline{\mathbf{P}}}(x) - \sum_v Q(v) \log C_Q b_{\overline{\mathbf{P}}}(x+v)$, be increasing in x.

Proof. **(of Theorem 2.5)** Assume, without loss of generality, that $n \geq 2$. If $p_1 > p_2$, then Proposition 4.2 says that, $E((p_1 - p_2)/2) \leq E(0)$, that is,

$$-\sum_{x=0}^{\infty} C_Q b_{\mathbf{P}}(x) \log C_Q b_{\overline{\mathbf{P}}}(x) \leq -\sum_{x=0}^{\infty} C_Q b_{\mathbf{P}^*}(x) \log C_Q b_{\overline{\mathbf{P}}}(x),$$

where $\mathbf{p}^* = ((p_1 + p_2)/2, (p_1 + p_2)/2, p_3, \ldots p_n)$ and $\overline{\mathbf{p}} = (\lambda/n, \ldots, \lambda/n)$. Since the expression in the above right-hand-side is invariant under permutations of the elements of the parameter vectors, we deduce that it is maximized by $\mathbf{p_t} = \overline{\mathbf{p}}$. Therefore, using, as before, the nonnegativity of the relative entropy,

$$
\begin{aligned}
H(C_Q b_{\mathbf{p}}) &\leq H(C_Q b_{\mathbf{p}}) + D(C_Q b_{\mathbf{p}} \| C_Q b_{\overline{\mathbf{p}}}) \\
&= -\sum_{x=0}^{\infty} C_Q b_{\mathbf{p}}(x) \log C_Q b_{\overline{\mathbf{p}}}(x) \\
&\leq -\sum_{x=0}^{\infty} C_Q b_{\overline{\mathbf{p}}}(x) \log C_Q b_{\overline{\mathbf{p}}}(x) \\
&= H(C_Q b_{\overline{\mathbf{p}}}) = H(\mathrm{CBin}(n, \lambda/n, Q)),
\end{aligned}
$$

as claimed. □

Clearly one can also state a slightly more general version of Theorem 2.5 analogous to Theorem 3.8.

5. Conditions for Log-Concavity

Theorems 2.4 and 2.5 state that log-concavity is a sufficient condition for compound binomial and compound Poisson distributions to have maximal entropy within a natural class. In this section, we discuss when log-concavity holds.

Recall that Steutel and van Harn [50, Theorem 2.3] showed that, if $\{iQ(i)\}$ is a decreasing sequence, then $\mathrm{CPo}(\lambda, Q)$ is a unimodal distribution, which is a necessary condition for log-concavity. Interestingly, the same condition provides a dichotomy of results in compound Poisson approximation bounds as developed by Barbour, Chen and Loh [3]: If $\{iQ(i)\}$ is decreasing, then the bounds are of the same form and order as in the Poisson case, otherwise the bounds are much larger. In a slightly different direction, Cai and Willmot [9, Theorem 3.2] showed that if $\{Q(i)\}$ is decreasing then the distribution function of the compound Poisson distribution $\mathrm{CPo}(\lambda, Q)$ is log-concave. Finally, Keilson and Sumita [33, Theorem 4.9] proved that, if Q is log-concave, then the ratio,

$$
\frac{C_Q \Pi_\lambda(n)}{C_Q \Pi_\lambda(n+1)},
$$

is decreasing in λ for any fixed n.

In the present context, we first show that a compound Bernoulli sum is log-concave if the compounding distribution Q is log-concave and the Bernoulli parameters are sufficiently large.

Lemma 5.1. *Suppose Q is a log-concave distribution on $\mathbb{N}$, and all the elements p_i of the parameter vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ satisfy $p_i \geq \frac{1}{1+Q(1)^2/Q(2)}$. Then the compound Bernoulli sum distribution $C_Q b_{\mathbf{p}}$ is log-concave.*

Proof. Observe that, given that Q is log-concave, the compound Bernoulli distribution $\mathrm{CBern}\,(p, Q)$ is log-concave if and only if,

$$
p \geq \frac{1}{1 + Q(1)^2/Q(2)}. \tag{30}
$$

Indeed, let Y have distribution CBern(p, Q). Since Q is log-concave itself, the log-concavity of CBern(p, Q) is equivalent to the inequality, $\Pr(Y = 1)^2 \geq \Pr(Y = 2)\Pr(Y = 0)$, which states that, $(pQ(1))^2 \geq (1-p)pQ(2)$, and this is exactly the assumption (30).

The assertion of the lemma now follows since the sum of independent log-concave random variables is log-concave; see, e.g., [30]. $\qquad\square$

Next we examine conditions under which a compound Poisson measure is log-concave, starting with a simple necessary condition.

Lemma 5.2. *A necessary condition for* $\mathrm{CPo}(\lambda, Q)$ *to be log-concave is that,*

$$\lambda \geq \frac{2Q(2)}{Q(1)^2}. \tag{31}$$

Proof. For any distribution P, considering the difference, $C_Q P(1)^2 - C_Q P(0)C_Q P(2)$, shows that a necessary condition for $C_Q P$ to be log-concave is that,

$$(P(1)^2 - P(0)P(2))/P(0)P(1) \geq Q(2)/Q(1)^2. \tag{32}$$

Now take P to be the Po(λ) distribution. $\qquad\square$

Similarly, for $P = b_{\mathbf{p}}$, a necessary condition for the compound Bernoulli sum $C_Q b_{\mathbf{p}}$ to be log-concave is that,

$$\sum_i \frac{p_i}{1-p_i} + \left(\sum_i \frac{p_i^2}{(1-p_i)^2}\right)\left(\sum_i \frac{p_i}{1-p_i}\right)^{-1} \geq \frac{2Q(2)}{Q(1)^2},$$

which, since the left-hand-side is greater than $\sum_i p_i/(1-p_i) \geq \sum_i p_i$, will hold as long as $\sum_i p_i \geq 2Q(2)/Q(1)^2$.

Note that, unlike for the Poisson distribution, it is not the case that every compound Poisson distribution $\mathrm{CPo}(\lambda, Q)$ is log-concave.

Next we show that for some particular choices of Q and general compound distributions $C_Q P$, the above necessary condition is sufficient for log-concavity.

Theorem 5.3. *Let Q be a geometric distribution on $\mathbb{N}$. Then $C_Q P$ is log-concave for any distribution P which is log-concave and satisfies the condition* (32).

Proof. If Q is geometric with mean $1/\alpha$, then, $Q^{*y}(x) = \alpha^y(1-\alpha)^{x-y}\binom{x-1}{y-1}$, which implies that,

$$C_Q P(x) = \sum_{y=0}^{x} P(y)\alpha^y(1-\alpha)^{x-y}\binom{x-1}{y-1}.$$

Condition (32) ensures that $C_Q P(1)^2 - C_Q P(0)C_Q P(2) \geq 0$, so, taking $z = y - 1$, we need only prove that the sequence,

$$C(x) := C_Q P(x+1)/(1-\alpha)^x = \sum_{z=0}^{x} P(z+1)\left(\frac{\alpha}{1-\alpha}\right)^{z+1}\binom{x}{z}$$

18

is log-concave. However, this follows immediately from [30, Theorem 7.3], which proves that if $\{a_i\}$ is a log-concave sequence, then so is $\{b_i\}$, defined by $b_i = \sum_{j=0}^{i} \binom{i}{j} a_j$. □

Theorem 5.4. *Let Q be a distribution supported on the set $\{1,2\}$. Then the distribution $C_Q P$ is log-concave for any ultra log-concave distribution P with support on $\{0,1,\ldots,N\}$ (where N may be infinite), which satisfies*

$$(x+1)P(x+1)/P(x) \geq 2Q(2)/Q(1)^2 \tag{33}$$

for all $x = 0,1,\ldots,N$.

In particular, if Q is supported on $\{1,2\}$, the compound Poisson distribution $\mathrm{CPo}(\lambda,Q)$ is log-concave for all $\lambda \geq \frac{2Q(2)}{Q(1)^2}$.

Note that the condition (33) is equivalent to requiring that $NP(N)/P(N-1) \geq 2Q(2)/Q(1)^2$ if N is finite, or that $\lim_{x\to\infty}(x+1)P(x+1)/P(x) \geq 2Q(2)/Q(1)^2$ if N is infinite.

The proof of Theorem 5.4 is based in part on some of the ideas in Johnson and Goldschmidt [27], and also in Wang and Yeh [53], where transformations that preserve log-concavity are studied. Since the proof is slightly involved and the compound Poisson part of the theorem is superseded by Theorem 5.5 below, we give it in the appendix.

Lemma 5.2 and Theorems 5.3 and 5.4, supplemented by some calculations of the quantities $C_Q\Pi_\lambda(x)^2 - C_Q\Pi_\lambda(x-1)C_Q\Pi_\lambda(x+1)$ for small x, suggest that compound Poisson measure $CPo(\lambda,Q)$ should be log-concave, as long as Q is log-concave and $\lambda Q(1)^2 \geq 2Q(2)$. Indeed, the following slightly more general result holds; see Section 7 for some remarks on its history. As per Definition 3.3, we use $Q^\#$ to denote the size-biased version of Q. Observe that log-concavity of $Q^\#$ is a weaker requirement than log-concavity of Q.

Theorem 5.5. *If $Q^\#$ is log-concave and $\lambda Q(1)^2 \geq 2Q(2)$, then the compound Poisson measure $\mathrm{CPo}(\lambda,Q)$ is log-concave.*

Proof. It is well-known that compound Poisson probability mass functions obey a recursion formula:

$$kC_Q\Pi_\lambda(k) = \lambda\sum_{j=1}^{k} jQ(j)C_Q\Pi_\lambda(k-j) \quad \text{for all } k \in \mathbb{N}. \tag{34}$$

(This formula, which is easy to prove for instance using characteristic functions, has been repeatedly rediscovered; the earliest reference we could find was to a 1958 note of Katti and Gurland mentioned by N. de Pril [15], but later references are Katti [31], Adelson [1] and Panjer [43]; in actuarial circles, the above is known as the Panjer recursion formula.) For notational convenience, we write μ_Q for the mean of Q, $r_j = \lambda(j+1)Q(j+1) = \lambda\mu_Q Q^\#(j)$, and $p_j = C_Q\Pi_\lambda(j)$ for $j \in \mathbb{Z}_+$. Then (34) reads,

$$(k+1)p_{k+1} = \sum_{j=0}^{k} r_j p_{k-j}$$

for all $k \in \mathbb{Z}_+$.

19

Theorem 5.5 is just a restatement using (34) of [20, Theorem 1]. For completeness, we sketch the proof of Hansen [20], which proceeds by induction. Note that one only needs to prove the following statement: If $Q^{\#}$ is strictly log-concave and $\lambda Q(1)^2 > 2Q(2)$, then the compound Poisson measure $\mathrm{CPo}(\lambda, Q)$ is strictly log-concave. The general case follows by taking limits.

By assumption, $\lambda Q(1)^2 > 2Q(2)$, which can be rewritten as $r_0^2 > r_1$, and hence,

$$p_1^2 - p_0 p_2 = \frac{p_0^2}{2}(r_0^2 - r_1) > 0.$$

This initializes the induction procedure by showing that the subsequence (p_0, p_1, p_2) is strictly log-concave. Hansen [20] developed the following identity, which can be verified using the recursion (34): Setting $p_{-1} = 0$,

$$m(m+2)[p_{m+1}^2 - p_m p_{m+2}] = p_{m+1}(r_0 p_m - p_{m+1})$$
$$+ \sum_{l=0}^{m} \sum_{k=0}^{l} (p_{m-l} p_{m-k-1} - p_{m-k} p_{m-l-1})(r_{k+1} r_l - r_{l+1} r_k). \tag{35}$$

Observe that each term in the double sum is positive as a consequence of the induction hypothesis (namely, that the subsequence $(p_0, p_1, \ldots, p_{m+1})$ is strictly log-concave), and the strict log-concavity of r. To see that the first term is also positive, note that the induction hypothesis implies that p_{k+1}/p_k is decreasing for $k \leq m+1$; hence,

$$r_0 = \frac{p_1}{p_0} > \frac{p_{m+1}}{p_m}.$$

Thus it is shown that $p_{m+1}^2 > p_m p_{m+2}$, which proves the theorem. $\qquad\square$

We note that Hansen's remarkable identity (35) is reminiscent of (although more complicated than) an identity that can be used to prove the well-known fact that the convolution of two log-concave sequences is log-concave. Indeed, as shown for instance in Liggett [38], if $c = a \star b$, then,

$$c_k^2 - c_{k-1} c_{k+1} = \sum_{i<j} (a_i a_{j-1} - a_{i-1} a_j)(b_{k-i} b_{k-j+1} - b_{k-i+1} b_{k-j}).$$

Observe that (34) can be interpreted as saying that the size-biased version of $C_Q \Pi_\lambda$ is the convolution of the sequence r with the sequence p.

It is interesting to note that the condition $Q(2) > 0$ (which would imply $Q(1) > 0$ under the other assumptions of Theorem 5.5) is not necessary. For instance, one can consider the case where Q is the unit mass at 3, which is the "Poisson" distribution supported at multiples of 3. In the continuous setting, there do not exist log-concave distributions with disconnected supports, but in our discrete setting, such compound Poissons are also log-concave.

6. Applications to Combinatorics

There are numerous examples of ultra-log-concave sequences in discrete mathematics, and also many examples of interesting sequences where ultra-log-concavity is conjectured. The above maximum entropy results for ultra-log-concave probability distributions on $\mathbb{Z}_+$ yield bounds on the "spread" of such ultra-log-concave sequences, as measured by entropy. Two particular examples are considered below.

6.1. Counting independent sets in a claw-free graph

Recall that for a graph $G = (V, E)$, an independent set is a subset of the vertex set V such that no two elements of the subset are connected by an edge in E. The collection of independent sets of G is denoted $\mathcal{I}(G)$.

Consider a graph G on a randomly weighted ground set, i.e., associate with each $i \in V$ the random weight X_i drawn from a probability distribution Q on $\mathbb{N}$, and suppose the weights $\{X_i : i \in V\}$ are independent. Then for any independent set $I \in \mathcal{I}(G)$, its weight is given by the sum of the weights of its elements,

$$w(I) = \sum_{i \in I} X_i \overset{\mathcal{D}}{=} \sum_{i=1}^{|I|} X_i',$$

where $\overset{\mathcal{D}}{=}$ denotes equality in distribution, and X_i' are i.i.d. random variables drawn from Q. For the weight of a random independent set $\mathbb{I}$ (picked uniformly at random from $\mathcal{I}(G)$), one similarly has,

$$w(\mathbb{I}) = \sum_{i \in \mathbb{I}} X_i \overset{\mathcal{D}}{=} \sum_{i=1}^{|\mathbb{I}|} X_i',$$

and the latter, by definition, has distribution $C_Q P$, where P is the probability distribution on $\mathbb{Z}_+$ induced by $|\mathbb{I}|$.

The following result of Hamidoune [19] (see also Chudnovsky and Seymour [12] for a generalization and different proof) connects this discussion with ultra-log-concavity. Recall that a claw-free graph is a graph that does not contain the complete bipartite graph $K_{1,3}$ as an induced subgraph.

Theorem 6.1 (Hamidoune [19]). *For a claw-free finite graph G, the sequence $\{I_k\}$, where I_k is the number of independent sets of size k in G, is ultra-log-concave.*

Clearly, Theorem 6.1 may be restated as follows: For a random independent set $\mathbb{I}$,

$$P(k) := \Pr\{|\mathbb{I}| = k\} \propto I_k,$$

is an ultra-log-concave probability distribution. This yields the following corollary.

Corollary 6.2. *Suppose the graph G on the ground set V is claw-free. Let $\mathbb{I}$ be a random independent set, and let the average cardinality of $\mathbb{I}$ be λ. Suppose the elements of the ground set are given i.i.d. weights drawn from a probability distribution Q on $\mathbb{N}$, where Q is log-concave and satisfies $\lambda Q(1)^2 \geq 2Q(2)$. If W is the random weight assigned to $\mathbb{I}$, then,*

$$H(W) \leq H(C_Q \Pi_\lambda).$$

If Q is the unit mass at 1, then $W = |\mathbb{I}|$, and Corollary 6.2 gives a bound on the entropy of the cardinality of a random independent set in a claw-free graph. That is,

$$H(|\mathbb{I}|) \leq H(\Pi_\lambda),$$

where $\lambda = E|\mathbb{I}|$. Observe that this bound is independent of n and depends *only* on the average size of a random independent set, which suggests that it could be of utility in studying sequences associated with

graphs on large ground sets. And, although the entropy of a Poisson (or compound Poisson) measure cannot easily expressed in closed form, there are various simple bounds [13, Theorem 8.6.5] such as,

$$H(\Pi_\lambda) \leq \frac{1}{2} \log \left[2\pi e \left(\lambda + \frac{1}{12} \right) \right],$$ (36)

as well as good approximations for large λ; see, e.g., [16, 24, 36]. One way to use this bound is via the following crude relaxation: Bound the average size λ of a random independent set by the independence number $\alpha(G)$ of G, which is defined as the size of a largest independent set of G. Then,

$$H(|\mathbb{I}|) \leq \frac{1}{2} \log \left[2\pi e \left(\alpha(G) + \frac{1}{12} \right) \right],$$ (37)

which can clearly be much tighter than the trivial bound, $H(|\mathbb{I}|) \leq \log \alpha(G)$, using the uniform distribution, when $\alpha(G) > 16$.

6.2. Mason's conjecture

Recall that a matroid M on a finite ground set $[n]$ is a collection of subsets of $[n]$, called "independent sets"[2], satisfying the following: (i) The empty set is independent. (ii) Every subset of an independent set is independent. (iii) If A and B are two independent sets and A has more elements than B, then there exists an element in A which is not in B and when added to B still gives an independent set.

Consider a matroid M on a randomly weighted ground set, i.e., associate with each $i \in [n]$ the random weight X_i drawn from a probability distribution Q on $\mathbb{N}$, and suppose the weights $\{X_i : i \in [n]\}$ are independent. As before, for any independent set $I \in M$, its weight is given by the sum of the weights of its elements, and the weight of a random independent set $\mathbb{I}$ (picked uniformly at random from M), is,

$$w(\mathbb{I}) = \sum_{i \in \mathbb{I}} X_i \overset{\mathcal{D}}{=} \sum_{i=1}^{|\mathbb{I}|} X_i',$$

where the X_i' are i.i.d. random variables drawn from Q. Then $w(\mathbb{I})$ has distribution $C_Q P$, where P is the probability distribution on $\mathbb{Z}_+$ induced by $|\mathbb{I}|$.

Conjecture 6.3 (Mason [41]). *The sequence $\{I_k\}$, where I_k is the number of independent sets of size k in a matroid on a finite ground set, is ultra-log-concave.*

Strictly speaking, Mason's original conjecture asserts ultra-log-concavity of some finite order (not defined in this paper) whereas this paper is only concerned with ultra-log-concavity of order infinity; however the slightly weaker form of the conjecture stated here is still difficult and open. The only special case in which Conjecture 6.3 is known to be true is for matroids whose rank (i.e., cardinality of the largest independent set) is 6 or smaller; this was proved by Zhao [55].

Conjecture 6.3 equivalently says that, for a random independent set $\mathbb{I}$, the distribution, $P(k) = \Pr\{|\mathbb{I}| = k\} \propto I_k$, is ultra-log-concave. This yields the following corollary.

---

[2]Note that although graphs have associated cycle matroids, there is no connection between independent sets of matroids and independent sets of graphs; indeed, the latter are often called "stable sets" in the matroid literature to distinguish the two.

Corollary 6.4. *Suppose the matroid M on the ground set $[n]$ satisfies Mason's conjecture. Let $\mathbb{I}$ be a random independent set of M, and let the average cardinality of $\mathbb{I}$ be λ. Suppose the elements of the ground set are given i.i.d. weights drawn from a probability distribution Q on $\mathbb{N}$, where Q is log-concave and satisfies $\lambda Q(1)^2 \geq 2Q(2)$. If W is the random weight assigned to $\mathbb{I}$, then,*

$$H(W) \leq H(C_Q \Pi_\lambda).$$

Of course, if Q is the unit mass at 1, Corollary 6.4 gives (modulo Mason's conjecture) a bound on the entropy of the cardinality of a random independent set in a matroid. That is,

$$H(|\mathbb{I}|) \leq H(\Pi_\lambda),$$

where $\lambda = E|\mathbb{I}|$. As in the case of graphs, this bound is independent of n and can be estimated in terms of the average size of a random independent set (and hence, more loosely, in terms of the matroid rank) using the Poisson entropy bound (36).

7. Extensions and Conclusions

The main results in this paper describe the solution of a discrete entropy maximization problem, under both shape constraints involving log-concavity and constraints on the mean. Different entropy problems involving log-concavity of continuous densities have also been studied by Cover and Zhang [14] and by Bobkov and Madiman [6], using different methods and motivated by different questions than those in this work.

The primary motivation for this work was the development of an information-theoretic approach to discrete limit laws, and specifically those corresponding to compound Poisson limits. Above we have shown that, under appropriate conditions, compound Poisson distributions have maximum entropy within a natural class. This is analogous to the maximum entropy property of the Gaussian and Poisson measures, and their corresponding roles in Gaussian and Poisson approximation, respectively. Moreover, the techniques introduced here – especially the introduction and analysis of a new score function that naturally connects with the compound Poisson family – turn out to play a central role in the development of an information-theoretic picture of compound Poisson limit theorems and approximation bounds [4].

After a preliminary version of this paper was made publicly available [28], Y. Yu [54] provided different proofs of our Theorems 2.4 and 2.5, under less restrictive conditions, and using a completely different mathematical approach. Also, in the first version of [28], motivated in part by the results of Lemma 5.2 and Theorems 5.3 and 5.4, we conjectured that the compound Poisson measure $\mathrm{CPo}(\lambda, Q)$ is log-concave, if Q is log-concave and $\lambda Q(1)^2 \geq 2Q(2)$. Y. Yu [54] subsequently established the truth of the conjecture by pointing out that it could be proved by an application of the results of Hansen in [20]. Theorem 5.5 in Section 5 is a slightly more general version of that earlier conjecture. Note that in order to prove the conjecture it is not necessary to reduce the problem to the strictly log-concave case (as done in the proof of Theorem 5.5), because the log-concavity of Q implies a bit more than log-concavity for $Q^{\#}$. Indeed, the following variant of Theorem 5.5 is easily proved: If Q is log-concave and $\lambda Q(1)^2 > 2Q(2) > 0$, then the compound Poisson measure $\mathrm{CPo}(\lambda, Q)$ is strictly log-concave.

In closing we mention another possible direction in which the present results may be extended. Suppose that the compounding distribution Q in the setup described in Section 2 is supported on $\mathbb{R}$ and has a

density with respect to Lebesgue measure. The definition of compound distributions $C_Q P$ (including compound Poissons) continues to make sense for probability distributions P on the nonnegative integers, but these now clearly are of mixed type, with a continuous component and an atom at 0. Furthermore, limit laws for sums converging to such mixed-type compound Poisson distributions hold exactly as in the discrete case. It is natural and interesting to ask for such 'continuous' analogs of the present maximum entropy results, particularly as neither their form nor method of proof are obvious in this case.

Acknowledgement

Appendix A. Appendix

Proof. Writing $R(y) = y! P(y)$, we know that $C_Q P(x) = \sum_{y=0}^{x} R(y) \left(Q^{*y}(x)/y! \right)$. Hence, the log-concavity of $C_Q P(x)$ is equivalent to showing that,

$$\sum_r \frac{Q^{*r}(2x)}{r!} \sum_{y+z=r} R(y) R(z) \binom{r}{y} \left(\frac{Q^{*y}(x) Q^{*z}(x)}{Q^{*r}(2x)} - \frac{Q^{*y}(x+1) Q^{*z}(x-1)}{Q^{*r}(2x)} \right) \geq 0, \qquad \text{(A.1)}$$

for all $x \geq 2$, since the case of $x = 1$ was dealt with previously by equation (32). In particular, for (i), taking $P = \text{Po}(\lambda)$, it suffices to show that for all r and x, the function,

$$g_{r,x}(k) := \sum_{y+z=r} \binom{r}{y} \frac{Q^{*y}(k) Q^{*z}(2x - k)}{Q^{*r}(2x)}$$

is unimodal as a function of k (since $g_{r,x}(k)$ is symmetric about x).

In the general case (ii), writing $Q(2) = p = 1 - Q(1)$, we have, $Q^{*y}(x) = \binom{y}{x-y} p^{x-y} (1-p)^{2y-x}$, so that,

$$\binom{r}{y} \frac{Q^{*y}(k) Q^{*z}(2x - k)}{Q^{*r}(2x)} = \binom{2x - r}{k - y} \binom{2r - 2x}{2y - k}, \qquad \text{(A.2)}$$

for any p. Now, following [27, Lemma 2.4] and [53, Lemma 2.1], we use summation by parts to show that the inner sum in (A.1) is positive for each r (except for $r = x$ when x is odd), by case-splitting according to the parity of r.

(a) For $r = 2t$, we rewrite the inner sum of equation (A.1) as,

$$\sum_{s=0}^{t} (R(t+s) R(t-s) - R(t+s+1) R(t-s-1)) \times$$

$$\left(\sum_{y=t-s}^{t+s} \left(\binom{2x - r}{x - y} \binom{2r - 2x}{2y - x} - \binom{2x - r}{x + 1 - y} \binom{2r - 2x}{2y - x - 1} \right) \right),$$

24

where the first term in the above product is positive by the ultra log-concavity of P (and hence log-concavity of R), and the second term is positive by Lemma Appendix A.1 below.

(b) Similarly, for $x \neq r = 2t + 1$, we rewrite the inner sum of equation (A.1) as,

$$\sum_{s=0}^{t}(R(t+s+1)R(t-s) - R(t+s+2)R(t-s-1))\times$$
$$\left(\sum_{y=t-s}^{t+1+s}\left(\binom{2x-r}{x-y}\binom{2r-2x}{2y-x} - \binom{2x-r}{x+1-y}\binom{2r-2x}{2y-x-1}\right)\right),$$

where the first term in the product is positive by the ultra log-concavity of P (and hence log-concavity of R) and the second term is positive by Lemma Appendix A.1 below.

(c) Finally, in the case of $x = r = 2t + 1$, substituting $k = x$ and $k = x + 1$ in (A.2), combining the resulting expression with (A.1), and noting that $\binom{2r-2x}{u}$ is 1 if and only if $u = 0$ (and is zero, otherwise), we see that the inner sum becomes, $-R(t+1)R(t)\binom{2t+1}{t}$, and the summands in (A.1) reduce to,

$$-\frac{p^x R(t)R(t+1)}{(t+1)!t!}.$$

However, the next term in the outer sum of equation (A.1), $r = x + 1$, gives

$$\frac{p^{x-1}(1-p)^2}{2(2t)!}\left[R(t+1)^2\left(2\binom{2t}{t} - \binom{2t}{t+1}\right) - R(t)R(t+2)\binom{2t}{t}\right]$$
$$\geq \frac{p^{x-1}(1-p)^2}{2(2t)!}R(t+1)^2\left(\binom{2t}{t} - \binom{2t}{t+1}\right) = \frac{p^{x-1}(1-p)^2}{2(t+1)!t!}R(t+1)^2.$$

Hence, the sum of the first two terms is positive (and hence the whole sum is positive) if $R(t+1)(1-p)^2/(2p) \geq R(t)$.

If P is Poisson(λ), this simply reduces to equation (31), otherwise we use the fact that $R(x+1)/R(x)$ is decreasing. $\qquad\square$

Lemma Appendix A.1. (a) *If $r = 2t$, for any $0 \leq s \leq t$, the sum,*

$$\sum_{y=t-s}^{t+s}\left(\binom{2x-r}{x-y}\binom{2r-2x}{2y-x} - \binom{2x-r}{x+1-y}\binom{2r-2x}{2y-x-1}\right) \geq 0.$$

(b) *If $x \neq r = 2t + 1$, for any $0 \leq s \leq t$, the sum,*

$$\sum_{y=t-s}^{t+1+s}\left(\binom{2x-r}{x-y}\binom{2r-2x}{2y-x} - \binom{2x-r}{x+1-y}\binom{2r-2x}{2y-x-1}\right) \geq 0.$$

Proof. The proof is in two stages; first we show that the sum is positive for $s = t$, then we show that there exists some S such that, as s increases, the increments are positive for $s \leq S$ and negative for $s > S$. The result then follows, as in [27] or [53].

For both (a) and (b), note that for $s = t$, equation (A.2) implies that the sum is the difference between the coefficients of T^x and T^{x+1} in $f_{r,x}(T) = (1+T^2)^{2x-r}(1+T)^{2r-2x}$. Since $f_{r,x}(T)$ has degree $2x$ and has coefficients which are symmetric about T^x, it is enough to show that the coefficients form a unimodal sequence. Now, $(1+T^2)^{2x-r}(1+T)$ has coefficients which do form a unimodal sequence. Statement S_1 of Keilson and Gerber [32] states that any binomial distribution is strongly unimodal, which means that it preserves unimodality on convolution. This means that $(1+T^2)^{2x-r}(1+T)^{2r-2x}$ is unimodal if $r - x \geq 1$, and we need only check the case $r = x$, when $f_{r,x}(T) = (1+T^2)^r$. Note that if $r = 2t$ is even, the difference between the coefficients of T^x and T^{x+1} is $\binom{2t}{t}$, which is positive.

In part (a), the increments are equal to $\binom{2x-2t}{x-t+s}\binom{4t-2x}{2t-2s-x}$ multiplied by the expression,

$$2 - \frac{(x-t-s)(2t-2s-x)}{(x+1-t+s)(2t+2s-x+1)} - \frac{(x-t+s)(2t+2s-x)}{(x+1-t-s)(2t-2s-x+1)},$$

which is positive for s small and negative for s large, since placing the term in brackets over a common denominator, the numerator is of the form $(a - bs^2)$.

Similarly, in part (b), the increments equal $\binom{2x-2t-1}{x-t+s}\binom{4t+2-2x}{2t-2s-x}$ times the expression,

$$2 - \frac{(x-t-s-1)(2t-2s-x)}{(x+1-t+s)(2t+2s-x+3)} - \frac{(x-t+s)(2t+2+2s-x)}{(x-t-s)(2t+1-2s-x)},$$

which is again positive for s small and negative for s large. $\qquad\square$

References

[1] R. M. Adelson. Compound Poisson distributions. *Operations Research Quarterly*, 17:73–75, 1966.

[2] S. Artstein, K. M. Ball, F. Barthe, and A. Naor. Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.*, 17(4):975–982 (electronic), 2004.

[3] A. Barbour, L. Chen, and W.-L. Loh. Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.*, 20(4):1843–1866, 1992.

[4] A. Barbour, O. T. Johnson, I. Kontoyiannis, and M. Madiman. Compound Poisson approximation via local information quantities. *In preparation*, 2009.

[5] A. R. Barron. Entropy and the Central Limit Theorem. *Ann. Probab.*, 14(1):336–342, 1986.

[6] S. G. Bobkov and M. Madiman. On the dispersion of the entropy per coordinate of convex measures. *In preparation*.

[7] F. Brenti. Unimodal, log-concave and Pólya frequency sequences in combinatorics. *Mem. Amer. Math. Soc.*, 81(413):viii+106, 1989.

[8] G. Brightwell and P. Tetali. The number of linear extensions of the boolean lattice. *Order*, 20:333–345, 2003.

[9] J. Cai and G. E. Willmot. Monotonicity and aging properties of random sums. *Statist. Probab. Lett.*, 73(4):381–392, 2005.

[10] D. Chafaï. Binomial-Poisson entropic inequalities and the M/M/∞ queue. *ESAIM Probability and Statistics*, 10:317–339, 2006.

[11] Z. Chi. Personal communication, 2006.

[12] M. Chudnovsky and P. Seymour. The roots of the independence polynomial of a clawfree graph. *J. Combin. Theory Ser. B*, 97(3):350–357, 2007.

[13] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[14] T. M. Cover and Z. Zhang. On the maximum entropy of the sum of two dependent random variables. *IEEE Trans. Information Theory*, 40(4):1244–1246, 1994.

[15] N. de Pril. Recursions for convolutions of arithmetic distributions. *ASTIN Bulletin*, 15(2):135–139, 1985.

[16] P. Flajolet. Singularity analysis and asymptotics of Bernoulli sums. *Theoret. Comput. Sci.*, 215(1-2):371–381, 1999.

[17] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[18] B. V. Gnedenko and V. Y. Korolev. *Random Summation: Limit Theorems and Applications*. CRC Press, Boca Raton, Florida, 1996.

[19] Y. O. Hamidoune. On the numbers of independent k-sets in a claw free graph. *J. Combin. Theory Ser. B*, 50(2):241–244, 1990.

[20] B. G. Hansen. On log-concave and log-convex infinitely divisible sequences and densities. *Ann. Probab.*, 16(4):1832–1839, 1988.

[21] P. Harremoës. Binomial and Poisson distributions as maximum entropy distributions. *IEEE Trans. Information Theory*, 47(5):2039–2041, 2001.

[22] 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*, pages 1491–1495, 2007.

[23] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[24] P. Jacquet and W. Szpankowski. Entropy computations via analytic de-Poissonization. *IEEE Trans. Inform. Theory*, 45(4):1072–1081, 1999.

[25] O. T. Johnson. *Information theory and the Central Limit Theorem*. Imperial College Press, London, 2004.

[26] O. T. Johnson. Log-concavity and the maximum entropy property of the Poisson distribution. *Stoch. Proc. Appl.*, 117(6):791–802, 2007.

[27] O. T. Johnson and C. A. Goldschmidt. Preservation of log-concavity on summation. *ESAIM Probability and Statistics*, 10:206–215, 2006.

[28] O. T. Johnson, I. Kontoyiannis, and M. Madiman. On the entropy and log-concavity of compound Poisson measures. *Preprint*, May 2008, `arXiv:0805.4112v1 [cs.IT]`.

[29] J. Kahn. Entropy, independent sets and antichains: a new approach to Dedekind's problem. *Proc. Amer. Math. Soc.*, 130(2):371–378, 2001.

[30] S. Karlin. *Total positivity. Vol. I.* Stanford University Press, Stanford, Calif, 1968.

[31] S. K. Katti. Infinite divisibility of integer-valued random variables. *Ann. Math. Statist.*, 38:1306–1308, 1967.

[32] J. Keilson and H. Gerber. Some results for discrete unimodality. *Journal of the American Statistical Association*, 66(334):386–389, 1971.

[33] J. Keilson and U. Sumita. Uniform stochastic ordering and related inequalities. *Canad. J. Statist.*, 10(3):181–198, 1982.

[34] D. J. Kleitman. Families of non-disjoint subsets. *J. Combinatorial Theory*, 1:153–155, 1966.

[35] D. Kleitman and G. Markowsky. On Dedekind's problem: the number of isotone Boolean functions. II. *Trans. Amer. Math. Soc.*, 213:373–390, 1975.

[36] C. Knessl. Integral representations and asymptotic expansions for Shannon and Renyi entropies. *Appl. Math. Lett.*, 11(2):69–74, 1998.

[37] I. Kontoyiannis, P. Harremoës, and O. T. Johnson. Entropy and the law of small numbers. *IEEE Trans. Inform. Theory*, 51(2):466–472, 2005.

[38] T. M. Liggett. Ultra logconcave sequences and negative dependence. *J. Combin. Theory Ser. A*, 79(2):315–325, 1997.

[39] M. Madiman and A. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory*, 53(7):2317–2329, 2007.

[40] M. Madiman, A. Marcus, and P. Tetali. Entropy and set cardinality inequalities for partition-determined functions, with applications to sumsets. *Preprint*, 2008, `arXiv:0901.0055v1 [cs.IT]`.

[41] J. H. Mason. Matroids: unimodal conjectures and Motzkin's theorem. In *Combinatorics (Proc. Conf. Combinatorial Math., Math. Inst., Oxford, 1972)*, pages 207–220. Inst. Math. Appl., Southend, 1972.

[42] P. Mateev. The entropy of the multinomial distribution. *Teor. Verojatnost. i Primenen.*, 23(1):196–198, 1978.

[43] H. H. Panjer. Recursive evaluation of a family of compound distributions. *Astin Bull.*, 12(1):22–26, 1981.

[44] R. Pemantle. Towards a theory of negative dependence. *J. Math. Phys.*, 41(3):1371–1390, 2000.

[45] M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[46] J. Radhakrishnan. An entropy proof of Bregman's theorem. *J. Combinatorial Theory, Ser. A*, 77:161–164, 1997.

[47] J. C. Sha and D. J. Kleitman. The number of linear extensions of subset ordering. *Discrete Math.*, 63(2-3):271–278, 1987. Special issue: ordered sets (Oberwolfach, 1985).

[48] L. A. Shepp and I. Olkin. Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution. In *Contributions to probability*, pages 201–206. Academic Press, New York, 1981.

[49] R. P. Stanley. Log-concave and unimodal sequences in algebra, combinatorics, and geometry. In *Graph theory and its applications: East and West (Jinan, 1986)*, volume 576 of *Ann. New York Acad. Sci.*, pages 500–535. New York Acad. Sci., New York, 1989.

[50] F. W. Steutel and K. van Harn. Discrete analogues of self-decomposability and stability. *Ann. Probab.*, 7(5):893–899, 1979.

[51] A. Tulino and S. Verdú. Monotonic decrease of the non-Gaussianness of the sum of independent random variables: a simple proof. *IEEE Trans. Inform. Theory*, 52(9):4295-4297, 2006.

[52] D. G. Wagner. Negatively correlated random variables and Mason's conjecture for independent sets in matroids. *Ann. Comb.*, 12(2):211–239, 2008.

[53] Y. Wang and Y.-N. Yeh. Log-concavity and LC-positivity. *J. Combin. Theory Ser. A*, 114(2):195–210, 2007.

[54] Y. Yu. On the entropy of compound distributions on nonnegative integers. *IEEE Trans. Inform. Theory*, 55(8):3645–3650, August 2009.

[55] C. K. Zhao. A conjecture on matroids. *Neimenggu Daxue Xuebao*, 16(3):321–326, 1985.