THE QUARTER OBSERVATION FOR HYPERGEOMETRIC TAILS: THE WHOLE BAND $1 < \mathbb{E}H \le 2$, THE SHARP CONSTANT, AND THE EXACT MINIMUM FOR $n \le 250$

KOREA SUPERINTELLIGENCE LABS

Abstract. Let $H \sim \mathrm{Hyp}(n, i, k)$ count the black marbles in a sample without replacement of size k drawn from an urn holding i black and $n - i$ white marbles, so that $\mathbb{E}H = ik/n$. In a study of lower bounds for $\mathbb{P}(H \ge \mathbb{E}H)$, Ai and Pelekis record a numerical observation: under the two hypotheses $ik/n \ge \frac{1}{2}$ and $(n-i)(n-k)/n \ge \frac{1}{2}$ one has $\mathbb{P}(H \ge \mathbb{E}H) \ge \frac{1}{4}$. They prove no case of it, and a later note of Ai, Kuželka and Pelekis states that the corresponding minimisation is not known below $n = 8k$. Our main theorem proves the inequality on the whole band $n < ik \le 2n$, that is $1 < \mathbb{E}H \le 2$, for all n, i and k and with neither hypothesis assumed; together with the bound $\frac{1}{3}$ that we prove when $\mathbb{E}H \le 1$, this settles every admissible triple with $\mathbb{E}H \le 2$. The band is where the observation is tight: it contains the family $k = 2$, n odd, $i = \frac{n+1}{2}$, whose tail is exactly $\frac{n+1}{4n}$, so no constant larger than $\frac{1}{4}$ can be used, and $\frac{1}{4}$ is the infimum there. The engine of the proof is a binomial inequality obtained from a five-term truncation of Vandermonde's convolution which, unlike the corresponding truncation of the elementary estimate $\binom{a}{\kappa}/\binom{b}{\kappa} \ge (a/b)^\kappa$, leaves no exceptional pairs at all, so no finite table of cases enters. We prove the observation for every n also when $\min(i, k) \le 3$ or $\min(n - i, n - k) \le 3$, the second half through a complementation invariance $T(n, i, k) = T(n, n - i, n - k)$ which exhibits the two hypotheses as the two halves of one symmetric pair; and we certify the whole admissible range $n \le 250$ by exhaustive exact computation, determining its exact minimum $\frac{125}{498} = \frac{249+1}{4\cdot 249}$, attained at $(249, 125, 2)$. Letting $n \to \infty$ inside the band degenerates the main theorem to the part of the binomial theorem of Greenberg and Mohri in which the threshold is 2, which is published; what is new is the statement at finite n, for sampling without replacement, and its independence of the two hypotheses. What remains open is $\mathbb{E}H > 2$ with $\min(i, k) \ge 4$ and $\min(n - i, n - k) \ge 4$; we explain why the published routes to a constant $\frac{1}{4}$ for this distribution cannot reach the band, and why the region that is left is no longer a question about the constant $\frac{1}{4}$. Every statement below is machine-checked in Lean 4, and all of it by the kernel alone except the lower-bound half of the $n \le 250$ result, which additionally rests on six compiled evaluations.

1. Introduction

1.1. The tail of a hypergeometric law. Fix integers $n \ge 1$ and $i, k \in \{1, \ldots, n\}$. An urn holds n marbles, i of them black; a sample of size k is drawn without replacement, and $H = H_{n,i,k}$ is the number of black marbles in it, so that

$$\mathbb{P}(H = j) = \frac{\binom{i}{j}\binom{n-i}{k-j}}{\binom{n}{k}}, \tag{1}$$

with the convention $\binom{a}{b} = 0$ for $b < 0$ and for $b > a$, so that j runs over $\max(0, k-(n-i)) \le j \le \min(i, k)$. We write $H \sim \mathrm{Hyp}(n, i, k)$. The first two moments are

$$\mathbb{E}H = \frac{ik}{n}, \qquad \mathrm{Var}\, H = k \cdot \frac{i}{n} \cdot \frac{n-i}{n} \cdot \frac{n-k}{n-1} \quad (n \ge 2). \tag{2}$$

The object of this paper is the quantity

$$T(n, i, k) \;:=\; \mathbb{P}\big(H \ge \mathbb{E}H\big) \;=\; \sum_{j \ge ik/n} \mathbb{P}(H = j) \;=\; \sum_{j \ge \lceil ik/n \rceil} \mathbb{P}(H = j), \tag{3}$$

the probability that a hypergeometric variable is at least its mean. Since H is integer-valued while $\mathbb{E}H = ik/n$ usually is not, the event is a comparison of an integer with a rational; the two right-hand sides of (3) agree, and this is the event written $\mathbb{P}(H \ge ik/n)$ throughout [1].

The quantity is not a curiosity. Ai and Pelekis [1] arrive at it from a smuggler-and-customs problem — a unit of contraband is spread equally over i of n couriers, exactly k of whom are searched, and the operation stays profitable, in the sense that the unseized amount exceeds its own expectation $\frac{n-k}{n}$, with probability $1 - T(n,i,k)$ — and from the Manickam–Miklós–Singhi conjecture: as [1] explains, if the conjecture of Aydinian and Blinovsky holds, then the hypergeometric tails $T(n,i,k)$, minimised over i, are exactly the extreme instances of that conjecture.

1.2. **What is proved, and what is observed.** Write $\mathrm{Bin}(k,p)$ for the binomial law. The binomial counterpart of (3) has a constant lower bound: Greenberg and Mohri [3, Theorem 1] proved that $\mathbb{P}(\mathrm{Bin}(k,p) \geq kp) > \frac{1}{4}$ for every positive integer k and every $p > \frac{1}{k}$, and their title records that the bound is tight; [1] quotes it, as its Theorem 3.2, in the weaker form $\geq \frac{1}{4}$ for $p \in (\frac{1}{k}, 1)$. The hypergeometric side is thinner. Ai and Pelekis [1] prove two lower bounds:

- $T(n,i,k) \geq k/n$ whenever $n \geq 8k$ ([1, Theorem 1.1]); and
- a variance-dependent bound ([1, Theorem 1.2]) which, as the paper itself remarks, "can never be larger than $\frac{e^{-1/12}}{4\sqrt{2}} \approx 0.162$".

They also record two bounds from the literature. A hypergeometric variable is a sum of independent Bernoulli variables (Vatutin and Mikhailov [7]), so Berry–Esseen theory applies; [1] quotes from Mattner and Schulz [6, Formula (17)] the estimate

$$T(n,i,k) \ \geq \ \frac{1}{2} - \frac{0.5583}{\sqrt{\operatorname{Var} H}}, \tag{4}$$

from which $T \geq \frac{1}{4}$ follows as soon as $\operatorname{Var} H \geq (4 \cdot 0.5583)^2 \geq 4.98$. And combining the same representation with Ehm's total-variation estimate [5, Theorem 2], [1] records that if $X \sim \mathrm{Bin}(k, \frac{i}{n})$ satisfies $\operatorname{Var} X \geq 1$ then

$$T(n,i,k) \ \geq \ \mathbb{P}(X \geq \mathbb{E}X) - \frac{k-1}{n-1}. \tag{5}$$

Section 1 of [1] then closes with a sentence we quote verbatim:

> "Furthermore, our numerical experiments suggest that under the assumptions $\frac{ik}{n} \geq \frac{1}{2}$ and $\frac{(n-i)(n-k)}{n} \geq \frac{1}{2}$ it holds $\mathbb{P}(H \geq \mathbb{E}(H)) \geq \frac{1}{4}$."

Nothing further is said: no case is proved, no range of the experiments is stated, the constant is not asserted to be sharp, and no family approaching it is named. That the region is genuinely open is the authors' own later position: in June 2026 Ai, Kuželka and Pelekis [2] write of the minimisation of these very tails that "the latter optimization problem can be solved under the assumption $n \geq 8k$ (see [1]) but, to the best of our knowledge, its solution is not known for smaller values of n."

Definition 1.1. A triple (n,i,k) of integers is *admissible* if $1 \leq i \leq n$, $1 \leq k \leq n$ and

$$\frac{ik}{n} \ \geq \ \frac{1}{2}, \qquad \frac{(n-i)(n-k)}{n} \ \geq \ \frac{1}{2},$$

equivalently, in integers, $n \leq 2ik$ and $n \leq 2(n-i)(n-k)$. We write $\mathcal{A}$ for the set of admissible triples. Admissibility already forces $n \geq 2$.

In this notation the observation of [1] reads: $T(n,i,k) \geq \frac{1}{4}$ for every $(n,i,k) \in \mathcal{A}$. We call it the *quarter observation.*

1.3. **Results.** The parameter square splits along the line $\mathbb{E}H = 1$ and along the size of $\min(i,k)$, which is the number of cells of (1) that the tail can contain. Our first theorem settles the whole boundary strip, at every n.

Theorem 1.2 (Theorems 3.2, 5.1 and 5.4)**.** *Let (n,i,k) be admissible.*

1. *If $ik \leq n$, i.e. if $\frac{1}{2} \leq \mathbb{E}H \leq 1$, then $T(n,i,k) \geq \frac{1}{3}$. Only the first hypothesis of Definition 1.1 is used.*
2. *If $\min(i,k) \leq 3$, then $T(n,i,k) \geq \frac{1}{4}$.*

Part (1) is stronger than the observation on its regime, and says where the truth is not: the constant $\frac{1}{4}$ is not the right one when the mean is at most one. It is where the mean exceeds one that the observation becomes delicate, and our second theorem says how delicate.

Theorem 1.3 (Theorem 6.1 and Corollary 6.2)**.** *For every odd $n \geq 5$ the triple $\big(n, \frac{n+1}{2}, 2\big)$ is admissible and*

$$T\Big(n, \tfrac{n+1}{2}, 2\Big) \;=\; \frac{n+1}{4n} \;=\; \frac{1}{4} + \frac{1}{4n} \qquad \textit{exactly.}$$

Consequently no constant larger than $\frac{1}{4}$ can replace it in the quarter observation: for every rational $c > \frac{1}{4}$ there is an admissible triple with $T < c$. At $k = 2$ the inequality $T \geq \frac{1}{4}$ is moreover strict for every admissible triple, so the value $\frac{1}{4}$ is not attained there; and on the region of Theorem 1.2(2) the infimum of T is exactly $\frac{1}{4}$.

The third theorem replaces the unquantified numerical experiments of [1] by an exhaustive exact computation on a stated range, and reports not a verified inequality but the exact minimum on that range.

Theorem 1.4 (Theorems 7.1 and 7.2)**.** *For every admissible triple with $n \leq 250$,*

$$T(n, i, k) \;\geq\; \frac{125}{498} \;=\; \frac{249+1}{4 \cdot 249} \;=\; 0.25100\ldots,$$

and the bound is attained: $(249, 125, 2)$ is admissible with $T(249, 125, 2) = \frac{125}{498}$. Thus $\frac{125}{498}$ is the exact minimum of the hypergeometric tail over the whole admissible range $n \leq 250$, and it is attained by the member of the family of Theorem 1.3 at the largest odd n in that range. In particular the quarter observation holds, strictly, for every admissible triple with $n \leq 250$.

The three theorems above cut the parameter square by $\min(i, k)$, the number of cells the tail can contain. Our main new result cuts it in the other direction, by the position of the mean. Theorem 1.2(1) is the region below the line $\mathbb{E}H = 1$; the next line is $\mathbb{E}H = 2$, and between the two the observation holds with neither of its hypotheses.

Theorem 1.5 (Theorem 9.5 and Corollary 9.6)**.** *Let n, i, k be integers with $1 \leq i, k \leq n$ and*

$$n < ik \leq 2n, \qquad \textit{that is,} \qquad 1 < \mathbb{E}H \leq 2.$$

Then $T(n, i, k) \geq \frac{1}{4}$; neither hypothesis of Definition 1.1 is assumed. Consequently, with Theorem 1.2(1), every admissible triple with $\mathbb{E}H \leq 2$ satisfies the quarter observation.

The band is transversal to the strip of Theorem 1.2(2), and it contains the extremal family of Theorem 1.3: it contains the shape at which the constant $\frac{1}{4}$ is sharp, so the theorem is tight and its proof has to be. It also contains triples that no theorem about a small value of $\min(i, k)$ or of $\min(n - i, n - k)$ reaches, such as $(45, 12, 4)$, and triples that are not admissible at all, such as $(3, 2, 2)$.

Two qualifications are owed. First, fix $k \geq 2$ and $p \in (\frac{1}{k}, \frac{2}{k})$ and put $i = \lceil pn \rceil$: the triple (n, i, k) is then in the band for all large n, and $\mathrm{Hyp}(n, i, k) \to \mathrm{Bin}(k, p)$, so Theorem 1.5 gives in the limit $\mathbb{P}(\mathrm{Bin}(k, p) \geq kp) \geq \frac{1}{4}$ for those p. That is the part of [3, Theorem 1] in which the threshold $\lceil kp \rceil$ equals 2, published in 2014 and in the stronger, strict form. What is new here is therefore the statement at finite n, for sampling without replacement, and its independence of the two hypotheses. Second, and for the same reason, no proof of the band can be softer than a proof of that part of [3]: the argument of Section 9 does reprove it, from scratch and for the hypergeometric law.

A second, smaller addition changes the shape of the strip that Theorem 1.2(2) settles. The parameter square carries a second involution besides the transpose $(i, k) \mapsto (k, i)$ of Lemma 4.1: passing simultaneously to the complement of the sample and to the complement of the set of black marbles gives $T(n, i, k) = T(n, n - i, n - k)$ (Lemma 4.2), and the two conditions of Definition 1.1 are exchanged by it — they are the two halves of one symmetric pair, which is why the observation has two hypotheses rather than one. Theorem 5.4 therefore doubles: the quarter observation holds, for every n, whenever $\min(i, k) \leq 3$ *or* $\min(n - i, n - k) \leq 3$ (Corollary 5.5).

1.4. **What is not claimed.** The quarter observation is *not* proved here in general. What remains is exactly the region

$$\mathbb{E}H > 2, \qquad \min(i,k) \geq 4, \qquad \min(n-i, n-k) \geq 4,$$

outside the range $n \leq 250$ of Theorem 1.4. Section 11 explains what can be said about it. Two things have changed there. The published routes to the constant $\frac{1}{4}$ for this distribution fail on the band of Theorem 1.5 by a margin larger than the margin the statement itself has, so that band was not reachable by importing a bound; and on the region that is left the constant $\frac{1}{4}$ is, on the numerical evidence recorded in Section 11, no longer the natural one, so what remains is a different question from the one the extremal family poses. We also do not prove that the minimiser of Theorem 1.4 is unique; we do not prove strictness on the whole region $\min(i,k) \leq 3$ (only at $k = 2$); the limiting statement discussed above, and recorded as Remark 9.7, is argued in prose and is not part of the verified development; and we do not address the second numerical remark of [1], that the constant $\frac{e^{-1/12}}{4\sqrt{2}}$ of its Theorem 1.2 may be replaceable by $\frac{1}{2\sqrt{2}}$.

1.5. **Changes from the first version.** This is a second version of the paper, and it is additive: no statement of the first version has been withdrawn, weakened or renumbered. Definition 1.1, Theorems 1.2, 1.3 and 1.4, Proposition 2.1, Lemma 3.1, Theorem 3.2, Lemma 4.1, Theorems 5.1 and 5.2, Remark 5.3, Theorem 5.4, Theorem 6.1, Corollary 6.2, Theorems 7.1 and 7.2 and Proposition 8.1 stand exactly as they stood, with the same numbers. What is new is: Theorem 1.5, proved as Theorem 9.5 with Corollary 9.6 in Section 9, together with the two lemmas it rests on (Lemmas 9.1 and 9.2) and the binomial inequality Lemma 9.3 that carries it; the complement symmetry Lemma 4.2 and the doubled strip Corollary 5.5; the telescoping identity and the order-statistic form of the tail (Section 10); and the controls Propositions 5.6, 9.8 and 10.4. Section 11 has been rewritten, because the region it discusses has changed: the first version left $\mathbb{E}H > 1$ with $\min(i,k) \geq 4$ open and recorded that the constant is sharp there; what is left now is $\mathbb{E}H > 2$ with $\min(i,k) \geq 4$ and $\min(n-i, n-k) \geq 4$, on which it is not. One item of that section's list of priced and unfinished work — the complementary invariance, which the first version stated as an observation and explicitly did not establish — is now Lemma 4.2. None of the new results uses a compiled evaluation of any kind.

2. Preliminaries

Throughout, $n \geq 1$ and $i, k \in \{1, \ldots, n\}$, and $H \sim \mathrm{Hyp}(n,i,k)$. It is convenient to clear denominators once and for all. Put

$$N(n,i,k) \;:=\; \sum_{j \geq \lceil ik/n \rceil} \binom{i}{j}\binom{n-i}{k-j} \;=\; \binom{n}{k} T(n,i,k) \;\in\; \mathbb{N}, \tag{6}$$

so that every inequality $T \geq p/q$ is the integer inequality $p\binom{n}{k} \leq q\, N(n,i,k)$. The sum in (6) is finite and runs over $\lceil ik/n \rceil \leq j \leq \min(i,k)$.

The two classical facts that make (1) a probability distribution with the mean (2) are the following; we record them because everything below is a statement about this normalisation.

Proposition 2.1. *For $1 \leq i, k \leq n$,*

$$\sum_{j=0}^{k} \frac{\binom{i}{j}\binom{n-i}{k-j}}{\binom{n}{k}} = 1 \qquad \textit{and} \qquad \sum_{j=0}^{k} j\,\frac{\binom{i}{j}\binom{n-i}{k-j}}{\binom{n}{k}} = \frac{ik}{n}.$$

Proof sketch. The first is Vandermonde's convolution $\sum_j \binom{i}{j}\binom{n-i}{k-j} = \binom{n}{k}$. For the second, the absorption identity $(d+1)\binom{i}{d+1} = i\binom{i-1}{d}$ turns the sum into $i\sum_d \binom{i-1}{d}\binom{n-i}{k-1-d} = i\binom{n-1}{k-1}$, again by Vandermonde, and $n\binom{n-1}{k-1} = k\binom{n}{k}$ finishes it. □

The variance in (2) is quoted from [1] and is used in this paper only to locate the regions in which the published bounds (4) and (5) apply; it is not part of the verified development described in Section 12.

3. The regime $\mathbb{E}H \le 1$

When $ik \le n$ the mean lies in $(0,1]$, so $\{H \ge \mathbb{E}H\} = \{H \ge 1\}$ and $T = 1 - \mathbb{P}(H=0)$. Everything then rests on an upper bound for the single cell $\mathbb{P}(H=0) = \binom{n-i}{k}/\binom{n}{k}$. For a binomial variable $\mathrm{Bin}(k,p)$ the corresponding bound is the elementary $(1-p)^k \le 1/(1+pk)$. The hypergeometric analogue is exact and has the same shape.

Lemma 3.1. *For all $k \ge 0$ and all $0 \le i \le n$,*

$$(n+ik)\binom{n-i}{k} \;\le\; n\binom{n}{k}, \qquad \text{that is} \qquad \mathbb{P}(H=0) \;\le\; \frac{n}{n+ik}.$$

Proof sketch. Induction on i, the case $i=0$ being an identity. Write $m = n-i$. In the step from i to $i+1$, if $k > m-1$ the left-hand side vanishes and there is nothing to prove; otherwise $\binom{m-1}{k}\,m = \binom{m}{k}\,(m-k)$, so it suffices that $(n+(i+1)k)(m-k) \le (n+ik)\,m$. Writing $m = k+r$ and $a = n+ik$ this reads $(a+k)r \le a(k+r)$, i.e. $kr \le ak$, i.e. $r \le n+ik$, and indeed $r = m-k \le n$. □

Theorem 3.2. *Let $1 \le i,k \le n$ with $ik \le n$ and $n \le 2ik$, i.e. $\frac{1}{2} \le \mathbb{E}H \le 1$. Then*

$$T(n,i,k) \;\ge\; \frac{ik}{n+ik} \;\ge\; \frac{1}{3}.$$

Proof sketch. Since $1 \le ik \le n$, every $j \ge 1$ satisfies $jn \ge n \ge ik$, so the tail (6) consists of all cells but $j=0$; hence $T = 1-\mathbb{P}(H=0)$, and Lemma 3.1 gives $T \ge ik/(n+ik)$. The hypothesis $n \le 2ik$ gives $n+ik \le 3ik$. □

Note that $\frac{1}{3} > \frac{1}{4}$: on this regime the quarter observation is true but not tight, which is the first piece of information about where its extremal cases can lie. Note also that only the first of the two hypotheses of Definition 1.1 was used.

4. The two symmetries of the parameter square

The transpose symmetry. Drawing k marbles from an urn with i black ones is the same experiment as drawing i marbles and counting how many lie in a fixed k-set: the joint law of the two sets is symmetric. We use the resulting invariance to turn each theorem about a small value of k into a theorem about the same small value of i.

Lemma 4.1. *For $1 \le i,k \le n$ we have $T(n,i,k) = T(n,k,i)$, and (n,i,k) is admissible if and only if (n,k,i) is.*

Proof sketch. For $j \le \min(i,k)$ the trinomial revision identity, applied twice, gives

$$\binom{n}{i}\binom{i}{j}\binom{n-i}{k-j} = \binom{n}{k}\binom{k}{j}\binom{n-k}{i-j},$$

and the guard $ik \le jn$ defining the range of summation in (6) is symmetric in i and k; summing gives $N(n,i,k)\binom{n}{i} = N(n,k,i)\binom{n}{k}$, which is the claim after dividing by $\binom{n}{i}\binom{n}{k}$. Both conditions of Definition 1.1 are visibly symmetric. (The terms of (6) with $j > \min(i,k)$ vanish, so the restriction of the range is harmless.) □

We remark, since it is a natural question, that [1] makes no use of this identity: it states nothing about $T(n,k,i)$, and the one place where it appeals to "the symmetry between the parameters i and k" is a single step in the proof of its Theorem 2.6, on the mean absolute deviation of H.

The complement symmetry. The square carries a second involution, and it is the one that explains why the observation has two hypotheses rather than one. Let B be the set of i black marbles and S the sample, so that $H = |S \cap B|$, and look at what is left over: the $n-k$ marbles outside S contain $|S^c \cap B^c| = H + n - i - k$ white ones. Replacing both S and B by their complements replaces H by $W = H + n - i - k$, which is hypergeometric with parameters $(n, n-i, n-k)$ and mean $(n-i)(n-k)/n$; and since $n-i-k$ is an integer, $H \ge ik/n$ holds if and only if $W \ge \mathbb{E}W$ does.

Lemma 4.2. *For $1 \le i, k \le n$,*

$$T(n,i,k) = T(n, n-i, n-k),$$

and admissibility is invariant: if (n,i,k) is admissible then so is $(n, n-i, n-k)$, and conversely, the map being an involution. The two conditions of Definition 1.1 are exchanged by it.

Proof sketch. In coefficients the argument above is the reindexing $j \mapsto j + (n-i-k)$ of (6), using $\binom{i}{j} = \binom{i}{i-j}$, $\binom{n-i}{k-j} = \binom{n-i}{n-i-k+j}$ and $\binom{n}{k} = \binom{n}{n-k}$. The guard $ik \le jn$ that cuts the range of (6) translates correctly because $(n-i)(n-k) = (n-i-k)n + ik$ and $\big(j + (n-i-k)\big)n = (n-i-k)n + jn$, so $(n-i)(n-k) \le \big(j + (n-i-k)\big)n$ is the same condition as $ik \le jn$. Over the natural numbers the shift exists only when $i + k \le n$; the identity is proved on that half, and the other half follows from it applied to the complementary pair, whose two parts do add up to at most n. For admissibility, the map carries $n \le 2ik$ to $n \le 2(n-i)(n-k)$ and back; an admissible triple has $i < n$ and $k < n$, since otherwise $(n-i)(n-k) = 0$, so $n-i$ and $n-k$ again lie in $\{1, \dots, n\}$. □

5. Every case with $\min(i,k) \le 3$

Above the line $\mathbb{E}H = 1$ the tail consists of few cells, and each of them can be written out. The pattern is the same at $k = 2$ and $k = 3$, with one more case each time.

Theorem 5.1. *For every n and every admissible triple $(n,i,2)$, $T(n,i,2) \ge \frac{1}{4}$.*

Proof sketch. If $2i \le n$ then $ik = 2i \le n$ and Theorem 3.2 gives $\frac{1}{3}$. Otherwise $\mathbb{E}H = 2i/n \in (1,2]$, so the tail is the single top cell,

$$T(n,i,2) = \mathbb{P}(H = 2) = \frac{\binom{i}{2}}{\binom{n}{2}} = \frac{i(i-1)}{n(n-1)},$$

and $2i \ge n+1$ gives $4i(i-1) = (2i)(2i-2) \ge (n+1)(n-1) = n^2 - 1 > n^2 - n = n(n-1)$. □

The final chain has relative slack $1/n$, and no more: at $i = \frac{n+1}{2}$ its first inequality is an equality, while $n^2 - 1$ exceeds $n(n-1)$ by $n-1$, which is $1/n$ of it. That is exactly the margin the extremal family of Section 6 occupies.

Theorem 5.2. *For every n and every admissible triple $(n,i,3)$, $T(n,i,3) \ge \frac{1}{4}$; and, by Lemma 4.1, the same holds for every admissible $(n,3,k)$.*

Proof sketch. Three cases. If $3i \le n$, Theorem 3.2 gives $\frac{1}{3}$. If $2n < 3i$ then $\mathbb{E}H \in (2,3]$ and $T = \binom{i}{3}/\binom{n}{3}$, so the claim is the polynomial inequality

$$n(n-1)(n-2) \ \le\ 4\,i(i-1)(i-2). \tag{7}$$

If $n < 3i \le 2n$ then $\mathbb{E}H \in (1,2]$, the tail has two cells, $T = \big(\binom{i}{2}(n-i) + \binom{i}{3}\big)/\binom{n}{3}$, and the claim is

$$n(n-1)(n-2) \ \le\ 12\,i(i-1)(n-i) + 4\,i(i-1)(i-2). \tag{8}$$

Both (7) and (8) are proved from the case condition and $i \le n$ alone — neither hypothesis of Definition 1.1 is needed. For (7), substituting $i = j+3$ and $n = i + r$ turns the case condition into $2r \le j+2$ and the claim into the termwise chain

$$8\,n(n-1)(n-2) \ \le\ (3j+8)\cdot 3(j+2)\cdot(3j+4) \ \le\ 32\,(j+3)(j+2)(j+1),$$

whose last step is $(j+2)(5j^2 + 20j) \ge 0$. For (8), substituting $i = b+2$ and $n = i + r$ turns the case conditions into $r \le 2b+3$ and $b+2 \le 2r$, after which the inequality is a positive combination of those two and of squares. □

Remark 5.3. The middle case (8) is the tight one. At $n = 3m$, $i = m+1$ the difference between its two sides is $m^3 + 39m^2 - 22m$, against $n(n-1)(n-2) \sim 27m^3$ on the left: the slack is positive but carries the leading coefficient 1 against 27. Accordingly

$$\frac{T(3m, m+1, 3)}{1/4} \ =\ 1 + \frac{m^3 + 39m^2 - 22m}{3m(3m-1)(3m-2)} \ \longrightarrow\ \frac{28}{27} = 1.037\ldots$$

as $m \to \infty$. So $k = 3$ already comes within four per cent of the constant, and $k = 2$, by Theorem 6.1, within $1/n$ of it.

Theorem 5.4. *Let (n, i, k) be admissible with $\min(i, k) \le 3$. Then $T(n, i, k) \ge \frac{1}{4}$.*

Proof sketch. If $\min(i, k) = 1$ then $ik \le n$ and Theorem 3.2 applies. The cases $k \in \{2, 3\}$ are Theorems 5.1 and 5.2, and the cases $i \in \{2, 3\}$ follow from them through Lemma 4.1. □

Corollary 5.5. *Let (n, i, k) be admissible with $\min(i, k) \le 3$ or $\min(n-i, n-k) \le 3$. Then $T(n, i, k) \ge \frac{1}{4}$.*

Proof. The first case is Theorem 5.4. In the second, $(n, n-i, n-k)$ is admissible by Lemma 4.2 and satisfies $\min(n-i, n-k) \le 3$, so Theorem 5.4 applies to it; and the two triples have the same tail, again by Lemma 4.2. □

Proposition 5.6.
(1) *$(9, 6, 4)$ is admissible, its complement is $(9, 3, 5)$ — a different point of the square, and not its transpose — and $T(9, 6, 4) = T(9, 3, 5) = \frac{25}{42}$.*
(2) *$(12, 6, 9)$ is admissible with $\min(i, k) = 6$ and $\min(n-i, n-k) = 3$, so Theorem 5.4 says nothing about it while Corollary 5.5 does.*

Proof. For (1), one evaluation of (6) and Lemma 4.2; the transpose of $(9, 6, 4)$ is $(9, 4, 6)$, not $(9, 3, 5)$. For (2), $\min(12-6, 12-9) = 3$ and Corollary 5.5. □

6. Sharpness: the extremal family

Theorem 6.1. *Let $n = 2t + 1$ with $t \ge 2$ be odd, and put $i = t + 1 = \frac{n+1}{2}$, $k = 2$. Then (n, i, k) is admissible and*

$$T\Big(n, \tfrac{n+1}{2}, 2\Big) = \frac{n+1}{4n}.$$

Consequently, for every rational $c > \frac{1}{4}$ there is an admissible triple with $T < c$; and for every admissible triple $(n, i, 2)$ one has $T(n, i, 2) > \frac{1}{4}$ strictly.

Proof sketch. Admissibility: $2ik = 2(n+1) \ge n$, and $2(n-i)(n-k) = (n-1)(n-2) \ge n$ exactly when $n^2 - 4n + 2 \ge 0$, which holds for $n \ge 4$. Since $\mathbb{E}H = \frac{n+1}{n} \in (1, 2]$, the computation in the proof of Theorem 5.1 gives

$$T = \frac{\binom{t+1}{2}}{\binom{2t+1}{2}} = \frac{(t+1)t}{(2t+1)(2t)} = \frac{t+1}{2(2t+1)} = \frac{n+1}{4n}.$$

As $\frac{n+1}{4n} = \frac{1}{4} + \frac{1}{4n}$ decreases to $\frac{1}{4}$, choosing t with $4(2t+1) > 1/(c - \frac{1}{4})$ gives the second claim. For strictness, the chain in the proof of Theorem 5.1 is strict: $4i(i-1) \ge n^2 - 1 > n(n-1)$ for $n \ge 2$. □

Corollary 6.2.

$$\inf\big\{\, T(n, i, k) \;:\; (n, i, k) \in \mathcal{A},\ \min(i, k) \le 3 \,\big\} \;=\; \frac{1}{4}.$$

Proof. The lower bound is Theorem 5.4; the family of Theorem 6.1 has $\min(i, k) = 2$ and values $\frac{1}{4} + \frac{1}{4n} \downarrow \frac{1}{4}$. □

So on the part of the region that we can settle, the constant $\frac{1}{4}$ of [1] is not merely correct but optimal, and the optimal shape is $k = 2$ with i just above $\frac{n}{2}$: the smallest sample size for which the mean can exceed one at all, taken at the point where it has just crossed an integer. This is the shape a proof of the general case must survive.

7. The exact minimum for $n \le 250$

Theorem 7.1. *For every admissible triple (n, i, k) with $n \le 250$,*

$$T(n, i, k) \;\ge\; \frac{125}{498} = \frac{249 + 1}{4 \cdot 249} = 0.251004\ldots,$$

and $(249, 125, 2)$ is admissible with $T(249, 125, 2) = \frac{125}{498}$. Hence $\frac{125}{498}$ is the exact minimum of T over the admissible triples with $n \le 250$.

Proof sketch. The attainment half is Theorem 6.1 at $t = 124$, so it involves no computation. The lower bound rests on exhaustive computation, and we state precisely what was run. For each n with $2 \le n \le 250$ and each pair (i,k) with $1 \le i,k \le n$ — 5 239 624 pairs in all — admissibility was tested in the integer form of Definition 1.1; for each of the 5 036 913 admissible triples the integer inequality

$$125 \cdot \binom{n}{k} \;\le\; 498 \cdot N(n,i,k)$$

was evaluated exactly, and it held in every case. This is the sharp threshold, not a rounded one: $\frac{125}{498}$ is the value the family of Theorem 6.1 takes at the largest odd n in the range. The evaluation is in exact integer arithmetic throughout; $N(n,i,k)$ is computed by walking the cells of (6) downwards from $j = \min(i,k)$, carrying $\binom{i}{j}$ and $\binom{n-i}{k-j}$ and updating each by one multiplication and one exact division, and $\binom{n}{k}$ by the ascending recursion. Both routines are proved equal to their definitions rather than assumed to compute them (see Section 12). The range was cut into the six blocks $[2,170]$, $[171,197]$, $[198,215]$, $[216,229]$, $[230,240]$, $[241,250]$, of roughly equal cost, and each block was evaluated separately. ☐

Theorem 7.2. *For every admissible triple (n,i,k) with $n \le 250$ one has $T(n,i,k) > \frac{1}{4}$. In particular the quarter observation of* [1] *holds, strictly, on that range.*

Proof. $\frac{1}{4} < \frac{125}{498}$, and Theorem 7.1 applies. ☐

We emphasise what Theorem 7.1 does and does not add to Theorem 7.2. It is not a numerical verification of an inequality: it determines the minimum of the tail on a stated range, with the minimiser, and that minimiser is the $k = 2$ family of Theorem 6.1. Together with Corollary 6.2 this is evidence of a definite kind for the shape of the general problem: everywhere we can compute or prove, the smallest values of T occur at $\min(i,k) = 2$.

8. Controls

An inequality on a region proved by cases invites two failure modes: a region so small that the statement is vacuous, and hypotheses that are not needed. We record the witnesses that exclude both. The triple in the first item is the minimiser over $n \le 120$, located by an exact-rational scan before any of the above was proved.

Proposition 8.1.
(1) $(119,60,2)$ *is admissible and* $T(119,60,2) = \frac{30}{119} = 0.252100\ldots$. *So the admissible region is not empty, the constant* $\frac{1}{4}$ *cannot be raised even to* 0.253*, and the bound* $\frac{1}{3}$ *of Theorem 3.2 does not extend to the whole admissible region.*
(2) $(5,1,1)$ *satisfies the second hypothesis of Definition 1.1 but not the first, and* $T(5,1,1) = \frac{1}{5} < \frac{1}{4}$.
(3) $(5,4,4)$ *satisfies the first hypothesis but not the second, and* $T(5,4,4) = \frac{1}{5} < \frac{1}{4}$.

In particular neither hypothesis of Definition 1.1 may be dropped.

Proof. Three evaluations of (6). For (1), $\binom{60}{2}/\binom{119}{2} = \frac{1770}{7021} = \frac{30}{119}$, which is $\frac{n+1}{4n}$ at $n = 119$, in agreement with Theorem 6.1. ☐

9. The band $1 < \mathbb{E}H \le 2$

On the band the position of the mean is forced. If $n < ik \le 2n$ then $1 < \mathbb{E}H \le 2$, so $\lceil ik/n \rceil = 2$ and the tail (6) misses exactly the two bottom cells: by Vandermonde's convolution,

$$N(n,i,k) + \left[\binom{n-i}{k} + i\binom{n-i}{k-1}\right] = \binom{n}{k}, \tag{9}$$

the bracket being $\binom{n}{k}\big(\mathbb{P}(H=0) + \mathbb{P}(H=1)\big)$, so that $T \ge \frac{1}{4}$ is the integer inequality

$$4\left[\binom{n-i}{k} + i\binom{n-i}{k-1}\right] \;\le\; 3\binom{n}{k}, \tag{10}$$

that is, $\mathbb{P}(H \le 1) \le \frac{3}{4}$. The bracket has a product closed form, and that is what makes the band tractable where the next band is not.

Lemma 9.1. *For $1 \le i \le n$ and $k \ge 1$,*

$$k\Big[\binom{n-i}{k} + i\binom{n-i}{k-1}\Big] = \binom{n-i}{k-1}\big(n+(i-1)(k-1)\big), \qquad k\binom{n}{k} = n\binom{n-1}{k-1},$$

so that

$$\mathbb{P}(H \le 1) = \frac{\big(n+(i-1)(k-1)\big)\binom{n-i}{k-1}}{n\binom{n-1}{k-1}}.$$

Proof sketch. The absorption identity $k\binom{w}{k} = \binom{w}{k-1}(w-k+1)$ at $w = n-i$ turns the bracket into $\binom{n-i}{k-1}\big((n-i-k+1)+ik\big)$, and $(n-i-k+1)+ik = n+(i-1)(k-1)$. The second identity is $k\binom{n}{k} = n\binom{n-1}{k-1}$. Both hold as stated over the natural numbers: where $n-i < k-1$ the two binomial coefficients on the left vanish together. ☐

Lemma 9.2. *For fixed $i \ge 1$ and $k \ge 1$ the two-cell mass $\mathbb{P}(H_{n,i,k} \le 1)$ is nondecreasing in n on $n \ge i$. Cleared of denominators, the one-step comparison is — when $n-i \ge k-1$, the remaining case being trivial because both sides then vanish — an identity with an explicit defect: writing $\kappa = k-1$, $e = i-1$, $s = ik$ and $u = n-i-\kappa \ge 0$,*

$$\big((u+1)+s\big)(u+\kappa+1)(e+u+1) = (u+1)(u+s)(e+u+\kappa+2) + i(i-1)k(k-1).$$

No relation between ik and n is assumed; only the threshold, not the two-cell mass, knows about the band.

Proof sketch. By Lemma 9.1 the comparison of $\mathbb{P}(H_{n,i,k} \le 1)$ with $\mathbb{P}(H_{n+1,i,k} \le 1)$ is, after clearing denominators, a comparison of two products of a binomial coefficient with a linear factor. The two binomial ratios that occur are both instances of $\binom{w}{\kappa}(w+1) = \binom{w+1}{\kappa}(w+1-\kappa)$, at $w = n-i$ and at $w = n-1$, and what is left after substituting them is the displayed identity, whose defect $i(i-1)k(k-1)$ is nonnegative. Iterating the step compares n with any larger n. ☐

Lemma 9.2 is what turns (10) into a statement at a single n. The band is $n < ik$, so its worst instance is $n = ik-1$, and there $n-i = ik-i-1$, $n-1 = ik-2$ and $n+(i-1)(k-1) = 2ik-i-k$. What has to be proved is the following inequality between two binomial coefficients; it is the whole content of Theorem 9.5.

Lemma 9.3. *For all integers $i \ge k \ge 2$,*

$$4\binom{ik-i-1}{k-1}(2ik-i-k) \;\le\; 3\,(ik-1)\binom{ik-2}{k-1}.$$

Proof sketch. Write $e = i-1$, $\kappa = k-1$, $b = ik-i-1$ and $a = ik-2 = b+e$, so that $1 \le \kappa \le e$ and, crucially,

$$b-(\kappa-1) = e\kappa \qquad \text{exactly.}$$

For $\kappa \le 3$ both binomial coefficients are products of at most three consecutive factors, so after multiplying through by $\kappa!$ the inequality is a polynomial inequality in the single variable e; the differences between the two sides of that cleared form are $2e$, e^3+9e^2+2e and $12e(e^3+6e^2+3e)$, at $\kappa = 1$, 2 and 3 respectively.

Let $\kappa \ge 4$, and put $P_j = \binom{b}{\kappa-j}$. Truncate Vandermonde's convolution at five terms,

$$\binom{a}{\kappa} = \binom{b+e}{\kappa} = \sum_{j\le\kappa}\binom{e}{j}P_j \;\ge\; \sum_{j\le 4}\binom{e}{j}P_j,$$

every omitted term being nonnegative. Because $b-\kappa+1 = e\kappa$ exactly, the four ratios between consecutive P_j are exact identities:

$$eP_1 = P_0, \qquad e(e\kappa+1)P_2 = (\kappa-1)P_0, \qquad e(e\kappa+1)(e\kappa+2)P_3 = (\kappa-1)(\kappa-2)P_0,$$
$$e(e\kappa+1)(e\kappa+2)(e\kappa+3)P_4 = (\kappa-1)(\kappa-2)(\kappa-3)P_0.$$

The first of them says that the terms $j = 0$ and $j = 1$ of the truncation are *equal*, and that is where the strength of this particular truncation comes from. Multiplying the truncated sum by $24e\Pi$, where $\Pi = (e\kappa+1)(e\kappa+2)(e\kappa+3)$, collapses it to $eP_0(48\Pi + R)$ with

$$\begin{aligned} R = {} & 12(e-1)(\kappa-1)(e\kappa+2)(e\kappa+3) + 4(e-1)(e-2)(\kappa-1)(\kappa-2)(e\kappa+3) \\ & + (e-1)(e-2)(e-3)(\kappa-1)(\kappa-2)(\kappa-3), \end{aligned}$$

so that after cancelling P_0 and e it suffices to prove

$$16\,\Pi\, e\kappa \;\le\; 16\,\Pi\,(e+\kappa) + (e\kappa + e + \kappa)\,R. \tag{11}$$

Substituting $e = f + 4$ and $\kappa = g + 4$ makes every subtraction in R honest, and the difference of the two sides of (11) is then a polynomial in (f, g) with 25 terms, all of whose coefficients are positive: the constant term is 208 800 and the smallest coefficient is 1. So (11) holds for all $e, \kappa \ge 4$ by expansion alone, with no case analysis and no search. □

Remark 9.4. Which truncation is used is not a matter of taste, and the margins say why. Asymptotically the two sides of (11) are $17(e\kappa)^4$ against $16(e\kappa)^4$: a margin of 6.25 per cent. The obvious alternative is to bound $\binom{a}{\kappa}/\binom{b}{\kappa}$ below by $(a/b)^\kappa$ — legitimate, since $(a-t)/(b-t) \ge a/b$ for $0 \le t < b \le a$ — and then to truncate the binomial expansion of $(1 + e/b)^\kappa$ at the same five terms. That route has asymptotic margin $2.7083\ldots$ against $2.6667\ldots$, that is 1.56 per cent, and the difference is decisive: over the range $2 \le k \le i$, $k \le 200$, $i \le 600$ exactly 181 pairs (i, k) fail the resulting polynomial inequality, all of them with $i, k \le 26$, and an independent enumeration over $2 \le k \le i \le 1200$, $k \le 400$ finds the same 181 pairs and no others. A proof along that route therefore has to close 181 exceptional cases by a finite table; the Vandermonde truncation has none to close, and Lemma 9.3 accordingly carries no finite case check. Before it was proved, (11) was tested exactly for all $1 \le e, \kappa \le 400$ and the inequality of Lemma 9.3 for all $2 \le i, k \le 400$, in both cases without a failure.

Theorem 9.5. *Let n, i, k be integers with $1 \le i, k \le n$ and $n < ik \le 2n$. Then*

$$T(n, i, k) \;\ge\; \frac{1}{4}.$$

Neither of the two hypotheses $ik/n \ge \frac{1}{2}$ and $(n-i)(n-k)/n \ge \frac{1}{2}$ of Definition 1.1 is used.

Proof sketch. Both hypotheses of the theorem are symmetric in i and k, so by Lemma 4.1 we may assume $k \le i$. If $k = 1$ then $ik = i \le n$, contradicting $n < ik$; so $k \ge 2$, and likewise $i \ge 2$. Since $1 < \mathbb{E}H \le 2$ the threshold is 2 and (9) holds, so the claim is (10), which by Lemma 9.1 reads

$$4\binom{n-i}{k-1}\big(n + (i-1)(k-1)\big) \;\le\; 3\,n\binom{n-1}{k-1}.$$

By Lemma 9.2 the left-hand side is worst, relative to the right, at the largest n the band allows, namely $n = ik-1$; and there the displayed inequality is exactly Lemma 9.3. Every step is an inequality between natural numbers, and no hypothesis beyond $i, k \le n$ and $n < ik \le 2n$ enters. □

Corollary 9.6. *Every admissible triple (n, i, k) with $ik \le 2n$, that is with $\mathbb{E}H \le 2$, satisfies $T(n, i, k) \ge \frac{1}{4}$.*

Proof. If $ik \le n$ this is Theorem 3.2, which gives the stronger $\frac{1}{3}$; otherwise $n < ik \le 2n$ and Theorem 9.5 applies. □

Remark 9.7. Theorem 9.5 is a statement about sampling without replacement, and its limit is not. Fix $k \ge 2$ and $p \in (\frac{1}{k}, \frac{2}{k})$, put $i = \lceil pn \rceil$ and let $n \to \infty$: the triples are eventually in the band, $\mathrm{Hyp}(n, i, k)$ converges to $\mathrm{Bin}(k, p)$ and the threshold $\lceil ik/n \rceil$ is eventually 2, so Theorem 9.5 yields $\mathbb{P}(\mathrm{Bin}(k, p) \ge kp) \ge \frac{1}{4}$ for those p. That is the part of [3, Theorem 1] in which the threshold equals 2, and [3] proves it in the stronger strict form and for every $p > \frac{1}{k}$. So the content of Theorem 9.5 that is not already in the literature is the statement at finite n; and, conversely, the proof above necessarily reproves that part of [3], which is why no argument for the band could be softer than one for the binomial theorem. This limiting statement is the one assertion in this section that is not part of the verified development described in Section 12; nothing else here depends on it.

Proposition 9.8.

(1) $(119, 60, 2)$ *lies in the band,* $119 < 120 \le 238$, *and* $T(119, 60, 2) = \frac{30}{119} = 0.2521\ldots$*; so the constant of Theorem 9.5 cannot be raised even to* 0.253.

(2) $(45, 12, 4)$ *lies in the band,* $45 < 48 \le 90$, *and has* $\min(i, k) = 4$ *and* $\min(n-i, n-k) = 33$*: neither Theorem 5.4 nor Corollary 5.5 says anything about it.*

(3) $(3, 2, 2)$ *lies in the band and is* not *admissible — its second hypothesis fails,* $2(n-i)(n-k) = 2 < 3 = n$ *— and* $T(3, 2, 2) = \frac{1}{3}$*. So dropping the hypotheses is not a vacuous strengthening. Of the triples in the band with* $n \le 79$*, exactly* 156 *are not admissible; all but this one have* $i = n$ *or* $k = n$.

(4) $(5, 4, 4)$ *has* $ik = 16 > 10 = 2n$ *and* $T(5, 4, 4) = \frac{1}{5} < \frac{1}{4}$*: the upper edge* $ik \le 2n$ *of the band cannot be removed.*

Proof. Items (1), (3) and (4) are evaluations of (6); item (2) is Theorem 9.5 together with two comparisons of integers. The count in (3) is an exhaustive enumeration of the triples in the band with $n \le 79$. □

10. The telescoping identity and the order statistic

The identities of this section are not needed for anything above. We record them because they say where the binding configurations of the quarter observation lie — the question Section 11 returns to — and because the last of them is the exact discrete analogue of the representation on which the binomial literature runs. For $m \ge 0$ put

$$N_m(n, i, k) := \sum_{j \ge m} \binom{i}{j}\binom{n-i}{k-j} = \binom{n}{k}\,\mathbb{P}(H \ge m), \tag{12}$$

so that $N(n, i, k)$ of (6) is N_m at the threshold $m = \lceil ik/n \rceil$. In (12) the threshold is held *fixed* while n, i and k vary; that is what makes the statements below clean, and Proposition 10.4(2) shows that T itself, whose threshold moves with the parameters, is not monotone in i.

Proposition 10.1. *Let* n, i, k, m *be integers with* $1 \le m \le k$*. Then, with the convention of* (1)*,*

$$N_m(n+1, i, k) = N_m(n, i, k) + N_m(n, i, k-1) \qquad (i \le n), \tag{13}$$

$$N_m(n+1, i+1, k) = N_m(n, i, k) + N_{m-1}(n, i, k-1) \qquad (i \le n), \tag{14}$$

$$N_m(n, i+1, k) - N_m(n, i, k) = \binom{i}{m-1}\binom{n-i-1}{k-m} \qquad (i+1 \le n). \tag{15}$$

Proof sketch. Identity (13) splits $\binom{n+1-i}{k-j} = \binom{n-i}{k-j} + \binom{n-i}{k-1-j}$ in each term, and (14) splits $\binom{i+1}{j} = \binom{i}{j} + \binom{i}{j-1}$ and reindexes the second sum. Subtracting the first from the second, the terms with sample size k cancel and the two remaining sums differ by exactly the bottom cell $j = m-1$ of the lower one, which is $\binom{i}{m-1}\binom{n-i}{k-m}$; rewriting for the common urn size gives (15). □

Corollary 10.2. *At a fixed threshold* $m \ge 1$, $N_m(n, i, k)$ *is nondecreasing in* i*. Hence inside a block of values of* i *on which* $\lceil ik/n \rceil$ *is constant,* $T(n, i, k)$ *is smallest at the smallest* i *of the block.*

Proof. The increment in (15) is a nonnegative integer, and $T = N_m/\binom{n}{k}$ with $\binom{n}{k}$ independent of i. □

This is what turns the observation that the binding configurations have the mean just above an integer — visible at $k = 2$ and $k = 3$ in Section 5, and at the top of the band in Section 9 — from a pattern into a statement.

Proposition 10.3. *Let* $1 \le i \le n$ *and suppose the threshold* $m = \lceil ik/n \rceil$ *satisfies* $1 \le m \le k$*. Then*

$$\binom{n}{k} T(n, i, k) \;=\; N_m(n, i, k) \;=\; \sum_{a=0}^{i-1} \binom{a}{m-1}\binom{n-1-a}{k-m}.$$

Proof sketch. Telescope (15) down to $i = 0$, where N_m vanishes because $m \ge 1$. □

Since $\binom{a}{m-1}\binom{n-1-a}{k-m}$ is the number of k-subsets of $\{1,\dots,n\}$ whose m-th smallest element is $a+1$, Proposition 10.3 reads

$$T(n,i,k) = \mathbb{P}\big(S_{(m)} \le i\big),$$

where $S_{(m)}$ is the m-th smallest element of a uniform random k-subset of $\{1,\dots,n\}$: the exact discrete analogue of $\mathbb{P}(\mathrm{Bin}(k,p) \ge m) = I_p(m, k+1-m)$, the incomplete-beta representation on which the binomial results quoted in Section 1 are proved. It is the counting identity that is verified; the reading in terms of $S_{(m)}$ is its interpretation.

Proposition 10.4. (1) *The increment in* (15) *is a genuine cell: at* $n = 12$, $k = 4$, $m = 2$ *and* $i = 5$ *it equals* $\binom{5}{1}\binom{6}{2} = 75$, *and indeed* $N_2(12,5,4) = 285 < 360 = N_2(12,6,4)$.

(2) *The monotonicity of Corollary 10.2 holds at a fixed threshold only:* $T(9,5,4) = \frac{5}{14} < \frac{9}{14} = T(9,4,4)$ *although* $4 < 5$, *because the threshold* $\lceil ik/n\rceil$ *jumps from* 2 *to* 3 *between the two.*

(3) *At* $(45,12,4)$ *the threshold is* 2 *and* $N(45,12,4) = \sum_{a<12}\binom{a}{1}\binom{44-a}{2}$.

11. Where the difficulty is

What is left of the quarter observation is the region

$$\mathbb{E}H > 2, \qquad \min(i,k) \ge 4, \qquad \min(n-i, n-k) \ge 4,$$

outside the range of Theorem 7.1. We collect what can be said about it, and first about why the band of Theorem 9.5 could not be reached by importing a known bound.

The published routes stop short, and stop short on the band. Of the four bounds recalled in Section 1, none reaches $\frac{1}{4}$ anywhere on the band $1 < \mathbb{E}H \le 2$; for the third of them that is arithmetic rather than a search.

- [1, Theorem 1.1] gives $T \ge k/n$ on $n \ge 8k$, hence at most $\frac{1}{8}$ there.
- [1, Theorem 1.2] has, by the paper's own remark, a value never exceeding $\frac{e^{-1/12}}{4\sqrt{2}} = 0.16264\ldots$.
- The Berry–Esseen bound (4) yields $\frac{1}{4}$ only once $\operatorname{Var} H \ge 4.98$. But by (2),

$$\operatorname{Var} H = k \cdot \frac{i}{n} \cdot \frac{n-i}{n} \cdot \frac{n-k}{n-1} \;\le\; \frac{ik}{n} = \mathbb{E}H \qquad (n \ge 2),$$

  since $\frac{n-i}{n} \le 1$ and $\frac{n-k}{n-1} \le 1$. On the band this gives $\operatorname{Var} H \le 2$, so (4) is silent at *every* triple Theorem 9.5 covers; over the triples of the band with $n \le 199$ the largest value of $\operatorname{Var} H$ is $1.6244\ldots$. More crudely, $\operatorname{Var} H \le \frac{k}{4}$, so (4) is silent for every $k \le 19$, and in particular at the minimising shape $k = 2$, where $\operatorname{Var} H < \frac{1}{2}$.
- The transfer (5) of the binomial bound is the closest analogue, and it is short by exactly the wrong amount. Greenberg–Mohri [3] give $\mathbb{P}(X \ge \mathbb{E}X) \ge \frac{1}{4}$, so (5) yields $T \ge \frac{1}{4} - \frac{k-1}{n-1}$, never $\frac{1}{4}$. At the extremal family of Theorem 6.1, where $k = 2$, the true excess of T over $\frac{1}{4}$ is $\frac{1}{4n}$, while the total-variation loss is $\frac{k-1}{n-1} = \frac{1}{n-1}$: the loss is about four times the whole margin. And in fact (5) does not even apply there, since its hypothesis $\operatorname{Var} X = k\frac{i}{n}\frac{n-i}{n} \ge 1$ fails at every member of the family, where $\operatorname{Var} X = \frac{n^2-1}{2n^2} < \frac{1}{2}$.

The binomial analogue is tight at the same shape. The reason the total-variation transfer cannot be repaired by a better constant on the binomial side is that there is no better constant. For $p > \frac{1}{2}$ one has $\mathbb{P}(\mathrm{Bin}(2,p) \ge 2p) = \mathbb{P}(\mathrm{Bin}(2,p) = 2) = p^2$, which decreases to $\frac{1}{4}$ as $p \downarrow \frac{1}{2}$; and this is what [3] itself records of its constant, that it "is never reached but is approached asymptotically when $m = 2$ as $p \to \frac{1}{2}$ from the right" (its m is our k). So the Greenberg–Mohri constant is approached at $k = 2$ with $p \downarrow \frac{1}{2}$, which is precisely the shape — $k = 2$, $i/n \downarrow \frac{1}{2}$ — at which our Theorem 6.1 shows the hypergeometric tail approaches $\frac{1}{4}$. The two extremal families sit on top of each other, and the hypergeometric one is *above* the binomial one by an amount, $\frac{1}{4n}$, far smaller than any available total-variation estimate between the two laws. A proof of the general case must therefore either be intrinsic to the hypergeometric law — as our Lemma 3.1 and the cell inequalities (7), (8) are — or must compare the two laws with an error term of size $o(1/n)$ at $k = 2$.

The proof of Theorem 9.5 is of the first kind: Lemmas 9.1, 9.2 and 9.3 are statements about binomial coefficients and never mention a binomial law. Nor, so far as we could determine, has the constant been carried across in the meantime: we have found no work that bounds a hypergeometric tail $T(n,i,k)$ below by an absolute constant. [1] itself cites [3], and imports it only through (5), which reaches at best $\frac{1}{4}-\frac{k-1}{n-1}$; and [4], which gives an elementary proof of a statement of the same kind, stays with the binomial law — its e-print mentions neither the hypergeometric law nor sampling without replacement. This is a search that did not find, and not an argument that nothing exists: the two citation indexes we consulted on 3 September 2026 do not even agree on the size of the set of works citing [3], one listing ten of them and the other seventy-five.

What is left is a different problem. On the band the constant $\frac{1}{4}$ is exactly right: the family of Theorem 6.1 lies in the band and drives T down to $\frac{1}{4}$, so Theorem 9.5 is tight and its proof had to be. On the region that remains this appears not to be so. The following is a numerical observation, made outside the formal development and stated as such. Over the triples with $4 \le k \le i \le n-k$, $\mathbb{E}H > 2$ and $251 \le n \le 900$, the smallest value of T is $T(899, 450, 4) = \frac{168225}{537602} = 0.312917\ldots$; and along the boundary families $ik = n(m-1)+1$ with $i \to \infty$, T tends to $\mathbb{P}\big(\mathrm{Bin}(k, \frac{m-1}{k}) \ge m\big)$, a quantity which, tabulated exactly over $2 \le m \le k \le 60$, increases in k at each fixed m but is *not* monotone in m — at $k = 5$ it is larger at $m = 4$ than at $m = 5$ — and whose smallest value over $m \ge 3$ and $k \ge 4$ is $\mathbb{P}\big(\mathrm{Bin}(4, \frac{1}{2}) \ge 3\big) = \frac{5}{16} = 0.3125$. If $\frac{5}{16}$ is indeed the infimum there, the region that is left carries a margin of $\frac{1}{16}$, where the band carries only the $O(1/n)$ margin of Theorem 6.1 — so what remains is not a question about the constant $\frac{1}{4}$ at all, and a proof of it may be allowed to be lossy in a way that no proof of the band could be. We claim none of this; we record it because it changes what the remaining work is.

What such a proof cannot do is repeat Section 9 verbatim. That argument turns on the product closed form of Lemma 9.1 for the two-cell mass $\mathbb{P}(H \le 1)$, and $\mathbb{P}(H \le m-1)$ has no such form for $m \ge 3$. The reduction to a single worst n does survive in general — it is the block-minimum statement of Corollary 10.2 together with the monotonicity in n — but the inequality it reduces to is no longer a comparison of two binomial coefficients.

Priced and not done. Three extensions were costed and deliberately not carried out; we record the prices so that they need not be re-measured. These are engineering measurements about the computation of Theorem 7.1 and lie outside the formal development.

- *A wider range.* The cost of the exhaustive evaluation grows like $N^{4.3}$ in the upper limit N (fitted from three measured points). The range $n \le 250$ is six block evaluations and about 18 minutes of processor time at a peak of 0.46 GB; the range $n \le 400$ would be about 2.1 processor-hours. Theorem 9.5 has made this markedly less interesting: every triple of the window with $ik \le 2n$ is now covered, at every n, by an argument that needs no computation at all.
- *Two halvings.* Restricting the search to $k \le i$ is legitimate by Lemma 4.1, and restricting it further to $i + k \le n$ is legitimate by Lemma 4.2; each halves the work, and with both, $n \le 350$ would cost what $n \le 250$ costs now. Neither restriction has been applied to the evaluation that Theorem 7.1 rests on, which is the one reported in Section 7.
- *The other numerical remark.* [1] also reports that the constant $\frac{e^{-1/12}}{4\sqrt{2}}$ of its Theorem 1.2 may be replaceable by $\frac{1}{2\sqrt{2}} = 0.35355\ldots$. That statement involves $\sqrt{\operatorname{Var} H}$ and so is a real-analytic rather than a rational inequality; it is not treated here.

12. Verification

Every statement above that is labelled as a proposition, lemma, theorem or corollary — Propositions 2.1, 8.1, 5.6, 9.8, 10.1, 10.3 and 10.4, Lemmas 3.1, 4.1, 4.2, 9.1, 9.2 and 9.3, Theorems 3.2, 5.1, 5.2, 5.4, 6.1, 7.1, 7.2 and 9.5, and Corollaries 6.2, 5.5, 10.2 and 9.6, of which Theorems 1.2, 1.3, 1.4 and 1.5 of the introduction are restatements — is machine-checked, in the form stated, in Lean 4 [8] over the Mathlib library [9]; the development is eleven modules, 2272 lines and 124 declarations, and contains no `sorry`. The one assertion of this paper that is outside it is the limiting statement of Remark 9.7, which is flagged there and used nowhere.

The development is arranged so that the computational part is not trusted: the exhaustive evaluation of Theorem 7.1 runs on a self-contained arithmetic layer (a linear recursion for $\binom{n}{k}$, a descending two-binomial recurrence for $N(n,i,k)$, and an integer comparison), and each of those routines is *proved* equal to its definition in terms of $\binom{\cdot}{\cdot}$, so that a positive answer from the evaluator is converted into a statement about (6) by a proof rather than by assumption. An axiom print of each headline declaration returns `propext`, `Classical.choice` and `Quot.sound` and nothing else for the whole general part — everything with $\min(i,k) \le 3$, both symmetries and the doubled strip, the extremal family and the sharpness statements, the whole of Section 9 including the polynomial inequality (11) at its heart, the identities of Section 10, every control, and the attainment half of Theorem 7.1 — so none of that part depends on any compiled evaluation; several of those declarations use strictly fewer axioms than three, and Lemma 9.1 and the polynomial step (11) need only `propext`. The lower-bound half of Theorem 7.1, and hence Theorem 7.2, additionally depends on the six compiled-evaluation axioms of the six range blocks, one per block; that is the precise and only sense in which those two statements rest on computation.

Independently of the kernel, the constants printed in [1] that this paper quotes ($\frac{e^{-1/12}}{4\sqrt{2}} \approx 0.162$, the threshold 4.98, and $\frac{1}{2\sqrt{2}}$) were recomputed from the definitions of [1] before anything above was claimed, and agree; and the exhaustive evaluation was reproduced outside Lean in exact rational arithmetic — on the smaller range $n \le 120$ twice, once with a Pascal table and once with the algorithm described in Section 7, both returning the minimum $\frac{30}{119}$ of Proposition 8.1, and then on the full range $n \le 250$ with a Pascal table, returning the counts 5 239 624 and 5 036 913 quoted in Section 7, no violation, and the minimum $\frac{125}{498}$ at $(249, 125, 2)$ and at its transpose $(249, 2, 125)$. Each statement proved after the first version was likewise checked by exact integer or rational computation before it was proved, on ranges we state: the inequality of Theorem 9.5 for every triple of the band with $n \le 199$; the inequality of Lemma 9.3 for all $2 \le i, k \le 400$ and the polynomial inequality (11) for all $1 \le e, \kappa \le 400$; the complement symmetry of Lemma 4.2 for all $n \le 40$, and independently for all $n \le 44$ and $1 \le i, k \le n-1$; the three identities of Proposition 10.1 for all $n \le 40$; the closed form of Lemma 9.1 for all $n \le 59$; and the monotonicity of Lemma 9.2 for $i, k \le 25$ and $n \le 89$, its defect identity for $u \le 29$ and $e, \kappa \le 19$. None of these checks failed, and none of them is part of any proof — every statement of Sections 4, 9 and 10 is proved for all parameters — but they are what the proofs were written against. The source of the development is currently held in a private repository: repository reference to be added at submission.

On the reference list. Entries [1], [2], [4], [3] and [6] were read here from their arXiv LaTeX e-prints, and every quotation above is from those files. The statement numbers of [1] are those of the v2 e-print as it compiles — its theorem environments are numbered by section, and Theorems 1.1, 1.2, 2.6 and 3.2 are the ones referred to above; the published version may number them differently. Likewise the Greenberg–Mohri statement quoted above is Theorem 1 of the v3 e-print of [3], and the sentence about $m = 2$, $p \to \frac{1}{2}$ is the remark that follows it there. Formula (17) of [6], the pointer [1] gives for (4), is in the v1 e-print of [6] the two-sided estimate $\frac{1}{2\sqrt{1+12\sigma^2}} \le \|F - G\|_\infty < \frac{0.5583}{\sigma}$ for the distribution function F of a Bernoulli convolution and that of the normal law G with the same mean and variance; [6] applies it to hypergeometric laws through the same representation of [7], which is how (4) arises. The bibliographic data of every entry above were confirmed against Crossref, and the series volume of [8] against dblp. Entries [5] and [7] were not read here, so the pointer to Theorem 2 of [5] is the one [1] gives, repeated as such. The single entry added in this version, [4], is cited only for the statement in Section 11 that the elementary re-proofs of the binomial bound stay with the binomial law; its journal data are those on which its arXiv record and Crossref agree.

References

[1] J. Ai and C. Pelekis, *On lower bounds for hypergeometric tails*, Statistics & Probability Letters **236** (2026), Article 110798, doi:10.1016/j.spl.2026.110798; preprint arXiv:2601.09485v2 [math.PR], 22 April 2026 (v1, 14 January 2026). Read here in the v2 e-print, which is what the statement numbers above refer to.

[2] J. Ai, O. Kuželka and C. Pelekis, *An entropic analogue of the MMS conjecture*, arXiv:2606.30486v1 [math.CO], 29 June 2026.

[3] S. Greenberg and M. Mohri, *Tight lower bound on the probability of a binomial exceeding its expectation*, Statistics & Probability Letters **86** (2014) 91–98, doi:10.1016/j.spl.2013.12.009; preprint arXiv:1306.1433v3 [cs.LG], 11 November 2013 (v1, 6 June 2013). Read here in the v3 e-print.

[4] B. Doerr, *An elementary analysis of the probability that a binomial random variable exceeds its expectation*, Statistics & Probability Letters **139** (2018) 67–74, doi:10.1016/j.spl.2018.03.016; preprint arXiv:1712.00519v4 [math.PR], 4 January 2018 (v1, 1 December 2017). Read here in the v4 e-print.

[5] W. Ehm, *Binomial approximation to the Poisson binomial distribution*, Statistics & Probability Letters **11** (1991), no. 1, 7–16, doi:10.1016/0167-7152(91)90170-V. Cited for attribution only, through the form in which [1] quotes it; the article itself was not read here.

[6] L. Mattner and J. Schulz, *On normal approximations to symmetric hypergeometric laws*, Transactions of the American Mathematical Society **370** (2018), no. 1, 727–748, doi:10.1090/tran/6986; preprint arXiv:1404.7657v1 [math.PR], 30 April 2014. Read here in the v1 e-print, for Formula (17), the pointer [1] gives for (4).

[7] V. A. Vatutin and V. G. Mikhailov, *Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles*, Theory of Probability and its Applications **27** (1983), no. 4, 734–743, doi:10.1137/1127084; Russian original in Teor. Veroyatnost. i Primenen. **27** (1982) 684–692. Cited for attribution only, for the representation of a hypergeometric variable as a sum of independent Bernoulli variables, in the form in which [1] and [6] quote it; the article itself was not read here.

[8] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science **12699**, Springer, 2021, 625–635, doi:10.1007/978-3-030-79876-5_37.

[9] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, 367–381, doi:10.1145/3372885.3373824.