Adv. Appl. Prob. **42,** 392–410 (2010)
Printed in Northern Ireland

STOCHASTIC ORDERING OF CLASSICAL DISCRETE DISTRIBUTIONS

ACHIM KLENKE,* *Johannes Gutenberg-Universität Mainz*
LUTZ MATTNER,** *Universität Trier*

Abstract

For several pairs (P, Q) of classical distributions on $\mathbb{N}_0$, we show that their stochastic ordering $P \le_{\mathrm{st}} Q$ can be characterized by their extreme tail ordering equivalent to $P(\{k_*\})/Q(\{k_*\}) \ge 1 \ge \lim_{k\to k^*} P(\{k\})/Q(\{k\})$, with k_* and k^* denoting the minimum and the supremum of the support of $P + Q$, and with the limit to be read as $P(\{k^*\})/Q(\{k^*\})$ for finite k^*. This includes in particular all pairs where P and Q are both binomial ($b_{n_1,p_1} \le_{\mathrm{st}} b_{n_2,p_2}$ if and only if $n_1 \le n_2$ and $(1-p_1)^{n_1} \ge (1-p_2)^{n_2}$, or $p_1 = 0$), both negative binomial ($b^-_{r_1,p_1} \le_{\mathrm{st}} b^-_{r_2,p_2}$ if and only if $p_1 \ge p_2$ and $p_1^{r_1} \ge p_2^{r_2}$), or both hypergeometric with the same sample size parameter. The binomial case is contained in a known result about Bernoulli convolutions, the other two cases appear to be new. The emphasis of this paper is on providing a variety of different methods of proofs: (i) half monotone likelihood ratios, (ii) explicit coupling, (iii) Markov chain comparison, (iv) analytic calculation, and (v) comparison of Lévy measures. We give four proofs in the binomial case (methods (i)–(iv)) and three in the negative binomial case (methods (i), (iv), and (v)). The statement for hypergeometric distributions is proved via method (i).

Keywords: Bernoulli convolution; binomial distribution; coupling; hypergeometric distribution; negative binomial distribution; monotone likelihood ratio; occupancy problem; Pascal distribution; Poisson distribution; stochastic ordering; waiting times

2010 Mathematics Subject Classification: Primary 60E15

1. Introduction

1.1. Stochastic ordering

For probability measures P and Q on the real numbers, the stochastic ordering is the partial ordering

$$P \le_{\mathrm{st}} Q \quad\Longleftrightarrow\quad P([x,\infty)) \le Q([x,\infty)) \quad \text{for all } x \in \mathbb{R}.$$

This condition is equivalent to the existence of two real-valued random variables X and Y with distributions P and Q, respectively, and such that $X \le Y$ almost surely. In fact, let F_P and F_Q denote the distribution functions of P and Q, respectively, and let F_P^{-1} and F_Q^{-1} be their left-continuous inverses. That is,

$$F_P^{-1}(t) := \inf\{x \in \mathbb{R} \colon F_P(x) \ge t\}.$$

Furthermore, let U be uniformly distributed on $(0, 1)$. Then $X := F_P^{-1}(U)$ and $Y := F_Q^{-1}(U)$ have the desired property. Such a pair (X, Y) is called a coupling.

Received 9 March 2009; revision received 7 March 2010.
* Postal address: Johannes Gutenberg-Universität Mainz, Institut für Mathematik, Staudingerweg 9, 55099 Mainz, Germany. Email address: math@aklenke.de
** Postal address: Universität Trier, FB IV - Mathematik, 54286 Trier, Germany. Email address: mattner@uni-trier.de

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

Recall that $P \leq_{\text{st}} Q$ is equivalent to the condition that, for any bounded and monotone increasing function $f : \mathbb{R} \to \mathbb{R}$, we have

$$\int f \, \mathrm{d}P \leq \int f \, \mathrm{d}Q.$$

If P and Q have finite expectations, then taking $f(x) = \max(-n, \min(x, n))$ and letting $n \to \infty$, it follows that $P \leq_{\text{st}} Q$ implies that $\int x P(\mathrm{d}x) \leq \int x Q(\mathrm{d}x)$. Thus, stochastic ordering implies ordering of the expected values but not vice versa.

There is a vast literature on stochastic orderings, but we only refer the reader to [5], [10], and [11].

Let $b_{n,p}$ denote the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$, let Poi_λ denote the Poisson distribution with parameter $\lambda > 0$, and let $b^-_{r,p}$ denote the negative binomial distribution (also known as the Pascal distribution) with parameters $r \in (0, \infty)$ and $p \in (0, 1]$. Recall that $b^-_{r,p}$ is the probability measure on $\mathbb{N}_0$ with weights

$$b^-_{r,p}(\{k\}) = \binom{-r}{k}(-1)^k p^r (1-p)^k = \binom{r+k-1}{k} p^r (1-p)^k \quad \text{for all } k \in \mathbb{N}_0.$$

Furthermore, we denote by

$$\mathrm{hyp}_{B,W,n}(\{k\}) = \binom{B}{k}\binom{W}{n-k} \bigg/ \binom{B+W}{n}, \qquad k = (n-W)^+, \ldots, B \wedge n,$$

the hypergeometric distribution with parameters $B, W \in \mathbb{N}_0$ and $n \in \mathbb{N}$, where $n \leq B + W$. The main goal of this paper is to prove necessary and sufficient conditions for $b_{n_1,p_1} \leq_{\text{st}} b_{n_2,p_2}$, $b^-_{r_1,p_1} \leq_{\text{st}} b^-_{r_2,p_2}$, and $\mathrm{hyp}_{B_1,W_1,n_1} \leq_{\text{st}} \mathrm{hyp}_{B_2,W_2,n_2}$ in terms of the parameters r_1, r_2, n_1, n_2, p_1, p_2, B_1, W_1, B_2, and W_2.

Since stochastic ordering implies ordering of expectations, $b_{n_1,p_1} \leq_{\text{st}} b_{n_2,p_2}$ implies that $p_1 n_1 \leq p_2 n_2$, but this condition is not sufficient for $b_{n_1,p_1} \leq_{\text{st}} b_{n_2,p_2}$. However, if $n := n_1 = n_2$ then

$$b_{n,p_1} \leq_{\text{st}} b_{n,p_2} \quad \iff \quad p_1 \leq p_2. \tag{1.1}$$

There are various proofs of this statement, the simplest being a coupling. Let $U_1, \ldots, U_n$ be independent and identically distributed random variables that are uniformly distributed on $[0, 1]$. For $i = 1, 2$, let

$$N_i = \#\{k : U_k \leq p_i\}.$$

Then $N_i \sim b_{n,p_i}$ and $N_1 \leq N_2$ almost surely. In Section 3 we present a more involved coupling, proving the sufficiency of a characterization of $b_{n_1,p_1} \leq_{\text{st}} b_{n_2,p_2}$ when $n_1 \neq n_2$ also.

1.2. The likelihood ratio order

Before we come to the statement of the main theorem of this paper, let us briefly discuss a stronger notion of ordering of two probability measures on $\mathbb{R}$, the so-called *monotone likelihood ratio* order. Let μ be any σ-finite measure such that P and Q are absolutely continuous with respect to μ and μ is absolutely continuous with respect to $P + Q$. Furthermore, define the respective densities

$$f = \frac{\mathrm{d}P}{\mathrm{d}\mu} \quad \text{and} \quad g = \frac{\mathrm{d}Q}{\mathrm{d}\mu}.$$

Here P is said to be smaller than or equal to Q in the monotone likelihood ratio order ($P \le_{\mathrm{lr}} Q$) if there exist versions of f and g such that the likelihood ratio

$$x \mapsto \ell(x) := \frac{f(x)}{g(x)} \text{ is monotone decreasing.} \tag{1.2}$$

Note that the ordering does not depend on the choice of μ; in particular, $\mu = P + Q$ is possible.

It is well known that $P \le_{\mathrm{lr}} Q$ implies that $P \le_{\mathrm{st}} Q$, but not vice versa. This will become even more obvious by the following characterization of the monotone likelihood ratio order. Let $\mathcal{B}(\mathbb{R})$ denote the Borel σ-algebra on $\mathbb{R}$. Then we have

$$P \le_{\mathrm{lr}} Q \quad \Longleftrightarrow \quad P(\cdot \mid B) \le_{\mathrm{st}} Q(\cdot \mid B) \quad \text{for all } B \in \mathcal{B}(\mathbb{R}),\ P(B) > 0,\ Q(B) > 0 \tag{1.3}$$

by any of [6, pp. 1217–1218], [13, Theorems 1.1, 1.3], or [8, pp. 50–52]. In fact, the $\Leftarrow$ implication is valid even if we replace $\mathcal{B}(\mathbb{R})$ by the class of all intervals in $\mathbb{R}$ (see [6]) or by any smaller class $\mathcal{C}$ of subsets of $\mathbb{R}$ such that, for any $r < s$, there exist an $\varepsilon > 0$ and a $B \in \mathcal{C}$ such that $[r - \varepsilon, r] \cup [s, s + \varepsilon] \in B$ (see [13, Theorem 1.3]). In particular, if P and Q live on a discrete subset of $\mathbb{R}$, then it suffices to check the right-hand side of (1.3) only for sets B of cardinality 2.

For the binomial distributions, we have $b_{n_1,p_1} \le_{\mathrm{lr}} b_{n_2,p_2}$ if and only if $p_1 = 0$ or

$$n_1 \le n_2 \quad \text{and} \quad \frac{n_1 p_1}{1 - p_1} \le \frac{n_2 p_2}{1 - p_2}. \tag{1.4}$$

(See [1, Theorem 1(iv)] for a result for a larger class of distributions that comprises the binomial distributions.) In fact, if we exclude the trivial case $p_1 = 0$, then $n_1 \le n_2$ is clearly necessary for $b_{n_1,p_1} \le_{\mathrm{lr}} b_{n_2,p_2}$. In order to see that (1.4) is sufficient, assume that $n_1 \le n_2$ and let f_1 and f_2 be the corresponding densities, say with respect to the counting measure on $\mathbb{N}_0$. Then $\ell = f_1/f_2$ is decreasing if and only if, for all $k = 0, \dots, n_1 - 1$,

$$1 \ge \frac{f_1(k+1)/f_2(k+1)}{f_1(k)/f_2(k)} = \frac{p_1}{1 - p_1} \frac{1 - p_2}{p_2} \frac{n_1 - k}{n_2 - k}.$$

Clearly, the expression on the right-hand side is maximal for $k = 0$ and in this case the inequality is equivalent to (1.4).

As the monotone likelihood ratio order is stronger than the stochastic order, it is clear that (1.4) is sufficient for $b_{n_1,p_1} \le_{\mathrm{st}} b_{n_2,p_2}$, but it is not necessary, as we will see.

Note that, for the Poisson distribution, we have

$$\mathrm{Poi}_\lambda \le_{\mathrm{st}} \mathrm{Poi}_\mu \quad \Longleftrightarrow \quad \mathrm{Poi}_\lambda \le_{\mathrm{lr}} \mathrm{Poi}_\mu \quad \Longleftrightarrow \quad \lambda \le \mu.$$

Hence, for this subclass of distributions, stochastic ordering and monotone likelihood ratio ordering coincide.

1.3. Main result

For distributions P and Q on $\mathbb{N}_0$, the likelihood ratio $\ell = f/g$ (see (1.2)) is given by $\ell(k) := P(\{k\})/Q(\{k\})$, $k \in \mathbb{N}_0$. Let

$$k_* := \min(\{k \colon (P + Q)(\{k\}) > 0\}) \quad \text{and} \quad k^* = \sup(\{k \colon (P + Q)(\{k\}) > 0\}).$$

If $k^* = \infty$, define $\ell(k^*) := \limsup_{k\to\infty} \ell(k)$, $\underline{\ell}(k^*) := \liminf_{k\to\infty} \ell(k)$, and the extreme right tail ratio

$$\varrho := \limsup_{k\to\infty} \frac{P(\{k, k+1, \ldots\})}{Q(\{k, k+1, \ldots\})}.$$

If $k^* < \infty$, define $\varrho := \ell(k^*)$. Note that $\underline{\ell}(k^*) \le \varrho \le \ell(k^*)$. In order that $P \le_{\text{st}} Q$ holds, it is clearly necessary that

$$\ell(k_*) \ge 1 \tag{1.5}$$

and

$$\varrho \le 1. \tag{1.6}$$

Clearly, (1.6) is implied by

$$\ell(k^*) \le 1. \tag{1.7}$$

We say that (P, Q) fulfills the left tail condition if (1.5) holds and the right tail condition if (1.7) holds.

While we have just argued that (at least if $k^* < \infty$ or if $\ell(k)$ converges as $k \to \infty$) both tail conditions are necessary for $P \le_{\text{st}} Q$, the next theorem shows that, for certain classes of distributions, the tail conditions (1.5) and (1.7) are in fact equivalent to $P \le_{\text{st}} Q$.

Theorem 1.1. *In each of the following seven cases we have $P_1 \le_{\text{st}} P_2$ if and only if the left and right tail conditions hold.*

(a) *Binomial distribution: $P_i = b_{n_i, p_i}$ with $p_i \in (0, 1)$ and $n_i \in \mathbb{N}$, $i = 1, 2$. The left tail condition is given by*

$$(1 - p_1)^{n_1} \ge (1 - p_2)^{n_2}. \tag{1.8}$$

The right tail condition is given by

$$n_1 \le n_2. \tag{1.9}$$

(b) *Negative binomial distribution: $P_i = b^-_{r_i, p_i}$ with $r_i > 0$ and $p_i \in (0, 1]$, $i = 1, 2$. The left tail condition is given by*

$$p_1^{r_1} \ge p_2^{r_2}. \tag{1.10}$$

The right tail condition is given by

$$p_1 \le p_2. \tag{1.11}$$

(c) *Hypergeometric distribution: $P_i = \text{hyp}_{B_i, W_i, n_i}$ with $B_1, B_2, W_1, W_2, n_i \in \mathbb{N}_0$, $B_i + W_i \ge n_i \ge 1$, $i = 1, 2$. Furthermore, assume that*

$$B_2 + W_2 \ge B_1 + W_1 \tag{1.12}$$

or

$$\{n_1, B_1, n_2 - W_2 - 1\} \cap \{n_2, B_2, n_1 - W_1 - 1\} \ne \varnothing. \tag{1.13}$$

Define

$$k_* = (n_1 - W_1)^+ \wedge (n_2 - W_2)^+ \quad \textit{and} \quad k^* = (n_1 \wedge B_1) \vee (n_2 \wedge B_2).$$

The left tail condition is given by

$$\text{hyp}_{B_1, W_1, n_1}(\{k_*\}) \ge \text{hyp}_{B_2, W_2, n_2}(\{k_*\}). \tag{1.14}$$

The right tail condition is given by

$$\text{hyp}_{B_1, W_1, n_1}(\{k^*\}) \le \text{hyp}_{B_2, W_2, n_2}(\{k^*\}). \tag{1.15}$$

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

(d) *Hypergeometric versus binomial:* $P_1 = \mathrm{hyp}_{B,W,m}$, $P_2 = b_{n,p}$ *with* $B, W, m, n \in \mathbb{N}$, $B + W \geq m$, *and* $p \in (0, 1]$. *The left tail condition is given by*

$$\binom{W}{m}\Big/\binom{B+W}{m} \geq (1-p)^n. \tag{1.16}$$

The right tail condition is given by

$$m \wedge B \leq n.$$

(e) *Binomial versus hypergeometric:* $P_1 = b_{m,p}$, $P_2 = \mathrm{hyp}_{B,W,m}$ *with* $B, W, m \in \mathbb{N}_0$, $B + W \geq m \geq 1$, *and* $p \in [0, 1]$. *The right tail condition is given by*

$$p^m \leq \binom{B}{m}\Big/\binom{B+W}{m}. \tag{1.17}$$

The left tail condition is implied by the right tail condition.

(f) *Binomial versus Poisson:* $P_1 = b_{n,p}$, $P_2 = \mathrm{Poi}_\lambda$ *with* $n \in \mathbb{N}$, $p \in [0, 1]$, *and* $\lambda > 0$. *The left tail condition is given by*

$$(1-p)^n \geq \mathrm{e}^{-\lambda}. \tag{1.18}$$

The right tail condition is trivially fulfilled.

(g) *Poisson versus negative binomial:* $P_1 = \mathrm{Poi}_\lambda$, $P_2 = b^-_{r,p}$ *with* $p \in (0, 1)$ *and* $r, \lambda > 0$. *The left tail condition is given by*

$$\mathrm{e}^{-\lambda} \geq p^r. \tag{1.19}$$

The right tail condition is trivially fulfilled.

For (a), it is obvious that (1.8) is the left tail condition and (1.9) is right tail condition.

For (b), (1.10) is obviously the left tail condition since $b_{r_i,p_i}(\{0\}) = p_i^{r_i}$. For the right tail condition, note that, for $k \in \mathbb{N}$, we have

$$\left|\binom{-r_i}{k}\right| = \prod_{l=1}^{k}\left(1 + \frac{r_i - 1}{l}\right) \leq \exp\left(r_i \sum_{l=1}^{k} \frac{1}{l}\right) \leq \mathrm{e}^{r_i} k^{r_i}.$$

Hence,

$$\lim_{k\to\infty} \frac{\log(b^-_{r_i,p_i}(\{k, k+1, \ldots\}))}{k} = \log(1 - p_i),$$

and the right tail condition (1.6) is equivalent to $p_1 \geq p_2$.

For (c), note that k_* and k^* are the minimum and maximum of the support of $\mathrm{hyp}_{B_1,W_1,n} + \mathrm{hyp}_{W_2,B_2,n}$, respectively. Furthermore, note that in the case $n := n_1 = n_2$, condition (1.13) is satisfied. In this case the left tail condition simplifies to

$$\binom{B_1 + W_1 - n}{B_1 - k_*}\binom{B_2 + W_2}{B_2} \geq \binom{B_2 + W_2 - n}{B_2 - k_*}\binom{B_1 + W_1}{B_1}$$

and the right tail condition becomes

$$\binom{B_1 + W_1 - n}{B_1 - k^*}\binom{B_2 + W_2}{B_2} \leq \binom{B_2 + W_2 - n}{B_2 - k^*}\binom{B_1 + W_1}{B_1}.$$

For (d), (e), (f), and (g), the statements are (almost) trivial. In particular, (d) is a consequence of (c) since $b_{n,p}$ is the limit of $\text{hyp}_{\lfloor pN\rfloor,\lfloor(1-p)N\rfloor,n}$ as $N \to \infty$ and, for sufficiently large N, condition (1.12) is satisfied. Taking a further limit we recover (a). Similarly, (e) can be inferred from (c) noting that condition (1.13) is satisfied. In Section 2 we give the short proofs, in order to demonstrate the flexibility of our Method 1, described below.

Part (a) of the theorem is not trivial, but it is not new either. However, in this paper we give new and elementary proofs using different methods.

Method 1. This method is based on likelihood ratio considerations. We show in Proposition 2.1, below, that the left and right tail conditions are sufficient for stochastic ordering whenever the likelihood ratio ℓ or $1/\ell$ is a unimodal function; that is, if ℓ is either first monotone increasing and then monotone decreasing or vice versa. In this case we say that P and Q have half monotone likelihood ratios.

Method 2. This method works for the binomial distribution only and relies on an explicit coupling of two random variables $N_i \sim b_{n_i,p_i}$, $i = 1, 2$, such that $N_1 \leq N_2$ almost surely.

Method 3. This method also works for the binomial distribution only. Similarly to Method 2, this method is based on the observation that $b_{n,p}$ can be represented as the number of nonempty boxes when we throw a certain random Poisson number of balls into n boxes. Unlike in Method 2, here we do not construct an explicit coupling of N_1 and N_2, but give a stochastic comparison of the Markov dynamics of subsequently throwing the balls.

Method 4. This method works for the binomial and negative binomial distributions, and relies on explicitly calculating the changes when we modify the parameter p continuously.

Method 5. This method uses infinite divisibility of the negative binomial distribution to give a proof for part (b).

1.4. Organization of the paper

In Section 1.5 we provide a brief review of stochastic orderings of Bernoulli convolutions. In Sections 2–6 we give proofs of Theorem 1.1 using the different methods presented above.

1.5. A review of Bernoulli convolutions

We give a brief review on a result concerning the stochastic ordering of *Bernoulli convolutions* (that comprises part (a) of Theorem 1.1) due to Proschan and Sethuraman [7]. Fix $n \in \mathbb{N}$, and let

$$\Delta_n = \{\boldsymbol{p} = (p_1, \dots, p_n) \in [0, 1]^n : p_1 \geq p_2 \geq \cdots \geq p_n\}.$$

Let $\boldsymbol{p} \in \Delta_n$, and let $X_1, \dots, X_n$ be independent random variables with

$$\mathrm{P}[X_i = 1] = 1 - \mathrm{P}[X_i = 0] = p_i.$$

Then the distribution of $X_1 + \cdots + X_n$ is said to be the Bernoulli convolution $BC_{\boldsymbol{p}}$ with parameter $\boldsymbol{p}$.

Let $\boldsymbol{p}, \boldsymbol{q} \in \Delta_n$. By [7, Corollary 5.2], for $BC_{\boldsymbol{p}} \leq_{\text{st}} BC_{\boldsymbol{q}}$, it is sufficient that

$$\prod_{j=1}^{k} p_j \leq \prod_{j=1}^{k} q_j \quad \text{for all } k = 1, \dots, n. \tag{1.20}$$

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

By the obvious symmetry in the problem (changing the roles of the 1s and 0s), it is also sufficient to have

$$\prod_{j=k}^{n}(1-p_j) \geq \prod_{j=k}^{n}(1-q_j) \quad \text{for all } k=1,\dots,n. \tag{1.21}$$

Note that (1.20) and (1.21) are in fact not equivalent.

Assume that $n_1, n_2 \leq n$, and $p_1 = \cdots = p_{n_1}$, $p_{n_1+1} = \cdots = p_n = 0$, $q_1 = \cdots = q_{n_2}$, and $q_{n_2+1} = \cdots = q_n = 0$. Then (1.21) is equivalent to $n_1 \leq n_2$ and $(1-p_1)^{n_1} \geq (1-q_1)^{n_2}$. Hence, Theorem 1.1(a) is a special case of the result of [7].

A special case of [7, Corollary 5.2] (which is still more general than our Theorem 1.1(a)) was investigated independently of Proschan and Sethuraman by Ma [4]. Ma stated [4, Theorem 1] that, if $\boldsymbol{q} \in \Delta_n$ and $p \in (0,1)$, then

$$BC_{\boldsymbol{q}} \leq_{\text{st}} b_{n,p} \quad\Longleftrightarrow\quad b_{n,p}(\{0\}) \leq BC_{\boldsymbol{q}}(\{0\}). \tag{1.22}$$

In fact, the condition on the right-hand side of (1.22) is (1.21) (with the roles of $\boldsymbol{p} = (p,\dots,p)$ and $\boldsymbol{q}$ interchanged). Again, by the obvious symmetry, this statement is equivalent to

$$b_{n,p} \leq_{\text{st}} BC_{\boldsymbol{q}} \quad\Longleftrightarrow\quad b_{n,p}(\{n\}) \leq BC_{\boldsymbol{q}}(\{n\}). \tag{1.23}$$

Since the hypergeometric distribution is a Bernoulli convolution (see [12]), Theorem 1.1(d) and (e) could be inferred from (1.22) and (1.23). A limiting case of (1.22), more general than the present Theorem 1.1(f), was given in [2, Equation (A.5)].

2. Method 1: half monotone likelihood ratios

In this section we provide a criterion which, together with the left and right tail conditions (see (1.5) and (1.7)) is sufficient for stochastic ordering. We first present this method in the general situation and then apply it to all seven cases, (a)–(g), of Theorem 1.1.

2.1. A special criterion for the stochastic order

Definition 2.1. Let P and Q be as in Section 1.2. Define the set $\mathcal{H}$ of pairs (P, Q) such that there exists a version ℓ of the likelihood ratio $(\mathrm{d}P/\mathrm{d}(P+Q))/(\mathrm{d}Q/\mathrm{d}(P+Q))$ with the following properties.

(i) There exists an $x_0 \in \mathbb{R}$ such that ℓ is monotone (increasing or decreasing) on $(-\infty, x_0]$ and is monotone on $[x_0, \infty)$.

(ii) The left and right tail conditions hold:

$$\lim_{x\to-\infty} \ell(x) \geq 1 \tag{2.1}$$

and

$$\lim_{x\to\infty} \ell(x) \leq 1. \tag{2.2}$$

If only (i) is fulfilled then we write $P \sim_{\text{hmlr}} Q$, and we say that P and Q have a half monotone likelihood ratio.

Remark 2.1. For distributions P and Q on $\mathbb{N}_0$, the quotient f/g in Definition 2.1 is the likelihood ratio $\ell(k) := P(\{k\})/Q(\{k\})$, $k \in \mathbb{N}_0$. In this case, for $P \sim_{\text{hmlr}} Q$, it is sufficient that

$$\frac{\ell(k+1)}{\ell(k)} \text{ is monotone (increasing or decreasing) for } k_* \leq k < k^*. \tag{2.3}$$

That is, (1.5), (1.7), and (2.3) imply that $(P, Q) \in \mathcal{H}$.

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

Note that the relation '$\sim_{\text{hmlr}}$' is symmetric and reflexive, but it is not transitive. Furthermore, note that (trivially) $P \leq_{\text{lr}} Q$ implies that $(P, Q) \in \mathcal{H}$.

Proposition 2.1. *If* $(P, Q) \in \mathcal{H}$ *then* $P \leq_{\text{st}} Q$.

Proof. For $P = Q$, the statement is trivial. Hence, now assume that $P \neq Q$. Let x_0 be as in the definition of $\mathcal{H}$. We will show that there exists an $x_1 \in \mathbb{R}$ such that $\ell(x) \geq 1$ for $x < x_1$ and $\ell(x) \leq 1$ for $x > x_1$. Clearly, this implies that $P((-\infty, x]) \geq Q((-\infty, x])$ for $x < x_1$ and $P((x, \infty)) < Q((x, \infty))$ for $x \geq x_1$. Combining these two inequalities, we get $P \leq_{\text{st}} Q$.

In order to establish the existence of such an x_1, we distinguish three cases.

Case 1. If ℓ is monotone decreasing then the statement is trivial.

Case 2. Assume that ℓ is monotone decreasing on $(-\infty, x_0]$ and monotone increasing on $[x_0, \infty)$. Hence, $\ell(x) \geq \ell(x_0)$ for all $x \in \mathbb{R}$, which implies that $\ell(x_0) < 1$ unless $\ell(x) = 1$ for all $x \in \mathbb{R}$, which was ruled out by the assumption that $P \neq Q$. By assumption (2.2) we have $\ell(x) \leq 1$ for all $x \geq x_0$. Now take $x_1 = \sup\{x \colon \ell(x) \geq 1\} \leq x_0$.

Case 3. Assume that ℓ is monotone increasing on $(-\infty, x_0]$. By assumption (2.1) we have $\ell(x_0) > 1$, $\ell(x) \geq 1$ for all $x \leq x_0$, and ℓ is monotone decreasing on $[x_0, \infty)$. Choose $x_1 = \inf\{x \geq x_0 \colon \ell(x) \leq 1\}$.

In Sections 2.2–2.4 we show that any two binomial distributions, negative binomial distributions, and hypergeometric distributions with the same sample size parameter, have half monotone likelihood ratios. For hypergeometric distributions, we can show this also under the assumptions of Theorem 1.1(c). For other distributions, this method is not applicable in such generality. For example, $\text{hyp}_{400,509,500}$ and $\text{hyp}_{310,710,700}$ do not have half monotone likelihood ratios. In fact, the likelihood ratio is increasing on $\{0, 1, 2\}$ (with maximal value greater than 1), decreasing on $\{2, \ldots, 150\}$, and increasing on $\{150, \ldots, 400\}$ (to values greater than 1). These distributions are not stochastically ordered as

$$\text{hyp}_{400,509,500}(\{0, \ldots, k\}) < \text{hyp}_{310,710,700}(\{0, \ldots, k\}) \quad \text{for } k \leq 44$$

and

$$\text{hyp}_{400,509,500}(\{0, \ldots, k\}) > \text{hyp}_{310,710,700}(\{0, \ldots, k\}) \quad \text{for } k \geq 45.$$

On the other hand, it is simple to check numerically that $\text{hyp}_{100,100,18} \leq_{\text{st}} \text{hyp}_{21,23,22}$ but that $\text{hyp}_{100,100,18} \not\sim_{\text{hmlr}} \text{hyp}_{21,23,22}$.

It is tempting to try this method to obtain a necessary condition for $b_{n,p} \leq \text{hyp}_{B,W,m}$ with $m \neq n$. However, here in general, we do not have $\text{hyp}_{B,W,m} \sim_{\text{hmlr}} b_{n,p}$, as the following example illustrates. Let $\ell(k) = \text{hyp}_{21,23,22}(\{k\})/b_{18,0.5106}(\{k\})$. Then $\ell(0) = 0.000\,004\,2$ and ℓ increases monotonically to $\ell(13) = 2.05$. Then it decreases to $\ell(17) = 0.997$ and finally takes the value $\ell(18) = 1.006$. Although condition (ii) of Definition 2.1 is fulfilled, we do not have $(\text{hyp}_{21,23,22}, b_{18,0.5106}) \in \mathcal{H}$. In fact, $\text{hyp}_{21,23,22}$ and $b_{18,0.5106}$ are not stochastically ordered since

$$\text{hyp}_{21,23,22}(\{0\}) - b_{18,0.5106}(\{0\}) = -2.5 \cdot 10^{-6} < 0$$

and

$$\text{hyp}_{21,23,22}(\{0, \ldots, 16\}) - b_{18,0.5106}(\{0, \ldots, 16\}) = 8.4 \cdot 10^{-8} > 0.$$

It is easy to check that $b_{18,1/2} \leq_{\text{st}} \text{hyp}_{21,23,22}$ although $\text{hyp}_{21,23,22} \not\sim_{\text{hmlr}} b_{18,1/2}$. In fact, the likelihood quotient $\ell(k)$ increases for $k \leq 13$, then decreases to $\ell(17) = 1.393$, and finally takes the value $\ell(18) = 1.467$.

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

2.2. Proof of Theorem 1.1(a): binomial distributions

Lemma 2.1. *Let* $n_1, n_2 \in \mathbb{N}$ *and* $p_1, p_2 \in [0, 1]$. *Then* $b_{n_1,p_1} \sim_{\text{hmlr}} b_{n_2,p_2}$*; that is,* b_{n_1,p_1} *and* b_{n_2,p_2} *have half monotone likelihood ratios.*

Proof. The cases $p_1 \in \{0, 1\}$ and $p_2 \in \{0, 1\}$ are trivial. Hence, now assume that p_1, $p_2 \in (0, 1)$. Furthermore, due to the symmetry of '$\sim_{\text{hmlr}}$', we may assume without loss of generality that $n_1 \le n_2$.

Denote by

$$\ell(k) := \frac{b_{n_1,p_1}(\{k\})}{b_{n_2,p_2}(\{k\})}, \qquad k = 0, \dots, n_1,$$

the likelihood ratio. We compute

$$\frac{\ell(k+1)}{\ell(k)} = \frac{n_1 - k}{n_2 - k}\frac{p_1(1-p_2)}{(1-p_1)p_2} \quad \text{for } k = 0, \dots, n_1 - 1.$$

Since $n_1 \le n_2$, we see that $k \mapsto \ell(k+1)/\ell(k)$ is monotone decreasing and, hence, $\ell(k)$ is first monotone increasing and then monotone decreasing.

Proof of Theorem 1.1(a). We only have to show sufficiency of the tail conditions (1.8) and (1.9) for $b_{n_1,p_1} \le_{\text{st}} b_{n_2,p_2}$. By Proposition 2.1 and Lemma 2.1, it remains to show (1.5) and (1.7). Since we have $n_2 \ge n_1$, we have $k_* = 0$ and $k^* = n_2$. Since $p_2 \ge p_1$, we get $\ell(n_2) = b_{n_1,p_1}(\{n_2\})/b_{n_2,p_2}(\{n_2\}) \le 1$; that is, (1.7) holds. Furthermore, by assumption we have

$$b_{n_1,p_1}(\{0\}) = (1-p_1)^{n_1} \ge (1-p_2)^{n_2} = b_{n_2,p_2}(\{0\}),$$

which implies (1.5).

2.3. Proof of Theorem 1.1(b): negative binomial distributions

Lemma 2.2. *Let* $r_1, r_2 > 0$ *and* $p_1, p_2 \in (0, 1]$. *Then* $b^-_{r_1,p_1} \sim_{\text{hmlr}} b^-_{r_2,p_2}$*; that is,* $b^-_{r_1,p_1}$ *and* $b^-_{r_2,p_2}$ *have half monotone likelihood ratios.*

Proof. The cases $p_1 = 1$ and $p_2 = 1$ are trivial. Hence, now assume that $p_1, p_2 \in (0, 1)$. Furthermore, due to the symmetry of '$\sim_{\text{hmlr}}$', we may assume without loss of generality that $r_1 \le r_2$.

Denote by

$$\ell(k) := \frac{b^-_{r_1,p_1}(\{k\})}{b^-_{r_2,p_2}(\{k\})}, \qquad k \in \mathbb{N}_0,$$

the likelihood ratio. We compute

$$\frac{\ell(k+1)}{\ell(k)} = \frac{r_1 + k}{r_2 + k}\frac{1-p_1}{1-p_2} \quad \text{for } k \in \mathbb{N}_0.$$

Since $r_1 \le r_2$, we see that $\ell(k+1)/\ell(k)$ is monotone increasing. This implies that $\ell(k)$ is first monotone decreasing and then monotone increasing; that is, $b^-_{r_1,p_1} \sim_{\text{hmlr}} b^-_{r_2,p_2}$.

Proof of Theorem 1.1(b). We only have to show sufficiency of the tail conditions (1.10) and (1.11) for $b^-_{r_1,p_1} \le_{\text{st}} b^-_{r_2,p_2}$. By Proposition 2.1 and Lemma 2.2, it remains to show (1.5) and (1.7). Since $p_1 \ge p_2$, we get

$$\lim_{k\to\infty}\left(\frac{b^-_{r_1,p_1}(\{k\})}{b^-_{r_2,p_2}(\{k\})}\right)^{1/k} = \frac{1-p_1}{1-p_2} \le 1,$$

which implies (1.7). Furthermore, by assumption we have

$$b^{-}_{r_1,p_1}(\{0\}) = (1-p_1)^{r_1} \geq (1-p_2)^{r_2} = b^{-}_{r_2,r_2}(\{0\}),$$

which implies (1.5).

2.4. Proof of Theorem 1.1(c): hypergeometric distributions

The left tail condition (1.14) implies that $(n_1 - W_1)^+ \leq (n_2 - W_2)^+$ and the right tail condition (1.15) implies that $n_1 \wedge B_1 \leq n_2 \wedge B_2$. Furthermore, trivially, we have $P_1 \leq_{\mathrm{st}} P_2$ and even $P_1 \leq_{\mathrm{lr}} P_2$ if

$$n_1 \wedge B_1 \leq (n_2 - W_2)^+ \tag{2.4}$$

(and, hence, this condition implies the left and right tail conditions). Hence, in this case, $(P_1, P_2) \in \mathcal{H}$. Since this proves the theorem in the case (2.4), we may henceforth exclude this case. That is, we assume that

$$(n_2 - W_2)^+ < n_1 \wedge B_1.$$

Lemma 2.3. *Assume that (1.13) holds or that (1.15) and (1.12) hold. Then we have* $\mathrm{hyp}_{B_1,W_1,n_1} \sim_{\mathrm{hmlr}} \mathrm{hyp}_{B_2,W_2,n_2}$*; that is,* $\mathrm{hyp}_{B_1,W_1,n_1}$ *and* $\mathrm{hyp}_{B_2,W_2,n_2}$ *have half monotone likelihood ratios.*

Proof. By the discussion preceeding this lemma we may assume that

$$(n_1 - W_1)^+ \leq (n_2 - W_2)^+ < n_1 \wedge B_1 \leq n_2 \wedge B_2.$$

Now let

$$k_+ := (n_2 - W_2)^+ \geq k_* = (n_1 - W_1)^+ \quad \text{and} \quad k^+ := B_1 \wedge n_1 \leq k^* = B_2 \wedge n_2.$$

Furthermore, let

$$\ell(k) := \frac{\mathrm{hyp}_{W_1,B_1,n_1}(\{k\})}{\mathrm{hyp}_{W_2,B_2,n_2}(\{k\})}$$

with the convention that $1/0 = \infty$. We have

$$\ell(k) \begin{cases} = \infty & \text{if } k_* \leq k < k_+, \\ \in (0,\infty) & \text{if } k_+ \leq k \leq k^+, \\ = 0 & \text{if } k^+ < k \leq k^*. \end{cases}$$

For $k \in I := \{k_* \vee (k_+ - 1), \dots, k^* \wedge (k^+ + 1)\}$, we have

$$q(k) := \frac{\ell(k+1)}{\ell(k)} = \frac{B_1 - k}{B_2 - k} \frac{W_2 - n_2 + 1 + k}{W_1 - n_1 + 1 + k} \frac{n_1 - k}{n_2 - k}.$$

(Note that $q(k_+ - 1) = 0$ if $k_+ > k_*$.)

We are done if we can show that $q(k) - 1$ changes sign at most once in I. In fact, this implies that $q(k)$ can cross 1 at most once. This in turn implies that ℓ is half monotone on I (in the sense of Definition 2.1(i)). Since ℓ is constant on $\{k_*, \dots, k_+ - 1\}$ (taking the value ∞) and constant on $\{k^+ + 1, \dots, k^*\}$ (taking the value 0), we infer that ℓ is half monotone on $\{k_*, \dots, k^*\}$. Hence, by Remark 2.1 we get $\mathrm{hyp}_{B_1,W_1,n_1} \sim_{\mathrm{hmlr}} \mathrm{hyp}_{B_2,W_2,n_2}$.

In order to show that $q(k) - 1$ changes sign at most once, we have to rely on assumption (1.12) or (1.13).

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

Assume first that (1.13) holds. There are nine cases to consider and we start with the case $n_1 = n_2$. Then, for $k \in I$, we have

$$q(k) - 1 = \frac{B_1(W_2 - n_2 + 1) - B_2(W_1 - n_1 + 1) + (B_1 + W_1 - B_2 - W_2)k}{(B_2 - k)(W_1 - n_1 + 1 + k)}.$$

Note that the numerator is affine linear and the denominator is positive for $k \in I$. Hence, $q(k) - 1$ changes its sign at most once. The other eight cases, $n_1 = B_2$, $n_2 = B_1$, $B_1 = B_2$, and so on, are similar, resulting in an affine numerator and a denominator without sign change.

Now assume that (1.12) holds, but (1.13) does not hold. Then $k^+ < k^*$ and

$$q(k) - 1 = \frac{p(k)}{r(k)} := \frac{a_2k^2 + a_1x + a_0}{(B_2 - k)(W_1 - n_1 + 1 + k)(n_2 - k)}$$

with

$$\begin{aligned} a_0 &= B_1n_1W_2 - B_1n_1n_2 + B_1n_1 - B_2n_2W_1 + B_2n_2n_1 - B_2n_2, \\ a_1 &= -B_1W_2 + B_1n_2 - B_1 + B_1n_1 - n_1W_2 - n_1 \\ &\quad + B_2W_1 - B_2n_1 + B_2 - B_2n_2 + n_2W_1 + n_2, \\ a_2 &= B_2 + W_2 - B_1 - W_1. \end{aligned}$$

For $k \in I$, the denominator is positive. We have

$$q(k^+) - 1 = -1,$$

and, hence, $p(k^+) < 0$. Since p is at most quadratic, condition (1.12) (that is, $a_2 \geq 0$) implies that p changes its sign at most once on $(-\infty, k^+]$. Hence, again $q(x) - 1$ changes its sign at most once.

Proof of Theorem 1.1(c). We only have to show sufficiency of the tail conditions (1.14) and (1.15) for $\mathrm{hyp}_{B_1,W_1,n} \leq_{\mathrm{st}} \mathrm{hyp}_{B_2,W_2,n}$. However, this is an immediate consequence of Proposition 2.1 and Lemma 2.2.

2.5. Proof of Theorem 1.1(d): hypergeometric versus binomial

For $m \wedge B > n$, the implications are clear. Hence, without loss of generality, we may and will assume that $m \leq n$ and $B \leq n$.

Denoting the likelihood ratio by $\ell(k) = \mathrm{hyp}_{B,W,m}(\{k\})/b_{n,p}(\{k\})$, we find that

$$\frac{\ell(k+1)}{\ell(k)} = \frac{(B-k)(m-k)}{(W-m+k)(n-k)} \frac{1-p}{p} \quad \text{for } k = 0, \ldots, (B \wedge m) - 1$$

is monotone decreasing and, hence, $\mathrm{hyp}_{B,W,m} \sim_{\mathrm{hmlr}} b_{n,p}$. It is a simple exercise to check that

$$\frac{\mathrm{hyp}_{B,W,m}(\{0\})}{b_{n,p}(\{0\})} \geq 1 \quad \Longrightarrow \quad \frac{\mathrm{hyp}_{B,W,m}(\{n\})}{b_{n,p}(\{n\})} \leq 1$$

and

$$\frac{\mathrm{hyp}_{B,W,m}(\{m\})}{b_{m,p}(\{m\})} \geq 1 \quad \Longrightarrow \quad \frac{\mathrm{hyp}_{B,W,m}(\{0\})}{b_{m,p}(\{0\})} \leq 1.$$

Hence, the left tail condition (1.16) implies that $(\mathrm{hyp}_{B,W,m}, b_{n,p}) \in \mathcal{H}$ and, thus,

$$\mathrm{hyp}_{B,W,m} \leq_{\mathrm{st}} b_{n,p}.$$

2.6. Proof of Theorem 1.1(e): binomial versus hypergeometric

The proof of Theorem 1.1(e) is quite similar to that of Theorem 1.1(d). In fact, it is easy to see that the right tail condition (1.17) implies that $(b_{m,p}, \mathrm{hyp}_{B,W,m}) \in \mathscr{H}$ and, thus,

$$\mathrm{hyp}_{B,W,m} \geq_{\mathrm{st}} b_{m,p}.$$

2.7. Proof of Theorem 1.1(f): binomial versus Poisson

Clearly, the left tail condition is necessary for $b_{n,p} \leq_{\mathrm{st}} \mathrm{Poi}_\lambda$. Hence, now assume that the left tail condition (1.18) holds. Let

$$\ell(k) = \frac{b_{n,p}(\{k\})}{\mathrm{Poi}_\lambda(\{k\})},$$

and compute

$$\frac{\ell(k+1)}{\ell(k)} = \frac{p}{(1-p)\lambda}(n-k) \quad \text{for } k = 0, \ldots, n.$$

Hence, $\ell(k+1)/\ell(k)$ is monotone decreasing and, thus, $b_{n,p} \sim_{\mathrm{hmlr}} \mathrm{Poi}_\lambda$. Since the right tail condition holds trivially and the left tail condition holds by assumption, we infer that $(b_{n,p}, \mathrm{Poi}_\lambda) \in \mathscr{H}$ and, thus, by Proposition 2.1 we obtain $b_{n,p} \leq_{\mathrm{st}} \mathrm{Poi}_\lambda$.

Of course, this result is trivial, since we can even easily derive a coupling. Let $\hat{\lambda} = -\log(1-p) \leq \lambda/n$, and let $X_0, X_1, \ldots, X_n$ be independent with $X_i \sim \mathrm{Poi}_{\hat{\lambda}}$ for $i = 1, \ldots, n$ and $X_0 \sim \mathrm{Poi}_{\lambda - n\hat{\lambda}}$. Then

$$S := X_0 + X_1 + \cdots + X_n \geq T := (X_1 \wedge 1) + \cdots + (X_n \wedge 1) \quad \text{almost surely}$$

and $S \sim \mathrm{Poi}_\lambda$, $T \sim b_{n,p}$.

2.8. Proof of Theorem 1.1(g): Poisson versus negative binomial

Clearly, the left tail condition is necessary for $\mathrm{Poi}_\lambda \leq_{\mathrm{st}} b^-_{r,p}$. Furthermore, it is easy to see that the right tail condition always holds. Hence, now assume that the left tail condition (1.19) holds. Let

$$\ell(k) = \frac{\mathrm{Poi}_\lambda(\{k\})}{b^-_{r,p}(\{k\})},$$

and compute

$$\frac{\ell(k+1)}{\ell(k)} = \frac{\lambda}{1-p}\frac{1}{k+1} \quad \text{for } k \in \mathbb{N}_0.$$

Hence, $\ell(k+1)/\ell(k)$ is monotone decreasing and, thus, $\mathrm{Poi}_\lambda \sim_{\mathrm{hmlr}} b^-_{r,p}$. Since the right tail condition holds trivially and the left tail condition holds by assumption, we infer that $(\mathrm{Poi}_\lambda, b^-_{r,p}) \in \mathscr{H}$ and, thus, by Proposition 2.1 we obtain $\mathrm{Poi}_\lambda \leq_{\mathrm{st}} b^-_{r,p}$.

3. Method 2: coupling

In this section we give a proof of Theorem 1.1(a) that provides an explicit coupling of two random variables $N_i \sim b_{n_i,p_i}$ such that $N_1 \leq N_2$ almost surely. Clearly, this implies that $b_{n_1,p_1} \leq_{\mathrm{st}} b_{n_2,p_2}$.

Proof of Theorem 1.1(a). We only have to show sufficiency of the tail conditions (1.8) and (1.9) for $b_{n_1,p_1} \leq_{\mathrm{st}} b_{n_2,p_2}$. Hence, assume that (1.8) and (1.9) hold. By (1.1), it suffices to consider the smallest p_2 such that (1.8) holds. That is, we may assume that

$$(1-p_1)^{n_1} = (1-p_2)^{n_2}.$$

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

Define

$$\lambda := -n_1 \log(1-p_1) = -n_2 \log(1-p_2).$$

For $i = 1, 2$, let $(X_i(l),\ l = 1, \dots, n_i)$ be a family of independent Poisson random variables with parameter λ/n_i. (Note that we do not require that $X_1(l_1)$ and $X_2(l_2)$ be independent.) Then

$$N_i = \#\{l\colon X_i(l) \geq 1\} \sim b_{n_i,p_i}.$$

The idea is to construct a coupling of the $X_i(l)$ such that

$$N_1 \leq N_2 \quad \text{almost surely.} \tag{3.1}$$

This clearly implies that $b_{n_1,p_1} \leq_{\text{st}} b_{n_2,p_2}$.

Let T be a Poisson random variable with parameter λ. Assume that, for $i = 1, 2$, the family $(F_{i,k},\ k \in \mathbb{N})$ of random variables is independent and independent of T, and each $F_{i,k}$ is uniformly distributed on $\{1, \dots, n_i\}$. Then

$$X_i(l) := \#\{k \leq T\colon F_{i,k} = l\}, \qquad l = 1, \dots, n_i,$$

are independent and Poisson distributed with parameter λ/n_i. The remaining task is to construct the families $(F_{i,k},\ k \in \mathbb{N})$ such that (3.1) holds.

For $A_i \subset \{1, \dots, n_i\}$, let $a_i = \#A_i$ and $A_i^c = \{1, \dots, n_i\} \setminus A_i$. For $r_1 \in \{1, \dots, n_1\}$ and $r_2 \in \{1, \dots, n_2\}$, define $q^{A_1,A_2}(r_1, r_2)$ depending on whether $a_1 < a_2$ or $a_1 \geq a_2$. If $a_1 < a_2$ then let

$$q^{A_1,A_2}(r_1, r_2) = \frac{1}{n_1 n_2}.$$

If $a_1 \geq a_2$ then let

$$q^{A_1,A_2}(r_1, r_2) = \begin{cases} \dfrac{1}{a_1 n_2} & \text{if } r_1 \in A_1 \text{ and } r_2 \in A_2, \\ \dfrac{a_1 n_2 - a_2 n_1}{a_1 n_1 n_2 (n_2 - a_2)} & \text{if } r_1 \in A_1 \text{ and } r_2 \in A_2^c, \\ \dfrac{1}{(n_2 - a_2) n_1} & \text{if } r_1 \in A_1^c \text{ and } r_2 \in A_2^c, \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$q_i^{A_1,A_2}(r_i) = \sum_{r_{3-i}=1}^{n_{3-i}} q^{A_1,A_2}(r_1, r_2)$$

denote the ith marginal of q^{A_1,A_2}. Clearly, for $a_1 < a_2$, we have $q_i^{A_1,A_2}(r_i) = 1/n_i$ for $i = 1, 2$ and $r_i \in \{1, \dots, n_i\}$. Now assume that $a_1 \geq a_2$. Then, for $r_1 \in A_1$,

$$q_1^{A_1,A_2}(r_1) = \frac{a_2}{a_1 n_2} + (n_2 - a_2) \frac{a_1 n_2 - a_2 n_1}{a_1 n_1 n_2 (n_2 - a_2)} = \frac{1}{n_1}.$$

On the other hand, for $r_1 \in A_1^c$,

$$q_1^{A_1,A_2}(r_1) = (n_2 - a_2) \frac{1}{(n_2 - a_2) n_1} = \frac{1}{n_1}.$$

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

Analogously, we obtain, for all $r_2 \in \{1, \dots, n_2\}$, $q_2^{A_1,A_2}(r_2) = 1/n_2$. Thus, independently of the choice of A_1 and A_2, the marginals of q^{A_1,A_2} are the uniform distributions on $\{1, \dots, n_1\}$ and $\{1, \dots, n_2\}$, respectively. Now, define $A_{0,1} = A_{0,2} = \varnothing$. Inductively, choose a pair $(F_{k,1}, F_{k,2}) \in \{1, \dots, n_1\} \times \{1, \dots, n_2\}$ at random according to $q^{A_{k-1,1},A_{k-1,2}}$ and define $A_{k,i} = A_{k-1,i} \cup \{F_{k,i}\}$. Clearly, $a_{T,i} = N_i$; hence, it is enough to show that

$$a_{k,1} \le a_{k,2} \quad \text{for all } k \in \mathbb{N}_0. \tag{3.2}$$

For $k = 0$, (3.2) holds trivially. Now we assume that (3.2) holds for $k - 1$, and we show that it also holds for k. If $a_{k-1,1} < a_{k-1,2}$ then $a_{k,1} \le a_{k-1,1} + 1 \le a_{k-1,2} \le a_{k,2}$. If $a_{k-1,1} = a_{k-1,2}$ then either $F_{k,1} \in A_{k-1,1}$, which implies that $a_{k,1} = a_{k-1,1} = a_{k-1,2} \le a_{k,2}$, or $F_{k,1} \in A^c_{k-1,1}$. In the latter case, according to the definition of q^{A_1,A_2}, we have $F_{k,2} \in A^c_{k-1,2}$; hence, $a_{k,1} = a_{k-1,1} + 1 = a_{k-1,2} + 1 = a_{k,2}$.

4. Method 3: Markov chains

The aim of this section is to give a proof of Theorem 1.1(a) that uses the interpretation of the binomial distribution as the distribution of nonempty boxes when we successively throw balls into n boxes. In contrast to Method 2, here we do not construct an explicit coupling of the random variables, but use Markov chains in order to get a very quick and elementary proof that could be taught in any first course on probability theory.

Let $n, t \in \mathbb{N}$. Assume that we throw t balls independently into n boxes with numbers $1, \dots, n$, and denote by $N_{n,t}$ the number of nonempty boxes. Let T be random and Poisson distributed with parameter $\lambda = -n \log(1 - p)$. Assume that T is independent of the numbers $N_{n,t}$, $t = 1, 2, \dots$. As indicated in Section 3, the number $N_{n,T}$ is binomially distributed with parameters n and p. Hence, in order to show Theorem 1.1(a), it is enough to show the following proposition.

Proposition 4.1. *For each $t \in \mathbb{N}$, the sequence $(N_{n,t})_{n\in\mathbb{N}}$ is stochastically increasing.*

Proposition 4.1 is in fact a special case of a more general result where the probabilities p_i for hitting box $i = 1, \dots, n$ differ from box to box (see [14]).

Proof of Proposition 4.1. For each n, $(N_{n,t})_{t=0,1,\dots}$ is a Markov chain on $\{0, \dots, n\}$ with transition matrix

$$p_n(k,l) = \begin{cases} k/n & \text{if } l = k, \\ 1 - k/n & \text{if } l = k + 1, \\ 0 & \text{otherwise.} \end{cases}$$

Define

$$h_{n,l}(k) = \sum_{j=l}^{n} p_n(k,j) = \begin{cases} 0 & \text{if } k < l - 1, \\ 1 - k/n & \text{if } k = l - 1, \\ 1 & \text{if } k > l - 1, \end{cases}$$

and note that $h_{n,l}(k)$ is increasing in k and n.

Let $m < n$, and note that, trivially, $N_{m,0} = 0$ is stochastically smaller than $N_{n,0} = 0$. By induction we show that $N_{m,t} \le_{\mathrm{st}} N_{n,t}$ for all $t \in \mathbb{N}_0$. Indeed, for every $\ell \in \{0, \dots, m\}$, by the induction hypothesis and due to the monotonicity of $(n, k) \mapsto h_{n,l}(k)$, we have

$$\mathrm{P}[N_{m,t+1} \ge l] = \mathrm{E}[h_{m,l}(N_{m,t})] \le \mathrm{E}[h_{m,l}(N_{n,t})] \le \mathrm{E}[h_{n,l}(N_{n,t})] = \mathrm{P}[N_{l,t+1} \ge l].$$

This however implies that $N_{m,t+1}$ is stochastically smaller than $N_{n,t+1}$.

5. Method 4: analytic proof

The aim of this section is to give proofs of Theorem 1.1(a) and (b) that rely on changing the parameter p of the distributions continuously and using calculus to compute the dependence of the distributions on this parameter. Although the proofs for (a) and (b) are rather similar, we felt that it was no loss in efficiency to give two separate proofs.

5.1. Proof of Theorem 1.1(a): binomial distributions

We only have to show sufficiency of the tail conditions (1.8) and (1.9) for $b_{n_1,p_1} \leq_{\mathrm{st}} b_{n_2,p_2}$. By (1.1) we only have to consider the case in which $n_1 < n_2$ and $(1-p_1)^{n_1} = (1-p_2)^{n_2}$.

Let $R := n_2/n_1 > 1$, and define the map

$$\pi\colon [0,1] \to [0,1], \qquad p \mapsto 1-(1-p)^R. \tag{5.1}$$

Denote by $\pi'(p) = R(1-p)^{R-1}$ the derivative of π.

For $n \in \mathbb{N}$, $p \in (0,1)$, and $A \subset \{0,\dots,n\}$, define

$$b'_{n,p}(A) = \frac{\mathrm{d}}{\mathrm{d}p} b_{n,p}(A).$$

Computing the derivative for $A = \{k\}$, $k = 0,\dots,n$, explicitly yields

$$b'_{n,p}(\{k\}) = -n[b_{n-1,p}(\{k\}) - b_{n-1,p}(\{k-1\})]$$

(where $b_{n-1,p}(\{-1\}) = 0$ and $b_{0,p}(\{k\}) = 1$ if and only if $k = 0$). Hence, building a telescope sum, we obtain

$$b'_{n,p}(\{0,\dots,k\}) = -nb_{n-1,p}(\{k\}) \quad \text{for } n \in \mathbb{N},\ p \in [0,1],\ k \in \mathbb{N}_0. \tag{5.2}$$

For $k \in \mathbb{N}_0$, define the map

$$f_k\colon [0,1] \to \mathbb{R}, \qquad p \mapsto b_{n_1,\pi(p)}(\{0,\dots,k\}) - b_{n_2,p}(\{0,\dots,k\}).$$

As $\pi(p_2) = p_1$, we have to show that $f_k(p_2) \geq 0$ for all k. Obviously, only the case $k \in \{1,\dots,n_1-1\}$ is nontrivial, and we fix such a k for the rest of this proof.

Since $\pi(0) = 0$ and $\pi(1) = 1$, we have $f_k(0) = f_k(1) = 0$. As f_k is differentiable in $(0,1)$ and continuous on $[0,1]$, it is enough to show that

$$f'_k(p) \text{ is strictly positive in a neighbourhood of } 0 \tag{5.3}$$

and

$$f'_k(p) = 0 \text{ for at most one } p \in (0,1). \tag{5.4}$$

Using (5.2), we compute the derivative

$$\begin{aligned}
f'_k(p) &= \pi'(p) b'_{n_1,\pi(p)}(\{0,\dots,k\}) - b'_{n_2,p}(\{0,\dots,k\}) \\
&= n_2 b_{n_2-1,p}(\{k\}) - n_1 \pi'(p) b_{n_1-1,\pi(p)}(\{k\}) \\
&= n_2 \binom{n_2-1}{k} p^k (1-p)^{n_2-1-k} \\
&\quad - n_1 R(1-p)^{R-1} \binom{n_1-1}{k} (1-(1-p)^R)^k (1-p)^{R(n_1-1-k)} \\
&= n_2(1-p)^{n_2-1}\left[\binom{n_2-1}{k}\left(\frac{p}{1-p}\right)^k - \binom{n_1-1}{k}\left(\frac{1-(1-p)^R}{(1-p)^R}\right)^k\right].
\end{aligned}$$

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

Hence, (5.3) follows from (recall that $R = n_2/n_1 > 1$)

$$\lim_{p\downarrow 0} \frac{f_k'(p)}{n_2 p^k} = \binom{n_2 - 1}{k} - \binom{n_1 - 1}{k} R^k > 0.$$

Now, for $p \in (0, 1)$, we have $f_k'(p) = 0$ if and only if

$$g(p) := p(1-p)^{R-1} - a(1 - (1-p)^R) = 0,$$

where

$$a := \left(\binom{n_1 - 1}{k} \Big/ \binom{n_2 - 1}{k}\right)^{1/k}.$$

Since $f_k'(p) > 0$ for sufficiently small $p > 0$, $g(p) > 0$ for these p also. Hence, in order to show that $g(p) = 0$ for at most one $p \in (0, 1)$, it is enough to show that $g'(p) = 0$ for at most one $p \in (0, 1)$. To this end, we compute

$$g'(p) = (1-p)^{R-2}[1 - aR - (1-a)Rp].$$

Hence, $g'(p) = 0$ exactly for $p = 1$ and $p = (1 - aR)/(1 - a)R$. As this shows (5.4), the proof is complete.

5.2. Proof of Theorem 1.1(b): negative binomial distributions

We only have to show sufficiency of the tail conditions (1.10) and (1.11) for $b^-_{r_1,p_1} \le_{\text{st}} b^-_{r_2,p_2}$. Recall that if $r_1 < r_2$ and X_1 and X_2 are independent random variables with distributions $b^-_{r_1,p}$ and $b^-_{r_2-r_1,p}$, respectively, then $X_1 + X_2$ has distribution $b_{r_2,p}$. That is, $b^-_{r_1,p} \le_{\text{st}} b^-_{r_2,p}$ if and only if $r_1 \le r_2$. Hence, we may assume that $r_1 \ge r_2$.

Step 1. By a simple computation we obtain

$$\frac{\mathrm{d}}{\mathrm{d}p} b^-_{r,p}(\{0, \dots, k-1\}) = k\binom{-r}{k}(-1)^k(1-p)^{k-1}p^{r-1}. \tag{5.5}$$

In fact, multiplying both sides of (5.5) by p^{-r}, as functions of r,

$$p^{-r}\frac{\mathrm{d}}{\mathrm{d}p} b^-_{r,p}(\{0, \dots, k-1\}) \quad \text{and} \quad k\binom{-r}{k}(-1)^k(1-p)^{k-1}p^{-1}$$

are polynomials. Hence, it is enough to check (5.5) for all $r \in \mathbb{N}$. Either by a direct computation or by appealing to the waiting time interpretation that in a Bernoulli chain with success probability p, $b^-_{n,p}$ is the distribution of the number of failures before the nth success occurs, we obtain

$$b^-_{n,p}(\{0, \dots, k-1\}) = b_{n+k-1,1-p}(\{0, \dots, k-1\}).$$

Hence, by (5.2) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}p} b^-_{n,p}(\{0, \dots, k-1\}) = (n+k-1)b_{n+k-2,1-p}(\{k-1\}) = k\binom{-n}{k}(-1)^k(1-p)^{k-1}p^{n-1},$$

as desired.

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

Step 2. Let $R = r_2/r_1 < 1$. Fix $k \in \mathbb{N}$. For $p \in (0, 1]$, define

$$f(p) := b^-_{r_1, p^R}(\{0, \ldots, k-1\}) - b^-_{r_2, p}(\{0, \ldots, k-1\}).$$

It is enough to show that $f(p) > 0$ for all $p \in (0, 1]$.

Clearly, $f(1) = 0$ and $\lim_{p \downarrow 0} f(p) = 0$. Hence, it is enough to show that $f'(p) > 0$ for sufficiently small p and $f'(p) = 0$ for at most one $p \in (0, 1)$. By (5.5) we have

$$f'(p) = kp^{r_2 - 1}\left[\binom{r_1 + k - 1}{k} R(1 - p^R)^{k-1} - \binom{r_2 + k - 1}{k}(1 - p)^{k-1}\right].$$

Hence,

$$\lim_{p \downarrow 0} \frac{f'(p)}{kp^{r_2 - 1}} = \binom{r_1 + k - 1}{k} R^k - \binom{r_2 + k - 1}{k} > 0.$$

Define

$$g(p) := c(1 - p^R) - (1 - p),$$

where

$$c := \left(\binom{r_1 + k - 1}{k} R \Big/ \binom{r_2 + k - 1}{k}\right)^{1/(k-1)}.$$

Clearly, $g(p) = 0$ if and only if $f'(p) = 0$. It is easy to see that $g(1) = 0$ and $g'(p) = 0$ if and only if $p = (cR)^{1/(1-R)}$. This implies that $g(p) = 0$ for at most one $p \in (0, 1)$.

6. Method 5: infinite divisibility

In this section, for infinitely divisible distributions on $[0, \infty)$, we derive a sufficient criterion (Lemma 6.1) for stochastic ordering in terms of the Lévy measures. We use this criterion to give a short proof of Theorem 1.1 for negative binomial distributions.

6.1. Stochastic ordering of infinitely divisible laws

An infinitely divisible distribution P on $[0, \infty)$ is characterized by the deterministic part $\alpha_P \in [0, \infty)$ and the Lévy measure ν_P on $(0, \infty)$. The connection is given via the Lévy–Khinchin formula (see, e.g. [3, Theorem 16.14])

$$-\log\left(\int \mathrm{e}^{-tx} P(\mathrm{d}x)\right) = \alpha_P t + \int (1 - \mathrm{e}^{-tx})\nu_P(\mathrm{d}x) \quad \text{for all } t \geq 0.$$

If μ and ν are two measures on arbitrary measurable spaces, we write $\mu \leq \nu$ if $\mu(A) \leq \nu(A)$ for every measurable set A. If X is a nonnegative infinitely divisible random variable with distribution P_X, we write $\nu_X = \nu_{\mathrm{P}_X}$ and $\alpha_X = \alpha_{\mathrm{P}_X}$ for the corresponding characteristics. If X and Y are independent then $\nu_{X+Y} = \nu_X + \nu_Y$ and $\alpha_{X+Y} = \alpha_X + \alpha_Y$. Thus, for infinitely divisible distributions P and Q, we have

$$\alpha_P \leq \alpha_Q \quad \text{and} \quad \nu_P \leq \nu_Q \quad \Longrightarrow \quad P \leq_{\mathrm{st}} Q. \tag{6.1}$$

(The opposite implication is not true.) For example, the negative binomial distribution $b^-_{r,p}$ is infinitely divisible with vanishing deterministic part and Lévy measure $\nu_{r,p}$ concentrated on $\mathbb{N}$, and given by

$$\nu_{r,p}(\{k\}) = \lim_{\lambda \downarrow 0} \lambda^{-1} b^-_{\lambda r, p}(\{k\}) = r\frac{(1-p)^k}{k} \quad \text{for } k \in \mathbb{N}. \tag{6.2}$$

In fact, a direct computation yields

$$-\log \sum_{k=0}^{\infty} b^{-}_{r,p}(\{k\})\mathrm{e}^{-tk} = r\log(p^{-1}(1-(1-p)\mathrm{e}^{-t})) = \sum_{k=1}^{\infty} \nu_{r,p}(\{k\})(1-\mathrm{e}^{-tk}).$$

Note that $\nu_{r_1,p_1} \le \nu_{r_2,p_2}$ if $p_1 = p_2$ and $r_1 \le r_2$ or if $r_1 = r_2$ and $p_1 \ge p_2$. Hence, similarly as for the binomial distribution, by (6.1), we obtain the two relations (for all $p \in (0,1)$ and $r > 0$)

$$b^{-}_{r_1,p} \le_{\mathrm{st}} b^{-}_{r_2,p} \iff r_1 \le r_2$$

and

$$b^{-}_{r,p_1} \le_{\mathrm{st}} b^{-}_{r,p_2} \iff p_1 \ge p_2.$$

For the case where both parameters differ, we need a more subtle criterion. Let μ_1 and μ_2 be two measures on $(0,\infty)$ with $\mu_i([x,\infty)) < \infty$ for all $x \in (0,\infty)$, $i = 1,2$. Extending the notion of stochastic ordering to such measures, we write $\mu_1 \le_{\mathrm{st}} \mu_2$ if

$$\mu_1([x,\infty)) \le \mu_2([x,\infty)) \quad \text{for } x \in (0,\infty).$$

Note that $\mu_1 \le \mu_2$ implies that $\mu_1 \le_{\mathrm{st}} \mu_2$.

For a multidimensional version of the following lemma, see [9, Theorem 2.2].

Lemma 6.1. *Let P_i, $i = 1,2$, be infinitely divisible distributions on $[0,\infty)$ with deterministic parts α_i and Lévy measures ν_i. Assume that $\alpha_1 \le \alpha_2$ and $\nu_1 \le_{\mathrm{st}} \nu_2$. Then there exist random variables Z_1 and Z_2 with distributions P_1 and P_2, respectively, such that $Z_1 \le Z_2$ almost surely. In particular, we have $P_1 \le_{\mathrm{st}} P_2$.*

Proof. It is enough to consider the situation $\alpha_1 = \alpha_2 = 0$. Let $G_i(x) = \nu_i([x,\infty))$ for $x \in (0,\infty)$, and define the inverse function $G_i^{-1}(y) := \inf\{x \ge 0 \colon G_i(x) \le y\}$ for $y \in (0,\infty)$. Let X be a Poisson point process on $(0,\infty)$ with rate 1. That is, X is an integer-valued random measure on $(0,\infty)$; for bounded measurable sets A, $X(A)$ is Poisson distributed with the Lebesgue measure of A as the parameter, and, for pairwise disjoint sets, the values of X are independent random variables. Define

$$Z_i := \int G_i^{-1}(y) X(\mathrm{d}y), \qquad i = 1,2.$$

Then, for $t \ge 0$, we have (compare [3, Theorem 24.14])

$$-\log \mathrm{E}[\mathrm{e}^{-tZ_i}] = \int_0^{\infty} (1-\mathrm{e}^{-tG_i^{-1}(y)})\,\mathrm{d}y = \int (1-\mathrm{e}^{-ty})\nu_i(\mathrm{d}y) = -\log \int \mathrm{e}^{-tx} P_i(\mathrm{d}x).$$

Thus, Z_i has distribution P_i. By the assumption that $G_1 \le G_2$, we have $G_1^{-1} \le G_2^{-1}$ and, hence, $Z_1 \le Z_2$ almost surely.

6.2. Proof of Theorem 1.1(b)

We only have to show sufficiency of the tail conditions (1.10) and (1.11) for $b^{-}_{r_1,p_1} \le_{\mathrm{st}} b^{-}_{r_2,p_2}$.

Recall from (6.2) that the negative binomial distribution $b^{-}_{r_i,p_i}$ is infinitely divisible with deterministic part $\alpha_{r_i,p_i} = 0$ and Lévy measure ν_{r_i,p_i} being concentrated on $\mathbb{N}$ and given by

$$\nu_{r_i,p_i}(\{k\}) = r_i \frac{(1-p_i)^k}{k} \quad \text{for all } k \in \mathbb{N}.$$

https://doi.org/10.1239/aap/1275055235 Published online by Cambridge University Press

As $p_1 \geq p_2$, we have $\nu_{r_1,p_1} \leq_{\mathrm{lr}} \nu_{r_2,p_2}$; that is, the map

$$k \mapsto \frac{\nu_{r_1,p_1}(\{k\})}{\nu_{r_2,p_2}(\{k\})} = \frac{r_1}{r_2}\frac{(1-p_1)^k}{(1-p_2)^k}$$

is monotone decreasing. This implies that

$$k \mapsto \phi(k) := \frac{\nu_{r_1,p_1}(\{k, k+1, \dots\})}{\nu_{r_2,p_2}(\{k, k+1, \dots\})} = \frac{r_1}{r_2}\frac{\sum_{l=k}^{\infty} l^{-1}(1-p_1)^l}{\sum_{l=k}^{\infty} l^{-1}(1-p_2)^l}$$

is monotone decreasing. By the assumption that $p_1^{r_1} \geq p_2^{r_2}$ we have

$$\phi(1) = \frac{r_1 \log(p_1)}{r_2 \log(p_2)} \leq 1.$$

Hence, $\phi(k) \leq 1$ for all $k \in \mathbb{N}$; that is, $\nu_{r_1,p_1} \leq_{\mathrm{st}} \nu_{r_2,p_2}$ and, thus, $b^-_{r_1,p_1} \leq_{\mathrm{st}} b^-_{r_2,p_2}$ by Lemma 6.1.

6.3. Proof of Theorem 1.1(g)

When viewed from the perspective of infinitely divisible distributions, the statement of Theorem 1.1(g) is trivial. In fact, $b^-_{r,p}$ is infinitely divisible and the Lévy measure $\nu_{r,p}$ has total mass $\nu_{r,p}(\mathbb{N}) = -r\log(p)$. Since Poi_λ is infinitely divisible with Lévy measure $\nu_\lambda = \lambda\delta_1$, we see that $\nu_\lambda \leq_{\mathrm{st}} \nu_{r,p}$ if and only if $\mathrm{e}^{-\lambda} \geq p^r$. Hence, the claim follows using Lemma 6.1.

Acknowledgements

We thank Abram M. Kagan and Sergei V. Nagaev for fruitful discussions which triggered the present work.

References

[1] Boland, P. J., Singh, H. and Cukic, B. (2002). Stochastic orders in partition and random testing of software. *J. Appl. Prob.* **39,** 555–565.
[2] Gastwirth. J. L. (1977). A probability model of a pyramid scheme. *Amer. Statist.* **31,** 79–82.
[3] Klenke, A. (2008). *Probability Theory*. Springer, London.
[4] Ma, C. (1997). A note on stochastic ordering of order statistics. *J. Appl. Prob.* **34,** 785–789.
[5] Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley, Chichester.
[6] Pfanzagl, J. (1964). On the topological structure of some ordered families of distributions. *Ann. Math. Statist.* **35,** 1216–1228.
[7] Proschan, F. and Sethuraman, J. (1976). Stochastic comparisons of order statistics from heterogeneous populations, with applications in reliability. *J. Multivariate Anal.* **6,** 608–616.
[8] Rüschendorf, L. (1991). On conditional stochastic ordering of distributions. *Adv. Appl. Prob.* **23,** 46–63.
[9] Samorodnitsky, G. and Taqqu, M. S. (1993). Stochastic monotonicity and Slepian-type inequalities for infinitely divisible and stable random vectors. *Ann. Prob.* **21,** 143–160.
[10] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer, New York.
[11] Szekli, R. (1995). *Stochastic Ordering and Dependence in Applied Probability* (Lecture Notes Statist. **97**). Springer, New York.
[12] Vatutin, V. A. and Mikhajlov, V. G. (1982). Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory Prob. Appl.* **27,** 734–743.
[13] Whitt, W. (1980). Uniform conditional stochastic order. *J. Appl. Prob.* **17,** 112–123.
[14] Wong, C. K. and Yue, P. C. (1973). A majorization theorem for the number of distinct outcomes in N independent trials. *Discrete Math.* **6,** 391–398.