

On the Entropy of Compound Distributions on Nonnegative Integers

Yaming Yu, *Member, IEEE*

Abstract—Some entropy comparison results are presented concerning compound distributions on nonnegative integers. The main result shows that, under a log-concavity assumption, two compound distributions are ordered in terms of Shannon entropy if both the “numbers of claims” and the “claim sizes” are ordered accordingly in the convex order. Several maximum/minimum entropy theorems follow as a consequence. Most importantly, two recent results of Johnson et al. (2008) on maximum entropy characterizations of compound Poisson and compound binomial distributions are proved under fewer assumptions and with simpler arguments.

Index Terms—compound binomial, compound Poisson, convex order, infinite divisibility, log-concavity, maximum entropy, minimum entropy, random sum, stochastic orders

I. INTRODUCTION AND MAIN RESULT

This paper delivers some entropy comparison results concerning compound random variables taking nonnegative integer values, i.e., variables of the form

$$\sum_{i=1}^M X_i \quad (1)$$

where M, X_1, X_2, \dots , are independent random variables on $\mathbf{Z}_+ = \{0, 1, \dots\}$, and $X_i, i = 1, 2, \dots$, are also identically distributed. Letting M have a Poisson distribution in (1) yields the special case of the compound Poisson, which plays an important role in limit theorems and approximation bounds for discrete random variables; see, for example, [2], [3]. Recently, Kontoyiannis and Madiman [18], Madiman et al. [20], and Johnson et al. [13] have explored compound Poisson approximation and limit theorems using information theoretic ideas, extending the results of [17] and [12] for the Poisson (see also [8], [9], [32]). As a first step toward a compound Poisson limit theorem with the same appealing “entropy increasing to the maximum” interpretation as the central limit theorem ([4], [1], [19], [27]), we need to identify a suitable class of distributions among which the compound Poisson has maximum entropy ([13]). This partly motivates our investigation.

On the other hand, compound distributions are extensively used in applied settings. In insurance risk theory [22], for example, the random sum (1) can be used to model the total claim amount, M being the number of claims and X_i 's being the claim sizes. Results here therefore yield qualitative comparisons between distributions of the total claim amount with respect to Shannon's entropy, regarded purely as a measure

of variability or dispersion. (Note, however, that we focus on integer-valued variables only.)

Our results are closely related to those of Johnson et al. [13], who extend the semigroup argument of [12] to maximum entropy characterizations of compound Poisson and compound binomial distributions. We take a different approach based on convex ordering; as in [13], log-concavity also plays a critical role. Convex ordering and log-concavity together yield a rather general entropy comparison result (Theorem 1; see [30] for related work).

As usual, the Shannon entropy of a random variable X on \mathbf{Z}_+ with probability mass function (pmf) $f = \{f_i, i \geq 0\}$ is defined as

$$H(X) = H(f) = \sum_{i=0}^{\infty} -f_i \log f_i$$

where $0 \log 0 = 0$ and we use the natural logarithm. We also recall the following definitions.

Definition 1: A nonnegative sequence $f = \{f_i, i \geq 0\}$ on \mathbf{Z}_+ is *log-concave*, if $\text{supp}(f) = \{i : f_i > 0\}$ is an interval of consecutive integers, and

$$f_{i+1}^2 \geq f_i f_{i+2}, \quad i \geq 0. \quad (2)$$

A positive sequence f on \mathbf{Z}_+ is *log-convex* if it satisfies (2) with the inequality reversed. We call a random variable on \mathbf{Z}_+ *log-concave* (log-convex), if its pmf is log-concave (log-convex).

Definition 2: For random variables X and Y on \mathbf{Z}_+ , X is smaller than Y in the *convex order* (written as $X \leq_{cx} Y$), if $E\phi(X) \leq E\phi(Y)$ for every convex function ϕ on \mathbf{Z}_+ (i.e., $2\phi(i+1) \leq \phi(i) + \phi(i+2)$, $i \in \mathbf{Z}_+$).

We focus on integer-valued random variables, although Definition 2 extends readily to the real-valued case. Also, we use \leq_{cx} with the pmfs as well as the random variables. Informally, the convex order compares the “spread” or variability of two distributions. In particular, if $X \leq_{cx} Y$ and both X and Y have finite means, then $EX = EY$. In such a case, by choosing $\phi(x) = (x - EX)^2$ in Definition 2, we obtain that $X \leq_{cx} Y$ implies $\text{Var}(X) \leq \text{Var}(Y)$. Further basic properties of the convex order can be found in Shaked and Shanthikumar ([24], Chapter 3).

Since entropy ordering also compares the variability of distributions, although in a difference sense, it is reasonable to expect some connection with convex ordering. For compound distributions, our main result (Theorem 1) clarifies such a connection while highlighting the role of log-concavity. Let us first specify a convenient notation.

Yaming Yu is with the Department of Statistics, University of California, Irvine, CA, 92697-1250, USA (e-mail: yamingy@uci.edu). This work is supported in part by a start-up fund from the Bren School of Information and Computer Sciences at the University of California, Irvine.

Definition 3: The compound pmf $c_g(f)$ is the pmf of $\sum_{i=1}^M X_i$ where M, X_1, X_2, \dots , are independent random variables on \mathbf{Z}_+ , M has pmf g , and each X_i has pmf f .

Our main result is as follows.

Theorem 1: Let f, f^*, g, g^* be pmfs on \mathbf{Z}_+ such that $f \leq_{cx} f^*$, $g \leq_{cx} g^*$, and $c_{g^*}(f^*)$ is log-concave. Then

$$H(c_g(f)) \leq H(c_{g^*}(f^*)).$$

Theorem 1 need not hold without the log-concavity assumption ([13]; see also Remark 3 in Section II). Theorem 1 is useful because directly proving an entropy inequality is often more difficult than establishing a convex ordering and verifying a log-concavity condition ([32]).

The proof of Theorem 1 is facilitated by a key lemma, the basic idea of which dates back to Karlin and Rinott [14]; see [31] for a continuous version, and see [32] for an application analyzing the monotonicity of entropy in a special Poisson limit theorem.

Lemma 1: Let g and g^* be pmfs on \mathbf{Z}_+ such that $g \leq_{cx} g^*$ and g^* is log-concave. Then $H(g) \leq H(g^*)$.

Proof: We have

$$\begin{aligned} H(g^*) &= - \sum_{i \geq 0} g_i^* \log g_i^* \\ &\geq - \sum_{i \geq 0} g_i \log g_i^* \end{aligned} \quad (3)$$

$$\begin{aligned} &\geq - \sum_{i \geq 0} g_i \log g_i \\ &= H(g) \end{aligned} \quad (4)$$

where (3) follows from $g \leq_{cx} g^*$ and (4) follows from the fact that $D(g||g^*) = \sum_{i \geq 0} g_i \log(g_i/g_i^*) \geq 0$. All sums are effectively over $\text{supp}(g^*)$ as $g \leq_{cx} g^*$ implies that g assigns zero mass outside of $\text{supp}(g^*)$ when $\text{supp}(g^*)$ is an interval on \mathbf{Z}_+ . ■

Another key result is the following property of the convex order. A proof of Lemma 2 is included in the Appendix; see Shaked and Shanthikumar ([24], Theorem 3.A.13) and Jean-Marie and Liu [11] for other versions.

Lemma 2: Let f, f^*, g, g^* be pmfs on \mathbf{Z}_+ such that $f \leq_{cx} f^*$ and $g \leq_{cx} g^*$, then

$$c_g(f) \leq_{cx} c_{g^*}(f^*).$$

Theorem 1 follows directly from Lemmas 1 and 2.

In Section II, we use Theorem 1 to derive several maximum (minimum) entropy characterizations, including two maximum entropy results of Johnson et al. [13] on compound binomial and compound Poisson distributions. In view of the crucial role of the log-concavity assumption, in Section III we discuss conditions for a compound Poisson pmf to be log-concave. Specifically, a conjecture in [13] is confirmed using a result of Hansen [6]. Some implications of the log-concavity requirement are also mentioned.

II. MAXIMUM AND MINIMUM ENTROPY CHARACTERIZATIONS

Let $B_i, i = 1, \dots, n$, be independent Bernoulli random variables, $B_i \sim \text{Bern}(p_i)$, $p_i \in (0, 1)$. A theorem of

Shepp and Olkin [25] states that, subject to a fixed mean $\sum_{i=1}^n p_i$, the entropy of $S = \sum_{i=1}^n B_i$ is maximized when all p_i 's are equal, i.e., when S has a binomial distribution. (The pmf of the binomial distribution $\text{Bi}(n, p)$ is denoted by $bi(n, p) = \binom{n}{i} p^i (1-p)^{n-i}$, $i = 0, \dots, n$.) We show that this maximum entropy property generalizes naturally to the compound binomial, provided that the compound binomial pmf is log-concave (Theorem 2). A similar result (Theorem 3) holds for the compound Poisson.

Theorems 2 and 3 strengthen Theorems 1.4 and 1.5, respectively, of Johnson et al. [13] in that we impose fewer assumptions. Specifically, in both Theorems 2 and 3 we relax their assumption that f itself is log-concave. (As in [13], we assume that the compound pmf, e.g., $c_{g^*}(f)$ in Theorem 2, is log-concave.) Our derivation, based on Theorem 1, is also simpler.

Theorem 2 ([13], Theorem 1.4): Let g be the pmf of $\sum_{i=1}^n B_i$ where $B_i \sim \text{Bern}(p_i)$ independently, with $p_i \in (0, 1)$. Let $g^* \equiv bi(n, \bar{p})$ with $\bar{p} = (1/n) \sum_{i=1}^n p_i$. Then, for any pmf f on \mathbf{Z}_+ such that $c_{g^*}(f)$ is log-concave, we have

$$H(c_g(f)) \leq H(c_{g^*}(f)).$$

Proof: In view of Theorem 1, we only need to show that $g \leq_{cx} g^*$, which is a classical inequality of Hoeffding ([10], Theorem 3). ■

Remark 1: By setting f as a point mass at 1 in Theorem 2, we recover the theorem of Shepp and Olkin [25] and Mateev [21], i.e.,

$$H\left(\sum_{i=1}^n B_i\right) \leq H(bi(n, \bar{p})). \quad (5)$$

See [14] [7] [12] [29] and [30] for ramifications.

To prepare for an analogous result for the compound Poisson, let us recall the notion of *ultra log-concavity*, which was introduced by Pemantle [23] in the study of negative dependence.

Definition 4: A random variable Y on \mathbf{Z}_+ is ultra log-concave, or ULC, if the support of Y is an interval on \mathbf{Z}_+ , and the pmf of Y , $f_i = \text{Pr}(Y = i)$, satisfies

$$(i+1)f_{i+1}^2 \geq (i+2)f_i f_{i+2}, \quad i \geq 0.$$

Equivalently, Y is ULC if the sequence $\{i!f_i, i \geq 0\}$ is log-concave. Ultra log-concavity can be conveniently defined in terms of the *relative log-concavity order* [28].

Definition 5: Let g and g^* be pmfs on \mathbf{Z}_+ . Then g is log-concave relative to g^* , written as $g \leq_{lc} g^*$, if

- 1) $\text{supp}(g)$ and $\text{supp}(g^*)$ are both intervals on \mathbf{Z}_+ ;
- 2) $\text{supp}(g) \subset \text{supp}(g^*)$; and
- 3) $\log(g_i/g_i^*)$ is concave on $i \in \text{supp}(g)$.

We use \leq_{lc} with the random variables as well as their pmfs. Evidently, a random variable Y being ULC is equivalent to $Y \leq_{lc} Z$ for any $Z \sim \text{Po}(\lambda)$, $\lambda > 0$. (The pmf of the Poisson distribution $\text{Po}(\lambda)$ is denoted by $po(\lambda) = \{\lambda^i e^{-\lambda}/i!, i = 0, 1, \dots\}$.)

A result of [12], closely related to (5), states that among all ULC distributions with a fixed mean, the Poisson achieves maximum entropy. An extension to the compound Poisson is

considered in [13]. Henceforth we use $CP(\lambda, f)$ to denote a compound Poisson distribution, i.e., the pmf is

$$cp(\lambda, f) \equiv c_{po(\lambda)}(f).$$

Theorem 3 ([13], Theorem 1.5): Let g be a ULC pmf on \mathbf{Z}_+ with mean $\lambda \in (0, \infty)$. Then for any pmf f on \mathbf{Z}_+ such that $cp(\lambda, f)$ is log-concave, we have

$$H(c_g(f)) \leq H(cp(\lambda, f)).$$

In other words, if a compound Poisson pmf $cp(\lambda, f)$ is log-concave, then it achieves maximum entropy among all compound distributions with the same “claim size” pmf f , but with a ULC distribution (mean λ) for the “number of claims.” The maximum entropy result of [12] corresponds to f being a point mass at 1.

The main step in our proof of Theorem 3 is Lemma 3, which connects the relative log-concavity order \leq_{lc} with the convex order \leq_{cx} ; the basic idea is due to Whitt [28] (see [30], Theorem 12). A proof of Lemma 3 is sketched.

Lemma 3: Suppose g and g^* are pmfs on \mathbf{Z}_+ such that

$$g \leq_{lc} g^* \quad \text{and} \quad \sum_{i \geq 0} i g_i = \sum_{i \geq 0} i g_i^* < \infty. \quad (6)$$

Then $g \leq_{cx} g^*$.

Proof: The assumption (6) implies that, unless $g = g^*$, the sign pattern of $\log(g_i/g_i^*)$, or that of $g_i - g_i^*$, is $-$, $+$, $-$ as i traverses the support of g . A classical inequality (Karlin and Studden [15], Lemma XI. 7.2) then yields

$$\sum_{i \geq 0} g_i \phi(i) \leq \sum_{i \geq 0} g_i^* \phi(i)$$

for any convex function ϕ on \mathbf{Z}_+ . ■

Proof of Theorem 3: We have $g \leq_{cx} po(\lambda)$ by Lemma 3. The claim then follows from Theorem 1. ■

Remark 2: Building on the works of [12], [8] and [20], Yu [32] has recently obtained an “entropy increasing to the maximum” result in a version of the law of small numbers. Theorem 3 may be seen as a first step toward extending such a result to the compound Poisson case.

So far we have compared compound distributions with the same “claim size” distribution. We also present some minimum entropy characterizations (Propositions 1 and 2) to illustrate Theorem 1 when the “claim size” distributions may differ. (Of course, we may also allow this in Theorems 2 and 3.)

Proposition 1: Let $\lambda \in (0, \infty)$ and let f be a pmf on \mathbf{Z}_+ with mean $\mu < \infty$. Suppose $cp(\lambda, f)$ is log-concave. Then

$$H(po(\lambda\mu)) \leq H(cp(\lambda, f)).$$

In other words, subject to a fixed mean, the Poisson achieves minimum entropy among compound Poisson distributions that are log-concave. Proposition 1 should be compared with the *maximum* entropy characterization of Johnson ([12], Theorem 2.5).

Proof of Proposition 1: Fix $\alpha > \max\{1, \mu\}$. A Poisson with mean $\lambda\mu$ is equivalently a compound variable where the

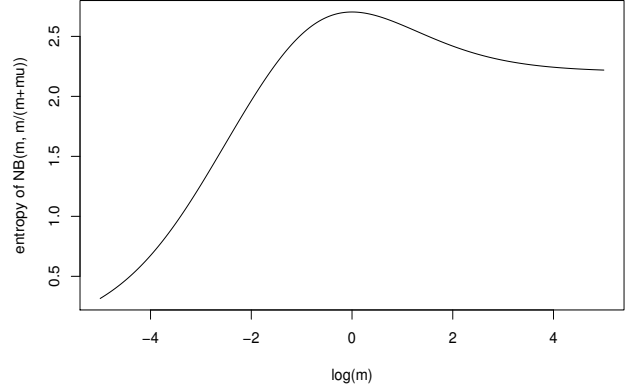


Fig. 1. Entropy of $nb(m, m/(m + \mu))$ as a function of $\log(m)$ for $\mu = 5$.

claim number is $Po(\alpha\lambda)$ and the claim size is $Bern(\alpha^{-1}\mu)$, i.e.,

$$po(\lambda\mu) = cp(\alpha\lambda, bi(1, \alpha^{-1}\mu)).$$

More generally, we have $cp(\lambda, f) = cp(\alpha\lambda, f^*)$, where f^* is the mixture

$$f^* = \alpha^{-1}f + (1 - \alpha^{-1})\delta^0$$

and δ^0 denotes a point mass at zero. To see this, note that a random variable X with pmf $cp(\lambda, f)$ can be obtained by drawing $Y \sim Po(\alpha\lambda)$, $Z|Y \sim Bin(Y, \alpha^{-1})$, and, given (Y, Z) , setting X as the sum of Z i.i.d. random variables, each with pmf f . Equivalently, X is a sum of Y i.i.d. random variables, each with pmf f^* .

It is easy to show that $bi(1, \alpha^{-1}\mu) \leq_{cx} f^*$. (Intuitively $bi(1, \alpha^{-1}\mu)$ is the least “spread out” distribution on \mathbf{Z}_+ with mean $\alpha^{-1}\mu$.) The claim follows from Theorem 1. ■

Remark 3: Proposition 1 need not hold without assuming that $cp(\lambda, f)$ is log-concave. Consider a negative binomial pmf $nb(m, p) = \{ \binom{m+i-1}{i} p^m (1-p)^i, i = 0, 1, \dots \}$, with $m > 0$ and $p = m/(m + \mu)$. It is a compound Poisson ([5], p. 271) with mean μ ; it is log-convex when $m \in (0, 1)$ and log-concave when $m > 1$. However, there are values of $m \in (0, 1)$ such that $H(nb(m, p)) < H(po(\mu))$, as is evident from Fig. 1, which displays $H(nb(m, p))$ as a function of $\log(m)$ for $\mu = 5$. Note that as m increases from 1 to ∞ , $H(nb(m, p))$ monotonically decreases to $H(po(\mu))$. (By a separate calculation $H(po(5)) \approx 2.20$.)

We mention an analogue of Proposition 1 for the binomial, which also follows from Theorem 1.

Proposition 2: Let f be a pmf on \mathbf{Z}_+ with mean $\mu < \infty$. Suppose $p \in (0, 1)$ satisfies $p\mu < 1$, and the compound binomial $c_{bi(n, p)}(f)$ is log-concave. Then

$$H(bi(n, p\mu)) \leq H(c_{bi(n, p)}(f)).$$

In other words, subject to a fixed mean and a fixed n (“the number of trials”), the binomial achieves minimum entropy among compound binomial distributions that are log-concave.

Proof of Proposition 2: Trivially, $bi(n, p\mu)$ is a compound pmf with the “claim number” fixed at n , and the

“claim size” as $bi(1, p\mu)$. On the other hand, $c_{bi(n,p)}(f)$ is equivalently the pmf of $\sum_{i=1}^n X_i B_i$ where X_i, B_i are independent, X_i 's have pmf f , and $B_i \sim \text{Bern}(p)$. That is, $c_{bi(n,p)}(f)$ is a compound pmf with the “claim number” fixed at n , and the “claim size” as a mixture $f^* = pf + (1-p)\delta^0$ where δ^0 is a point mass at zero. Similar to the Poisson case, the claim follows from $bi(1, p\mu) \leq_{cx} f^*$ and Theorem 1. ■

III. THE LOG-CONCAVITY ASSUMPTION

Since log-concavity is a crucial assumption in Theorem 1, it is natural to ask for conditions that ensure the log-concavity of a compound variable $\sum_{i=1}^M X_i$, e.g., in terms of the log-concavity properties of X_i . In general this appears to be a difficult combinatorial problem; in the compound Poisson case, a sufficient condition is summarized as Theorem 4, which was stated as a conjecture in [13].

Theorem 4 ([13], Conjecture 4.5): Suppose a pmf $f = \{f_i, i \in \mathbf{Z}_+\}$ is log-concave. Then the compound Poisson $cp(\lambda, f)$, $\lambda > 0$, is log-concave if and only if $\lambda f_1^2 \geq 2f_2$.

We point out that Theorem 4 can be deduced from Hansen ([6], Theorem 1), where the results are phrased in terms of infinite divisibility. A random variable X is *infinitely divisible*, if for each $n \geq 2$ there exist i.i.d. random variables Y_i , $i = 1, \dots, n$, such that X has the same distribution as $\sum_{i=1}^n Y_i$. A well-known connection between infinite divisibility and compound Poisson distributions is stated in Theorem 5; further results can be found in Steutel and Van Harn [26].

Theorem 5 ([5] [16] [26]): Let X be a random variable on \mathbf{Z}_+ with pmf $g = \{g_i, i \geq 0\}$ such that $g_0 > 0$. Then the following statements are equivalent.

- $X \sim \text{CP}(\lambda, f)$ for some $\lambda > 0$ and pmf f on \mathbf{Z}_+ .
- X is infinitely divisible.
- There exist $r_i \geq 0$, $i = 0, 1, \dots$, such that

$$(n+1)g_{n+1} = \sum_{i=0}^n r_i g_{n-i}, \quad n \geq 0. \quad (7)$$

Proof: For a \Leftrightarrow b, see Feller ([5], Chapter XII); for b \Leftrightarrow c, see Katti [16]. ■

Note that in Theorem 5, the $f = \{f_i, i \geq 0\}$ in (a) and the $r = \{r_i, i \geq 0\}$ in (c) obey the relation

$$r_i = \lambda(i+1)f_{i+1}, \quad i \geq 0. \quad (8)$$

Relation (7) (together with (8)) is known as Panjer's recursion [22] in actuarial literature.

Proof of Theorem 4: Write $g \equiv cp(\lambda, f)$, and let $r = \{r_i, i \geq 0\}$ be specified through (8), so that r and g are related by (7). Hansen ([6], Theorem 1) showed that, if r is log-concave, then g is log-concave if and only if $r_0^2 \geq r_1$. In view of (8), however, we know that (i) f being log-concave implies that r is log-concave, and (ii) $\lambda f_1^2 \geq 2f_2$ is equivalent to $r_0^2 \geq r_1$. Thus $\lambda f_1^2 \geq 2f_2$ is necessary and sufficient.

Hansen's argument for the sufficiency is short but nontrivial, and will not be presented here. The necessity of $\lambda f_1^2 \geq 2f_2$, however, is easy to prove, as noted in [13]. If g is log-concave then $g_1^2 \geq g_0 g_2$, and, using (7), we may express g_1, g_2 in terms of g_0 and r to obtain

$$(r_0 g_0)^2 \geq g_0 \frac{r_0^2 g_0 + r_1 g_0}{2}$$

which simplifies to $r_0^2 \geq r_1$, or equivalently $\lambda f_1^2 \geq 2f_2$. ■

Theorem 4 leads to the following result, which was also mentioned in [13] as a conjecture.

Corollary 1: If a pmf f on \mathbf{Z}_+ is log-concave and $cp(\lambda, f)$ is log-concave for some $\lambda > 0$, then $cp(\mu, f)$ is log-concave for all $\mu > \lambda$.

Of course, it would be interesting to see if Theorem 4 can be adapted to other compound distributions, e.g., compound binomial distributions.

We also explore the implications of requiring a compound pmf to be log-concave. In the compound Poisson case, an interesting observation (Proposition 3) is that such a pmf g must be “less log-concave” than the Poisson itself, i.e., the Poisson must be log-concave relative to g . This result, though elementary, does not seem well known, and hence a proof is provided.

Proposition 3: Let $\lambda > 0$ and let f be a pmf on \mathbf{Z}_+ . If the compound Poisson $g \equiv cp(\lambda, f)$ is non-degenerate and log-concave, then $po(\lambda) \leq_{lc} g$.

Proof: The assumptions imply $g_i > 0$ for all $i \in \mathbf{Z}_+$. From log-concavity, i.e., $g_i/g_{i-1} \geq g_{i+1}/g_i$, we deduce $g_{n-i}/g_{n-i-1} \geq g_n/g_{n-1}$, $0 \leq i < n$. Putting this in (7), we get

$$\begin{aligned} (n+1)g_{n+1} &= \sum_{i=0}^n r_i g_{n-i} \\ &\geq \sum_{i=0}^{n-1} r_i \frac{g_n g_{n-i-1}}{g_{n-1}} \\ &= \frac{ng_n^2}{g_{n-1}} \end{aligned}$$

where the last step uses (7) again with $n-1$ in place of n . Thus the sequence $n!g_n$, $n = 0, 1, \dots$, is log-convex, as required. ■

Proposition 3 shows that the log-concavity assumption places further nontrivial constraints on $g = cp(\lambda, f)$:

$$\frac{g_{i+1}}{g_i} \leq \frac{g_i}{g_{i-1}} \leq \frac{(i+1)g_{i+1}}{ig_i}, \quad i > 0.$$

Although these may seem stringent, they are not enough to ensure that g is a compound Poisson. A counter-example is a zero-inflated Poisson with a suitable probability at zero, as can be verified using (7).

Proposition 3 is also closely related to Proposition 1, our minimum entropy characterization of the Poisson. In fact, Proposition 3 can be used to give another derivation of Proposition 1.

Alternative Proof of Proposition 1: Because $cp(\lambda, f)$ is log-concave, Proposition 3 gives $po(\lambda\mu) \leq_{lc} cp(\lambda, f)$. Lemma 3 then implies $po(\lambda\mu) \leq_{cx} cp(\lambda, f)$, and the claim follows from Lemma 1. ■

Proposition 4 below is the counterpart of Proposition 3 for the compound binomial; see the Appendix for its proof.

Proposition 4: Let $p \in (0, 1)$ and let $f = \{f_i, i \geq 0\}$ be a pmf on \mathbf{Z}_+ . If the compound binomial $g \equiv c_{bi(n,p)}(f)$ is non-degenerate and log-concave, then $bi(n, p) \leq_{lc} g$.

Similar to the compound Poisson case, Proposition 4 shows that the log-concavity assumption places the following constraints on $g = c_{bi(n,p)}(f)$:

$$\frac{g_{i+1}}{g_i} \leq \frac{g_i}{g_{i-1}} \leq \frac{(i+1)(n-i+1)g_{i+1}}{i(n-i)g_i}, \quad 0 < i < n.$$

Proposition 4 also leads to an alternative proof of Proposition 2. The argument (omitted) completely parallels the compound Poisson case.

ACKNOWLEDGMENT

The author would like to thank the associate editor and two referees, whose comments greatly improved this paper.

APPENDIX

PROOF OF LEMMA 2

We recall that the convex order is closed under convolution ([24], Chapter 3).

Proposition 5: If X_i , $i = 1, \dots, n$, are independent variables, Y_i , $i = 1, \dots, n$, are another set of independent variables, and $X_i \leq_{cx} Y_i$ for each i , then

$$\sum_{i=1}^n X_i \leq_{cx} \sum_{i=1}^n Y_i.$$

Proof of Lemma 2: Let $\phi(x)$ be a convex function on \mathbf{Z}_+ . Then for any $a, b, c \in \mathbf{Z}_+$, $\phi(a+x) - \phi(x)$ is nondecreasing in x , and

$$\phi(a+b+c) - \phi(b+c) \geq \phi(a+c) - \phi(c). \quad (9)$$

Let $M, X_1, X_2, \dots, N, Y_1, Y_2, \dots$, be independent variables on \mathbf{Z}_+ such that M has pmf g , N has pmf g^* , X_i 's have pmf f , and Y_i 's have pmf f^* . The assumption says that $X_i \leq_{cx} Y_i$ and $M \leq_{cx} N$. Define

$$S_k = \sum_{i=1}^k Y_i \quad \text{and} \quad \psi(k) = E\phi(S_k).$$

Note that ψ is well defined because ϕ is convex on \mathbf{Z}_+ . We may assume that ψ is finite; the general case follows by a standard limiting argument. By symmetry and (9), we have

$$\begin{aligned} 2\psi(k+1) &= E[\phi(S_k + Y_{k+1}) + \phi(S_k + Y_{k+2})] \\ &\leq E[\phi(S_k) + \phi(S_k + Y_{k+1} + X_{k+2})] \\ &= \psi(k) + \psi(k+2), \end{aligned}$$

i.e., $\psi(k)$ is convex in k . Hence

$$\begin{aligned} E\phi\left(\sum_{i=1}^M X_i\right) &= E\left[E\phi\left(\sum_{i=1}^M X_i\right) \middle| M\right] \\ &\leq E[E\phi(S_M) | M] \end{aligned} \quad (10)$$

$$\begin{aligned} &= E\psi(M) \\ &\leq E\psi(N) \\ &= E\phi(S_N) \end{aligned} \quad (11)$$

where we apply Proposition 5 in (10), and use $M \leq_{cx} N$ and the convexity of ψ in (11). The claim readily follows. ■

PROOF OF PROPOSITION 4

The assumptions imply $g_i > 0$ for $0 \leq i \leq n$. Define

$$f_0^* = pf_0 + 1 - p \quad \text{and} \quad f_i^* = pf_i, \quad i = 1, 2, \dots$$

for convenience. Analogous to the Poisson case, we have the recursion ([22])

$$(k+1)g_{k+1}f_0^* = \sum_{i=0}^k [(n+1)i + n - k]g_{k-i}f_{i+1}^* \quad (12)$$

for $k = 0, 1, \dots$. From the log-concavity of g we obtain $g_{k-i}/g_{k-i-1} \geq g_k/g_{k-1}$, $0 \leq i < k$. Hence, for $1 \leq k \leq n$ we have

$$\begin{aligned} (k+1)g_{k+1}f_0^* &\geq \sum_{i=0}^k [(n+1)i + n - k] \frac{g_k g_{k-i-1}}{g_{k-1}} f_{i+1}^* \\ &\geq \sum_{i=0}^{k-1} [(n+1)i + n - k + 1] \frac{n-k}{n-k+1} \\ &\quad \times \frac{g_k g_{k-i-1}}{g_{k-1}} f_{i+1}^* \\ &= \frac{n-k}{n-k+1} \frac{kg_k^2 f_0^*}{g_{k-1}} \end{aligned}$$

where the last step uses (12) with $k-1$ in place of k . After rearrangement we get

$$\frac{g_{k+1}}{\binom{n}{k+1}} \frac{g_{k-1}}{\binom{n}{k-1}} \geq \frac{g_k^2}{\binom{n}{k}}, \quad 0 < k < n,$$

i.e., the sequence $\log(g_k/\binom{n}{k})$ is convex on $k \in \{0, \dots, n\}$, as required.

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, 2004.
- [2] A. D. Barbour, L. Chen, and W.-L. Loh, "Compound Poisson approximation for nonnegative random variables via Steins method," *Ann. Probab.*, vol. 20, no. 4, pp. 1843–1866, 1992.
- [3] A. D. Barbour and O. Chryssaphinou, "Compound Poisson approximation: a user's guide," *Ann. Appl. Prob.* vol. 11, pp. 964–1002, 2001.
- [4] A. R. Barron, "Entropy and the central limit theorem," *Ann. Probab.*, vol. 14, pp. 336–342, 1986.
- [5] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, New York, Wiley, 1966.
- [6] B. G. Hansen, "On log-concave and log-convex infinitely divisible sequences and densities," *Ann. Probab.*, vol. 16, no. 4, pp. 1832–1839, Oct. 1988.
- [7] P. Harremoës, "Binomial and Poisson distributions as maximum entropy distributions," *IEEE Trans. Information Theory*, vol. 47, no. 5, pp. 2039–2041, July 2001.
- [8] P. Harremoës, O. Johnson, and I. Kontoyiannis, "Thinning and the law of small numbers," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, Jun. 2007.
- [9] P. Harremoës, O. Johnson, and I. Kontoyiannis, "Thinning and information projections," in *Proc. IEEE International Symposium on Information Theory*, Toronto, Canada, Jul. 2008.
- [10] W. Hoeffding, "On the distribution of the number of successes in independent trials," *Ann. Math. Statist.*, vol. 27, pp. 713–721, 1956.
- [11] A. Jean-Marie and Z. Liu, "Stochastic comparisons for queueing models via random sums and intervals," *Advances in Applied Probability*, vol. 24, no. 4, pp. 960–985, 1992.
- [12] O. Johnson, "Log-concavity and the maximum entropy property of the Poisson distribution," *Stochastic Processes and their Applications*, vol. 117, no. 6, pp. 791–802, June 2007.

- [13] O. Johnson, I. Kontoyiannis, and M. Madiman, "On the entropy and log-concavity of compound Poisson measures," *Preprint*, 2008, <http://arxiv.org/abs/0805.4112v1>
- [14] S. Karlin and Y. Rinott. "Entropy inequalities for classes of probability distributions I. the univariate case," *Advances in Applied Probability*, vol. 13, pp. 93–112, 1981.
- [15] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Interscience, New York, 1966.
- [16] S. K. Katti, "Infinite divisibility of integer-valued random variables," *Ann. Math. Stat.*, vol. 38, pp. 1306–1308, 1967.
- [17] 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, Feb. 2005.
- [18] I. Kontoyiannis and M. Madiman, "Entropy, compound Poisson approximation, log-Sobolev inequalities and measure concentration," *2004 IEEE Information Theory Workshop*, San Antonio, TX, October 2004.
- [19] 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, Jul. 2007.
- [20] M. Madiman, O. Johnson and I. Kontoyiannis, "Fisher information, compound Poisson approximation, and the Poisson channel," *2007 IEEE International Symposium on Information Theory*, Nice, France, June 2007.
- [21] P. Mateev, "The entropy of the multinomial distribution," *Teor. Veroyatnost. i Primenen.*, vol. 23, no. 1, 196–198, 1978.
- [22] H. H. Panjer, "Recursive evaluation of a family of compound distributions," *ASTIN Bulletin*, vol. 12, no. 1, pp. 22–26, 1981.
- [23] R. Pemantle, "Towards a theory of negative dependence," *J. Math. Phys.*, vol. 41, no. 3, pp. 1371–1390, 2000.
- [24] M. Shaked and J. G. Shanthikumar. *Stochastic Orders*, Springer, New York, 2007.
- [25] L. A. Shepp and I. Olkin, "Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution," in *Contributions to probability*, pp. 201–206, Academic Press, New York, 1981.
- [26] F. W. Steutel and K. Van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*, New York: Marcel-Dekker, 2003.
- [27] A. M. 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, Sep. 2006.
- [28] W. Whitt, "Uniform conditional variability ordering of probability distributions," *Journal of Applied Probability*, vol. 22, pp. 619–633, 1985.
- [29] Y. Yu, "On the maximum entropy properties of the binomial distribution," *IEEE Trans. Inform. Theory*, vol. 54, pp. 3351–3353, 2008.
- [30] Y. Yu, "Relative log-concavity and a pair of triangle inequalities," *Technical Report*, Department of Statistics, University of California, Irvine, 2008, <http://www.ics.uci.edu/~yamingy/nb.ps>
- [31] Y. Yu, "On an inequality of Karlin and Rinott concerning weighted sums of i.i.d. random variables," *Adv. Appl. Prob.*, vol. 40, no. 4, pp. 1223–1226, 2008.
- [32] Y. Yu, "Monotonic convergence in an information-theoretic law of small numbers," *Preprint*, 2009, <http://arxiv.org/abs/0810.5203>

Yaming Yu received the B.S. degree in mathematics from Beijing University, Beijing, China, in 1999, and the Ph.D. degree in statistics from Harvard University, MA, in 2005. Since 2005 he has been an Assistant Professor in the Department of Statistics at the University of California, Irvine.