Bernoulli **23**(3), 2017, 1911–1950
DOI: 10.3150/15-BEJ800

Exponential bounds for the hypergeometric distribution

EVAN GREENE[1] and JON A. WELLNER[2]

[1] *Department of Statistics, University of Washington, Seattle, WA 98195-4322, USA.*
E-mail: egreene@uw.edu
[2] *Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322, USA.*
E-mail: jaw@stat.washington.edu; *url:* http://www.stat.washington.edu/jaw/

We establish exponential bounds for the hypergeometric distribution which include a finite sampling correction factor, but are otherwise analogous to bounds for the binomial distribution due to León and Perron (*Statist. Probab. Lett.* **62** (2003) 345–354) and Talagrand (*Ann. Probab.* **22** (1994) 28–76). We also extend a convex ordering of Kemperman's (*Nederl. Akad. Wetensch. Proc. Ser. A* *76* = *Indag. Math.* **35** (1973) 149–164) for sampling without replacement from populations of real numbers between zero and one: a population of all zeros or ones (and hence yielding a hypergeometric distribution in the upper bound) gives the extreme case.

Keywords: binomial distribution; convex ordering; exponential bound; finite sampling correction factor; hypergeometric distribution; sampling without replacement

1. Introduction and overview

In this paper, we derive several exponential bounds for the tail of the hypergeometric distribution. This distribution emerges as an extreme case in the setting of sampling without replacement from a finite population. We begin with a description of this setting. Consider a population C containing N elements, $C := \{c_1, \dots, c_N\}$, with $c_i \in \mathbb{R}$. Let $N = |C|$ denote the cardinality of this set, a the value of the minimum element, b the value of the maximum element, and $\mu := (N^{-1})(\sum_{i=1}^N c_i)$, the population mean. Let $1 \le i \le n \le N$, and X_i denote the ith draw without replacement from this population. Finally, let $S_n := \sum_{i=1}^n X_i$ denote the sum of this sampling procedure, and let $\bar{X}_n := S_n/n$ denote the sample mean.

R.J. Serfling obtained the following bound.

Finite Sampling Bound 1 (Serfling [20]). *For* $1 \le n \le N$, S_n *the sum in sampling without replacement, and* $\lambda > 0$:

$$P\bigl(\sqrt{n}(\bar{X}_n - \mu) \ge \lambda\bigr) \le \exp\biggl(-\frac{2\lambda^2}{(1 - f_n^*)(b-a)^2}\biggr), \tag{1.1}$$

where $f_n^* := (n-1)/N$.

This result applies to sampling without replacement from any finite bounded population. Let $D, N \in \mathbb{N}$ such that $D < N$. Then as a special case we may apply the bound to a population of N elements containing D 1's and $N - D$ 0's. Note that in this specific case $S_n =: S_{n,D,N} \sim$ Hypergeometric(n, D, N).

For the hypergeometric distribution, the following facts are well known:

$$\begin{aligned} P(S_n = k) &= \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}, \qquad \max\{0, n-(N-D)\} \le k \le \min\{D, n\}, \\ E(S_n) &= n\left(\frac{D}{N}\right), \\ \mathrm{Var}(S_n) &= n\left(\frac{D}{N}\right)\left(1-\frac{D}{N}\right)\left(1-\frac{n-1}{N-1}\right) =: n\mu_{D,N}(1-\mu_{D,N})(1-f_n) \end{aligned} \tag{1.2}$$

with the final line defining $\mu_{D,N} := D/N$ and $f_n := (n-1)/(N-1)$. Applying Serfling's result to the case of the hypergeometric distribution immediately gives

$$P\bigl(\sqrt{n}(\bar{X}_n - \mu_{D,N}) \ge \lambda\bigr) \le \exp\left(-\frac{2\lambda^2}{(1-f_n^*)}\right) \tag{1.3}$$

since $(b-a)^2 = (1-0)^2 = 1$. Comparison of the factor $1 - f_n^*$ in Serfling's bound to the factor $1 - f_n$ in (1.2) suggests the following question: can Serfling's bound be improved to

$$P\bigl(\sqrt{n}(\bar{X}_n - \mu) \ge \lambda\bigr) \le \exp\left(-\frac{2\lambda^2}{(1-f_n)(b-a)^2}\right) \tag{1.4}$$

in general, or at least in the special case of the hypergeometric distribution?

To date, the improvement conjectured in (1.4) has not been obtained. For the special case of the hypergeometric, Hush and Scovel derived the following bound by extending an argument given by Vapnik. See [13] and [24].

Hypergeometric Bound 1 (Hush and Scovel [13]). *Suppose* $S_n \sim$ Hypergeometric(n, D, N). *Then for all* $\lambda > 0$ *we have*

$$P\bigl(\sqrt{n}(\bar{X}_n - \mu_{D,N}) \ge \lambda\bigr) \le \exp\bigl(-2\alpha_{n,D,N}\bigl(n\lambda^2 - 1\bigr)\bigr), \tag{1.5}$$

where

$$\alpha_{n,D,N} := \left(\frac{1}{n+1} + \frac{1}{N-n+1}\right) \vee \left(\frac{1}{D+1} + \frac{1}{N-D+1}\right).$$

More recently, Bardenet and Maillard have improved a deficiency in Serfling's inequality that occurs when more than half the population is sampled without replacement by using a reverse-martingale argument. The statement here is a specialization of their Theorem 2.4 to the hypergeometric case. See [1] for additional discussion.

Hypergeometric Bound 2 (Bardenet and Maillard [1]). *Suppose* $S_n \sim \text{Hypergeometric}(n, D, N)$*. Then for all* $\lambda > 0$ *and* $n < N$ *we have*

$$P\big(\sqrt{n}(\bar{X}_n - \mu_{D,N}) \geq \lambda\big) \leq \exp\left(-\frac{2\lambda^2}{(1-n/N)(1+1/n)}\right).$$

We will justify the special consideration given to the hypergeometric distribution relative to the goal of obtaining (1.4) by adapting a result of Kemperman [15] to derive a convex order between samples without replacement from populations consisting of elements in [0, 1] and the hypergeometric distribution. We will then demonstrate how one may use this convex order to obtain exponential bounds for the more general problem of sampling without replacement from a bounded, finite population. In doing so, we return to the setting of Serfling. In this setting, we will consider the variance of the population as well. Anticipating this, we conclude the introduction with a specialization of a bound of Bardenet and Maillard which incorporates information about the population variance into the bound.

Finite Sampling Bound 2 (Bardenet and Maillard [1]). *For* $1 \leq n < N$, S_n *the sum in sampling without replacement from a population* $\mathbf{c} := \{c_1, \ldots, c_N\}$, $\delta \in [0, 1]$, *and* $\lambda > 0$, *we have*

$$P\big(\sqrt{n}(\bar{X}_n - \mu) \geq \lambda\big) \leq \exp\left(-\frac{\lambda^2}{2(\gamma^2 + (2/3)(b-a)(\lambda/\sqrt{n}))}\right) + \delta, \tag{1.6}$$

where

$$a := \min_{1 \leq i \leq N} c_i, \qquad b := \max_{1 \leq i \leq N} c_i, \qquad f_n^* := (n-1)/N,$$

$$\mu := (1/N)\sum_{i=1}^{N} c_i, \qquad \sigma^2 := (1/N)\sum_{i=1}^{N}(c_i - \mu)^2,$$

$$\gamma^2 := \big(1 - f_n^*\big)\sigma^2 + f_n^* c_{n-1}(\delta) \quad \textit{and} \quad c_n(\delta) := \sigma(b-a)\sqrt{\frac{2\log(1/\delta)}{n}}.$$

2. Exponential bounds

Binomial distributions arise when sampling with replacement from a population consisting only of 0's and 1's. As we saw in the Introduction, hypergeometric distributions arise when sampling without replacement from such populations. Intuitively, sampling without replacement is more informative than sampling with replacement: when items are not replaced, eventually, when $n = N$, the entire population is sampled. This being the case, it is natural to guess that upper bounds which apply to binomial tail probabilities will also apply to the hypergeometric tail probabilities.

Hoeffding [11] proved that this guess is true for exponential bounds derived via the Cramér–Chernoff method. This is because a convex order exists between samples with and without replacement (Hoeffding proves this order in his Theorem 4). Convex orders between a variety of sampling plans were subsequently explored by Kemperman [15] and Karlin [14].

Note that by Ehm (Theorem 2 [4]; see also Holmes, Theorem 3.2 [12]), the total variation distance between the hypergeometric distribution $P^{\text{hyper}}_{n,D,N}$ and the binomial distribution $P^{\text{bin}}_{n,D/N}$ satisfies

$$d_{\text{TV}}\big(P^{\text{hyper}}_{n,D,N}, P^{\text{bin}}_{n,D/N}\big) \le \frac{n-1}{N-1},$$

so we expect that the binomial bounds will be essentially optimal when $(n-1)/(N-1) \to 0$.

Here we are interested in sampling scenarios in which $(n-1)/(N-1) \nrightarrow 0$. Given the similarity in scenarios that produce binomial and hypergeometric probabilities, one might expect that the binomial exponential bounds provide a clue to the form that hypergeometric exponential bounds will take: hypergeometric bounds look like binomial bounds, with a finite sampling correction factor included. Indeed, this is the case when we compare the bound of Serfling (1.1) to Hoeffding's uniform bound (Theorem 2.1, [11]), since the only difference between the two is the quantity $1 - f_n^*$. We therefore state several exponential bounds which apply to the binomial distribution.

Binomial Bound 1 (León and Perron [16]). *Let* $n \ge 1$, $p \in (0,1)$, $\lambda < \sqrt{n}/2$, *and* $X_1, \dots, X_n$ *be independent* Bernoulli(p) *random variables. Then*

$$P\big(\sqrt{n}(\bar{X}_n - p) \ge \lambda\big) \le \sqrt{\frac{1}{2\pi\lambda^2}}\left(\frac{1}{2}\right)\sqrt{\frac{\sqrt{n}+2\lambda}{\sqrt{n}-2\lambda}}\, e^{-(2\lambda^2)}. \tag{2.1}$$

The second bound was established by Talagrand [22], pages 48–50. The statement here is taken from van der Vaart and Wellner [23], pages 460–462:

Binomial Bound 2 (Talagrand [22]). *Fix* p_0 *and consider* p *such that* $0 < p_0 \le p \le 1 - p_0 < 1$. *Suppose for* $n \in \mathbb{N}$ *that* $X_1, \dots, X_n$ *are i.i.d.* Bernoulli(p) *random variables. Then there exist constants* K_1 *and* K_2 *depending only on* p_0, *such that*:

(i) *For all* $\lambda > 0$, $$P\big(\sqrt{n}(\bar{X}_n - p) = \lambda\big) \le \frac{K_1}{\sqrt{n}} \exp\left(-\left[2\lambda^2 + \frac{\lambda^4}{4n}\right]\right).$$

(ii) *For all* $0 < t < \lambda$,
$$P\big(\sqrt{n}(\bar{X}_n - p) \ge t\big) \le \frac{K_2}{\lambda} \exp\left(-\left[2\lambda^2 + \frac{\lambda^4}{4n}\right]\right)\exp\big(5\lambda[\lambda - t]\big). \tag{2.2}$$

(iii) *For all* $\lambda > 0$, $$P\big(\sqrt{n}(\bar{X}_n - p) \ge \lambda\big) \le \frac{K_2}{\lambda} \exp\left(-\left[2\lambda^2 + \frac{\lambda^4}{4n}\right]\right).$$

Another well-known exponential bound which applies to sums of independent random variables (and consequently the binomial distribution) was discovered by Bennett [2]. Bennett's bound incorporates information about the population variance, and so obtains notable improvements when the population variance is small. This statement of Bennett's inequality specialized to the binomial setting is adapted from Shorack and Wellner [21].

Binomial Bound 3 (Bennett [2]). *Let* $X_1, \dots, X_n$ *i.i.d.* Bernoulli(μ), *with* $0 < \mu \le 1/2$. *Then for all* $\lambda \ge 0$

$$P\big(\sqrt{n}(\bar{X}_n - \mu) \ge \lambda\big) \le \exp\left(-\frac{\lambda^2}{2\mu(1-\mu)}\psi\left(\frac{\lambda}{\sqrt{n}\mu(1-\mu)}\right)\right), \tag{2.3}$$

where $\psi(\lambda) := (2/\lambda^2)h(1+\lambda)$ *where* $h(\lambda) := \lambda(\log\lambda - 1) + 1$.

Inspecting the form of (2.1), (2.2), and (2.3), we notice that when these bounds are compared to the hypergeometric tail bound (1.3) obtained from Serfling's bound they do not take advantage of the finite sampling setting. Our hope then is we can derive probability bounds which look like the preceding binomial expressions, but improved by a finite-sampling correction factor. Such improved bounds exist and are the main results of this paper. Their statements follow.

Theorem 1. *Suppose* $S_n \sim$ Hypergeometric(n, D, N). *Define* $\mu := D/N$, *and suppose* $N > 4$ *and* $2 \le n < D \le N/2$. *Then for all* $0 < \lambda < \sqrt{n}/2$ *we have*

$$\begin{aligned} P\big(\sqrt{n}(\bar{X}_n - \mu) \ge \lambda\big) \le{}& \sqrt{\frac{1}{2\pi\lambda^2}}\left(\frac{1}{2}\right)\sqrt{\left(\frac{N-n}{N}\right)\left(\frac{\sqrt{n}+2\lambda}{\sqrt{n}-2\lambda}\right)\left(\frac{N-n+2\sqrt{n}\lambda}{N-n-2\sqrt{n}\lambda}\right)} \\ &\times \exp\left(-\frac{2}{1-n/N}\lambda^2\right)\exp\left(-\frac{1}{3}\left(1+\frac{n^3}{(N-n)^3}\right)\frac{\lambda^4}{n}\right). \end{aligned} \tag{2.4}$$

Theorem 2. *Suppose* $\sum_{i=1}^n X_i \sim$ Hypergeometric(n, D, N). *Define* $\psi := n/N$ *and* $\mu := D/N$, *and let* $n < D$. *Fix* $\mu_0, \psi_0 > 0$ *such that* $0 < \mu_0 \le \mu \le 1-\mu_0 < 1$ *and* $0 < \psi_0 \le \psi \le 1-\psi_0 < 1$. *Then there exist constants* K_1, K_2 *depending only on* μ_0 *and* ψ_0 *such that*:

(i) *For all* $\lambda > 0$,

$$P\big(\sqrt{n}(\bar{X}_n - \mu) = \lambda\big) \le \frac{K_1}{\sqrt{n}}\exp\left(-\frac{2\lambda^2}{1-n/N}\right)\exp\left(-\left(\frac{1}{4}+\frac{1}{3}\left(\frac{n}{N-n}\right)^3\right)\frac{\lambda^4}{n}\right).$$

(ii) *For all* $0 < t < \lambda$,

$$P\big(\sqrt{n}(\bar{X}_n - \mu) \ge t\big) \le \frac{K_2}{\lambda}\left(\begin{array}{c} \exp\left(-\frac{2\lambda^2}{1-n/N}\right)\exp\left(-\left(\frac{1}{4}+\frac{1}{3}\left(\frac{n}{N-n}\right)^3\right)\frac{\lambda^4}{n}\right) \\ \times\exp\left(\lambda(\lambda-t)\left(\frac{4}{1-n/N}+1+\frac{4n^3}{3(N-n)^3}\right)\right) \end{array}\right). \tag{2.5}$$

(iii) *For all* $\lambda > 0$,

$$P\big(\sqrt{n}(\bar{X}_n - \mu) \ge \lambda\big) \le \frac{K_2}{\lambda}\exp\left(-\frac{2\lambda^2}{1-n/N}\right)\exp\left(-\left(\frac{1}{4}+\frac{1}{3}\left(\frac{n}{N-n}\right)^3\right)\frac{\lambda^4}{n}\right).$$

We are also able to obtain an analogue of Bennett's inequality by using an important representation of the hypergeometric distribution as a sum of independent Bernoulli random variables with different means. This representation results from a special case of results established by Vatutin and Mikhaĭlov [25] (also see Ehm [4], Theorem A, and Pitman [18]).

Hypergeometric Representation Theorem 1. *If* $1 \le n \le D \wedge (N - D)$, *then*

$$S_{n,D,N} =_d \sum_{i=1}^{n} Y_i, \tag{2.6}$$

where $Y_i \sim \text{Bernoulli}(\pi_i)$ *are independent.*

We may use this representation along with Bennett's inequality to obtain a Bennett-type exponential bound for Hypergeometric random variables (this bound was also discussed earlier in [7], though without proof). The proof of this claim is short, so we will provide it here.

Theorem 3. *Suppose* $S_{n,D,N} \sim \text{Hypergeometric}(n, D, N)$ *with* $1 \le n \le D \wedge (N - D)$. *Define* $\mu_N := D/N$, $\sigma_N^2 := \mu_N(1 - \mu_N)$, *and* $1 - f_n := 1 - (n - 1)/(N - 1)$ *is the finite-sampling correction factor. Then for all* $\lambda > 0$

$$P\big(\sqrt{n}(\bar{X}_{n,D,N} - \mu_N) > \lambda\big) \le \exp\left(-\frac{\lambda^2}{2\sigma_N^2(1 - f_n)}\psi\left(\frac{\lambda}{\sqrt{n}\sigma_N^2(1 - f_n)}\right)\right), \tag{2.7}$$

where $\psi(\lambda) := (2/\lambda^2)h(1 + \lambda)$ *and* $h(\lambda) := \lambda(\log \lambda - 1) + 1$.

Proof. Under the hypotheses it follows from (2.6) that

$$\begin{aligned}
&P\big(\sqrt{n}(\bar{X}_{n,D,N} - \mu_N) > \lambda\big) \\
&\quad = P\left(n^{-1/2}\sum_{i=1}^{n}(Y_i - \mu_i) > \lambda\right) \\
&\quad \le \exp\left(-\frac{\lambda^2}{2(\sum_1^n \pi_i(1 - \pi_i)/n)}\psi\left(\frac{\lambda \cdot n^{-1/2}}{(\sum_1^n \pi_i(1 - \pi_i)/n)}\right)\right) \qquad (2.8) \\
&\quad = \exp\left(-\frac{\lambda^2}{2(n\mu_N(1 - \mu_N)(1 - f_n)/n)}\psi\left(\frac{\lambda \cdot n^{-1/2}}{(n\mu_N(1 - \mu_N)(1 - f_n)/n)}\right)\right) \\
&\quad = \exp\left(-\frac{\lambda^2}{2\sigma_N^2(1 - f_n)}\psi\left(\frac{\lambda}{\sqrt{n}\sigma_N^2(1 - f_n)}\right)\right).
\end{aligned}$$

Note that (2.8) follows by applying Bennett's inequality (his general inequality, rather than the binomial specialization), which is applicable since each Y_i is independent Bernoulli(μ_i) and hence $Y_i - \mu_i \le 1$ a.s. for $1 \le i \le n$. This gives the bound. □

Since $\psi(v) \ge 1/(1 + v/3)$ for all $v \ge 0$ (Shorack and Wellner [21], Proposition 1, page 441), Theorem 3 immediately yields following Bernstein type tail bound.

Corollary 1. *With the same assumptions and notation as in Theorem 3,*

$$P\left(\sqrt{n}(\bar{X}_{n,D,N} - \mu_N) > \lambda\right) \le \exp\left(-\frac{\lambda^2/2}{\sigma_N^2(1-f_n) + \lambda/3\sqrt{n}}\right). \tag{2.9}$$

Detailed proofs of the bounds (2.4) and (2.5) are provided in Section 4. The proofs of these two bounds are complicated and do not proceed by the Cramér–Chernoff method. The proof of (2.4) adapts the argument of León and Perron for the binomial distribution to the hypergeometric case. In adapting the argument, we derive an analogue of a well-known binomial tail probability bound going back to at least Feller [5], pages 150–151: see Lemma 7 for details. The proof of (2.5) adapts Talagrand's argument to the hypergeometric setting. The tools developed in the course of the proofs are specialized to the analysis of binomial coefficients. As such, they may prove useful in understanding how to analyze the tail of distributions such as the multinomial and multivariate hypergeometric by providing guidance for parametrizations which could appear in those settings after the application of Stirling's formula.

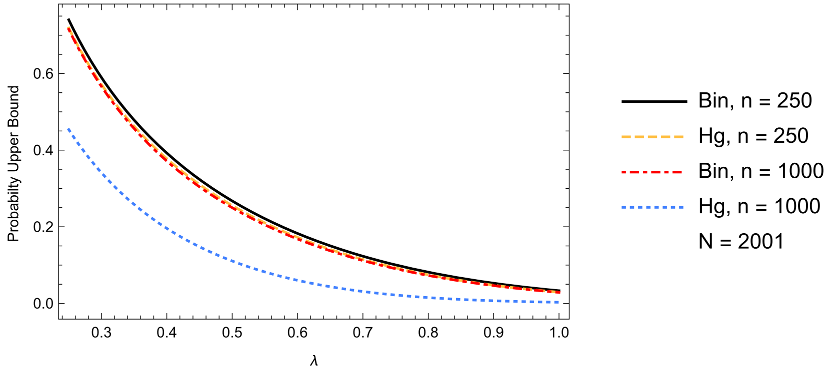
Note that if $N \nearrow \infty$ with n fixed, (2.4) yields a slight improvement of (2.1), the bound of León and Perron, since it contains a quartic term in the exponential. Recovery of this sort is exactly the behavior we would expect in the limit, since (2.1) bounds binomial probabilities and as $N \nearrow \infty$ with $n/N \to 0$ the hypergeometric law converges to the binomial. A similar limiting argument shows we may recover (2.2) from (2.5) as well as (2.3) from (2.7).

Also observe that the bounds (2.4) and (2.5) contain terms involving $1 - n/N$, which incorporates information about the proportion of the population sampled into the bound. This sampling fraction is sharper than the improvement conjectured in Serfling's bound: $1 - n/N < 1 - (n-1)/(N-1) < 1 - (n-1)/N$. For $\lambda > (\sqrt{n}(N-n))/(2(2N-n))$, the expression outside the exponential terms in (2.4) exceeds the non-exponential expression in (2.1). However, for such λ the increase in magnitude is compensated for by the $1 - n/N$ term appearing in the exponent.

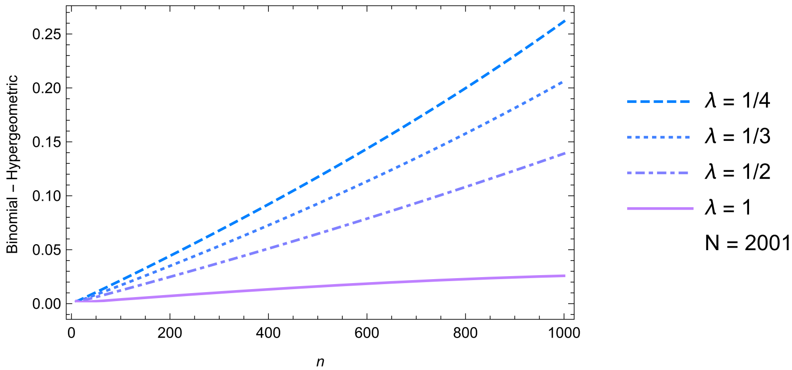
Figure 1 demonstrates the benefit of including a finite-sampling correction factor inside the exponential term: when enough of the population is sampled, the difference between the binomial and hypergeometric bounds can differ by as much as 1/4 for specific deviation values. Figure 2 compares the performance of the new hypergeometric bounds to each other and to the bounds of Serfling (1.1) and Hush and Scovel (1.5). It also provides some insight as to when (2.4) outperforms (2.7) and vice-versa. The finite-but-unspecified constants appearing in (2.5) prevent its inclusion in the figures. Additionally, the constants are not immediately comparable to those in (2.2) because they depend on how one chooses to truncate the sampling fraction and population proportion. The bound (2.5) demonstrates that the factor $1 - n/N$ in the exponential may apply for all $\lambda > 0$ as long as a suitable leading constant is selected.

Chatterjee [3] used Stein's method to derive very general concentration bounds for statistics based on random permutations. For example, here is a restatement of his Proposition 1.1: let $\{a_{i,j} : 1 \le i, j \le N\}$ be a collection of numbers in $[0, 1]$ and let $S \equiv \sum_{i=1}^{N} a_{i,\pi(i)}$ where $\pi \sim$ uniformly on all permutations of $\{1, \ldots, N\}$. Then

$$P\left(\left|S - E(S)\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{4E(S) + 2t}\right) \qquad \text{for all } t > 0.$$



(a) 1a



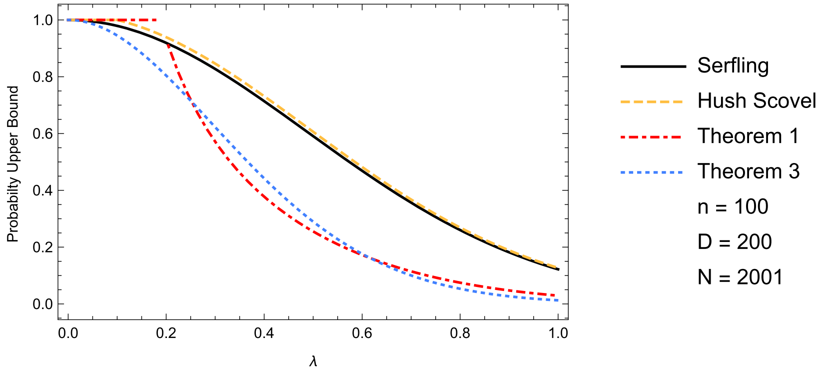
(b) 1b

Figure 1. Comparison of León and Perron's binomial bound (2.1) to the new hypergeometric bound (2.4). In Figure 1(a), the sample size n is set to 250 and 1000 for both bounds. The population size N is taken to be 2001 in both cases. In the legend, lines with the description "Bin" correspond to the binomial bound of León and Perron (2.1), while lines with the description "Hg" correspond to the new hypergeometric bound (2.4). In Figure 1(b), we plot the difference between León and Perron's binomial bound (2.1) to the new hypergeometric bound (2.4) at the fixed deviation-values $\lambda \in \{1/4, 1/3, 1/2, 1\}$. We let n vary between 10 and 1000 to illustrate the impact of introducing the finite-sampling correction factor into the exponential term of the probability bound.

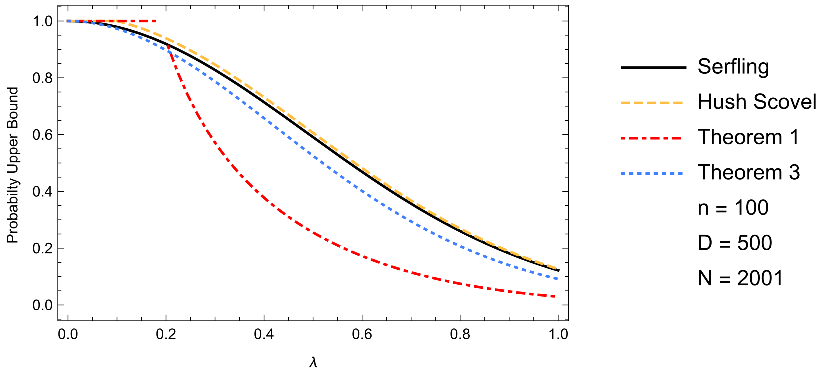
The statistic S was first studied by Hoeffding [10]. The special case which yields the setting of Serfling's inequality is $a_{i,j} := 1_{[i \le n]} c_j$ for $1 \le i, j \le N$ where $1 \le n < N$. Then $S = \sum_{i=1}^{n} c_{\pi(i)} \stackrel{d}{=} S_n$ where $S_n = \sum_{i=1}^{n} X_i$ is as defined in the first paragraph of Section 1 above. In this special case $E(S) = n\bar{c}_N = n\mu$ and Chatterjee's (Bernstein type) bound becomes

$$P\left(n^{-1/2}(S_n - n\mu) \ge \lambda\right) \le \exp\left(-\frac{\lambda^2}{4\bar{c}_N + 2\lambda/\sqrt{n}}\right) \tag{2.10}$$

for all $\lambda > 0$.



(a) 2a



(b) 2b

Figure 2. These plots compare the various exponential bounds for the hypergeometric distribution. In these plots we fix the population to $N = 2001$, and the sample size to $n = 100$. The plots consider a setting with smaller variances by setting $D = 200$ in the first plot (so $D/N = 1/10$) and $D = 500$ in the second (so $D/N = 1/4$). We see that the bound of Theorem 1 (2.4) performs comparably with the bound of Theorem 3 (2.7) in the setting $D = 200$ (2a), and surpasses it when $D = 500$ (2b). This suggests that when $D/N < 1/10$, the bound of Theorem 3 will perform better than the bound of Theorem 1, and when $1/10 \le D/N < 1/2$, the converse.

Goldstein and Işlak [6] recently used a variant of Stein's method to give another inequality for the tails of Hoeffding's statistic S:

$$P\big(\big|S - E(S)\big| > t\big) \le 2\exp\left(-\frac{t^2}{2(\sigma_A^2 + 8\|a\|t)}\right), \tag{2.11}$$

where $\|a\| \equiv \max_{i,j\le N} |a_{i,j} - a_{i\cdot}|$,

$$a_{i\cdot} = \frac{1}{N}\sum_{j=1}^{N} a_{ij}, \qquad a_{\cdot j} = \frac{1}{N}\sum_{i=1}^{N} a_{ij}, \qquad a_{\cdot\cdot} = \frac{1}{N^2}\sum_{i,j=1}^{N} a_{ij}, \quad \text{and}$$

$$\sigma_A^2 = \frac{1}{N-1} \sum_{i,j \le N} (a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot})^2.$$

Specializing (2.11) to the setting of Serfling's inequality (with $a_{i,j} := 1_{[i \le n]} c_j$) yields

$$P\big(n^{-1/2}|S_n - n\bar{c}_N| > \lambda\big) \le 2\exp\bigg(-\frac{\lambda^2/2}{\sigma_c^2(1-f_n) + 8\|c\|\lambda/\sqrt{n}}\bigg), \tag{2.12}$$

where $\sigma_c^2 = N^{-1}\sum_{j=1}^N (c_j - \bar{c}_N)^2$ and $\|c\| \equiv \max_{j \le N} |c_j - \bar{c}_N|$. This Bernstein type bound is in the same setting as Serfling's inequality, but the bound has an explicit dependence on σ_c^2. This is similar to the bound of Bardenet and Maillard (1.6) which incorporates variance information through the parameter γ^2.

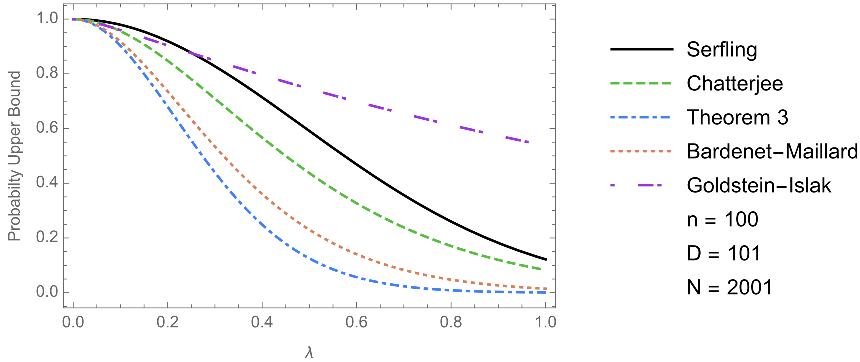
Further specialization of (2.12) to the (one-sided) hypergeometric setting (with $c_j = 1\{j \le D\}$ for $j = 1, \dots, N$) yields

$$\begin{aligned} &P\big(n^{-1/2}\big(S_n - n(D/N)\big) > \lambda\big) \\ &\quad \le \exp\bigg(-\frac{\lambda^2/2}{(D/N)(1-D/N)(1-f_n) + 8\{(D/N) \vee (1-D/N)\}\lambda/\sqrt{n}}\bigg) \\ &\quad = \exp\bigg(-\frac{\lambda^2/2}{\sigma_N^2(1-f_n) + 8\{\mu_N \vee (1-\mu_N)\}\lambda/\sqrt{n}}\bigg). \end{aligned} \tag{2.13}$$

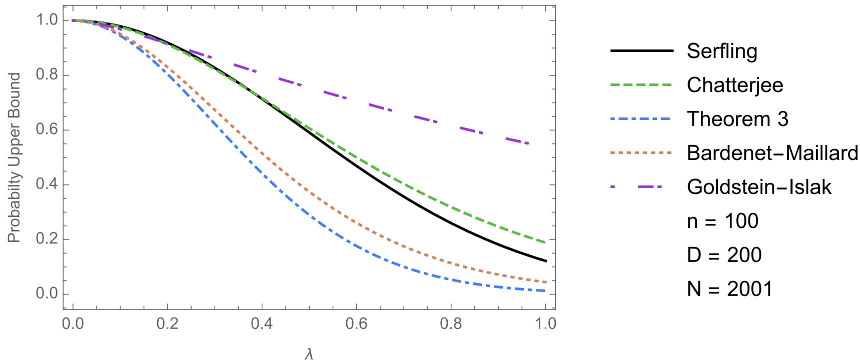
This bound differs from the bound given in (2.9) (the Bernstein type corollary of Theorem 3) only through the second term in the denominator inside the exponential: note that $8\{\mu_N \vee (1-\mu_N)\} \ge 4 > 1/3$.

Comparing the Bernstein type bounds (2.12) and (2.13) to Serfling's inequality (1.3) with $b = 1$ and $a = 0$, we see that the bound of Goldstein and Işlak is smaller than Serfling's bound when $\lambda \le \sqrt{n}/(32(\bar{c}_N \vee (1-\bar{c}_N)))(1 - 4\sigma_N^2 + 4\sigma_N^2 f_n - f_n^*)$. Similarly, we see that Chatterjee's bound (2.10) is smaller than Serfling's bound only if $\bar{c}_N \le (1-(n-1)/N)/8$ and then $\lambda \le \sqrt{n}(1-(n-1)/N - 8\bar{c}_N)/4$. Figure 3 gives a comparison of Serfling's bound, Chatterjee's bound, Bardenet and Maillard's bound (1.6), Goldstein and Işlak's bound (2.13), and the Bennett type bound (2.7) in the further hypergeometric special case with $n = 100$, $N = 2001$, and $D \in \{101, 200\}$; note that in the case $D = 200$, $\bar{c}_N = D/N \approx 0.10$ so $8\bar{c}_N \approx 0.8$ while $1 - (n-1)/N) \approx 1 - 0.05$ so the first condition holds and then Chatterjee's bound should win approximately when $\lambda \le \sqrt{n}(0.15)/4 \approx 1.5/4$.

Comparing (2.13) to (2.10), we find the Goldstein–Işlak bound improves Chatterjee's bound when $\lambda \le \sqrt{n}(2\bar{c}_N - \sigma_N^2(1-f_n))/(8(\bar{c}_N \vee (1-\bar{c}_N)) - 1)$. In Figure 3, this region is approximately equal to $\lambda \le 0.08$ when $D = 101$ and $\lambda \le 0.18$ when $D = 200$. From Figure 3(b), we see that the improvement of Goldstein and Işlak's bound to those of Chatterjee and Serfling is very small in this region. From Figure 3(a), we see that Chatterjee's bound is smaller than both the Goldstein and Işlak bound (2.13) as well as Serfling's bound, when D/N is small and $0.08 \le \lambda \le \sqrt{n}(0.55)/4 \approx 1.37$, but that all three are improved by Bardenet and Maillard's bound (1.6) and the Bennett type bound (2.7).



(a) 1a



(b) 1b

Figure 3. Comparison of Serfling's bound (1.3), Chatterjee's bound (2.10), and the bound of Goldstein and Işlak (2.13), Bardenet and Maillard's bound (1.6), and Theorem 3. In Figure 3(a), the sample size is $n = 100$, the population size is $N = 2001$, and the number of successes is $D = 101$. In Figure 3(b), the sample size remains $n = 100$, the population size remains $N = 2001$, but $D = 200$.

3. Convex order for the hypergeometric distribution

When sampling without replacement from a finite population concentrated on $[0, 1]$, the hypergeometric distribution occupies an extreme position with respect to convex order. This extreme position offers additional reason to give the hypergeometric distribution special consideration, since we might hope to adapt bounds for its tail to the tails of the random variables it dominates through the convex order.

The extreme position of the hypergeometric distribution was essentially proved by Kemperman [15]. In his paper, Kemperman studied (among many other things) finite populations majorized by nearly Rademacher populations; through transformation, this describes the hypergeometric setting. We say nearly Rademacher since Kemperman's analysis resulted in majorizing populations consisting entirely of -1's and 1's with the exception of a single exceptional element α with $-1 < \alpha < 1$.

Here, we revisit his argument, modified so it applies to a population with elements between 0 and 1. We then provide an extension of the argument in order to obtain a hypergeometric population which sub-majorizes this initial population. Since the extension follows naturally from Kemperman's majorization result, we begin with his procedure here. We start with relevant definitions from Marshall, Olkin, and Arnold [17].

Definition 1. *For a vector* $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$, *let*

$$x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[N]}$$

denote the components of $\mathbf{x}$ *in decreasing order.*

Definition 2. *For* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$,

$$\mathbf{x} \prec \mathbf{y} \qquad \textit{if} \begin{cases} \sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, & k = 1, \ldots, N-1, \\ \sum_{i=1}^{N} x_{[i]} = \sum_{i=1}^{N} y_{[i]}, \end{cases}$$

where $\mathbf{x} \prec \mathbf{y}$ *is read as "*$\mathbf{x}$ *is majorized by* $\mathbf{y}$*".*

Definition 3. *For* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$,

$$\mathbf{x} \prec_w \mathbf{y} \qquad \textit{if} \sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \qquad k = 1, \ldots, N,$$

where $\mathbf{x} \prec_w \mathbf{y}$ *is read as "*$\mathbf{x}$ *is weakly sub-majorized by* $\mathbf{y}$*" or, more briefly, "*$\mathbf{x}$ *is sub-majorized by* $\mathbf{y}$*".*

Figure 4 provides an illustration of these definitions. In the following lemma, we restate Kemperman's procedure so it constructs a majorizing hypergeometric population. See Section 4, pages 165–168 in [15] for the original Rademacher argument.

Lemma 1 (Kemperman [15]). *For any finite population* $\mathbf{x} \in \mathbb{R}^N$, *such that* $0 \leq x_i \leq 1$ *for all* $1 \leq i \leq N$, *there exists a population* $\mathbf{c} \in \mathbb{R}^N$, *consisting only of* 0*'s,* 1*'s, and at most a single element between* 0 *and* 1, *which majorizes the original population. In fact,* $\mathbf{c}$ *consists of* D 1*'s,* $N - D - 1$ 0*'s, and a number* $\alpha \in [0, 1)$ *where* D *and* α *are determined by* $D = \lfloor N\bar{\mathbf{x}}_N \rfloor$, *and* $\alpha = N\bar{\mathbf{x}}_N - D$.

Proof. For an updated version of Kemperman's argument see the longer arXiv version, [8]. □

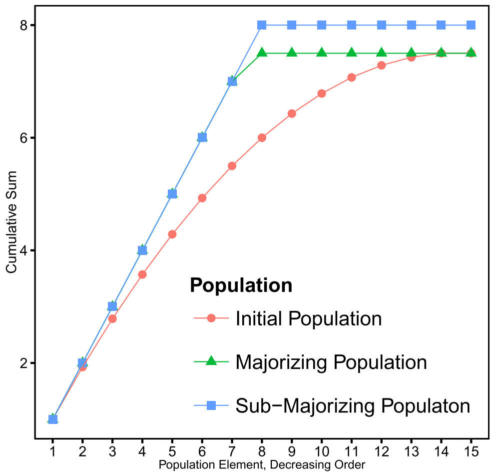


Figure 4. The initial population line in the display coresponds to $\mathbf{c} = \{0, 1/14, 2/14, \ldots, 13/14, 1\}$. The majorizing population contains seven 0's, seven 1's, and a single exceptional element of $1/2$. The sub-majorizing population contains seven 0's and eight 1's. In the display, each population is sorted in decreasing order; the corresponding lines show the cumulative sum of the ordered population elements.

Lemma 2. *For any finite population* $\mathbf{x} \in \mathbb{R}^N$, *such that* $0 \le x_i \le 1$ *for all* $1 \le i \le N$, *there exists a population* $\mathbf{z} \in \mathbb{R}^N$, *consisting only of* 0*'s and* 1*'s, which sub-majorizes the original population.*

Proof. Consider a finite population $\mathbf{x} \in \mathbb{R}^N$ which obeys the hypotheses. Using Lemma 1, we may construct a population $\mathbf{y} \in \mathbb{R}^N$ which majorizes $\mathbf{x}$. By Lemma 1, we know that $\mathbf{y}$ consists only of 0's, 1's, and at most a single exceptional element y_N between 0 and 1.

If the exceptional element is either exactly 0 or exactly 1, we are done. So, suppose $0 < y_N < 1$. Create a new population $\mathbf{z} \in \mathbb{R}^N$ such that $z_i = y_i$ for $1 \le i \le N - 1$, and $z_N = 1$. This population $\mathbf{z}$ then sub-majorizes $\mathbf{y}$ and hence sub-majorizes $\mathbf{x}$, completing the proof. □

Lemma 3. *Suppose* $\mathbf{x} \in \mathbb{R}^N$ *is a population consisting only of* 0*'s,* 1*'s, and a single exceptional element,* x_1, *such that* $0 < x_1 < 1$. *Suppose* $\mathbf{y} \in \mathbb{R}^N$ *is a population whose elements are the same as those in* $\mathbf{x}$, *except* $y_1 = 1$ *and so* $y_1 > x_1$. *Let* $X_1, \ldots, X_n$ *denote a sample without replacement from* $\mathbf{x}$, *and* $Y_1, \ldots, Y_n$ *denote a sample without replacement from* $\mathbf{y}$, $1 \le n \le N$. *Finally, suppose* ϕ *is a continuous convex increasing function on* $\mathbb{R}$. *Then*

$$E\phi\left(\sum_{i=1}^{n} X_i\right) \le E\phi\left(\sum_{i=1}^{n} Y_i\right). \tag{3.1}$$

Proof. We adapt Kemperman's (1973) argument for Rademacher populations to the current setting of hypergeometric sub-majorization.

Observe that

$$E\phi\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{\binom{N}{n}} \sum \phi(x_{i_1} + \cdots + x_{i_n}),$$

where the sum is over all sets of indices $1 \le i_1 < i_2 < \cdots < i_n \le N$. Note the same holds for sampling without replacement from $\mathbf{y}$, with suitable substitution. Therefore,

$$\binom{N}{n}\left[E\phi\left(\sum_{i=1}^{n} Y_i\right) - E\phi\left(\sum_{i=1}^{n} X_i\right)\right]$$
$$= \sum \bigl(\phi(y_1 + y_{i_2} + \cdots + x_{i_n}) - \phi(x_1 + x_{i_2} + \cdots + x_{i_n})\bigr),$$

where the sum is over all distinct indices $2 \le i_2 < i_3 < \cdots < i_n \le N$. Note that sets of indices with $i_1 > 1$ cancel out by definition of the two populations. Since ϕ is assumed convex increasing, each term of the sum is non-negative. Hence, the entire sum is non-negative as well. This gives the claim. □

We next specialize a proposition stated in Marshall, Olkin, and Arnold [17], page 455, to the current problem. Proof of the general statement given in the text is credited to Karlin; proof for the specific cases of sampling with and without replacement to Kemperman. As proof is given in Marshall, Olkin, and Arnold, we simply state the result here.

Lemma 4. *Let $\mathbf{x} \in \mathbb{R}^N$ be an arbitrary finite population. Let $\mathbf{y} \in \mathbb{R}^N$ be a finite population which majorizes $\mathbf{x}$. Let $X_1, \ldots, X_n$ denote a sample without replacement from $\mathbf{x}$, and $Y_1, \ldots, Y_n$ denote a sample without replacement from $\mathbf{y}$, $1 \le n \le N$. Finally, suppose ϕ is a continuous convex increasing function on $\mathbb{R}$. Then*

$$E\phi\left(\sum_{i=1}^{n} X_i\right) \le E\phi\left(\sum_{i=1}^{n} Y_i\right).$$

Note that Lemma 4 requires majorization between populations. We may combine the preceding lemmas to demonstrate the following claim.

Theorem 4. *For any finite population $\mathbf{x} \in \mathbb{R}^N$, such that $0 \le x_i \le 1$ for all $1 \le i \le N$, there exists a population $\mathbf{y} \in \mathbb{R}^N$, consisting only of 0's and 1's which sub-majorizes the original population. Let $X_1, \ldots, X_n$ denote a sample without replacement from $\mathbf{x}$, and $Y_1, \ldots, Y_n$ denote a sample without replacement from $\mathbf{y}$, $1 \le n \le N$. Finally, suppose ϕ is a continuous convex increasing function on $\mathbb{R}$. Then*

$$E\phi\left(\sum_{i=1}^{n} X_i\right) \le E\phi\left(\sum_{i=1}^{n} Y_i\right).$$

Proof. Suppose $\mathbf{x} \in \mathbb{R}^N$ is a finite population which satisfies the hypotheses. We may use Lemma 1 to construct a population $\mathbf{z} \in \mathbb{R}^N$ such that $\mathbf{z}$ majorizes $\mathbf{x}$, and $\mathbf{z}$ consists only of 0's, 1's, and at most a single exceptional element between 0 and 1. For $1 \le n \le N$, let $Z_1, \dots, Z_n$ denote a sample without replacement from $\mathbf{z}$. By Lemma 4, we then have the order

$$E\phi\left(\sum_{i=1}^{n} X_i\right) \le E\phi\left(\sum_{i=1}^{n} Z_i\right). \tag{3.2}$$

Next, by Lemma 2 we may construct a population $\mathbf{y} \in \mathbb{R}^N$ consisting only of 0's and 1's that sub-majorizes $\mathbf{z}$. Then by (3.1) we have

$$E\phi\left(\sum_{i=1}^{n} Z_i\right) \le E\phi\left(\sum_{i=1}^{n} Y_i\right). \tag{3.3}$$

Combining (3.2) and (3.3) proves the claim. □

Inequality (2.7) provides an opportunity to apply Theorem 4. Recalling the notation of the Introduction, let $\mathbf{c} := \{c_1, \dots, c_N\}$ be a population such that $0 \le c_i \le 1$ for $1 \le i \le N$, $a = 0$, $b = 1$, and $\bar{c}_N + 1/N \le 1/2$. Using Kemperman's algorithm, as stated in Lemma 1, there exists a population $\mathbf{m} := \{m_1, \dots, m_N\}$ which majorizes $\mathbf{c}$ such that $m_i \in \{0, 1\}$ for $1 \le i \le N - 1$, and $0 \le m_N \le 1$. In the following, suppose $0 < m_N < 1$, since if $m_N = 0$ or $m_N = 1$ we can apply (2.7) directly.

Since $\mathbf{c} \prec \mathbf{m}$, we have $\bar{m}_N = \bar{c}_N$. Using Lemma 2, there is a population $\{h_1, \dots, h_N\}$ with $h_i \in \{0, 1\}$ for $1 \le i \le N$ that sub-majorizes $\mathbf{c}$. By the preceding construction, we have $h_i = m_i$ for $1 \le i \le N - 1$, and $h_N \equiv 1 > m_N$.

Without loss of generality, relabel $\mathbf{m}$ and $\mathbf{h}$ so that: for $1 \le i \le D - 1$ we have $h_i = m_i = 1$; for $i = D$ we have $h_D = 1 > m_D > 0$; for $D + 1 \le i \le N$ we have $h_i = m_i = 0$. Denote the exceptional element of $\mathbf{m}$ by $\alpha := m_D$. With the populations so modified, we derive bounds for the difference between the populations means:

$$\frac{1}{N} \ge \bar{h}_N - \bar{m}_N = \frac{h_D - m_D}{N} = \frac{1 - \alpha}{N} \ge 0. \tag{3.4}$$

By construction, we thus have

$$\bar{m}_N \le \bar{h}_N \le \bar{m}_N + \frac{1}{N} \le \frac{1}{2}.$$

Suppose then that we sample $n < D$ items without replacement from $\mathbf{c}$. Let X_i denote the sample without replacement from $\mathbf{c}$, and let H_i denote a corresponding sample without replacement from $\mathbf{h}$. Then for $t > 0$

$$\begin{aligned} P\left(\sum_{i=1}^{n} X_i - n\mu_c \ge t\right) &\le \inf_{r>0} \frac{E\exp(r\sum_{i=1}^{n} X_i)}{\exp(rt + rn\mu_c)} \\ &\le \inf_{r>0} \frac{E\exp(r\sum_{i=1}^{n} H_i)}{\exp(rt + rn\mu_c)} \end{aligned} \tag{3.5}$$

$$
\begin{aligned}
&= \inf_{r>0} \exp\big(rn(\mu_h - \mu_c)\big) \frac{E \exp(r \sum_{i=1}^n (H_i - \mu_h))}{\exp(rt)} \\
&\le \inf_{r>0} \exp\left(r \frac{n}{N}\right) \frac{E \exp(r \sum_{i=1}^n (H_i - \mu_h))}{\exp(rt)} \qquad (3.6)\\
&= \inf_{r>0} \exp\left(r \frac{n}{N}\right) \frac{E \exp(r \sum_{i=1}^n (Y_i - \pi_i))}{\exp(rt)} \\
&\le \inf_{r>0} \exp\left(r \frac{n}{N} - rt + n\left(\frac{\sum_{i=1}^n \pi_i(1-\pi_i)}{n}\right)\big(e^r - 1 - r\big)\right) \qquad (3.7)\\
&= \inf_{r>0} \exp\left(r \frac{n}{N} - rt + n\gamma^2\big(e^r - 1 - r\big)\right), \qquad (3.8)
\end{aligned}
$$

where in the final line we write $\gamma^2 := (1/n)\sum_{i=1}^n \pi_i(1-\pi_i)$. The inequality at (3.6) follows by (3.4). The inequality at (3.5) follows by Theorem 4. The inequality at line (3.7) follows by Shorack and Wellner, page 852, display (b) [21].

At this point, we may continue from (3.8) and optimize over r. Doing so yields an optimal choice of

$$
r^* = \log\left(1 + \frac{Nt - n}{nN\gamma^2}\right).
$$

Using this value, however, yields an exponential bound that is somewhat difficult to compare to (2.7). If instead we simply choose

$$
r_2^* = \log\left(1 + \frac{t}{n\gamma^2}\right),
$$

we obtain a bound similar in performance to the bound we find using r^*, but has the benefit of easy comparison to (2.7). The choice r_2^* corresponds to the optimal value of r when the original population is majorized by a hypergeometric population. We continue from (3.8) using r_2^*, and obtain

$$
\begin{aligned}
&P\left(\sum_{i=1}^n X_i - n\mu_c \ge t\right) \\
&\quad \le \exp\left(\frac{n}{N} \log\left(1 + \frac{t}{n\gamma^2}\right)\right) \cdot \exp\left(-t\left[\left(1 + \frac{n\gamma^2}{t}\right)\log\left(1 + \frac{t}{n\gamma^2}\right) - 1\right]\right) \qquad (3.9)\\
&\quad = \exp\left(\frac{n}{N} \log\left(1 + \frac{t}{n\gamma^2}\right)\right) \cdot \exp\left(-\frac{t^2}{2n\gamma^2}\psi\left(\frac{t}{n\gamma^2}\right)\right).
\end{aligned}
$$

Writing $\lambda = t/\sqrt{n}$, and substituting $\gamma^2 = (D/N)(1 - D/N)(1 - f_n) \equiv \sigma_N^2(1 - f_n)$, we obtain the following bound:

$$
\begin{aligned}
&P\big(\sqrt{n}(\bar{X}_n - \mu_c) \ge \lambda\big) \\
&\quad \le \left(1 + \frac{\lambda}{\sqrt{n}\sigma_N(1 - f_n)}\right)^{n/N} \exp\left(-\frac{\lambda^2}{2\sigma_N^2(1 - f_n)}\psi\left(\frac{\lambda}{\sqrt{n}\sigma_N^2(1 - f_n)}\right)\right).
\end{aligned}
$$

In this form, the cost of sub-majorization is clear when compared to (2.7): we incur the leading term outside the exponent. By shifting and scaling the population, we may use this bound to obtain the following theorem for the general problem of sampling without replacement:

Theorem 5. *Let* $\mathbf{c} := \{c_1, \ldots, c_N\}$ *be a population with* $a = \min_{1 \le i \le N} c_i$ *and* $b = \max_{1 \le i \le N} c_i$ *both finite. Let* $\mathbf{d} := \{(c_1 - a)/(b - a), \ldots, (c_N - a)/(b - a)\}$*. Suppose first that* $\mathbf{d}$ *is majorized by a Hypergeometric population such that* $D/N \le 1/2$*. From this Hypergeometric population define* $\sigma_N^2 := (D/N)(1 - D/N)$*. Then the following bound holds for a sample without replacement of* $n < D$ *items from* $\mathbf{c}$:

$$P\big(\sqrt{n}(\bar{X}_n - \mu_c) \ge \lambda\big) \le \exp\biggl(-\frac{\lambda^2}{2(b-a)^2\sigma_N^2(1-f_n)}\psi\biggl(\frac{\lambda}{\sqrt{n}(b-a)\sigma_N^2(1-f_n)}\biggr)\biggr). \tag{3.10}$$

If instead $\bar{d}_N + 1/N \le 1/2$*, then the following bound holds for a sample without replacement of* $n < D$ *items from* $\mathbf{c}$:

$$\begin{aligned} &P\big(\sqrt{n}(\bar{X}_n - \mu_c) \ge \lambda\big) \\ &\quad\le \biggl(1 + \frac{\lambda}{\sqrt{n}(b-a)\sigma_N(1-f_n)}\biggr)^{n/N} \\ &\qquad\times \exp\biggl(-\frac{\lambda^2}{2(b-a)^2\sigma_N^2(1-f_n)}\psi\biggl(\frac{\lambda}{\sqrt{n}(b-a)\sigma_N^2(1-f_n)}\biggr)\biggr). \end{aligned} \tag{3.11}$$

Two-sample rank tests provide an opportunity to explore the behavior of the bounds of Theorem 5. Following the exposition in Chapter 4 of Hájek, Šidák, and Sen [9] (with the notation modified), let $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_m$ be random samples with continuous distributions F_Y and F_Z. Form the pooled sample $Y_{n+j} = Z_j$, $j = 1, \ldots, m$, and $N = n + m$. Let R_i $(i = 1, \ldots, N)$ denote the rank of the observation Y_i in the ordered sequence $Y_{(1)} < Y_{(2)} < \cdots < Y_{(N)}$. To test the null hypothesis $H_0 : F_Y = F_Z$ against alternatives of shifts in location, one may use the Wilcoxon test (see page 96, [9]). The test statistic, expectation under the null, and variance under the null are

$$S_W := \sum_{i=1}^{n} R_i, \qquad ES_W = \frac{1}{2}n(n+m+1) \quad \text{and} \quad \operatorname{Var}(S_W) = \frac{1}{12}nm(n+m+1).$$

Under the null, S_W may be viewed as the sum in a sample without replacement from the population $\mathbf{c}_W := \{1, 2, \ldots, N\}$, where $a = 1$ and $b = N$. Shifting and scaling the population produces $\mathbf{d}_W := \{0, 1/(N-1), \ldots, (N-2)/(N-1), 1\}$. If N is even, then $\mathbf{d}_W$ is majorized by a Hypergeometric population containing $N/2$ 1's and $N/2$ 0's, and hence $\sigma_N^2 = (D/N)(1 - D/N) = 1/4$. If we additionally assume $n \le m$, we may use (3.10) to study its finite sample behavior. Doing so we find for $\lambda > 0$

$$P\biggl(\sqrt{n}\biggl(\bar{X}_n - \frac{(N+1)}{2}\biggr) \ge \lambda\biggr) \le \exp\biggl(-\frac{2\lambda^2}{(n+m-1)^2(1-f_n)}\psi\biggl(\frac{2\lambda}{\sqrt{n}(n+m-1)(1-f_n)}\biggr)\biggr).$$

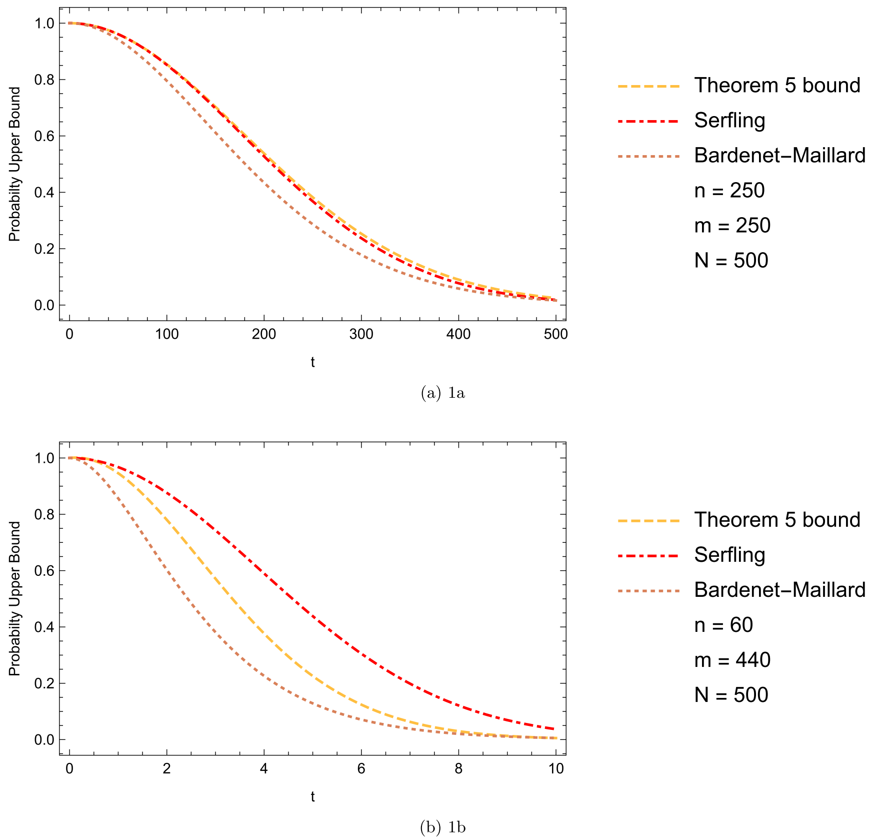


Figure 5. Comparison of the bounds of Theorem 5 to Bardenet and Maillard's bound (1.6) and Serfling's bound. The first Figure 5(a), corresponds to the Wilcoxon example. The second Figure 5(b), corresponds to the Klotz example.

Serfling's bound (1.1) may be applied in this case as well; through its application we find

$$P\left(\sqrt{n}\left(\bar{X}_n - \frac{(N+1)}{2}\right) \geq \lambda\right) \leq \exp\left(-\frac{2\lambda^2}{(n+m-1)^2(1-(n-1)/(n+m))}\right).$$

Finally, we may use Bardenet and Maillard's bound (1.6) with $\delta = \delta_f = 1 \times 10^{-7}$ to analyze the situation as well. Figure 5 compares the performance of these three bounds when $n = m = 250$. In this case, we see that the bounds are comparable, with Bardenet and Maillard's bound performing the best, and Serfling's performance superior to (3.10). This occurs because the variance component $(D/N)(1 - D/N)$ (which is close to $1/4$ when $n = m = 250$) that appears in the bound is the variance of the majorizing hypergeometric population rather than the variance of the shifted and scaled population $\mathbf{d}_w$ (which is close to $1/12$ when $n = m = 250$). Bardenet and

Maillard's bound performs well because it incorporates information about the variance of the untransformed population into its bound.

Another example is found in the Klotz test, which is used to test the null $H_0 : F_Y = F_Z$ against alternatives of differences in scale (see [9], page 104). Recalling $N := n + m$, the test statistic, expectation under the null, and variance under the null are

$$S_K := \sum_{i=1}^{n}\left[\Phi^{-1}\left(\frac{R_i}{N+1}\right)\right]^2, \qquad ES_K = \frac{n}{N}\sum_{i=1}^{N}\left[\Phi^{-1}\left(\frac{i}{N+1}\right)\right]^2,$$

and

$$\operatorname{Var} S_K = \frac{nm}{N(N-1)}\sum_{i=1}^{N}\left[\Phi^{-1}\left(\frac{i}{N+1}\right)\right]^4 - \frac{m}{n(N-1)}(ES_K)^2.$$

Defining the population

$$\mathbf{c}_K := \left\{c_i := \left[\Phi^{-1}\left(\frac{i}{N+1}\right)\right]^2, 1 \le i \le N\right\},$$

we may view S_K under the null as the sum in a sample without replacement from $\mathbf{c}_K$. If $n + m = 500$, we may compute $\mathbf{c}_K$, and find $a \approx 6.26 \times 10^{-6}$ and $b \approx 8.29$. Shifting and scaling the population produces $\mathbf{d}_K$, which is bounded by 0 and 1. This population is majorized by a population containing 59 1's, 440 0's, and a single exceptional element approximately equal to 0.044. Hence, it is sub-majorized by a population containing 60 1's and 440 0's. Supposing that $n = 60$ and $m = 440$, we may use (3.11) to analyze this scenario since the mean of the sub-majorizing population is 3/25 (also note $(D/N)(1 - D/N) = 66/625$ in this case). Doing so (with the conservative approximation that $b - a \approx 8.29$), we find for $\lambda \ge 0$ that

$$\begin{aligned} &P\left(\sqrt{60}(\bar{X}_{60} - \mu_K) \ge \lambda\right) \\ &\quad \le \left(1 + \frac{\lambda}{\sqrt{60}(8.29)(66/625)(440/499)}\right)^{3/25} \\ &\qquad \times \exp\left(-\frac{\lambda^2}{2(8.29)^2(66/625)(440/499)}\psi\left(\frac{\lambda}{\sqrt{60}(8.29)(66/625)(440/499)}\right)\right). \end{aligned}$$

Once again, we may apply Serfling's uniform bound. Doing so here, we find

$$P\left(\sqrt{60}(\bar{X}_{60} - \mu_K) \ge \lambda\right) \le \exp\left(-\frac{2\lambda^2}{(441/500)(8.29)^2}\right). \tag{3.12}$$

As in the Wilcoxon example, we may use Bardenet and Maillard's bound (1.6) with $\delta = \delta_f = 1 \times 10^{-7}$ to analyze the situation. Figure 5 also compares the performance of these three bounds for the special case $n = 60$ and $m = 440$. In this case, we again see that Bardenet and Maillard's bound performs the best, but that the bound obtained via sub-majorization now improves on Serfling's result. This is because the variance component of the sub-majorizing hypergeometric population, $(D/N)(1-D/N) = 66/625 = 0.1056 < 1/4$ reflects some of the variability of the untransformed population. However, the untransformed population d_K has variance $\sigma^2 \approx 0.0258$; this variability is captured in the bound of Bardenet and Maillard, and so we see the improved performance.

Thus we see that (sub-)majorization, as a strategy for finding exponential bounds which incorporate information about the population variance in the problem of sampling without replacement from a bounded finite population, can produce sub-optimal results. As we saw, this is because the (sub-)majorizing hypergeometric population can be more variable than the underlying population from which we sample. However, if our goal is to find uniform exponential bounds, this information loss is immaterial: such bounds apply to all underlying populations, regardless of their variability. Hence the analysis of the hypergeometric distribution which produced Theorems 1 and 2. We turn to the proofs of these bounds in the concluding section.

4. Proofs of the bounds

Our proofs depend on a version of Stirling's formula from Robbins [19].

Lemma 5. *For* $n \in \mathbb{N}_0$

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n+1)} \le n! \le \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n)}. \tag{4.1}$$

To prove (2.4), we will need some additional tools. We start with the following lemma.

Lemma 6. *Suppose* $S_n \sim \mathrm{Hypergeometric}(n, D, N)$ *with* $1 \le n < D \le \lfloor N/2 \rfloor$ *and* $1 \le k \le n-1$. *Then for* $k \ge n(D/N)$ *we have*

$$\begin{aligned} P(S_n = k) \le{}& \frac{1}{\sqrt{2\pi}} \frac{\sqrt{D(N-D)n(N-n)}}{\sqrt{k(D-k)(n-k)(N-D-(n-k))N}} \\ &\times \exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right). \end{aligned} \tag{4.2}$$

Proof. The proof follows by direct analysis. Using Stirling's formula (4.1), we have

$$\begin{aligned} P(S_n = k) &= \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \\ &\le \frac{1}{\sqrt{2\pi}} \frac{\sqrt{D(N-D)n(N-n)}}{\sqrt{k(D-k)(n-k)(N-D-(n-k))N}} \\ &\quad \times \frac{D^D(N-D)^{N-D}n^n(N-n)^{N-n}}{k^k(D-k)^{D-k}(n-k)^{n-k}(N-D-(n-k))^{N-D-(n-k)}N^N} \end{aligned} \tag{4.3}$$

$$\times \frac{\exp(\frac{1}{12D} + \frac{1}{12(N-D)} + \frac{1}{12n} + \frac{1}{12(N-n)})}{\exp(\frac{1}{12k+1} + \frac{1}{12(D-k)+1} + \frac{1}{12(n-k)+1} + \frac{1}{12(N-D-(n-k))+1} + \frac{1}{12N+1})}$$
$$=: A \cdot B \cdot C.$$

We consider the B term first. Define $u := k/n - \mu$, recalling $\mu := D/N$. We then have

$$\begin{aligned}
B &= \frac{D^D(N-D)^{N-D}/N^N}{(\frac{k}{n})^k \cdot (1-\frac{k}{n})^{n-k} \cdot (\frac{D-k}{N-n})^{D-k}(1-\frac{D-k}{N-n})^{N-n-(D-k)}} \\
&= \frac{(\frac{D}{N})^D(1-\frac{D}{N})^{N-D}}{(\frac{k}{n})^k(1-\frac{k}{n})^{n-k}(\frac{D-k}{N-n})^{D-k}(1-\frac{D-k}{N-n})^{N-n-(D-k)}} \\
&= \frac{(\frac{D}{N})^D}{(\frac{k}{n})^k(\frac{D-k}{N-n})^{D-k}} \cdot \frac{(1-\frac{D}{N})^{N-D}}{(1-\frac{k}{n})^{n-k}(1-\frac{D-k}{N-n})^{N-n-(D-k)}} \\
&= \frac{(\frac{D}{N})^k(\frac{D}{N})^{D-k}}{(\frac{k}{n})^k(\frac{D-k}{N-n})^{D-k}} \cdot \frac{(1-\frac{D}{N})^{n-k}(1-\frac{D}{N})^{N-D-(n-k)}}{(1-\frac{k}{n})^{n-k}(1-\frac{D-k}{N-n})^{N-n-(D-k)}} \\
&= \left(\frac{\mu}{u+\mu}\right)^k \left(\frac{\frac{N-n}{N}}{\frac{D-k}{D}}\right)^{D-k} \cdot \left(\frac{1-\mu}{1-(u+\mu)}\right)^{n-k} \left(\frac{\frac{N-n}{N}}{\frac{N-n-(D-k)}{N-D}}\right)^{N-D-(n-k)} \\
&= \exp\bigl(-n\Psi(u,\mu)\bigr) \cdot \frac{(\frac{N-n}{N})^{N-n}}{(\frac{D-k}{D})^{D-k}(\frac{N-D-(n-k)}{N-D})^{N-D-(n-k)}} \\
&= \exp\bigl(-n\Psi(u,\mu)\bigr) \cdot B_2,
\end{aligned}$$

where the first factor corresponds to the same function as in Talagrand's argument for the binomial distribution [22], pages 48–50, and we recall

$$\Psi(u,\mu) := (u+\mu)\log\left(\frac{u+\mu}{\mu}\right) + \bigl(1-(u+\mu)\bigr)\log\left(\frac{1-(u+\mu)}{1-\mu}\right).$$

Now, we can further rewrite B_2 as

$$B_2 = \left(\frac{\frac{N-n}{N}}{\frac{D-k}{D}}\right)^{D-k} \cdot \left(\frac{\frac{N-n}{N}}{\frac{N-n-(D-k)}{N-D}}\right)^{N-D-(n-k)} =: \exp(-\Gamma),$$

where

$$\begin{aligned}
\Gamma &= -\log(B_2) \\
&= (D-k)\log\left[\frac{(\frac{D-k}{D})}{(\frac{N-n}{N})}\right] + \bigl[N-n-(D-k)\bigr]\log\left[\frac{(\frac{N-n-(D-k)}{N-D})}{(\frac{N-n}{N})}\right] \\
&= (N-n)\left(\frac{D-k}{N-n}\right)\log\left[\frac{(D-k)/D}{(N-n)/N}\right] + \left(1-\frac{D-k}{N-n}\right)\log\left[\frac{[N-n-(D-k)]/(N-D)}{(N-n)/N}\right].
\end{aligned}$$

Now $k = n(u+\mu)$, so

$$\frac{D-k}{N} = \mu - \frac{n}{N}(u+\mu) = \mu(1-n/N) - (n/N)u$$

and

$$\frac{(D-k)/N}{(N-n)/N} = \frac{\mu(1-(n/N)) - (n/N)u}{1-n/N} = \mu - \frac{n/N}{1-n/N}u.$$

Thus we also have

$$1 - \frac{(D-k)/N}{(N-n)/N} = 1 - \mu + \frac{n/N}{1-(n/N)}u.$$

Thus it follows that, with $f = f_N := n/N$, $\bar{f} = \bar{f}_N := 1 - f_N$,

$$\begin{aligned}\frac{\Gamma}{N-n} &= \left(\mu - \frac{f}{\bar{f}}u\right)\log\left[\left(\mu - \frac{f}{\bar{f}}u\right)\frac{1}{\mu}\right] + \left(1-\mu+\frac{f}{\bar{f}}u\right)\log\left[\left(1-\mu+\frac{f}{\bar{f}}u\right)\frac{1}{1-\mu}\right]\\ &= \Psi\left(\frac{f}{\bar{f}}u, 1-\mu\right),\end{aligned}$$

where Ψ is as defined above. Thus the B term can be rewritten as

$$B = \exp\left(-n\Psi(u,\mu) - (N-n)\Psi\left(\frac{f}{\bar{f}}u, 1-\mu\right)\right).$$

Now Ψ satisfies $\Psi(0,\mu) = 0$, $\frac{\partial}{\partial u}\Psi(0,\mu) = 0$, and, as in Talagrand (as well as van der Vaart and Wellner [23], pages 460–461),

$$\frac{\partial^2}{\partial u^2}\Psi(u,\mu) = \frac{4}{1-4(u-(1/2-\mu))^2} \geq 4\bigl(1+4\bigl(u-(1/2-\mu)\bigr)^2\bigr).$$

Thus

$$\begin{aligned}&\frac{\partial^2}{\partial u^2}\left[n\Psi(u,\mu) + (N-n)\Psi\left(\frac{f}{\bar{f}}u, 1-\mu\right)\right]\\ &\quad = n\frac{4}{1-4(u-(1/2-\mu))^2} + (N-n)\frac{4(f/\bar{f})^2}{1-4((f/\bar{f})u-(\mu-1/2))^2}\\ &\quad \geq 4n\bigl(1+4\bigl(u-(1/2-\mu)\bigr)^2\bigr) + 4(N-n)(f/\bar{f})^2\left(1+4\left(\frac{f}{\bar{f}}u-(\mu-1/2)\right)^2\right).\end{aligned}$$

Integration across this inequality yields

$$\begin{aligned}&\frac{\partial}{\partial u}\left[n\Psi(u,\mu) + (N-n)\Psi\left(\frac{f}{\bar{f}}u, 1-\mu\right)\right]\\ &\quad \geq 4n\left(u+\frac{1}{3}u^3\right) + 4(N-n)\left(\frac{f}{\bar{f}}\right)^2\left(u+\frac{1}{3}\left(\frac{f}{\bar{f}}\right)^2u^3\right)\end{aligned}$$

$$
\begin{aligned}
&= 4\left(n + (N-n)\left(\frac{n}{N-n}\right)^2\right)u \\
&\quad + \frac{4}{3}\left(n + (N-n)\left(\frac{n}{N-n}\right)^4\right)u^3 \\
&= \frac{4nN}{N-n}u + \frac{4}{3}n\left(1 + \frac{n^3}{(N-n)^3}\right)u^3.
\end{aligned} \tag{4.4}
$$

Here we used

$$
\begin{aligned}
\int_0^u \bigl(1 + 4\bigl(v - (1/2-\mu)\bigr)^2\bigr)\,dv &= u + \frac{4}{3}\bigl(v - (1/2-\mu)\bigr)^3\Big|_0^u \\
&= u + \frac{4}{3}\bigl[u - (1/2-\mu)^3 - \bigl(-(1/2-\mu)\bigr)^3\bigr] \\
&= u + \frac{4}{3}\bigl[\bigl(u - (1/2-\mu)\bigr)^3 + (1/2-\mu)^3\bigr] \\
&\ge u + \frac{4}{3}\bigl[(u/2)^3 + (u/2)^3\bigr] \\
&= u + (1/3)u^3,
\end{aligned}
$$

where the inequality follows since the function $\beta \mapsto (u-\beta)^3 + \beta^3$ is minimized by $\beta = u/2$: with $h_u(\beta) \equiv (u-\beta)^3 + \beta^3$,

$$
\begin{aligned}
h_u'(\beta) &= 3(u-\beta)^2(-1) + 3\beta^2 = 3\bigl\{\beta^2 - \bigl(\beta^2 - 2u\beta + u^2\bigr)\bigr\} \\
&= 3u\{2\beta - u\} = 0 \qquad \text{if } \beta = u/2,
\end{aligned}
$$

while $h_u''(\beta) = 6u > 0$. Similarly,

$$
\begin{aligned}
\int_0^u \left(1 + 4\left(\frac{f}{\bar f}v - (\mu - 1/2)\right)^2\right)dv &= u + \frac{4}{3}\left(\frac{f}{\bar f}v - (\mu-1/2)\right)^3 \frac{\bar f}{f}\bigg|_0^u \\
&= u + \frac{4}{3}\frac{\bar f}{f}\left[\frac{f}{\bar f}u - (\mu-1/2)^3 - \bigl(-(\mu-1/2)\bigr)^3\right] \\
&= u + \frac{4}{3}\frac{\bar f}{f}\left[\left(\frac{f}{\bar f}u - (\mu-1/2)\right)^3 + (\mu-1/2)^3\right] \\
&\ge u + \frac{4}{3}\frac{\bar f}{f}\left[\left(\frac{fu}{\bar f 2}\right)^3 + \left(\frac{fu}{\bar f 2}\right)^3\right] \\
&= u + \frac{1}{3}\left(\frac{f}{\bar f}\right)^2 u^3.
\end{aligned}
$$

Integrating across (4.4) yields

$$n\Psi(u,\mu)+(N-n)\Psi\left(\frac{f}{\bar{f}}u,1-\mu\right)\geq\frac{2nN}{N-n}u^2+(1/3)n\left(1+\frac{n^3}{(N-n)^3}\right)u^4.$$

Thus the B term in (4.3) has the following bound:

$$B\leq\exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right). \tag{4.5}$$

We next analyze the C term in (4.3). We have

$$\begin{aligned}
C &= \frac{\exp(\frac{1}{12D}+\frac{1}{12(N-D)}+\frac{1}{12n}+\frac{1}{12(N-n)})}{\exp(\frac{1}{12k+1}+\frac{1}{12(D-k)+1}+\frac{1}{12(n-k)+1}+\frac{1}{12(N-D-(n-k))+1}+\frac{1}{12N+1})}\\
&=\exp\left(\frac{1}{12D}-\frac{1}{12(D-k)+1}\right)\exp\left(\frac{1}{12(N-D)}-\frac{1}{12(N-D-(n-k))+1}\right)\\
&\quad\times\exp\left(\frac{1}{12n}-\frac{1}{12k+1}\right)\exp\left(\frac{1}{12(N-n)}-\frac{1}{12(n-k)+1}\right)\exp\left(-\frac{1}{12N+1}\right)\\
&=\exp\left(\frac{-12k+1}{[12D][12(D-k)+1]}\right)\exp\left(\frac{-12[n-k]+1}{[12(N-D)][12([N-D]-[n-k])+1]}\right)\\
&\quad\times\exp\left(\frac{1-12(n-k)}{12(12k+1)n}\right)\exp\left(\frac{1-12(N-2n+k)}{12(12(n-k)+1)(N-n)}\right)\exp\left(-\frac{1}{12N+1}\right)\\
&\leq 1,
\end{aligned} \tag{4.6}$$

where the final inequality follows since $k\in[\lceil n\mu\rceil,\dots,n-1]$ and $n\leq D\leq\lfloor N/2\rfloor$ which implies that each exponential argument preceding the inequality is negative. This gives a bound of 1 on the product. As the A term in (4.3) is already in the claimed form, combining (4.5) and (4.6) proves the claim. □

Next, we develop an upper bound for hypergeometric tail probabilities. This bound is similar to that discussed by Feller for the binomial [5], pages 150–151. To our knowledge this result is new.

Lemma 7. *Suppose* $S_{n,D,N}\sim\text{Hypergeometric}(n,D,N)$, $N>4$ *and* $1\leq n,D\leq N-1$. *For* $k>(nD)/N$, *we have*

$$P(S_{n,D,N}\geq k)\leq P(S_{n,D,N}=k)\left(\frac{k(N-D-n+k)}{Nk-nD}\right). \tag{4.7}$$

Proof. Suppose first that $n\leq D$ and $k=n$. Then (4.7) becomes

$$\begin{aligned}
P(S_{n,D,N}\geq n)&\leq P(S_{n,D,N}=n)\left(\frac{n(N-D-n+n)}{Nn-nD}\right)\\
&=P(S_{n,D,N}=n)\left(\frac{n(N-D)}{n(N-D)}\right)=P(S_{n,D,N}=n).
\end{aligned}$$

Since $P(S_{n,D,N} \geq n) = P(S_{n,D,N} = n)$, the result holds in this case. Next, suppose $D < n$ and $k = D$. Then (4.7) becomes

$$P(S_{n,D,N} \geq D) \leq P(S_{n,D,N} = D)\left(\frac{D(N-D-n+D)}{ND-nD}\right)$$
$$= P(S_{n,D,N} = D)\left(\frac{D(N-n)}{D(N-n)}\right) = P(S_{n,D,N} = D).$$

Since $P(S_{n,D,N} \geq D) = P(S_{n,D,N} = D)$, the result holds in this case too.

If (n, D, N) is a population such that $\lfloor (nD)/N \rfloor + 1 = n \wedge D$, we are done. Supposing this is not the case, let $\lfloor (nD)/N \rfloor + 1 \leq (j-1) < j \leq n \wedge D$. Assume the result holds when $k = j$. We will show this implies the result holds for $k = j-1$. We have

$$\begin{aligned}
&P(S_{n,D,N} \geq j-1) \\
&\quad = P(S_{n,D,N} = j-1) + P(S_{n,D,N} \geq j) \\
&\quad \leq P(S_{n,D,N} = j-1) + P(S_{n,D,N} = j)\left[\frac{j(N-D-n+j)}{Nj-nD}\right] \qquad \text{(by induction hypothesis)} \\
&\quad = P(S_{n,D,N} = j-1)\left[1 + \frac{P(S_{n,D,N} = j)}{P(S_{n,D,N} = j-1)}\left[\frac{j(N-D-n+j)}{Nj-nD}\right]\right] \\
&\quad = P(S_{n,D,N} = j-1)\left[1 + \frac{(D-j+1)(n-j+1)}{j(N-D-n+j)}\left[\frac{j(N-D-n+j)}{Nj-nD}\right]\right] \\
&\quad = P(S_{n,D,N} = j-1)\left[1 + \frac{(D-j+1)(n-j+1)}{Nj-nD}\right].
\end{aligned}$$

Under the current assumption, the right-hand side equals

$$P(S_{n,D,N} = j-1)\left[\left(\frac{(j-1)(N-D-n+j-1)}{N(j-1)-nD}\right)\right]$$

so we see it is enough to show

$$\left[\left(\frac{(j-1)(N-D-n+j-1)}{N(j-1)-nD}\right)\right] - \left[1 + \frac{(D-j+1)(n-j+1)}{Nj-nD}\right] \geq 0.$$

Combining terms and simplifying, we find this equivalent to showing

$$\frac{N(D-j+1)(n-j+1)}{(Nj-nD)(N(j-1)-nD)} \geq 0.$$

Since we assume $\lfloor (nD)/N \rfloor + 1 \leq (j-1) < j \leq n \wedge D$, we see that each term in parentheses in the fraction is non-negative. In particular, since $j \geq \lfloor (nD)/N \rfloor + 2 > (nD)/N + 1$, we have

$$N(j-1) - nD > N\big((nD)/N\big) - nD = 0.$$

Thus, the expression is non-negative. This implies the claim. □

We next prove a technical lemma.

Lemma 8. *Fix $N > 4$. Suppose that $n < D \leq \lfloor N/2 \rfloor$ and that $\gamma := (N-n)/n$. For all triples*

$$(\mu, u, \gamma) \in \left[\frac{n+1}{N}, \frac{1}{2}\right] \times \left(0, \frac{1}{2}\right] \times (1, \infty)$$

we have

$$\frac{\mu(1-\mu)(u+\mu)(\gamma(1-\mu)+u)}{(1-u-\mu)(\gamma\mu-u)} \leq \frac{1}{4}\frac{(u+(1/2))(\gamma(1-(1/2))+u)}{(1-u-(1/2))(\gamma(1/2)-u)}. \tag{4.8}$$

We pause to outline the strategy used to prove this statement, since the proof requires a rather detailed algebraic argument. We break the quantity into two functions, f and g, the second of which, g, is parabolic on $\mu \in [(n+1)/N, 1/2]$. We demonstrate that f is maximized at $\mu = 1/2$. We do this by obtaining the only root which falls in the interval, determining that it yields a local minimum, and finally showing the function is larger at the upper boundary of $\mu = 1/2$.

We then show that g has a local maximum in the interior of the interval (for $0 < u < 1/2$). Using the quadratic g function as a scaling function, we then define an upper envelope to the function of interest in terms of f, along with a second function that agrees with the function of interest at $\mu = 1/2$. By defining the two new functions in terms of f (scaled by positive numbers, which are obtained at fixed-points of g), we are still able to claim these functions are maximized at $\mu = 1/2$.

We then demonstrate the function of interest increases monotonically between the value of μ where it intersects its envelope and $\mu = 1/2$. We finally show that at the right endpoint of $\mu = 1/2$, the quantity of interest exceeds its envelope at the left end-point. This will prove the claim; the details now follow.

Proof of Lemma 8. With the previous comments in mind, define the following functions:

$$\begin{aligned} f(\mu) &:= \frac{\mu(1-\mu)}{(1-u-\mu)(\gamma\mu-u)} \quad \text{and} \\ g(\mu) &:= (u+\mu)\big(\gamma(1-\mu)+u\big). \end{aligned} \tag{4.9}$$

Note that the product $f(\mu)g(\mu)$ gives the quantity on the left-hand side of (4.8). We first analyze $f(\mu)$. Taking its derivative, we find

$$f'(\mu) = \frac{u((\gamma-1)\mu^2 + 2\mu(1-u) - (1-u))}{(1-u-\mu)^2(u-\gamma\mu)^2}.$$

Seeking critical points, we find $f'(\mu)$ has the following roots:

$$\frac{\pm\sqrt{(1-u)(\gamma-u)} + u - 1}{\gamma - 1}.$$

Since $\mu \in (0, 1/2)$, only the positive root is of potential interest. Since $\gamma > 1$ under the current restrictions, we have

$$\frac{\sqrt{(1-u)(\gamma-u)}+u-1}{\gamma-1} \geq \frac{\sqrt{(1-u)^2}+u-1}{\gamma-1} = 0.$$

Additionally, we can see

$$\frac{\sqrt{(1-u)(\gamma-u)}+u-1}{\gamma-1} \leq \frac{1}{2}$$

since, after algebra, it is equivalent to showing

$$0 \leq \frac{(\gamma-1)^2}{4}$$

which follows under the assumptions. A similar argument shows that the corresponding root with the negative radical is always negative, and therefore does not affect the current investigation. Next, differentiate again and evaluate the second derivative at the root. We then find

$$\begin{aligned}
&f''(\mu)\Big|_{(\frac{\sqrt{(1-u)(\gamma-u)}+u-1}{\gamma-1})} \\
&\quad = \frac{[2(\gamma-1)^4(1-u)u(\gamma-u)][(\gamma^2+1)u+2\gamma\sqrt{(1-u)(\gamma-u)}-\gamma^2-\gamma]}{[\sqrt{(1-u)(\gamma-u)}-\gamma(1-u)]^3[\gamma(\sqrt{(1-u)(\gamma-u)}-1)+u]^3} \\
&\quad =: \frac{[a(u,\gamma)][b(u,\gamma)]}{[c(u,\gamma)]^3[d(u,\gamma)]^3}.
\end{aligned}$$

We next show that this quantity is positive for any $(u, \gamma) \in (0, 1/2) \times (1, \infty)$. It is clear that $a(u, \gamma)$ is always positive under the current assumption, since each term in the product is positive.

We next claim $b(u, \gamma) < 0$ for all $(u, \gamma) \in (0, 1/2) \times (1, \infty)$. This claim is equivalent to showing

$$2\gamma\sqrt{(1-u)(\gamma-u)} < \gamma^2(1-u) + (\gamma-u).$$

Since both sides are positive, we square both sides and simplify to find that the claim is equivalent to showing

$$0 < (\gamma-1)^2(-\gamma+\gamma u+u)^2.$$

As this last claim follows for any admissible pair, we conclude that $b(u, \gamma) < 0$ for all $(u, \gamma) \in (0, 1/2) \times (1, \infty)$.

We next show that $c(u, \gamma) < 0$ for all $(u, \gamma) \in (0, 1/2) \times (1, \infty)$. This claim is equivalent to

$$(1-u)(\gamma-u) < \gamma^2(1-u)^2$$

which, after expanding and re-arranging, is equivalent to the claim

$$0 < (\gamma-1)(1-u)(\gamma-\gamma u-u)$$

for all $(u, \gamma) \in (0, 1/2) \times (1, \infty)$. On this set, it is clear $\gamma - 1$ and $1 - u$ are positive for any admissible pair. Hence, we need only show $(\gamma - \gamma u - u) > 0$ on this set. But this is equivalent to claiming $\gamma(1 - u) > u$ for any pair on this set, which is true because $\gamma > 1$ and $u < 1/2$. Thus we conclude $c(u, \gamma) < 0$ for all $(u, \gamma) \in (0, 1/2) \times (1, \infty)$.

We finish this sub-argument by showing $d(u, \gamma) > 0$ for $(u, \gamma) \in (0, 1/2) \times (1, \infty)$. This claim is equivalent to

$$\gamma\sqrt{(1-u)(\gamma-u)} > \gamma - u$$

for all admissible pairs. Since both sides are positive, we square and simplify to find the claim equivalent to

$$p(u) := \gamma^2 - \gamma^2 u - \gamma + u > 0.$$

Viewing the left-hand side as a function of u, we differentiate to see that $p'(u) = 1 - \gamma^2 < 0$ for any choice of $\gamma > 1$. So, $p(u)$ decreases in u for any $\gamma > 1$. Hence

$$p(u) > \gamma^2 - \frac{\gamma^2}{2} - \gamma + \frac{1}{2} = \frac{\gamma^2 - 2\gamma + 1}{2} = \frac{(\gamma-1)^2}{2} > 0.$$

Thus, we conclude $d(u, \gamma) > 0$ for $(u, \gamma) \in (0, 1/2) \times (1, \infty)$.

To summarize: we have shown that for all $(u, \gamma) \in (0, 1/2) \times (1, \infty)$, $a(u, \gamma) > 0$, $b(u, \gamma) < 0$, $c(u, \gamma) < 0$ and $d(u, \gamma) > 0$. This means that

$$f''(\mu)\Big|_{(\frac{\sqrt{(1-u)(\gamma-u)}+u-1}{\gamma-1})} = \frac{[a(u,\gamma)][b(u,\gamma)]}{[c(u,\gamma)]^3[d(u,\gamma)]^3} > 0.$$

Therefore we have found a local minimum of $f(\mu)$ that falls in $[(n + 1)/N, 1/2]$. Therefore, the maximum must be achieved at one of the endpoints.

We next show that the maximum is in fact achieved at $\mu = 1/2$. To do this, we compare the difference. Plugging in the definition $\gamma = (N - n)/n$, and simplifying, we find:

$$f\left(\frac{1}{2}\right) - f\left(\frac{n+1}{N}\right) = \frac{nu(N-2n-2)(nu(N-2n-2)+N)}{(1-2u)(N(1-u)-n-1)(N-2nu-n)((n+1)(N-n)-nNu)}.$$

Each term in this expression is positive for all $u \in (0, 1/2)$ and hence the entire expression is positive. To see this, first observe that the restriction $n < D \le \lfloor N/2 \rfloor$ means that the maximum value n can attain is $\lfloor N/2 \rfloor - 1$. This implies $N - 2n - 2 \ge 0$. Since we also restrict $u \in (0, 1/2)$, we also have $(N(1 - u) - n - 1) \ge 0$ and $(N - 2nu - n) \ge 0$. Finally note that

$$(n+1)(N-n) - nNu \ge (n+1)(N-n) - \frac{nN}{2} = \frac{n(N-2n-2)}{2} + N \ge 0.$$

We conclude that $f(\mu)$ is maximized at $\mu = 1/2$ over all choice of $(u, \gamma) \in (0, 1/2) \times (1, \infty)$.

We next consider the function $g(\mu)$, defined in (4.9). We write it again, its first two derivatives, and its critical point μ^* for subsequent discussion. As this function is much simpler than $f(\mu)$, we present these quantities without comment.

$$\begin{aligned} g(\mu) &= (u+\mu)\big(\gamma(1-\mu)+u\big), \\ g'(\mu) &= -2\gamma\mu+\gamma-\gamma u+u, \\ g''(\mu) &= -2\gamma \quad \text{and} \\ \mu^* &= \frac{\gamma(1-u)+u}{2\gamma}. \end{aligned}$$

Since $g''(\mu) < 0$ for any choice of $(u, \gamma) \in (0, 1/2) \times (1, \infty)$, we see that μ^* is a local maximum. For any $\gamma > 1$, we also see the critical point decreases for $u \in (0, 1/2)$, from a value of $1/2$ at $u = 0$ to a value of $(1/4) + (1/(4\gamma))$. As $\gamma \nearrow \infty$, this approaches $1/4$ asymptotically. Hence for any $(u, \gamma) \in (0, 1/2) \times (1, \infty)$, the maximum of the function is attained for $\mu \in (0, 1/2)$. Since we are ultimately interested in understanding the product $f(\mu)g(\mu)$, we next show that the maximum of g occurs at a value greater than the local minimum of f. We do this by comparing their difference to zero. The claim

$$\left[\frac{\gamma(1-u)+u}{2\gamma}\right] - \left[\frac{\sqrt{(1-u)(\gamma-u)}+u-1}{\gamma-1}\right] > 0$$

is equivalent to the claim

$$(\gamma-1)\big(\gamma(1-u)+u\big)+2\gamma(1-u) > 2\gamma\sqrt{(1-u)(\gamma-u)}.$$

Both sides of this inequality are positive. So, we square them and simplify to find that the claim is equivalent to the claim

$$(\gamma-1)^2\big(\gamma(1-u)-u\big)^2 > 0.$$

The claim follows by the final form, since the square each quantity positive. We now define three related functions.

$$\begin{aligned} ue(\mu) &:= g\left(\frac{\gamma(1-u)+u}{2\gamma}\right)f(\mu) = \frac{(1-\mu)\mu(\gamma+\gamma u+u)^2}{4\gamma(1-\mu-u)(\gamma\mu-u)}, \\ t(\mu) &:= g(\mu)f(\mu) = \frac{\mu(1-\mu)(u+\mu)(\gamma(1-\mu)+u)}{(1-u-\mu)(\gamma\mu-u)} \quad \text{and} \\ ep(\mu) &:= g(1/2)f(\mu) = \frac{(1-\mu)\mu(1+2u)(\gamma+2u)}{4(1-\mu-u)(\gamma\mu-u)}. \end{aligned}$$

First notice that $t(\mu)$ is the quantity of interest, which we wish to show is maximized at $\mu = 1/2$. As defined, the function $ue(\mu)$ is an upper envelope of $t(\mu)$, with agreement at $\mu = (\gamma(1-u)+u)/(2\gamma)$. $ep(\mu)$ is defined so that $ep(1/2) = t(1/2)$, that is ep agrees with t at the end-point of the μ-interval. Consider the behavior of $t(\mu)$ on $\mu \in [(\gamma(1-u)+u)/(2\gamma), 1/2]$. We have

$$t'(\mu) = 1-2\mu+\frac{(\gamma+1)u^2(\gamma\mu^2-\mu^2+2\mu-2\mu u+u-1)}{(1-\mu-u)^2(\gamma\mu-u)^2}. \tag{4.10}$$

Since $\mu \le 1/2$, the sign of $t'(\mu)$ may be determined by the behavior of the third term in the numerator. We consider its behavior separately. Let

$$\begin{aligned} a(\mu) &:= \gamma\mu^2 - \mu^2 + 2\mu - 2\mu u + u - 1, \\ a'(\mu) &= 2\bigl(1 - u + \mu(\gamma - 1)\bigr) > 0 \quad \text{and} \\ a\left(\frac{\gamma(1-u)+u}{2\gamma}\right) &= \frac{(\gamma-1)(u-\gamma(1-u))^2}{4\gamma^2} > 0. \end{aligned}$$

We see then that $a(\mu)$ will be non-negative for $\mu \in [(\gamma(1-u)+u)/(2\gamma), 1/2]$. Therefore, $t'(\mu) > 0$ on the same interval. Hence, $t(\mu)$ is increasing on the same interval. Finally, consider the difference

$$\begin{aligned} &ep(1/2) - ue\left(\frac{n+1}{N}\right) \\ &\quad = \frac{Nu\left(\begin{array}{c} 2u^2(N-2n)(N-2n-1)(2n(N-n-1)+N) \\ +Nu(N-n)(N-2n-3) + (N-n)^2(N-2n-2) - 4nNu^3(N-2n-1) \end{array}\right)}{4(1-2u)(N-n)(N-n-2nu)(N(1-u)-n-1)(N-n+nN(1-u)-n^2)}, \end{aligned} \tag{4.11}$$

where we have again substituted the definition $\gamma = (N-n)/n$. We will now argue that this quantity is positive for all $n \in \{1, \dots, \lfloor N/2 \rfloor - 2\}$. This is sufficient to demonstrate $t(\mu)$ is maximized at $\mu = 1/2$, since we are supposing $n < D \le \lfloor N/2 \rfloor$. This restriction is necessary to handle the sign-change implicit in the term $(N - 2n - 3)$. There, for $n = \lfloor N/2 \rfloor - 2$ it equals (for integer values of $N/2$) 1, while it flips signs for $N/2 - 1$. However, this sign-change is not problematic since our assumptions imply at $n = N/2 - 1$ that $D = N/2$, which is the value we are trying to demonstrate maximizes $t(\mu)$.

We will demonstrate positivity by analyzing the terms in the expression. For simplicity, we will assume $N/2$ is an integer, though the same analysis will hold for odd values of N. We will consider some of the denominator terms first. We have, using the assumptions,

$$(N-n) + \bigl(nN(1-u) - n^2\bigr) \ge \frac{n(N-2n-2)}{2} + N > 0.$$

We also have

$$N - n - 2nu \ge N - 2n \ge N - N + 4 > 0.$$

So we see all terms in the denominator are positive for any choice of (u, n). Hence, it is enough to show that under our assumptions

$$\begin{aligned} z(u) := {}& 2u^2(N-2n)(N-2n-1)\bigl(2n(N-n-1)+N\bigr) \\ &+ Nu(N-n)(N-2n-3) - 4nNu^3(N-2n-1) \ge 0. \end{aligned}$$

First viewing the left-hand-side as a function of u, we observe the following computations:

$$\begin{aligned} z'(u) = {}& 4u(N-2n)(N-2n-1)\bigl(2n(N-n-1)+N\bigr) \\ &+ N(N-n)(N-2n-3) - 12nNu^2(N-2n-1), \end{aligned}$$

$$z''(u) = 4(N-2n)(N-2n-1)\big(2n(N-n-1)+N\big) - 24nNu(N-2n-1),$$
$$z'''(u) = -24nN(N-2n-1) \le 0.$$

From the third derivative, we see $z''(u)$ is decreasing in u. Since $z''(0) = 4(N-2n-1)(N+2n(N-n-1))(N-2n) > 0$, we calculate the value of the second derivative at $u = 1/2$ to find

$$z''(u)|_{u=1/2} = 4(N-2n-1)\big(4n^3+4n^2+2nN^2+N^2-6n^2N-7nN\big) =: 4(N-2n-1)\phi(n),$$

where we define the function $\phi(n)$ in-line. We analyze the sign of $\phi(n)$ for $n \in \{1,\dots,\lfloor N/2\rfloor - 2\}$. Treating n as continuous temporarily, we differentiate twice to find

$$\phi''(n) := 24n - 12N + 8.$$

Since we assume $n \in \{1,\dots,\lfloor N/2\rfloor - 2\}$, we see

$$\phi''(n) = 24n - 12N + 8 \le 24\left(\frac{N}{2} - 2\right) - 12N + 8 = -40 \le 0.$$

This implies $\phi(n)$ is concave in n. Evaluating at the admissible endpoints, we find

$$\phi(1) = 8 + N(3N-13) \quad \text{and}$$
$$\phi\big((N/2)-2\big) = \frac{N^2+12N-32}{2}.$$

For $N \ge 4$, both of these expressions are positive. By concavity we conclude $\phi(n) \ge 0$. Therefore, we have that

$$z''(u)|_{u=1/2} > 0,$$

and so we conclude $z''(u) > 0$ for all $u \in (0, 1/2]$. But since

$$z'(u)|_{u=0} = N(N-n)(N-2n-3) > 0,$$

we infer that $z'(u) > 0$ for all $u \in (0, 1/2]$. Finally, since $z(0) = 0$, we conclude that $z(u) > 0$ for all $u \in (0, 1/2]$. But this implies that

$$ep(1/2) - ue\left(\frac{n+1}{N}\right) > 0. \tag{4.12}$$

Therefore, we can define the following function

$$\text{maj}(\mu) := \begin{cases} ue(\mu), & \text{if } \mu \in \left[\frac{n+1}{N}, \frac{\gamma(1-u)+u}{2\gamma}\right], \\ t(\mu), & \text{if } \mu \in \left(\frac{\gamma(1-u)+u}{2\gamma}, \frac{1}{2}\right]. \end{cases}$$

Observe that for all $\mu \in [(n+1)/N, 1/2]$, we have $\text{maj}(\mu) \ge t(\mu)$. Additionally, we know that $\text{maj}(\mu)$ is maximized at $\mu = 1/2$: the argument following (4.10) shows for μ such that $\text{maj}(\mu) =$

$t(\mu)$, $\text{maj}(\mu)$ strictly increases; the argument following (4.12) shows $\text{maj}(\mu)$ increases to its maximum on the interval. Finally, since we know $\text{maj}(1/2) = ep(1/2) = t(1/2)$, we conclude $t(\mu)$ is maximized at $\mu = 1/2$ for all choice of $(u, \gamma) \in (0, 1/2] \times (1, \infty)$. This completes the proof. □

We are now ready to prove (2.4).

Proof of Theorem 1. Pick $\lambda, k > 0$ such that $k = \sqrt{n}\lambda + n\mu$, $k \geq n(D/N)$. We then have

$$
\begin{aligned}
&P\bigl(\sqrt{n}(\bar{X} - \mu) \geq \lambda\bigr) \\
&\quad = P\left(\sum_{i=1}^{n} X_i \geq k\right) \\
&\quad \leq \left[P\left(\sum_{i=1}^{n} X_i = k\right)\right]\left(\frac{k(N-D-n+k)}{Nk-nD}\right) \qquad \text{by (4.7)} \\
&\quad \leq \left[\frac{1}{\sqrt{2\pi}}\frac{\sqrt{D(N-D)n(N-n)}}{\sqrt{k(D-k)(n-k)(N-D-(n-k))N}}\left(\frac{k(N-D-n+k)}{Nk-nD}\right)\right] \\
&\qquad \times \left[\exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right)\right] \qquad \text{by (4.2)} \\
&\quad = [A]\cdot\left[\exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right)\right].
\end{aligned} \tag{4.13}
$$

Recall that $u := (k/n) - (D/N)$ in the previous bound. Define $f := n/N$, $\bar{f} := 1 - f_N = (N - n)/N$, $\mu := D/N$, and furthermore, define the ratio

$$
\gamma := \frac{\bar{f}}{f} = \frac{N-n}{n}.
$$

We may then write:

$$
D - k = \left(N\frac{D}{N} - n\frac{k}{n}\right) = n\left(\frac{N}{n}\mu - \frac{k}{n}\right) = n\left(\frac{N}{n}\mu - u - \mu\right) = n(\gamma\mu - u).
$$

Similarly, we have

$$
N - n - (D - k) = N - n - n(\gamma\mu - u) = n(\gamma - \gamma\mu + u) = n\bigl(\gamma[1-\mu] + u\bigr).
$$

Using these parametrizations, we may write

$$
\begin{aligned}
[A] &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{D(N-D)n(N-n)}{k(D-k)(n-k)(N-D-(n-k))N}}\left(\frac{k(N-n-(D-k))}{Nn((k/n)-(D/N))}\right) \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{D(N-D)n(N-n)k^2(N-n-(D-k))^2}{k(D-k)(n-k)(N-n-(D-k))N^3n^2}}\left(\frac{1}{u}\right)
\end{aligned}
$$

$$= \sqrt{\frac{(N-n)}{2\pi nNu^2}} \sqrt{\frac{D}{N}\left(1-\frac{D}{N}\right)\frac{(k/n)}{(1-k/n)}\frac{(N-n-(D-k))}{(D-k)}} \tag{4.14}$$

$$= \sqrt{\frac{(N-n)}{2\pi nNu^2}} \sqrt{\mu(1-\mu)\frac{(u+\mu)}{(1-u-\mu)}\frac{\gamma(1-\mu)+u}{\gamma\mu-u}}$$

$$\leq \sqrt{\frac{(N-n)}{2\pi nNu^2}} \sqrt{\frac{1}{4}\frac{(u+(1/2))(\gamma(1-(1/2))+u)}{(1-u-(1/2))(\gamma(1/2)-u)}}$$

with the last inequality following by (4.8) established in Lemma 8. Observe under these parametrizations $u = \lambda/\sqrt{n}$. Hence, if we use (4.14) to provide an upper bound for (4.13), substitute $\lambda/\sqrt{n}$ for u, and then simplify, the claim is proved. □

Some of the machinery developed in the preceding lemmas will be adapted to prove (2.5). The argument follows.

Proof of Theorem 2. Suppose now that $1 \leq n < D \leq N-1$. We consider k such that $0 \vee n + D - N < k \leq n$. The decomposition of a Hypergeometric probability into A, B, and C terms stated in (4.3) still applies. For $k \geq n(D/N)$, the bound on the B term in (4.5) still holds. Thus, we may write

$$\begin{aligned} B &\leq \exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right) \\ &= \exp\left(-\frac{2n}{1-n/N}u^2\right)\exp\left(-\frac{n}{4}u^4\right)\exp\left(-\frac{n}{12}u^4\right)\exp\left(-\left[\frac{n^4}{3(N-n)^3}\right]u^4\right). \end{aligned} \tag{4.15}$$

Also recall we showed that $C \leq 1$ at (4.6) when $n \leq D \leq N/2$. In fact, the expression at (4.6) shows $C \leq 1$ under the current assumptions. When $n \leq N/2$, all exponential arguments may be determined to be negative by inspection. When $n > N/2$, the only fraction whose sign is unclear is

$$\frac{1-12(N-2n+k)}{12(12(n-k)+1)(N-n)}.$$

However, this remains negative under the current assumptions since $n > N/2$ implies $k \geq n + D - N$. Therefore, $N + k \geq n + D$ and so $N + k - 2n \geq D - N \geq 0$. We thus conclude $C \leq 1$. Here though, we provide a new analysis of the A term under the current assumptions.

Case 1. First restrict k so that $\mu_0 < \frac{k}{n} < 1 - \frac{\mu_0}{2}$. We then have

$$\begin{aligned} A &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{D(N-D)n(N-n)}{k(D-k)(n-k)(N-D-(n-k))N}} \\ &= \frac{N}{\sqrt{2\pi n}}\sqrt{\frac{D/N(1-D/N)(1-n/N)}{k/n(D-k)(1-k/n)(N-D-n+k)}} \end{aligned}$$

$$\begin{aligned}
&\le \frac{N}{\sqrt{2\pi n}}\sqrt{\frac{(1/4)(1-\psi_0)}{\mu_0(D-n+n\mu_0/2)(\mu_0/2)(N-D-n+n\mu_0)}} \qquad (4.16)\\
&\le \frac{N}{\sqrt{2\pi n}}\sqrt{\frac{(1/4)(1-\psi_0)}{\mu_0(n\mu_0/2)(\mu_0/2)(n\mu_0)}}\\
&= \frac{1}{(n/N)\sqrt{n}}\sqrt{\frac{2(1/4)(1-\psi_0)}{\pi\mu_0^4}} \le \frac{1}{\sqrt{n}}\frac{\sqrt{(1-\psi_0)}}{\psi_0\sqrt{2\pi\mu_0^4}}.
\end{aligned}$$

Combining (4.16) with (4.15) and (4.6), we have the bound

$$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \le \frac{K_{c1}}{\sqrt{n}}\exp\left(-\frac{2n}{1-n/N}u^2\right)\exp\left(-\frac{n}{4}u^4\right)\exp\left(-\frac{n}{12}u^4\right)\exp\left(-\left[\frac{n^4}{3(N-n)^3}\right]u^4\right),$$

where

$$K_{c1} = \left[\frac{\sqrt{(1-\psi_0)}}{\psi_0\sqrt{2\pi\mu_0^4}}\right].$$

Case 2. Next, suppose that $1-\frac{\mu_0}{2} \le \frac{k}{n} < 1$. This implies that

$$u = \frac{k}{n} - \frac{D}{N} \ge 1 - \frac{\mu_0}{2} - \frac{D}{N} \ge 1 - \frac{\mu_0}{2} - (1-\mu_0) = \frac{\mu_0}{2}.$$

We can bound the A term by

$$\begin{aligned}
A &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{D(N-D)n(N-n)}{k(D-k)(n-k)(N-D-(n-k))N}}\\
&= \frac{N}{\sqrt{2\pi}}\sqrt{\frac{D/N(1-D/N)(1-n/N)}{k/n(D-k)(n-k)(N-D-n+k)}}\\
&\le \frac{N}{\sqrt{2\pi}}\sqrt{\frac{(1/4)(1-\psi_0)}{(1-\mu_0/2)(D-n+1)(n-n+1)(N-D-n+n(1-\mu_0/2))}}\\
&\le \frac{N}{\sqrt{2\pi}}\sqrt{\frac{(1/4)(1-\psi_0)}{(1-\mu_0/2)(n(1-\mu_0/2))}} = \frac{n(N/n)}{\sqrt{n}}\sqrt{\frac{(1/4)(1-\psi_0)}{2\pi(1-\mu_0/2)^2}}\\
&\le \frac{n(1/\psi_0)}{\sqrt{n}}\sqrt{\frac{(1/4)(1-\psi_0)}{2\pi(1-\mu_0/2)^2}} = \frac{n}{\sqrt{n}}\sqrt{\frac{(1/4)(1-\psi_0)}{2\pi\psi_0^2(1-\mu_0/2)^2}}.
\end{aligned}$$

Taking the $\exp(-\frac{n}{12}u^4)$ term from (4.15) we have

$$n\exp\left(-\frac{n}{12}u^4\right) \le n\exp\left(-\frac{n}{12}\left(\frac{\mu_0}{2}\right)^4\right) = n\exp\left(-\frac{\mu_0^4}{192}n\right).$$

This is maximized at

$$n = \frac{192}{\mu_0^4},$$

and so

$$n\exp\left(-\frac{m}{12}u^4\right) \le \frac{192}{\mu_0^4 e}.$$

Combining the remaining terms in (4.5) together with this bound of the A term and the C bound of 1 yields

$$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \le \frac{K_{c2}}{\sqrt{n}}\exp\left(-\frac{2n}{1-n/N}u^2\right)\exp\left(-\frac{n}{4}u^4\right)\exp\left(-\left[\frac{n^4}{3(N-n)^3}\right]u^4\right),$$

where

$$K_{c2} = \sqrt{\frac{(1/4)(1-\psi_0)}{2\pi\psi_0^2(1-\mu_0/2)^2}}\left(\frac{192}{\mu_0^4 e}\right).$$

Case: $k = n$. When $k = n$ there are only two binomial coefficients to consider in the hypergeometric probability. Therefore, we must derive a new bound via Stirling's formula. Doing so yields

$$\begin{aligned}\frac{\binom{D}{n}\binom{N-D}{0}}{\binom{N}{n}} &= \frac{D!(N-n)!}{(D-n)!N!} \\ &\le \sqrt{\frac{D(N-n)}{(D-n)N}}\frac{D^D(N-n)^{(N-n)}}{(D-n)^{(D-n)}N^N} \\ &\quad \times \exp\left(\frac{1}{12D}+\frac{1}{12(N-n)}-\frac{1}{12(D-n)+1}-\frac{1}{12N+1}\right) \\ &=: A'B'C'.\end{aligned}$$

We can bound C' by

$$\begin{aligned}C' &= \exp\left(\frac{12(N-D)+1}{(12D)(12N+1)}-\frac{12(N-D)+1}{(12(N-n))(12(D-n)+1)}\right) \\ &= \exp\left(\frac{[12(N-D)+1]([(12(N-n))(12(D-n)+1)]-[(12D)(12N+1)])}{[(12D)(12N+1)][(12(N-n))(12(D-n)+1)]}\right) \\ &\le 1\end{aligned}$$

with the final bound following since $(N - D) > 0$ and $[(12(N - n))(12(D - n) + 1)] < [(12D)(12N + 1)]$. Continuing with B' we have

$$B' = \frac{D^D (N-n)^{(N-n)}}{(D-n)^{(D-n)} N^N} = \frac{(D/N)^D ((N-n)/N)^{(N-D)}}{((D-n)/(N-n))^{D-n}}$$
$$= \left(\frac{(N-n)/N}{(D-n)/D}\right)^{D-n} \left(\frac{(N-n)/N}{(N-n-(D-n))/(N-D)}\right)^{(N-D)} \left(\frac{D}{N}\right)^n$$
$$= \exp\bigl(-\Gamma + n\log(\mu)\bigr),$$

where, as before, we have

$$\Gamma = (N-n)\left[\left(\frac{D-n}{N-n}\right)\log\left(\frac{(D-n)/D}{(N-n)/N}\right) + \left(1 - \frac{D-n}{N-n}\right)\log\left(\frac{[N-n-(D-n)]/(N-D)}{(N-n)/N}\right)\right].$$

Using the previous analysis, we can write

$$B' = \exp\left(-(N-n)\Psi\left(\frac{f}{\bar{f}}u, 1-\mu\right) + n\log(\mu)\right) = \exp\bigl(-(N-n)\Psi(\gamma, u) + n\log(1-u)\bigr),$$

where we define $\gamma := \frac{f}{\bar{f}}u$, $f := f_N = \frac{n}{N}$ and $\bar{f} := \bar{f}_N = 1 - f_N = \frac{N-n}{N}$ and use the equality $u = 1 - \mu$ under the current hypothesis. Using the analysis from van der Vaart and Wellner, page 461, re-parametrized to the situation at hand, we obtain

$$\Psi(\gamma, u) \geq 2\gamma^2 + \gamma^4/3.$$

We also have the bound via the Taylor expansion:

$$\log(1-u) = -\left[\sum_{k=1}^{\infty}\frac{u^k}{k}\right] \leq -\left[\sum_{k=1}^{7}\frac{u^k}{k}\right].$$

Hence,

$$B' \leq \exp\left(-(N-n)\left[2\left(\frac{f}{\bar{f}}u\right)^2 + \frac{((f/\bar{f})u)^4}{3}\right] + n\log(1-u)\right)$$
$$= \exp\left(-2\left(\frac{n^2}{N-n}\right)u^2 - \frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4 + n\log(1-u)\right)$$
$$\leq \exp\left(-2\left(\frac{n^2}{N-n}\right)u^2 - \frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4 - n\left[\sum_{k=1}^{7}\frac{u^k}{k}\right]\right)$$

$$= \exp\left(-\left(\frac{2nN}{N-n}\right)u^2 - \frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4 - nu + \frac{3nu^2}{2} - \frac{nu^3}{3} - \frac{nu^6}{6} - \frac{nu^7}{7}\right)$$
$$\times \exp\left(-\frac{nu^4}{4}\right)\exp\left(-\frac{nu^5}{5}\right)$$
$$\leq \exp\left(-\frac{2n}{1-m/N}u^2\right)\exp\left(-\frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4\right)\exp\left(-\frac{nu^4}{4}\right)\exp\left(-\frac{nu^5}{5}\right),$$

where the last inequality follows since for $x > 0$

$$x + \frac{x^3}{3} + \frac{x^6}{6} + \frac{x^7}{7} - \frac{3}{2}x^2 > 0.$$

For $x \geq 0$ this polynomial has a global minimum at 0 and local minimum at $x \approx 0.851662$ with a value of approximately 0.0796078. Finally, we have

$$A' = \sqrt{\frac{D(N-n)}{(D-n)N}} = \frac{n}{\sqrt{n}}\sqrt{\frac{(D/N)(1-n/N)}{(D-n)n/N}} \leq \frac{n}{\sqrt{n}}\sqrt{\frac{(1-\mu_0)(1-\psi_0)}{\psi_0}},$$

where the final inequality uses the fact that $D - n \geq 1$ in this case. Taking the expression $\exp(-\frac{nu^5}{5})$ from the bound on B', and observing $u = 1 - \frac{D}{N} \geq \mu_0$ we have

$$n\exp\left(-\frac{nu^5}{5}\right) \leq n\exp\left(-\frac{\mu_0^5}{5}n\right) \leq \frac{5}{\mu_0^5 e}$$

since $xe^{-x} \leq e^{-1}$ for $x > 0$. Combining the bounds on A', B', and C', we have shown

$$\frac{\binom{D}{n}\binom{N-D}{0}}{\binom{N}{n}} \leq \frac{K_{c3}}{\sqrt{n}}\exp\left(-\frac{2n}{1-n/N}u^2\right)\exp\left(-\frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4\right)\exp\left(-\frac{nu^4}{4}\right),$$

where

$$K_{c3} = \sqrt{\frac{(1-\mu_0)(1-\psi_0)}{\psi_0}}\left(\frac{5}{\mu_0^5 e}\right).$$

Hence if we set $K_1 = \max(K_{c1}, K_{c2}, K_{c3})$, we have the bound

$$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \leq \frac{K_1}{\sqrt{n}}\exp\left(-\frac{2n}{1-n/N}u^2\right)\exp\left(-\frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4\right)\exp\left(-\frac{nu^4}{4}\right). \quad (4.17)$$

Plugging in the definitions $k = \sqrt{n}\lambda + n\mu$ and $u = k/n - \mu$

$$P\left(\sum_{i=1}^{n} X_i = k\right) = P\big(\sqrt{n}(\bar{X} - \mu) = \lambda\big)$$
$$\leq \frac{K_1}{\sqrt{n}}\exp\left(-\frac{2\lambda^2}{1-n/N}\right)\exp\left(-\frac{1}{3}\left(\frac{n}{N-n}\right)^3\frac{\lambda^4}{n}\right)\exp\left(-\frac{\lambda^4}{4n}\right).$$

This gives inequality (i). To obtain inequality (ii), define, for any n, N pair subject to our conditions,

$$h(x) = \left(\frac{2}{1-n/N}\right)x^2 + \left(\frac{1}{3}\left(\frac{n}{N-n}\right)^3 + \frac{1}{4}\right)x^4 =: ax^2 + bx^4$$

with $a, b > 0$ since $N > n$. Hence h is convex. Therefore, as in the Talagrand argument, we also have $h(x) \geq h(u) - (x-u)h'(u)$ for all x. Also for $0 \leq x \leq 1$ we see $h'(x) = 2ax + 4bx^3$ has linear envelopes

$$2ax \leq h'(x) \leq (2a + 4b)x.$$

Let $0 < t < \lambda \leq \sqrt{n}$. Let $k_0 = \lceil n\frac{D}{N} + \sqrt{n}t\rceil = \lceil n\mu + \sqrt{n}t\rceil$. Using the bound at (4.17), we have

$$\begin{aligned}
\sum_{k\geq k_0} \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} &\leq \sum_{k\geq k_0} \frac{K_1}{\sqrt{n}} \exp\left(-nh(u) - n\left[\frac{k}{n} - \frac{D}{N} - u\right]h'(u)\right) \\
&= \frac{K_1}{\sqrt{n}} \exp\bigl(-nh(u)\bigr) \sum_{k\geq k_0} \exp\bigl(\bigl[nu - (k - n\mu)\bigr]h'(u)\bigr) \\
&\leq \frac{K_1}{\sqrt{n}} \exp\bigl(-nh(u)\bigr)\left[\frac{\exp([nu - (k_0 - n\mu)]h'(u))}{1 - \exp(-h'(u))}\right] \\
&\leq \frac{K_1}{\sqrt{n}} \exp\bigl(-nh(u)\bigr)\left[\frac{K_{ab}}{h'(u)} \exp\bigl(\bigl[nu - (k_0 - n\mu)\bigr]h'(u)\bigr)\right] \\
&\leq \frac{K_1}{\sqrt{n}} \exp\bigl(-nh(u)\bigr)\left[\frac{K_{ab}}{2au} \exp\bigl([nu - \sqrt{n}t][2a + 4b]u\bigr)\right] \\
&= \frac{K_2}{\sqrt{n}u} \exp\bigl(-nh(u)\bigr)\left[\exp\left(nu\left(u - \frac{t}{\sqrt{n}}\right)[2a + 4b]\right)\right],
\end{aligned}$$

where K_{ab} is a constant that depends on a and b, and hence n and N (which we further explain below), and

$$K_2 = \frac{K_1 K_{ab}}{2}.$$

We determine K_{ab} by observing $1 - e^{-v} \geq v/M$ for $0 \leq v \leq v_0$ where $M = M_{v_0} = v_0/(1 - e^{-v_0})$ together with

$$\begin{aligned}
h'(u) &\leq (2a + 4b)u \leq 2a + 4b \qquad \text{(since } u \leq 1\text{)} \\
&= \frac{4}{1 - n/N} + \left(1 + \frac{4}{3}\left(\frac{n/N}{1 - n/N}\right)^3\right) \equiv v_N \\
&\leq \frac{4}{\psi_0} + \frac{4}{3}\frac{(1 - \psi_0)^2}{\psi_0^3} \equiv v_0.
\end{aligned}$$

Therefore $K_{a,b}$ can be taken to be $M = v_0/(1 - e^{-v_0})$ or $M_N = v_N/(1 - e^{-v_N})$ depending on how much dependence on n and N we leave in the bounds. Again by definition we have that

$$2a + 4b = \left(\frac{4}{1 - n/N}\right) + \left(1 + \frac{4}{3}\left(\frac{n}{N-n}\right)^3\right).$$

Therefore we have for all $0 < t < \lambda$

$$\begin{aligned}
&P\big(\sqrt{n}(\bar{X}_n - \mu) \geq t\big) \\
&= \sum_{k \geq k_0} \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \\
&\leq \frac{K_2}{\sqrt{nu}} \exp\big(-nh(u)\big) \exp\left(nu\left(u - \frac{t}{\sqrt{n}}\right)\left[\left(\frac{4}{1 - n/N}\right) + \left(1 + \frac{4}{3}\left(\frac{n}{N-n}\right)^3\right)\right]\right) \\
&= \frac{K_2}{\lambda} \exp\left(-nh\left(\frac{\lambda}{\sqrt{n}}\right)\right) \exp\left(\lambda(\lambda - t)\left[\left(\frac{4}{1 - n/N}\right) + \left(1 + \frac{4}{3}\left(\frac{n}{N-n}\right)^3\right)\right]\right)
\end{aligned}$$

which gives inequality (ii). Inequality (iii) is obtained by setting $t = \lambda$. This completes the proof. □

Acknowledgements

The second author owes thanks to Werner Ehm for several helpful conversations and to Martin Wells for pointing out the Pitman reference. We also thank the referee and the Editors for a number of constructive queries and suggestions, for pointing out the recent paper [6], and for encouraging the inclusion of several graphical comparisons of the bounds.

E. Greene was supported in part by NSF Grant DMS-11-04832. J.A. Wellner was supported in part by NSF Grant DMS-11-04832 and NI-AID Grant 2R01 AI291968-04.

References

[1] Bardenet, R. and Maillard, O.-A. (2015). Concentration inequalities for sampling without replacement. *Bernoulli* **21** 1361–1385. MR3352047

[2] Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[3] Chatterjee, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072

[4] Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16. MR1093412

[5] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, 3rd ed. New York: Wiley. MR0228020

[6] Goldstein, L. and Işlak, Ü. (2014). Concentration inequalities via zero bias couplings. *Statist. Probab. Lett.* **86** 17–23. MR3162712
[7] Greene, E. and Wellner, J.A. (2015). Finite sampling inequalities: An application to two-sample Kolmogorov–Smirnov statistics. Preprint. Available at arXiv:1502.00342.
[8] Greene, E. and Wellner, J.A. (2015). Exponential bounds for the hypergeometric distribution. Preprint. Available at arXiv:1507.08298.
[9] Hájek, J., Šidák, Z. and Sen, P.K. (1999). *Theory of Rank Tests*, 2nd ed. *Probability and Mathematical Statistics*. San Diego, CA: Academic Press. MR1680991
[10] Hoeffding, W. (1951). A combinatorial central limit theorem. *Ann. Math. Stat.* **22** 558–566. MR0044058
[11] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363
[12] Holmes, S. (2004). Stein's method for birth and death chains. In *Stein's Method: Expository Lectures and Applications. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **46** 45–67. Beachwood, OH: IMS. MR2118602
[13] Hush, D. and Scovel, C. (2005). Concentration of the hypergeometric distribution. *Statist. Probab. Lett.* **75** 127–132. MR2206293
[14] Karlin, S. (1974). Inequalities for symmetric sampling plans. I. *Ann. Statist.* **2** 1065–1094. MR0373085
[15] Kemperman, J.H.B. (1973). Moment problems for sampling without replacement. II. *Nederl. Akad. Wetensch. Proc. Ser. A* **76** = *Indag. Math.* **35** 165–180. MR0345260
[16] León, C.A. and Perron, F. (2003). Extremal properties of sums of Bernoulli random variables. *Statist. Probab. Lett.* **62** 345–354. MR1973309
[17] Marshall, A.W., Olkin, I. and Arnold, B.C. (2011). *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2759813
[18] Pitman, J. (1997). Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Combin. Theory Ser. A* **77** 279–303. MR1429082
[19] Robbins, H. (1955). A remark on Stirling's formula. *Amer. Math. Monthly* **62** 26–29. MR0069328
[20] Serfling, R.J. (1974). Probability inequalities for the sum in sampling without replacement. *Ann. Statist.* **2** 39–48. MR0420967
[21] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838963
[22] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76. MR1258865
[23] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. New York: Springer. MR1385671
[24] Vapnik, V.N. (1998). *Statistical Learning Theory. Adaptive and Learning Systems for Signal Processing, Communications, and Control*. New York: Wiley. MR1641250
[25] Vatutin, V.A. and Mikhaĭlov, V.G. (1982). Limit theorems for the number of empty cells in an equiprobable scheme for the distribution of particles by groups. *Teor. Veroyatnost. i Primenen.* **27** 684–692. MR0681461

Received July 2015 and revised December 2015