Finite Sampling Exponential Bounds

Evan Greene

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Jon A. Wellner, Chair

Mathias Drton

Thomas Richardson

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Finite Sampling Exponential Bounds

Evan Greene

Chair of the Supervisory Committee:
Professor Jon A. Wellner
Department of Statistics

This dissertation develops new exponential bounds for the tail of the hypergeometric distribution. It is organized as follows.

- In Chapter 1, it reviews existing exponential bounds used to control the hypergeometric tail.
- In Chapter 2, it extends several bounds used to control the binomial tail to the hypergeometric case.
- In Chapter 3, it describes a basic method to obtain upper bounds for the tail of discrete distributions.
- In Chapters 3 and 4, it applies this method to the Poisson tail and the hypergeometric tail.
- In Chapter 5, it proves an improvement to Serfling's inequality in the case of the hypergeometric distribution under constraints on the population proportion and sampling fraction.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First, I give sincere thanks to my advisor Jon Wellner for his time, support, and guidance. Several years ago he raised the question which led to this dissertation: I am grateful for his sustained interest, his enthusiasm and insights, and his encouragement throughout the research process. I also wish to thank my committee members Mathias Drton and Thomas Richardson as well as my GSR Professor Krzysztof Burdzy. Their interest in the underlying problem, questions, and comments throughout the development of this dissertation has greatly improved the quality of the work. Additionally, I give general thanks to the Department of Statistics at the University of Washington: the faculty, graduate students, and staff have created an environment in which it is possible to research many aspects of statistical theory, and I am happy to have been a part of it. Finally, I thank my friends and family for their constant support over the past five years (and, of course, the lifetime before that).

DEDICATION

To Erin and our daughter: with gratitude, anticipation, and love.

Chapter 1

INTRODUCTION

> Just as deduction should be supplemented by intuition, so the impulse to progressive generalization must be tempered and balanced by respect and love for colorful detail. The individual problem should not be degraded to the rank of special illustration of lofty general theories.
>
> Richard Courant [47]

This dissertation develops new exponential bounds for the tail of the hypergeometric distribution. The hypergeometric distribution arises as a special case in the problem of sampling without replacement from a finite, bounded population. This problem occurs often in statistical applications. For example, simple linear rank statistics such as the two-sample Wilcoxon can be understood (under the null hypothesis) as a sample without replacement from the pooled ranks [45, 21, 19].

The hypergeometric distribution occurs in many other contexts as well. In developing statistical learning theory, Vapnik needed bounds on hypergeometric tail behavior [53]. The hypergeometric tail also appears in computer science applications, such as analyzing the behavior of distributed protocols [1] and establishing the efficiency of order-preserving encryption schemes [4]. Control of the hypergeometric tail is also useful when studying certain stochastic processes: hypergeometric tail control helps determine the convergence of specific population models [44, Theorem 4], and may also be used to approximate types of bivariate uniform empirical processes [7].

Hence, good bounds for the tail of the hypergeometric distribution have many applications. The main result of this dissertation is to find and prove finite sample Gaussian upper bounds for the hypergeometric tail. These bounds are stated formally in Chapter 4, and proved there as well. The bounds of Chapter 4 place some restrictions on the hypergeometric parameters. However, the method has natural extensions to other sets of parameters. The bounds of Chapter 4 are related to those obtained by Castelle and Laurent-Bonvalot [7], insofar as our approach is also motivated by the analysis of the Binomial$(n, 1/2)$ tail in [6, 30]. However, we differ here in attempting to bound the hypergeometric tail over both small and large variance populations simultaneously; the approach yielding the bounds of [7] truncates away from the small variance case. That the small variance case turns out to entail detailed argument in order to elicit bounds for the hypergeometric tail is unsurprising: in the related case of the binomial tail, complicated arguments [32, 34] are employed to analyze small variance populations.

In this dissertation, we will motivate the study of the hypergeometric tail from the point of view of sampling without replacement from bounded finite populations. In Chapter 1 we will describe this basic problem, show how the hypergeometric distribution arises naturally when studying this problem, and then review exponential bounds which apply to various sampling schemes. In Chapter 2, we will present several initial bounds we developed for the hypergeometric tail. These bounds are interesting in their own right, and provide a useful comparison to the bounds developed in the later chapters. The forthcoming discussion in Chapter 1, and the bounds presented in Chapter 2, closely follow the paper by Greene and Wellner [19], to be published in the *Bernoulli* journal. In Chapter 3, we will present a basic method for obtaining tail bounds for discrete distributions, apply the method to the Poisson distribution, and then lay the groundwork to analyze the hypergeometric tail. In Chapter 4, we will state and prove the main results of the dissertation: Gaussian bounds for the hypergeometric tail. In Chapter 5, we conclude by showing these bounds imply an improvement to existing hypergeometric exponential bounds, discuss computational considerations, conjectures, and future work.

1.1 Sampling With and Without Replacement

Consider a population C consisting of N elements, $C := \{c_1, \ldots, c_N\}$, where each $c_i \in \mathbb{R}$. 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 $1 \leq i \leq n \leq N$, and X_i denote the i^{th} draw with or without replacement from this population. Visualizing the population as an urn, the difference between these two procedures is that when an item c_i is drawn from the urn, it is returned to the urn before the $i+1^{(st)}$ draw occurs when sampling with replacement, while it is set aside from the urn before the $i+1^{(st)}$ draw occurs when sampling without replacement. Finally, let $S_n := \sum_{i=1}^{n} X_i$ denote the sum of either sampling procedure, and let $\bar{X}_n := S_n/n$ denote the sample mean.

It is often of interest in statistical problems to characterize the probability of deviations of the sample mean from the population mean μ by a given amount $\lambda > 0$. That is, to bound $P(\sqrt{n}(\bar{X}_n - \mu) > \lambda)$. Frequently, it is difficult to compute this probability exactly, and in situations where we allow the number of samples n and the population size N to increase, it is useful to have simple, sharp upper bounds to estimate the magnitude of this quantity. The simplicity of the bounds can be important for analytical purposes; the sharpness of the upper bounds is useful when trying to assess situations with a finite amount of data.

Many good inequalities exist which bound this probability when sampling with replacement. Fewer exist for the specific problem of sampling without replacement. It turns out that many of the inequalities used to bound this probability when sampling with replacement also bound the probability when sampling without replacement. However, the resulting bounds are sub-optimal: tighter bounds that account for the proportion of the population sampled are possible when sampling without replacement.

R. J. Serfling obtained one such bound in 1974 [45]. For $1 \leq n \leq N$, S_n the sum in sampling without replacement, and $\lambda > 0$, he found :

$$P\left(\sqrt{n}(\bar{X}_n - \mu) \geq \lambda\right) \leq \exp\left(-\frac{2\lambda^2}{(1-f_n^*)(b-a)^2}\right) \tag{1.1}$$

where $f_n^* := (n-1)/N$. This result applies to sampling without replacement from any finite bounded population. In particular, if we suppose $D \in \mathbb{N}$ and $0 \leq D \leq N$, it applies to a population of N elements, consisting of 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). Here and in the rest of this document we will interchangeably use the notation S_n and $S_{n,D,N}$ for hypergeometric random variables, to emphasize that hypergeometric random variables arise when sampling n elements without replacement from finite populations consisting only of D 1s and $N - D$ 0s. Hence in the hypergeometric context we may also use the already established notation $\bar{X}_n = S_n/n$ without any confusion (as well as $\bar{X}_{n,D,N} = (S_{n,D,N})/n$ when we wish to emphasize parameters).

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

$$\begin{aligned} P(S_{n,D,N} = k) &= \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \;,\; k \in \{\max(0, n+D-N), \ldots, \min(n, D)\} \;, \\ E(S_{n,D,N}) &= n\left(\frac{D}{N}\right), \\ \text{and } Var(S_{n,D,N}) &= n\left(\frac{D}{N}\right)\left(1-\frac{D}{N}\right)\left(1-\frac{n-1}{N-1}\right) =: n\mu_N(1-\mu_N)(1-f_n) \end{aligned} \tag{1.2}$$

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

$$P\left(\sqrt{n}(\bar{X}_n - \mu_N) \geq \lambda\right) \leq \exp\left(-\frac{2\lambda^2}{(1-f_n^*)}\right)$$

since $(b-a)^2 = (1-0)^2 = 1$. Now, comparing 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

$$P\left(\sqrt{n}(\bar{X}_n - \mu_N) \geq \lambda\right) \leq \exp\left(-\frac{2\lambda^2}{(1-f_n)(b-a)^2}\right) \tag{1.3}$$

in general, or at least in the special case of the hypergeometric distribution? Serfling himself raised the former question, commenting "[it] is also of interest to obtain [the bound] with the usual sampling fraction f_n instead of f_n^*" among his concluding remarks [45, p. 47]. Theorem 8 of this dissertation finds a stronger improvement is possible for the hypergeometric distribution when $31 \leq n \leq D \leq N/2$: $1 - f_n^*$ can be replaced by $1 - n/N$. That this improvement holds for the hypergeometric distribution suggests it might hold for the general problem of sampling without replacement from bounded finite populations, due to the special status occupied by the hypergeometric distribution in this problem. We discuss this special status in the next section.

1.2 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 [28]. 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 [33].

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* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$,

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

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

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

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

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* $\mathbf{y}$*".*

In the following Lemma, we re-state Kemperman's procedure so it constructs a majorizing hypergeometric population. See section 4, pages 165–168 in [28] for the original Rademacher argument.

Lemma 1. *(**Kemperman** [28]) 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* $\alpha = N\bar{\mathbf{x}}_N - D$*.*

Lemma 2. *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* $\mathbf{z} \in \mathbb{R}^N$*, 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 $1 \leq i \leq N-1$, 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* $\mathbf{y}$*,* $1 \leq n \leq N$*. Finally, suppose* ϕ *is a continuous convex increasing function on* $\mathbb{R}$*. Then*

$$E\phi\left(\sum_{i=1}^{n} X_i\right) \leq E\phi\left(\sum_{i=1}^{n} Y_i\right) \quad . \tag{1.4}$$

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

Observe that

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

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

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

where the sum is over all distinct indices $2 \leq i_2 < i_3 < \cdots < i_n \leq N$. 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 [33, p. 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 a 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* $\mathbf{y}$*,* $1 \leq n \leq N$*. Finally, suppose* ϕ *is a continuous convex increasing function on* $\mathbb{R}$*. Then*

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

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

Theorem 1. *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* $\mathbf{y} \in \mathbb{R}^N$*, 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* $\mathbf{y}$*,* $1 \leq n \leq N$*. Finally, suppose* ϕ *is a continuous convex increasing function on* $\mathbb{R}$*. Then*

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

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 $1 \leq n \leq N$, let $Z_1, \ldots, Z_n$ denote a sample without replacement from $\mathbf{z}$. By Lemma 4, we then have the order

$$E\phi\left(\sum_{i=1}^{n} X_i\right) \leq E\phi\left(\sum_{i=1}^{n} Z_i\right) . \tag{1.5}$$

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 (1.4) we have

$$E\phi\left(\sum_{i=1}^{n} Z_i\right) \leq E\phi\left(\sum_{i=1}^{n} Y_i\right) . \tag{1.6}$$

Combining (1.5) and (1.6) proves the claim. □

Theorem 1 shows that bounding the expectation of hypergeometric random variables provides control of the expectation in a general bounded sample without replacement (up to a convex composition). In the next section, we will discuss the Cramér-Chernoff method. When combined with Theorem 1, this will provide a procedure for obtaining exponential bounds for the general problem of sampling without replacement from a bounded finite population. Control of the hypergeometric distribution is central to this procedure. This provides reason to examine its tail behavior in detail.

1.3 Survey of Exponential Bounds for Gaussian, Binomial, and Hypergeometric Tail Probabilities

In this section, we provide an overview of existing exponential bounds useful in the analysis of Gaussian, binomial, and hypergeometric tails.

1.3.1 Gaussian Tail Probabilities

The following Gaussian tail bounds will serve as a point of comparison for the exponential bounds to follow. The statement is taken from Dudley, with modification [10, p. 67]

Gaussian Bound 1. *Let $X \sim N(\mu, \sigma^2)$. Then for all $\lambda > 0$ we have*

$$P(|X-\mu| > \lambda) \leq \exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \ . \tag{1.7}$$

Furthermore, if $\lambda \geq \sigma$ then

$$\sqrt{\frac{\sigma^2}{2\pi\lambda^2}}\exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \leq P(|X-\mu| > \lambda) \leq \sqrt{\frac{2\sigma^2}{\pi\lambda^2}}\exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \ . \tag{1.8}$$

Note by symmetry we can obtain one-sided bounds by dividing the exponential terms by 2. The second bound (1.8) also presents a large-deviation inequality for the Gaussian tail: when λ exceeds the standard deviation, we are able to sharpen the bound by the leading factor of $1/\lambda$.

In the following sections we will discuss exponential bounds for binomial and hypergeometric tails. The Gaussian bounds given here are informative because they provide information about the asymptotic form of those bounds. For example, consider the classical setting of an i.i.d sequence $X_1, \dots, X_n$, with $X_k \sim$ Bernoulli(1/2) for $1 \leq k \leq n$. Letting $n \nearrow \infty$, we know by the central limit theorem that

$$\sqrt{n}(\bar{X}_n - (1/2)) \to_d N(0, 1/4) \ .$$

By (1.7), we know for $X \sim N(0, 1/4)$ that for $\lambda > 0$

$$P(X > \lambda) \leq \frac{1}{2} \exp\left(-2\lambda^2\right) \ . \tag{1.9}$$

Similarly, by (1.8), we know for $X \sim N(0, 1/4)$ that for $\lambda \geq 1/2$ we have the bound

$$P(X > \lambda) \leq \frac{1}{\sqrt{8\pi\lambda^2}} \exp\left(-2\lambda^2\right) \ . \tag{1.10}$$

Note for $X \sim$ Bernoulli(p), Var(X) $= p(1-p)$ is maximized at $p = 1/2$. Hence, these asymptotic bounds provide control in the ($\sqrt{n}$-scaled) limit of the worst-case (most-variable) scenario. Thus these bounds provide a clue as to the form exponential bounds ought to take for binomial random variables: for finite n, we expect to see terms that yield the right constants in the limit.

1.3.2 Binomial Tail Probabilities

Here we review some exponential bounds applicable to the analysis of binomial tails. These bounds are of two kinds: the first are bounds that apply to independent sums of bounded random variables and the second are bounds that apply specifically to binomial random variables. Of course, the first kind can be specialized so that they too apply to binomial tails. For ease of reference, we collect both kinds under the heading "Binomial Bound".

Hoeffding's Exponential Bound

The following bound is due to Hoeffding [24, (p. 15)]:

Binomial Bound 1. *Let* $X_1, \ldots, X_n$ *be independent and* $0 \le X_i \le 1$ *for* $i \in \{1, \ldots, n\}$, *and define both* $\mu_i := EX_i$ *and* $\mu := (\sum EX_i)/n$. *Then for* $0 < \lambda < 1 - \mu$

$$\begin{aligned} P\left(\bar{X}_n - \mu \ge \lambda\right) &\le \left[\left(\frac{\mu}{\mu+\lambda}\right)^{(\mu+\lambda)} \left(\frac{1-\mu}{1-\mu-\lambda}\right)^{(1-\mu-\lambda)}\right]^n \\ &\le \exp\left(-n\lambda^2 g(\mu)\right) \\ &\le \exp\left(-2n\lambda^2\right) \ . \end{aligned}$$

where

$$g(\mu) := \begin{cases} \frac{1}{1-2\mu} \log\left(\frac{1-\mu}{\mu}\right) & \text{for } 0 < \mu < 1/2 \\ \frac{1}{2\mu(1-\mu)} & \text{for } 1/2 \le \mu < 1 \end{cases} \ .$$

On the $\sqrt{n}$ *-scale, we may summarize this bound as*

$$P\left(\sqrt{n}(\bar{X}_n - \mu) \ge \lambda\right) \le \exp(-2\lambda^2) \ . \tag{1.11}$$

This bound is well-known. Serfling's inequality can be viewed as translating this bound to the context of sampling without replacement in the following sense. Suppose $X_1, \ldots, X_n$ are independent and uniformly bounded such that $a \le X_i \le b$ for all $1 \le i \le n$. Then $0 \le (X - a)/(b - a) \le 1$. Hence if apply (1.11) to these standardized random variables, we obtain the following bound

$$P\left(\sqrt{n}(\bar{X}_n - \mu) \ge \lambda\right) \le \exp\left(-\frac{2\lambda^2}{(b-a)^2}\right) \ .$$

This is the same bound Serfling obtained at (1.1), absent the factor of $(1 - f_n^*)$. Intuitively this makes sense: sampling without replacement from a finite population should be more informative than sampling with replacement from the same population. The factor $1 - f_n^*$ quantifies this intuition, by inflating the exponent by something close to the finite sampling correction factor. We will provide an overview of Hoeffding's proof. To see how Hoeffding obtains the first bound consider the function class

$$\mathcal{E} := \left\{e^{kx} : k \ge 0\right\} \ .$$

Each element of $\mathcal{E}$ is non-decreasing and non-negative on $\mathbb{R}$. Hence, using Markov's inequality, he finds

$$P\left(\bar{X}_n - \mu \geq \lambda\right) = P\left(S_n \geq n\lambda + n\mu\right) \leq \inf_{\phi\in\mathcal{E}} \frac{E\phi(S_n)}{\phi(n\lambda + n\mu)} = \inf_{k\geq 0} \frac{Ee^{kS_n}}{e^{kn\lambda+kn\mu}} = \inf_{k\geq 0} \frac{\prod_{i=1}^{n} Ee^{kX_i}}{e^{kn\lambda+kn\mu}} . \tag{1.12}$$

Note the final equality follows by the independence of the summands. The Cramér-Chernoff method describes the act of bounding tail probabilities by applying Markov's inequality in conjunction with the class $\mathcal{E}$, and then minimizing the resulting expression. To minimize this quantity in his proof, Hoeffding uses an approximation lemma [24, (p. 21)], which we generalize here for further discussion.

Lemma 5. *Suppose X is a random variable such that $a \leq X \leq b$ and that $\phi : [a,b] \to \mathbb{R}$ is a convex function. Then*

$$E\phi(X) \leq \left(\frac{b - EX}{b-a}\right)\phi(a) + \left(\frac{EX - a}{b-a}\right)\phi(b) . \tag{1.13}$$

The proof of this lemma employs Hoeffding's exact argument.

Proof. Since $\phi(x)$ is convex on $[a,b]$, its graph lies under the line joining the points $(a,\phi(a))$ and $(b,\phi(b))$. Therefore

$$\phi(X) \leq \left(\frac{b - X}{b-a}\right)\phi(a) + \left(\frac{X - a}{b-a}\right)\phi(b) .$$

Take expectations to finish the proof. □

In his paper, Hoeffding specializes (1.13) for random variables $0 \leq X \leq 1$ and $\phi \in \mathcal{E}$. Such specialization implies

$$Ee^{kX} \leq 1 - \mu + \mu e^{k} = Ee^{kX_0}$$

where $\mu = EX = EX_0$ and X_0 is a Bernoulli(μ) random variable. Hence the right-hand-side of the inequality is the moment generating function of a Bernoulli(μ) random variable. Therefore Hoeffding's approximation lemma implies that the moment generating function of a random variable X with $P(X \in [0,1]) = 1$ is dominated by the moment generating function of a Bernoulli(EX) random variable. Indeed, the statement of the approximation lemma given here in (1.13) implies that a convex order exists between random variables concentrated on $[0,1]$ and Bernoulli random variables. See León and Perron section 1 [31] for more details on this ordering. Note that (1.13) provides an alternative proof of Proposition 3 from their paper.

Returning to Hoeffding, we now have all the pieces needed to obtain his first bound. From (1.12) we have, for arbitrary $k > 0$, that

$$\begin{aligned} e^{-kn\lambda - kn\mu} \prod_{i=1}^{n} Ee^{kX_i} &\leq e^{-kn\lambda - kn\mu} \left(\left[\prod_{i=1}^{n} (1 - \mu_i + \mu_i e^k) \right]^{\frac{1}{n}} \right)^n \\ &\leq e^{-kn\lambda - kn\mu} \left(\frac{1}{n} \left[\sum_{i=1}^{n} (1 - \mu_i + \mu_i e^k) \right] \right)^n \\ &= \left[e^{-k\lambda - k\mu} \left(1 - \mu + \mu e^k \right) \right]^n . \end{aligned}$$

The second inequality follows by the arithmetic-geometric mean inequality. The first inequality in (1.11) is found by minimizing over $k \geq 0$. The next inequality at (1.11) is obtained by writing the first bound as

$$\exp\left(-n\lambda^2 \left[\frac{\mu + \lambda}{\lambda^2} \log\left(\frac{\mu + \lambda}{\mu} \right) + \frac{1 - \mu - \lambda}{\lambda^2} \log\left(\frac{1 - \mu - \lambda}{1 - \mu} \right) \right] \right) = \exp^{-n\lambda^2 h(\lambda,\mu)}$$

and then minimizing over λ to find $g(\mu)$ minimizes $h(\mu, \lambda)$ for $0 < \lambda < 1 - \mu$. The final inequality in (1.11) is achieved by demonstrating $g(\mu)$ is minimized at $\mu = 1/2$. This completes the summary of Hoeffding's argument.

Bennett's Exponential Bound

This statement of Bennett's inequality is from Shorack and Wellner, with minor modification [46, p. 851].

Binomial Bound 2. *Let $X_1, \ldots, X_n$ be independent with $X_k \leq b$, $EX_k = \mu_k$, and $Var(X_k) = \sigma_k^2$ for $1 \leq k \leq n$. Define $\mu := (\mu_1 + \cdots + \mu_n)/n$ and $\sigma^2 := (\sigma_1^2 + \cdots + \sigma_n^2)/n$. Then for all*

$\lambda \geq 0$

$$P(\sqrt{n}(\bar{X}_n - \mu) \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{2\sigma^2}\psi\left(\frac{\lambda b}{\sigma^2\sqrt{n}}\right)\right) \tag{1.14}$$

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

Kiefer's Binomial Bound

This statement of Kiefer's inequality is from van der Vaart and Wellner, modified so the variance rather than the mean appears in the upper bound [52, pp. 460]. Since $0 \leq \mu(1-\mu) \leq 1/4$ for $\mu \in (0,1)$, no constraint on μ is needed in the statement.

Binomial Bound 3. *Suppose* $X_1, \ldots, X_n$ *are independent Bernoulli*(μ) *random variables. Then for all* n *and all* $\lambda > 0$*, we have*

$$P(\sqrt{n}\left(\bar{X}_n - \mu\right) \geq \lambda) \leq \exp\left(-\lambda^2\left(\log\left(\frac{1}{\mu(1-\mu)}\right) - 1\right)\right) \ . \tag{1.15}$$

Proof. We expand the argument given by van der Vaart and Wellner. Without loss of generality, we assume $\lambda \leq \sqrt{n}$, since the probability is zero for $\lambda > \sqrt{n}$. Consider the function $\psi(\lambda)$ in the statement of Bennett's inequality. ψ is decreasing in λ; hence, $\psi(\lambda/(\sqrt{n}\mu)) \geq \psi(1/\mu)$. Also observe for $\theta > 0$ that

$$\begin{aligned}
\frac{1}{2\theta}\psi\left(\frac{1}{\theta}\right) &= \left(\frac{1}{2\theta}\right)2\theta^2 h(1+\frac{1}{\theta}) \\
&= \theta\left(\frac{\theta+1}{\theta}(\log\left(\frac{\theta+1}{\theta}\right) - 1) + 1\right) \\
&= (\theta+1)\log\left(1+\frac{1}{\theta}\right) - 1 \\
&\geq \log\left(\frac{1}{\theta}\right) - 1
\end{aligned} \tag{1.16}$$

with the final inequality following since $\theta > 0$ and log is increasing. Finally, we have by

Bennett's inequality (1.14) with $b = 1$ that

$$\begin{aligned} P(\sqrt{n}(\bar{X}_n - \mu) \geq \lambda) &\leq \exp\left(-\frac{\lambda^2}{2\mu(1-\mu)}\psi\left(\frac{\lambda}{\sigma^2\sqrt{n}}\right)\right) \\ &\leq \exp\left(-\frac{\lambda^2}{2\sigma^2}\psi\left(\frac{1}{\sigma^2}\right)\right) \\ &\leq \exp\left(-\lambda^2\left(\log\left(\frac{1}{\mu(1-\mu)}\right) - 1\right)\right) \end{aligned}$$

with the final inequality following by (1.16). This proves the bound. □

León and Perron's Binomial Bound

The following large-deviation bound for binomial tails is due to León and Perron [31].

Binomial Bound 4. *Let $n \geq 1$, $p \in (0,1)$, $\lambda < \sqrt{n}/2$, and $X_1, \ldots, X_n$ be independent Bernoulli(μ) random variables. Then*

$$P(\sqrt{n}(\bar{X}_n - \mu) \geq \lambda) \leq \frac{1}{\sqrt{2\pi\lambda^2}}\left(\frac{1}{2}\right)\sqrt{\frac{\sqrt{n}+2\lambda}{\sqrt{n}-2\lambda}e^{-\left(2\lambda^2\right)}}\ . \tag{1.17}$$

Observe that the factor of 1/2 can be brought under the radical so the bound can be written

$$\sqrt{\frac{\sqrt{n}+2\lambda}{\sqrt{n}-2\lambda}}\left[\frac{1}{\sqrt{8\pi\lambda^2}}e^{-\left(2\lambda^2\right)}\right]\ .$$

The term in the brackets recovers the terms in (1.10) exactly. The leading quantity acts as a finite-sample penalty. Note that for fixed $\lambda > 0$, $(\sqrt{n}+2\lambda)/(\sqrt{n}-2\lambda) \to 1$ as $n \nearrow \infty$; this continues to hold if $\lambda = \lambda_n = o(\sqrt{n})$.

Talagrand's Binomial Bound

We next state the bound Talagrand discovered for binomial random variables [49, pp. 48–50]. The statement here is taken from van der Vaart and Wellner [52, pp. 460–462]:

Binomial Bound 5. *Fix p_0 and consider p such that $0 < p_0 \leq p \leq 1 - p_0 < 1$. Suppose for $n \in \mathbb{N}$ that $X_1, \ldots, X_n$ 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* $\lambda > 0$, $\quad P\left(\sqrt{n}(\bar{X}_n - p) = \lambda\right) \leq \frac{K_1}{\sqrt{n}} \exp\left(-\left[2\lambda^2 + \frac{\lambda^4}{4n}\right]\right)$.

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

(*iii*) *For all* $\lambda > 0$, $\quad P\left(\sqrt{n}(\bar{X}_n - p) \geq \lambda\right) \leq \frac{K_2}{\lambda} \exp\left(-\left[2\lambda^2 + \frac{\lambda^4}{4n}\right]\right)$. (1.18)

We refer to either cited source for proofs of these bounds, since a subsequent bound for hypergeometric random variables involves similar considerations, and will be proved in full.

1.3.3 Hypergeometric Tail Probabilities

As we observed in the introduction, many of the bounds that can control binomial tail behavior can also control the hypergeometric tail. However, the binomial bounds are not optimal, since they do not account for the proportion of the population sampled in the hypergeometric setting. In this section, we present several bounds which may be used to control tail probabilities in the case of sampling without replacement from a finite, bounded population. We will specialize these bounds to the hypergeometric tail, since providing control of this tail probability is the main focus of this work.

As we will see, the relative performance of the bounds depends on the hypergeometric parameters n, D, and N and the magnitude of the deviation λ. It is thus natural to ask: is there a single bound presented here which should be used in favor of all other bounds in all cases? Since different bounds demand different knowledge about the population – Serfling's bound only requires knowing the sample size n and the population size N, while Chatterjee's bound, for example, additionally requires knowing the number of 1's D – no such recommendation is possible: the utility of each bound depends on the details of the problem they are addressing and the assumptions one is willing to make. In addition, several of the bounds were developed to study problems more general than sampling without replacement. Hence, even if a given bound appears looser than the others in the setting of the hypergeometric tail, it is still extremely useful due to the breadth of problems to which it applies.

We begin by restating Serfling's inequality, specialized to the case of the hypergeometric distribution.

Hypergeometric Bound 1. *(**Serfling** [45]) Suppose $S_n \sim$ Hypergeometric(n,D,N). Let $\mu_N := D/N$. Then for all $\lambda > 0$ we have*

$$P\left(\sqrt{n}(\bar{X}_n - \mu_N) \geq \lambda\right) \leq \exp\left(-\frac{2\lambda^2}{(1-f_n^*)}\right) . \tag{1.19}$$

Note the correction factor $1 - f_n^*$ inflates the constant 2 by the ratio $(N-n+1)/N$. Therefore as n increases (that is, as more of the population N is sampled) this bound can decrease much faster than Hoeffding's bound (1.11).

Chatterjee [8] derives very general concentration bounds for statistics based on random permutations by way of Stein's method. His Proposition 1.1 obtains bounds for a permutation statistic S, which was first studied by Hoeffding [23]. This bound may be specialized to provide exponential control of the hypergeometric tail. We provide the specialization in the next bound.

Hypergeometric Bound 2. *(**Chatterjee** [8]) Suppose $S_n \sim$ Hypergeometric(n,D,N). Let $\mu_N := D/N$. Then for all $\lambda > 0$ we have*

$$P\left(\sqrt{n}(\bar{X}_n - \mu_N) \geq \lambda\right) \leq \exp\left(-\frac{\lambda^2}{4\mu_N + 2\frac{\lambda}{\sqrt{n}}}\right) . \tag{1.20}$$

Bardenet and Maillard have recently 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 [2] for additional discussion.

Hypergeometric Bound 3. *(**Bardenet and Maillard** [2]) Suppose $S_n \sim$ Hypergeometric(n,D,N). Let $\mu_N := D/N$. Then for all $\lambda > 0$ and $n < N$ we have*

$$P\left(\sqrt{n}(\bar{X}_n - \mu_N) \geq \lambda\right) \leq \exp\left(-\frac{2\lambda^2}{(1-n/N)(1+1/n)}\right) . \tag{1.21}$$

Later in this dissertation we will see a further refinement of Serfling's inequality is possible: the factor $1 + 1/n$ appearing in (1.21) may be eliminated under light conditions on n and D. To achieve this improvement, we will need to consider bounds which account for the population variance. In this thesis, we will consider populations when the sample size n and proportion of D's in the population are both much smaller than the population size N. These populations have very small variance, and are highly skewed: the skewness of a hypergeometric population may be written

$$\left[\frac{N-2D}{N-2}\right]\sqrt{\frac{N-2n}{N-D}}\sqrt{\frac{N-2n}{N-n}}\sqrt{\frac{N-1}{nD}}\;.$$

When $n, D \lll N$, we see the first three terms in the skewness are approximately 1, while the final $(N-1)/(nD)$ is unbounded above. In such populations, it is very unlikely for a sample to deviate from the population mean D/N (which approaches 0), and so we need special arguments to account for the tiny variance and extreme skewness. Exponential bounds which take the variance into account and are valid for all hypergeometric populations can decrease much faster than bounds such as Serfling's in small variance populations. Thus we can show improvements to Serfling's bound are possible in some cases by first developing exponential bounds which account for the population variance, and then comparing such bounds to the conjectured improvement under the assumption that the population variance is small.

Bardenet and Maillard also obtained the following bound which incorporates information about the population variance into the exponential.

Hypergeometric Bound 4. *(**Bardenet and Maillard** [2]) 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*

$$P\left(\sqrt{n}(\bar{X}_n - \mu) \geq \lambda\right) \leq \exp\left(-\frac{\lambda^2}{2(\gamma^2 + (2/3)(b-a)(\lambda/\sqrt{n}))}\right) + \delta \qquad (1.22)$$

where

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

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

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

We next present a bound derived by Hush and Scovel. They obtained the following bound by extending an argument given by Vapnik. See [26] and [53].

Hypergeometric Bound 5. *(**Hush and Scovel** [26]) Suppose $S_n \sim$ Hypergeometric(n,D,N). Then for all $\lambda > 0$ we have*

$$P\left(\sqrt{n}(\bar{X}_{n,D,N} - \mu_{D,N}) \ge \lambda\right) \le \exp\left(-2\alpha_{n,D,N}(n\lambda^2 - 1)\right)$$

where

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

Goldstein and Işlak [17] recently used a variant of Stein's method to give another inequality for the tails of Hoeffding's statistic S, one which accounts for the population variance. As with Chatterjee's bound, the bound of Goldstein and Işlak may be specialized to provide control of the hypergeometric tail. The following statement of their bound makes this specialization.

Hypergeometric Bound 6. *(**Goldstein and Işlak** [17]) Suppose $S_n \sim$ Hypergeometric(n,D,N). Let $\mu_N := D/N$, $\sigma_N^2 := \mu_N(1-\mu_N)$, and $1 - f_n := 1 - (n-1)/(N-1)$. Then for all $\lambda > 0$ we have*

$$P\left(\sqrt{n}(\bar{X}_n - \mu_N) \ge \lambda\right) \le \exp\left(\frac{-\lambda^2}{2\left[\sigma_N^2\,(1-f_n) + \frac{8}{\sqrt{n}}\left[\mu_N \vee (1-\mu_N)\right]\lambda\right]}\right) \ . \qquad (1.23)$$

Rohde obtained the following bound in her study of the properties of a high-dimensional multiple randomization test [43].

Hypergeometric Bound 7. *(**Rohde** [43]) Suppose $S_n \sim$ Hypergeometric(n,D,N). Let $\mu_N := D/N$, $\sigma_N^2 := \mu_N(1-\mu_N)$, and $1 - f_n := 1 - (n-1)/(N-1)$. Then for all $\lambda > 0$ we have*

$$P\left(\sqrt{n}(\bar{X}_n - \mu_N) \geq \lambda\right) \leq$$
$$\exp\left(\frac{-\lambda^2}{2\left(\sigma_N^2\left(1-f_n\right)\left(1-\left(\sqrt{\frac{n}{N(N-n)}} \vee \sqrt{\frac{N-n}{nN}}\right)\right) + \frac{2}{3}\left(\frac{1}{\mu_N}\right)\left(\frac{1}{1-f_n}\right)\frac{\lambda}{\sqrt{n}}\left(\frac{n}{N} \vee 1 - \frac{n}{N}\right)\right)}\right) \tag{1.24}$$

Recently, an analogue of Bennett's inequality was obtained in [19] by using a representation of the hypergeometric distribution established by Vatutin and Mikhaĭlov [54] (also see Ehm [11], Theorem A, and Pitman [40]). This representation along with Bennett's inequality yields the following Bennett-type exponential bound for hypergeometric random variables. See [19] for a discussion and proof.

Theorem 2. *(**Greene and Wellner** [19]) Suppose $S_{n,D,N} \sim$ Hypergeometric(n, D, N) with $1 \leq n \leq D \wedge (N-D)$. Let $\mu_N := D/N$, $\sigma_N^2 := \mu_N(1-\mu_N)$, and $1-f_n := 1-(n-1)/(N-1)$. Then for all $\lambda > 0$*

$$P(\sqrt{n}(\bar{X}_{n,D,N} - \mu_N) > \lambda) \leq \exp\left(-\frac{\lambda^2}{2\sigma_N^2(1-f_n)}\psi\left(\frac{\lambda}{\sqrt{n}\sigma_N^2(1-f_n)}\right)\right) \tag{1.25}$$

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

Since $\psi(v) \geq 1/(1+v/3)$ for all $v \geq 0$ (Shorack and Wellner [46], proposition 1, page 441), Theorem 2 immediately yields following Bernstein type tail bound.

Corollary 1. *(**Greene and Wellner** [19]) With the same assumptions and notation as in Theorem 2,*

$$P(\sqrt{n}(\bar{X}_{n,D,N} - \mu_N) > \lambda) \leq \exp\left(-\frac{\lambda^2/2}{\sigma_N^2(1-f_n) + \frac{\lambda}{3\sqrt{n}}}\right). \tag{1.26}$$

We conclude this chapter with a figure that compares these bounds in the small variance setting. There we see that the bound (1.25) provides significant improvement over Serfling's bound for moderate deviations. We exploit this improvement in Chapter 5.

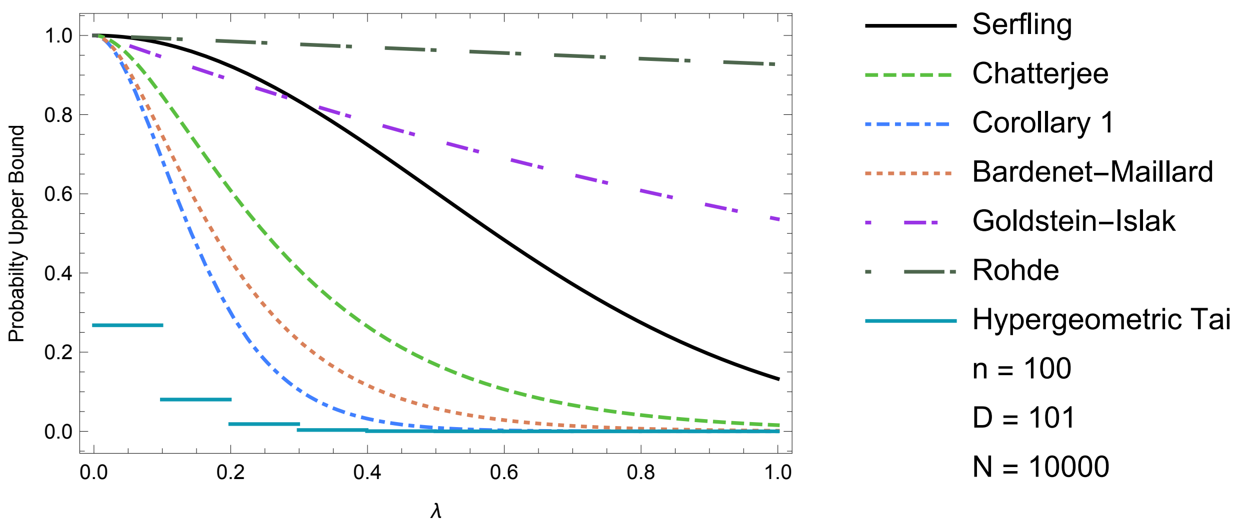

Figure 1.1: Comparison of Serfling's bound (1.19), Chatterjee's bound (1.20), Goldstein and Işlak's bound (1.23), Rohde's bound (1.24), Bardenet and Maillard's bound (1.22), and Corollary 1 . In the figure the sample size is $n = 100$, the population size is $N = 10000$, and the number of successes is $D = 101$.

Chapter 2

HYPERGEOMETRIC ADAPTATION OF BINOMIAL EXPONENTIAL BOUNDS

> If we have a sample of size 100, what would happen with a sample of size $100,000$ is not a decisive consideration.
>
> ---
> David A. Freedman [15]

In this chapter we develop two bounds for the tail of the hypergeometric distribution. These bounds are analogous to the binomial bounds of León and Perron (Binomial Bound 4) and Talagrand (Binomial Bound 5) presented in chapter 1. We begin with a statement of the two bounds. We first state the analogue of León and Perron's bound.

Theorem 3. *(**Greene and Wellner** [19]) Suppose $S_n \sim Hypergeometric(n, D, N)$. Define $\mu := D/N$, and suppose $N > 4$ and $2 \leq n < D \leq N/2$. Then for all $0 < \lambda < \sqrt{n}/2$ we have*

$$P\left(\sqrt{n}(\bar{X}_n - \mu) \geq \lambda\right) \leq \sqrt{\frac{1}{2\pi\lambda^2}}\left(\frac{1}{2}\right)\sqrt{\left(\frac{N-n}{N}\right)\left(\frac{\sqrt{n}+2\lambda}{\sqrt{n}-2\lambda}\right)\left(\frac{N-n+2\sqrt{n}\lambda}{N-n-2\sqrt{n}\lambda}\right)}$$
$$\cdot \quad \exp\left(-\frac{2}{1-\frac{n}{N}}\lambda^2\right)\exp\left(-\frac{1}{3}\left(1+\frac{n^3}{(N-n)^3}\right)\frac{\lambda^4}{n}\right) . \qquad (2.1)$$

We next state the analogue of Talagrand's bound.

Theorem 4. *(**Greene and Wellner** [19]) 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 $0 < \mu_0 \leq \mu \leq 1 - \mu_0 < 1$ and $0 < \psi_0 \leq \psi \leq 1 - \psi_0 < 1$. Then there exist constants K_1, K_2 depending only* *on* μ_0 *and* ψ_0 *such that:*

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

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

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

$$P(\sqrt{n}(\overline{X}_n - \mu) \geq t) \leq \frac{K_2}{\lambda} \begin{pmatrix} \exp\left(-\frac{2\lambda^2}{1-\frac{n}{N}}\right) \exp\left(-\left(\frac{1}{4} + \frac{1}{3}\left(\frac{n}{N-n}\right)^3\right)\frac{\lambda^4}{n}\right) \\ \cdot \exp\left(\lambda(\lambda - t)\left(\frac{4}{1-\frac{n}{N}} + 1 + \frac{4n^3}{3(N-n)^3}\right)\right) \end{pmatrix} .$$

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

$$P(\sqrt{n}(\overline{X}_n - \mu) \geq \lambda) \leq \frac{K_2}{\lambda} \exp\left(-\frac{2\lambda^2}{1-\frac{n}{N}}\right) \exp\left(-\left(\frac{1}{4} + \frac{1}{3}\left(\frac{n}{N-n}\right)^3\right)\frac{\lambda^4}{n}\right) . \quad (2.2)$$

The proofs of these two bounds do not proceed by the Cramér - Chernoff method. This means we cannot use these bounds to control the general sampling problem through Theorem 1. The proof of (2.1) 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 [13, pp. 150-151]: see Lemma 8 for details. The proof of (2.2) 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. A similar analysis is carried out in chapter 3, which underpins the proof presented in chapter 4: compare Lemma 7 to Lemma 14. The key difference between these two Lemmas is that Lemma 7 produces a bound in terms of a variable $u \in [0, 1-D/N-1/n]$. Lemma 14 yields a bound in terms of a variables $x \in [0, 1]$. The parameters appearing in the interval $[0, 1 - D/N - 1/n]$ make it difficult to extend the results presented in this chapter to deviations larger than $\sqrt{n}/2$ (in the León and Perron analogue), and other parameter combinations of the hypergeometric table.

On the other hand, the parametrization $x \in [0, 1]$ used in chapter 3 has natural extensions to other parameter combinations beyond those considered in this work. This is because the x-scale standardizes the analysis across many different combinations of parameters. The structure of this document follows history: the results in this chapter preceded the results of the later chapters. Indeed, it was the desire to generalize to other parameter combinations and larger deviations that led to developing the x-scale for the hypergeometric distribution.

The proofs in this chapter depend on a version of Stirling's formula from Robbins [42].

Lemma 6. *(**Robbins** [42]) For $n \in \mathbb{N}_0$*

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \ . \tag{2.3}$$

2.1 Upper Bound on Hypergeometric Probabilities

We begin by finding upper bounds on individual hypergeometric probabilities.

Lemma 7. *Suppose $S_n \sim$ Hypergeometric(n, D, N) with $1 \leq n < D \leq \lfloor N/2 \rfloor$ and $1 \leq k \leq n-1$. Then for $k \geq n(D/N)$ we have*

$$\begin{aligned} P(S_n = k) \leq \ & \frac{1}{\sqrt{2\pi}} \frac{\sqrt{D(N-D)n(N-n)}}{\sqrt{k(D-k)(n-k)(N-D-(n-k))N}} \\ & \cdot \exp\left(-\frac{2nN}{N-n}u^2\right) \exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right) \ , \end{aligned} \tag{2.4}$$

with $u := k/n - D/N$.

Proof. The proof follows by direct analysis. Using Stirling's formula (2.3), we have

$$\begin{aligned} P(S_n = k) &= \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \\ &\leq \frac{1}{\sqrt{2\pi}} \frac{\sqrt{D(N-D)n(N-n)}}{\sqrt{k(D-k)(n-k)(N-D-(n-k))N}} \\ &\quad \cdot \frac{D^D(N-D)^{N-D}n^n(N-n)^{N-n}}{k^k(D-k)^{D-k}(n-k)^{n-k}(N-D-(n-k))^{N-D-(n-k)}N^N} \\ &\quad \cdot \frac{\exp\left(\frac{1}{12D}+\frac{1}{12(N-D)}+\frac{1}{12n}+\frac{1}{12(N-n)}\right)}{\exp\left(\frac{1}{12k+1}+\frac{1}{12(D-k)+1}+\frac{1}{12(n-k)+1}+\frac{1}{12(N-D-(n-k))+1}+\frac{1}{12N+1}\right)} \\ &=: A \cdot B \cdot C \ . \end{aligned} \tag{2.5}$$

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

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

where the first factor corresponds to the same function as in Talagrand's argument for the binomial distribution [49, pp. 48–50] and we recall

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

Now, we can further re-write 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

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

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 $f = f_N := n/N$, $\overline{f} = \overline{f}_N := 1 - f_N$,

$$\frac{\Gamma}{N-n} = \left(\mu - \frac{f}{\overline{f}}u\right)\log\left[\left(\mu - \frac{f}{\overline{f}}u\right)\frac{1}{\mu}\right] + \left(1-\mu+\frac{f}{\overline{f}}u\right)\log\left[\left(1-\mu+\frac{f}{\overline{f}}u\right)\frac{1}{1-\mu}\right]$$

$$= \Psi\left(\frac{f}{\overline{f}}u, 1-\mu\right)$$

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

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

Recall that we assume $D > k \geq n(D/N)$. Notice the assumptions of the Lemma imply

$$0 \leq u < 1-\mu \text{ and } 0 \leq \frac{f}{\overline{f}}u < \mu \ .$$

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 [52, pp. 460-461]),

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

Thus

$$\frac{\partial^2}{\partial u^2}\left[n\Psi(u,\mu) + (N-n)\Psi\left(\frac{f}{\overline{f}}u, 1-\mu\right)\right]$$
$$= n\frac{4}{1-4(u-(1/2-\mu))^2} + (N-n)\frac{4(f/\overline{f})^2}{1-4\left(\frac{f}{\overline{f}}u-(\mu-1/2)\right)^2}$$
$$\geq 4n\left(1+4(u-(1/2-\mu))^2\right) + 4(N-n)(f/\overline{f})^2\left(1+4\left(\frac{f}{\overline{f}}u-(\mu-1/2)\right)^2\right) \ .$$

Integration across this inequality yields

$$\frac{\partial}{\partial u}\left[n\Psi(u,\mu)+(N-n)\Psi\left(\frac{f}{\overline{\overline{f}}}u,1-\mu\right)\right] \geq 4n\left(u+\frac{1}{3}u^3\right)+4(N-n)\left(\frac{f}{\overline{\overline{f}}}\right)^2\left(u+\frac{1}{3}\left(\frac{f}{\overline{\overline{f}}}\right)^2u^3\right)$$
$$= 4\left(n+(N-n)\left(\frac{n}{N-n}\right)^2\right)u$$
$$+\ \frac{4}{3}\left(n+(N-n)\left(\frac{n}{N-n}\right)^4\right)u^3$$
$$= \frac{4nN}{N-n}u+\frac{4}{3}n\left(1+\frac{n^3}{(N-n)^3}\right)u^3\ . \qquad (2.6)$$

Here we used

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

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

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

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

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

Integrating across (2.6) yields

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

Thus the B term in (2.5) has the following bound:

$$B \leq \exp\left(-\frac{2nN}{N - n}u^2\right)\exp\left(-\frac{n}{3}\left(1 + \frac{n^3}{(N - n)^3}\right)u^4\right) \ . \tag{2.7}$$

We next analyze the C term in (2.5). We have

$$\begin{aligned}
C &= \frac{\exp\left(\frac{1}{12D} + \frac{1}{12(N-D)} + \frac{1}{12n} + \frac{1}{12(N-n)}\right)}{\exp\left(\frac{1}{12k+1} + \frac{1}{12(D-k)+1} + \frac{1}{12(n-k)+1} + \frac{1}{12(N-D-(n-k))+1} + \frac{1}{12N+1}\right)} \\
&= \exp\left(\frac{1}{12D} - \frac{1}{12(D - k) + 1}\right)\exp\left(\frac{1}{12(N - D)} - \frac{1}{12(N - D - (n - k)) + 1}\right) \\
&\quad \cdot \exp\left(\frac{1}{12n} - \frac{1}{12k + 1}\right)\exp\left(\frac{1}{12(N - n)} - \frac{1}{12(n - k) + 1}\right)\exp\left(-\frac{1}{12N + 1}\right) \\
&= \exp\left(\frac{-12k + 1}{[12D][12(D - k) + 1]}\right)\exp\left(\frac{-12[n - k] + 1}{[12(N - D)][12([N - D] - [n - k]) + 1]}\right) \\
&\quad \cdot \exp\left(\frac{1 - 12(n - k)}{12(12k + 1)n}\right)\exp\left(\frac{1 - 12(N - 2n + k)}{12(12(n - k) + 1)(N - n)}\right)\exp\left(-\frac{1}{12N + 1}\right) \\
&\leq 1
\end{aligned} \tag{2.8}$$

where the final inequality follows since $k \in [\lceil n\mu \rceil, \ldots, n-1]$ and $n \le D \le \lfloor N/2 \rfloor$ 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 (2.5) is already in the claimed form, combining (2.7) and (2.8) proves the claim. □

2.2 Upper Bound on Hypergeometric Tail

To determine an analogue of Léon and Perron's bound (see (1.17) in chapter 1), we need an upper bound for hypergeometric tail probabilities similar to that bound discussed by Feller for the binomial [13, pp. 150-151].

Lemma 8. *Suppose* $S_{n,D,N} \sim Hypergeometric(n, D, N)$*,* $N > 4$ *and* $1 \le n, D \le N-1$*. For* $k > (nD)/N$*, we have*

$$P\left(S_{n,D,N} \ge k\right) \le P\left(S_{n,D,N} = k\right)\left(\frac{k(N-D-n+k)}{Nk-nD}\right). \tag{2.9}$$

Proof. Suppose first that $n \le D$ and $k = n$. Then (2.9) becomes

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

Since $P\left(S_{n,D,N} \ge n\right) = P\left(S_{n,D,N} = n\right)$, the result holds in this case. Next, suppose $D < n$ and $k = D$. Then (2.9) becomes

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

Since $P(S_{n,D,N} \ge D) = P(S_{n,D,N} = D)$, 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 $\lfloor (nD)/N \rfloor + 1 \le (j-1) < j \le n \wedge D$. Assume the result holds when

$k = j$. We will show this implies the result holds for $k = j - 1$. We have

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

Under the current assumption, the right-hand side of (2.9) equals

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

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

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

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

2.3 Extending León and Perron to the Hypergeometric

We now develop an upper bound for the hypergeometric tail in the spirit of León and Perron (again, see (1.17) in chapter 1). The proof relies on the following technical Lemma.

Lemma 9. *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

$$\frac{\mu(1-\mu)(u+\mu)(\gamma(1-\mu)+u)}{(1-u-\mu)(\gamma\mu-u)} \leq \frac{1}{4}\frac{(u+(1/2))(\gamma(1-(1/2))+u)}{(1-u-(1/2))(\gamma(1/2)-u)} \ . \tag{2.10}$$

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 of its derivative 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 9. With the previous comments in mind, define the following functions:

$$f(\mu) := \frac{\mu\,(1-\mu)}{(1-u-\mu)\,(\gamma\mu-u)}$$

$$\text{and} \quad g(\mu) := (u+\mu)\,(\gamma(1-\mu)+u) \ . \tag{2.11}$$

Note that the product $f(\mu)g(\mu)$ gives the quantity on the left-hand side of (2.10). 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

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

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 $b(u,\gamma) < 0$ 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 $p'(u) = 1 - \gamma^2 < 0$ 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

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

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 $n < D \leq \lfloor N/2 \rfloor$ means that the maximum value n can attain is $\lfloor N/2 \rfloor - 1$. This implies $N - 2n - 2 \geq 0$. Since we also restrict $u \in (0, 1/2)$, we also have $(N(1-u) - n - 1) \geq 0$ and $(N - 2nu - n) \geq 0$. Finally note

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

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 (2.11). 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.

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

Since $g''(\mu) < 0$ for any choice of $(u,\gamma) \in (0,1/2)\times(1,\infty)$, we see that μ^* is a maximum within $(0,1/2)\times(1,\infty)$. 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 of μ 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

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

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

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

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

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

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

$$t'(\mu) = 1 - 2\mu + \frac{(\gamma+1)u^2(\gamma\mu^2 - \mu^2 + 2\mu - 2\mu u + u - 1)}{(1-\mu-u)^2(\gamma\mu - u)^2} \ . \tag{2.12}$$

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

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

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

$$\begin{aligned} & ep\,(1/2) - ue\left(\frac{n+1}{N}\right) \\ &= \frac{Nu\left(\begin{array}{c} 2u^2(N-2n)(N-2n-1)(2n(N-n-1)+N) \\ +Nu(N-n)(N-2n-3) + (N-n)^2(N-2n-2) - 4nNu^3(N-2n-1) \end{array}\right)}{4(1-2u)(N-n)(N-n-2nu)(N(1-u)-n-1)(N-n+nN(1-u)-n^2)} \end{aligned} \tag{2.13}$$

where we have again substituted the definition $\gamma = (N-n)/n$. We will now argue that this quantity is positive for all $n \in \{1, \ldots, \lfloor N/2 \rfloor - 2\}$. This is sufficient to demonstrate $t(\mu)$ is maximized at $\mu = 1/2$, since we are supposing $n < D \leq \lfloor N/2 \rfloor$. 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,

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

We also have

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

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

$$z(u) := 2u^2(N-2n)(N-2n-1)(2n(N-n-1)+N) + Nu(N-n)(N-2n-3) - 4nNu^3(N-2n-1) \geq 0 \ .$$

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

$$\begin{aligned}
z'(u) &= 4u(N-2n)(N-2n-1)(2n(N-n-1)+N) \\
&\quad + N(N-n)(N-2n-3) - 12nNu^2(N-2n-1) \ , \\
z''(u) &= 4(N-2n)(N-2n-1)(2n(N-n-1)+N) - 24nNu(N-2n-1) \ , \\
z'''(u) &= -24nN(N-2n-1) \leq 0 \ .
\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

$$ep\,(1/2) - ue\left(\frac{n+1}{N}\right) > 0 \quad . \tag{2.14}$$

Therefore, we can define the following function

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

Observe that for all $\mu \in [(n+1)/N, 1/2]$, we have $\mathrm{maj}(\mu) \geq t(\mu)$. Additionally, we know that $\mathrm{maj}(\mu)$ is maximized at $\mu = 1/2$: the argument following (2.12) shows for μ such that $\mathrm{maj}(\mu) = t(\mu)$, $\mathrm{maj}(\mu)$ strictly increases; the argument preceding (2.14) shows $\mathrm{maj}(\mu)$ increases to its maximum on the interval $[(n+1)/N, 1/2]$. Finally, since we know $\mathrm{maj}(1/2) = ep(1/2) = t(1/2)$, 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 Theorem 3.

Proof of Theorem 3. Pick $\lambda, k > 0$ such that $k = \sqrt{n}\lambda + n\mu$, $k \geq n(D/N)$. We then have

$$\begin{aligned}
&P\left(\sqrt{n}(\bar{X} - \mu) \geq \lambda\right) \\
&= P\left(\sum_{i=1}^{n} X_i \geq k\right) \\
&\leq \left[P\left(\sum_{i=1}^{n} X_i = k\right)\right]\left(\frac{k(N-D-n+k)}{Nk-nD}\right) && \text{by (2.9)} \\
&\leq \left[\frac{1}{\sqrt{2\pi}}\frac{\sqrt{D(N-D)n(N-n)}}{\sqrt{k(D-k)(n-k)(N-D-(n-k))N}}\left(\frac{k(N-D-n+k)}{Nk-nD}\right)\right] \\
&\cdot \left[\exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right)\right] && \text{by (2.4)} \\
&= [A]\cdot\left[\exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right)\right] \ . && (2.15)
\end{aligned}$$

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

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

We may then write:

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

Similarly we have

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

Using these parametrizations, we may write

$$
\begin{aligned}
[A] &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{D(N-D)n(N-n)}{k(D-k)(n-k)(N-D-(n-k))N}}\left(\frac{k(N-n-(D-k))}{Nn((k/n)-(D/N))}\right) \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{D(N-D)n(N-n)k^2(N-n-(D-k))^2}{k(D-k)(n-k)(N-n-(D-k))N^3n^2}}\left(\frac{1}{u}\right) \\
&= \sqrt{\frac{(N-n)}{2\pi nNu^2}}\sqrt{\frac{D}{N}\left(1-\frac{D}{N}\right)\frac{\left(\frac{k}{n}\right)}{\left(1-\frac{k}{n}\right)}\frac{(N-n-(D-k))}{(D-k)}} \\
&= \sqrt{\frac{(N-n)}{2\pi nNu^2}}\sqrt{\mu\,(1-\mu)\,\frac{(u+\mu)}{(1-u-\mu)}\frac{\gamma(1-\mu)+u}{\gamma\mu-u}} \\
&\leq \sqrt{\frac{(N-n)}{2\pi nNu^2}}\sqrt{\frac{1}{4}\frac{(u+(1/2))(\gamma(1-(1/2))+u)}{(1-u-(1/2))(\gamma(1/2)-u)}}
\end{aligned}
\tag{2.16}
$$

with the last inequality following by (2.10) established in Lemma 9. Observe under these parametrizations $u=\lambda/\sqrt{n}$. Hence, if we use (2.16) to provide an upper bound for (2.15), substitute $\lambda/\sqrt{n}$ for u, and then simplify, the claim is proved. □

2.4 Extending Talagrand to the Hypergeometric

Some of the machinery developed in the preceding lemmas will be adapted to prove (2.2). The argument follows.

Proof of Theorem 4. 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 (2.5) still applies. For $D > k \geq n(D/N)$, the bound on the B term in (2.7) still holds. Thus we may write

$$
\begin{aligned}
B &\leq \exp\left(-\frac{2nN}{N-n}u^2\right)\exp\left(-\frac{n}{3}\left(1+\frac{n^3}{(N-n)^3}\right)u^4\right) \\
&= \exp\left(-\frac{2n}{1-\frac{n}{N}}u^2\right)\exp\left(-\frac{n}{4}u^4\right)\exp\left(-\frac{n}{12}u^4\right)\exp\left(-\left[\frac{n^4}{3(N-n)^3}\right]u^4\right) .
\end{aligned}
\tag{2.17}
$$

Also recall we showed that $C \leq 1$ at (2.8) when $n \leq D \leq N/2$. In fact, the expression at (2.8) 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

$$\begin{aligned}
A &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{D(N-D)n(N-n)}{k(D-k)(n-k)(N-D-(n-k))N}} \\
&= \frac{N}{\sqrt{2\pi n}} \sqrt{\frac{\frac{D}{N}(1-\frac{D}{N})(1-\frac{n}{N})}{\frac{k}{n}(D-k)(1-\frac{k}{n})(N-D-n+k)}} \\
&\leq \frac{N}{\sqrt{2\pi n}} \sqrt{\frac{(1/4)(1-\psi_0)}{\mu_0(D-n+n\frac{\mu_0}{2})(\frac{\mu_0}{2})(N-D-n+n\mu_0)}} \leq \frac{N}{\sqrt{2\pi n}} \sqrt{\frac{(1/4)(1-\psi_0)}{\mu_0(n\frac{\mu_0}{2})(\frac{\mu_0}{2})(n\mu_0)}} \\
&= \frac{1}{\frac{n}{N}\sqrt{n}} \sqrt{\frac{2(1/4)(1-\psi_0)}{\pi\mu_0^4}} \leq \frac{1}{\sqrt{n}} \frac{\sqrt{(1-\psi_0)}}{\psi_0\sqrt{2\pi\mu_0^4}}. \qquad (2.18)
\end{aligned}$$

Combining (2.18) with (2.17) and (2.8), we have the bound

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

where

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

Case 2

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

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

We can bound the A term by

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

Taking the $\exp\left(-\frac{n}{12}u^4\right)$ term from (2.17) we have

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

This is maximized at

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

and so

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

Combining the remaining terms in (2.7) together with this bound of the A term and the C bound of 1 yields

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

where

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

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

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

We can bound C' by

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

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

$$\begin{aligned}
B' &= \frac{D^D(N-n)^{(N-n)}}{(D-n)^{(D-n)}N^N} = \frac{\left(\frac{D}{N}\right)^D\left(\frac{N-n}{N}\right)^{(N-D)}}{\left(\frac{D-n}{N-n}\right)^{D-n}} \\
&= \left(\frac{\frac{N-n}{N}}{\frac{D-n}{D}}\right)^{D-n}\left(\frac{\frac{N-n}{N}}{\frac{N-n-(D-n)}{N-D}}\right)^{(N-D)}\left(\frac{D}{N}\right)^n \\
&= \exp\left(-\Gamma + n\log(\mu)\right)
\end{aligned}$$

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

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

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

$$\begin{aligned}
B' &\leq \exp\left(-(N-n)\left[2\left(\frac{f}{\bar{f}}u\right)^2 + \frac{\left(\frac{f}{\bar{f}}u\right)^4}{3}\right] + n\log(1-u)\right) \\
&= \exp\left(-2\left(\frac{n^2}{N-n}\right)u^2 - \frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4 + n\log(1-u)\right) \\
&\leq \exp\left(-2\left(\frac{n^2}{N-n}\right)u^2 - \frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4 - n\left[\sum_{k=1}^{7}\frac{u^k}{k}\right]\right) \\
&= \exp\left(-\left(\frac{2nN}{N-n}\right)u^2 - \frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4 - nu + \frac{3nu^2}{2} - \frac{nu^3}{3} - \frac{nu^6}{6} - \frac{nu^7}{7}\right) \\
&\cdot \exp\left(-\frac{nu^4}{4}\right)\exp\left(-\frac{nu^5}{5}\right) \\
&\leq \exp\left(-\frac{2n}{1-\frac{m}{N}}u^2\right)\exp\left(-\frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right)u^4\right)\exp\left(-\frac{nu^4}{4}\right)\exp\left(-\frac{nu^5}{5}\right) ,
\end{aligned}$$

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

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

where the final inequality uses the fact that $D - n \geq 1$ in this case. Taking the expression $\exp\left(-\frac{nu^5}{5}\right)$ 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

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

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

$$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \leq \frac{K_1}{\sqrt{n}} \exp\left(-\frac{2n}{1-\frac{n}{N}} u^2\right) \exp\left(-\frac{1}{3}\left(\frac{n^4}{(N-n)^3}\right) u^4\right) \exp\left(-\frac{nu^4}{4}\right) . \tag{2.19}$$

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

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

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

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

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 (2.19) we have

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

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

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

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

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

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

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

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

Chapter 3
THE DE MOIVRE-TUSNÁDY METHOD

The central limit theorem, too, is found / In Abe de Moivre's book, *Doctrine of Chances*, and if / It's only for binomial distributions, well / Now only after Fourier and then by Paul Lévy / Is rendered easy such a proof.

Richard M. Dudley [9]

3.1 Method Description

The method used to obtain upper bounds on the tail of the hypergeometric distribution is simple in outline. Because of its simplicity, it is applicable to many discrete distributions. We provide an overview of the method here. We will refer to it as the *de Moivre-Tusnády method*.

3.1.1 The de Moivre-Tusnády method

Suppose D is a discrete random variable such that $P(D \in \mathbb{N}) = 1$ and $E(D) = \mu < \infty$. Let $n \in \mathbb{N}$ be such that $P(D = n) > 0$ and $P(D > n) = 0$ (we will abuse notation and allow that n is potentially infinite). Also suppose C is a continuous random variable with density $f_C(x)$. We wish to compare the tail of D to the tail of C. In the case where direct comparison is difficult, we proceed as follows.

1. Obtain upper bounds on individual discrete probabilities for $P(D = k)$. Specifically, we find functions $U_I(\cdot)$ and $U_T(\cdot)$ defined on $\mathbb{R}^+$ such that:

- If n is finite, find upper bounds for $k \geq \mu$, $k < n$ so that

$$P(D = k) \leq U_I(k) \tag{3.1}$$

and when $k = n$

$$P(D = n) \leq U_T(n)\ . \tag{3.2}$$

Notice $U_I(\cdot)$ and $U_T(\cdot)$ are distinguished since in application they may take different forms.

- If $n = \infty$, find upper bounds for all $k \geq \mu$ so that

$$P(D = k) \leq U_I(k)\ . \tag{3.3}$$

2. Obtain lower bounds on special intervals related to the continuous distribution C. Here, we find functions $L_I(\cdot)$ and $L_T(\cdot)$ defind on $\mathbb{R}^+$ such that:

- If n is finite, when $k \geq \mu$ and $k < n$, find constants a_k and b_k (with $a_k < b_k$) so that

$$P(a_k \leq C \leq b_k) = \int_{a_k}^{b_k} f_C(t)\ dt \geq L_I(k) \tag{3.4}$$

and also $b_k = a_{k+1}$. When $k = n$, set $a_n \equiv b_{n-1}$ and find a lower bound for

$$P(C \geq a_n) = \int_{a_n}^{\infty} f_C(t)\ dt \geq L_T(n)\ . \tag{3.5}$$

Notice $L_I(\cdot)$ and $L_T(\cdot)$ are distinguished since in application they may take different forms.

- If $n = \infty$, for all $k \geq \mu$ find constants a_k and b_k (with $a_k < b_k$) so that

$$P(a_k \leq C \leq b_k) = \int_{a_k}^{b_k} f_C(t)\ dt \geq L_I(k)\ . \tag{3.6}$$

3. Parametrize the upper bounds $U_I(\cdot), U_T(\cdot)$ and the lower bounds $L_I(\cdot), L_T(\cdot)$ on a common scale x.

4. Compare the bounds on an interval that contains the support of D on the x-scale. Show $U_I(x) \leq L_I(x)$ and $U_T(x) \leq L_T(x)$ for any choice of x in this interval.

The upper bounds appearing in (3.1), (3.2), and (3.3) arise naturally when the discrete probability contains factorial terms: we apply refinements of Stirling's inequality to obtain these bounds.

It is less obvious how to find the lower bounds appearing in (3.4), (3.5), and (3.6). Asymptotic theory and existing results on exponential bounds can provide clues. We will illustrate how this occurs by considering the example of the Poisson distribution in the next section.

Determining the common scale x depends on the goal of the analysis. If we wish to find an upper bound that holds for a specific parameter choice for a given discrete distribution, the common scale may differ from that used if our goal is to find an upper bound that holds for many (if not all) possible parameter values.

The comparison in the final point may be carried out by taking the continuous relaxation of the scale x. Once relaxed, analytical arguments may demonstrate the bound holds on a set that contains the discrete values attainable by the distribution.

Supposing this method works for a discrete distribution D, we find it yields a continuous upper bound on the discrete tail since for $k \geq \mu$ we have

$$P(D \geq k) = \sum_{j=k}^{n} P(D = k) \leq \sum_{j=k}^{n} \int_{a_j}^{b_j} f_C(t)\ dt = \int_{a_k}^{\infty} f_C(t)\ dt = P(C \geq a_k)\ ,$$

with the penultimate equality following by the linearity of the integral.

We illustrate this method in the next section through its application to the Poisson distribution.

3.2 Polynomial Notation

Often in this thesis, a proof relies on determining the sign of a multivariate polynomial on some set. These polynomials arise when comparing various quantities that emerge when applying the de Moivre-Tusnády method in section 3.1.1. They do not have much interest outside the scope of the proof.

The sign arguments usually proceed by fixing all variables except one, and then repeatedly taking the partial derivatives of these polynomials until the sign of a partial derivative may be determined by inspection (according to the ambient assumptions). Then, monotonicity, convexity, and concavity are used to characterize the behavior of the intermediate partial derivatives. Surprisingly, the original polynomial often turns out to be monotone increasing in the variable under consideration. This lets us eliminate the variable under consideration, by obtaining a bound that holds for any admissible choice of the fixed variables.

We introduce here notation to help us work with these polynomials. Suppose that $f(w,x,y,z)$ is a multivariate polynomial in w,x,y, and z. We write

$$p_1(w) := f(w,x,y,z)$$

to indicate that we are viewing x,y, and z in $f(w,x,y,z)$ as **fixed**. The symbol ":=" is read "is defined to be". We then use standard prime notation as short-hand for partial derivatives with respect to the free variable. For example,

$$p_1^{(k)}(w) := \frac{\partial^k}{\partial w^k} p_1(w) = \frac{\partial^k}{\partial w^k} f(w,x,y,z) \ .$$

We also will often want to evaluate a partial derivative at a given value, and then treat an additional coordinate as fixed. We use conditional notation to address this situation. For example, we write

$$p_2(z) := p_2''(z|w=1) = \frac{\partial^2}{\partial w^2} p_1(w)\Big|_{w=1} \ .$$

The context determines that $p_2(z)$ is defined in terms of the partial derivative of $p_1(w)$. Where ambiguity arises in the hypergeometric analysis, we will add clarifying remarks.

Finally, since these polynomials occur so frequently, we use numbers to refer to them within the scope of a single argument. Therefore, $p_i(x)$ in Lemma A is different from $p_i(w)$ in Proposition B (and both differ from $p_i(w)$ in Theorem C). We now turn to the Poisson example.

3.3 The Poisson Tail

We illustrate the de Moivre-Tusnády method 3.1.1 by considering Poisson probabilities. In this section, we suppose $X \sim \text{Poisson}(\lambda)$ and that $k \geq \lambda$.

Inequalities for Poisson probabilities have been long considered. Teicher obtained lower bounds on the left tail [51]; Glynn obtained upper bounds on the Poisson tail through comparison to the Normal tail [16]; more recently, Philips and Nelson found upper bounds for the Poisson tail by comparing to the moment bound [37]. In fact, Philips and Nelson's approach yields the forthcoming (3.9) for integer values of λ.

Philips and Nelson also compare their result to an exponential bound obtained via the Cramér - Chernoff method. This bound contains a function $h(x)$, which we define

$$h(x) := x(\log(x) - 1) + 1 \ . \tag{3.7}$$

For $k \geq \lambda$, we have $h(\lambda/\lambda) = 0$ and $h(k/\lambda)$ is a monotone increasing function [46, pg. 440]. Boucheron, Lugosi, and Massart [5, pg. 23] show how to apply the Cramér - Chernoff method to obtain the following bound (note that their function h is our $h(1 + x)$).

Proposition 1. *(MGF Bound) Let $X \sim$ Poisson(λ). Then*

$$P(X \geq k) \leq \exp\left(-\lambda h(k/\lambda)\right) \tag{3.8}$$

for all $k \geq \lambda$.

The following theorem summarizes the bounds obtained in this section.

Theorem 5. *Suppose $1/\sqrt{2} \leq \lambda \leq 4$ and $X \sim$ Poisson(λ). Then for all $k \geq \lambda$ with $k \in \mathbb{N}$ we have*

$$P(X \geq k) \leq \sqrt{\frac{\lambda}{k}} \exp\left(-\lambda \cdot h\left(\frac{k}{\lambda}\right)\right) \tag{3.9}$$

where

$$h(x) := x(\log(x) - 1) + 1 \ .$$

If we suppose instead that $\lambda \geq 4$ *and* $k \geq 2\lambda$ *with* $k \in \mathbb{N}$ *then we have*

$$P(X \geq k) \leq \sqrt{\frac{\lambda}{2\pi k}} \exp\left(-\lambda \cdot h\left(\frac{k}{\lambda}\right)\right) \ . \tag{3.10}$$

Finally, if we suppose $c \geq 2\pi$, $\alpha > 1$, *and*

$$\lambda \geq \frac{c \cdot \alpha^2}{2\pi(\alpha-1)^2} \ ,$$

then for $k \geq \alpha\lambda$ *with* $k \in \mathbb{N}$ *we have*

$$P(X \geq k) \leq \sqrt{\frac{\lambda}{ck}} \exp\left(-\lambda \cdot h\left(\frac{k}{\lambda}\right)\right) \ . \tag{3.11}$$

Proof. We first prove the bound (3.9). Under the assumptions we have for any $k \geq \lambda$ that

$$\begin{aligned} P(X \geq k) &= \sum_{j=k}^{\infty} P(X = j) \\ &\leq \sum_{j=k}^{\infty} \left(\int_j^{j+1} \left[\frac{1}{2t}\sqrt{\frac{\lambda}{t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) + \sqrt{\frac{\lambda}{t}} \log\left(\frac{t}{\lambda}\right) \exp\left(-\lambda \cdot h(t/\lambda)\right) \right] dt \right) \\ &= \int_k^{\infty} \left[\frac{1}{2t}\sqrt{\frac{\lambda}{t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) + \sqrt{\frac{\lambda}{t}} \log\left(\frac{t}{\lambda}\right) \exp\left(-\lambda \cdot h(t/\lambda)\right) \right] dt \ , \end{aligned}$$

with the inequality following by Lemma 10 . But the claim follows since

$$\begin{aligned} &\int_k^{\infty} \left[\frac{1}{2t}\sqrt{\frac{\lambda}{t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) + \sqrt{\frac{\lambda}{t}} \log\left(\frac{t}{\lambda}\right) \exp\left(-\lambda \cdot h(t/\lambda)\right) \right] dt \\ &= -\sqrt{\frac{\lambda}{t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) \Big|_k^{\infty} \\ &= \sqrt{\frac{\lambda}{k}} \exp\left(-\lambda \cdot h(k/\lambda)\right) \ . \end{aligned}$$

To prove the bound (3.10), we assume that $\lambda \geq 4$, $k \geq 2\lambda$, and repeat the argument using Lemma 11 in place of Lemma 10. Similarly, to prove (3.11), we assume the relevant assumptions hold and use Lemma 12 in place of Lemma 10. □

Theorem 5 is in the spirit of Talagrand's method to elicit "the missing factors in Hoeffding's inequality" [50]. The constants found in Theorem 5 for the special case of the Poisson

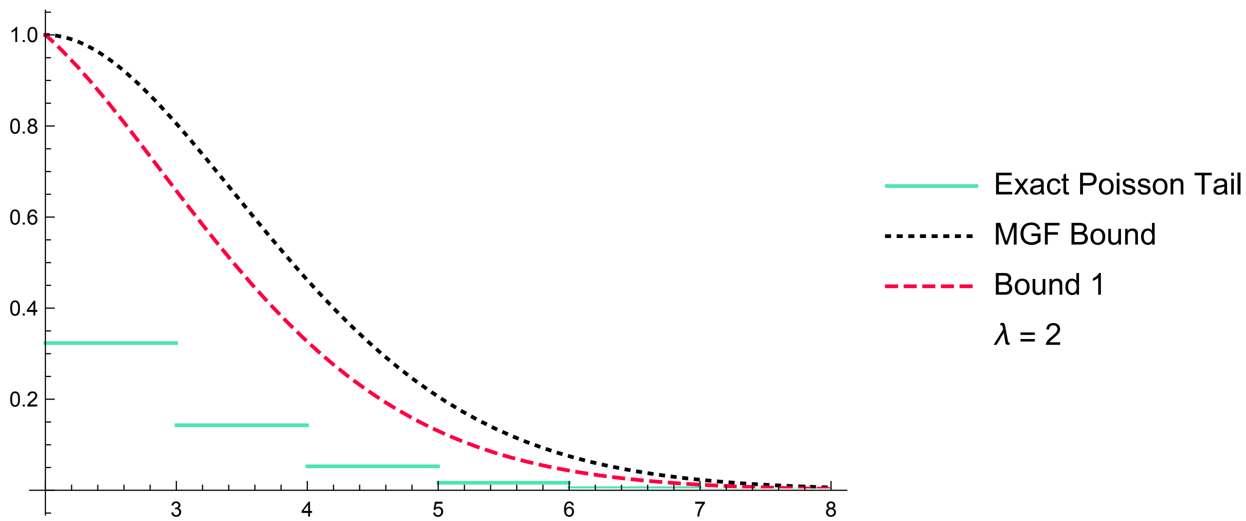

Figure 3.1: A plot of exact Poisson tail probabilities for $X \sim \text{Poisson}(2)$. The bound (3.12) is called the "MGF Bound" in the plot. The bound (3.9) is called "Bound 1". We plot $\lambda = 2 \leq k \leq 8 = 4\lambda$.

are more explicit than those appearing in [50], and the de Moivre-Tusnády method 3.1.1 differs from Talagrand's.

We prove Theorem 5 using the de Moivre-Tusnády method. The de Moivre-Tusnády method also differs from Philips and Nelson's [37] (which relies on factorial moments), and (3.9) extends the bound appearing in their paper to all values of $\lambda \in [1/\sqrt{2}, 4]$. The Poisson analysis is meant as prelude to the more difficult hypergeometric tail problem, where we also use the de Moivre-Tusnády method. The hypergeometric case will be treated in detail in chapters 4 and 5.

In the following figures, we compare the bounds (3.9) and (3.10) to the bound

$$b(\lambda) = \exp(-\lambda h(k/\lambda)) \ , \tag{3.12}$$

and exact Poisson tail probabilities.

We see that the proof of Theorem 5 relies on three Lemmas. Those Lemmas are the fruits of the de Moivre-Tusnády method outlined in section 3.1.1, and are used to obtain upper bounds on the Poisson tail. We now show how the method 3.1.1 yields the Lemmas.

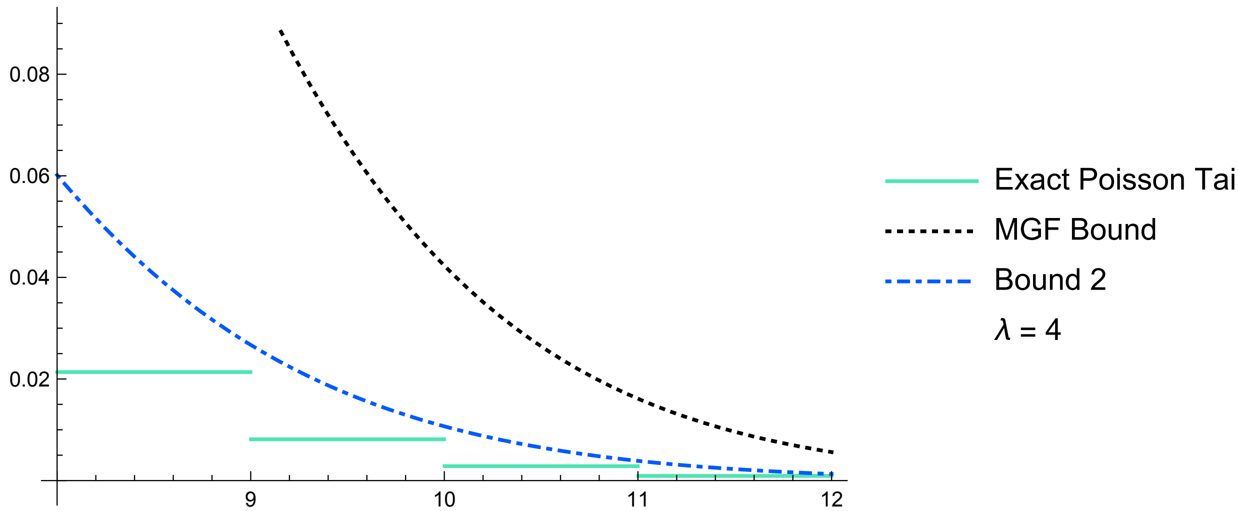

Figure 3.2: A plot of exact Poisson tail probabilities for $X \sim \text{Poisson}(4)$. The bound (3.12) is called the "MGF Bound" in the plot. The bound (3.10) is called "Bound 2". We plot $2\lambda = 8 \leq k \leq 12 = 3\lambda$.

Recall that an individual Poisson probability is

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \ .$$

We seek an upper bound on this probability that removes the $k!$ term. We apply Stirling's formula in Nanjundiah's form (see Lemma 13 stated ahead in section 3.4) to obtain

$$\begin{aligned}
P(X = k) &= e^{-\lambda}\frac{\lambda^k}{k!} \\
&\leq e^{-\lambda}\frac{\lambda^k}{\sqrt{2\pi k}(k/e)^k \exp\left(\frac{1}{12k} - \frac{1}{360k^3}\right)} \qquad \text{by Nanjundiah (Lemma 13)} \\
&= \frac{1}{\sqrt{2\pi k}} \exp\left(k - \lambda + k\log(\lambda) - k\log(k) - \frac{1}{12k} + \frac{1}{360k^3}\right) \\
&\leq \frac{1}{\sqrt{2\pi k}} \exp\left(k - \lambda + k\log(\lambda) - k\log(k)\right) \\
&= \frac{1}{\sqrt{2\pi k}} \exp\left(-\lambda\left[\left(\frac{k}{\lambda}\right)\left(\log\left(\frac{k}{\lambda}\right) - 1\right) + 1\right]\right) \\
&= \frac{1}{\sqrt{2\pi k}} \exp\left(-\lambda\left[h(k/\lambda)\right]\right) \ , \qquad (3.13)
\end{aligned}$$

where $h(x)$ is as defined in (3.7). We have thus obtained an upper bound satisfying the first component of the method (3.3). Namely, when $X \sim \text{Poisson}(\lambda)$, for $k \geq \lambda$ we have by (3.13)

$$P(X = k) \leq \frac{1}{\sqrt{2\pi k}} \exp\left(-\lambda \cdot h(k/\lambda)\right) \tag{3.14}$$

where $h(x)$ is defined by (3.7) .

To implement the next step of the method, (3.6), we need to find a continuous quantity to which we may compare the upper bound (3.14).

Comparing (3.8) to (3.14), we might hope to obtain a factor outside the exponential by pursuing a direct analysis, something along the lines of

$$P(X \geq k) \leq \sqrt{\frac{\lambda}{c \cdot k}} \exp\left(-\lambda h(k/\lambda)\right) \tag{3.15}$$

for a constant $c > 1$. Thinking of the right side of (3.15) as a continuous survival function, treating k as continuous, and λ, c as fixed, we differentiate to obtain

$$g_{\lambda,c}(k) := \frac{1}{2k}\sqrt{\frac{\lambda}{c \cdot k}} \exp\left(-\lambda \cdot h(k/\lambda)\right) + \sqrt{\frac{\lambda}{c \cdot k}} \log\left(\frac{k}{\lambda}\right) \exp\left(-\lambda \cdot h(k/\lambda)\right) \ .$$

Then for $\lambda \leq a < b$, we have

$$\int_a^b g_{\lambda,c}(t)\ dt = G_{\lambda,c}(b) - G_{\lambda,c}(a)\ ,$$

where

$$G_{\lambda,c}(t) := -\sqrt{\frac{\lambda}{c \cdot t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) \ .$$

In particular, for $a = k$ and $b = k + 1$ this implies

$$\begin{aligned} \int_k^{k+1} g_{\lambda,c}(t)\ dt &= G_{\lambda,c}(k+1) - G_{\lambda,c}(k) \\ &= \sqrt{\frac{\lambda}{c \cdot k}} \exp\left(-\lambda \cdot h(k/\lambda)\right) - \sqrt{\frac{\lambda}{c(k+1)}} \exp\left(-\lambda \cdot h([k+1]/\lambda)\right) \ . \end{aligned} \tag{3.16}$$

We see (3.16) provides the quantity (3.6) described by the method 3.1.1. For Poisson probabilities, the bounds of integration are determined by $a_k = k$ and $b_k = k + 1$. These bounds are simple to determine for a Poisson random variable, since it has support on $\mathbb{N}$. If we treat k as continuous, the common scale "x" referred to by the method is simply the k-scale. Observe for any choice of $k \in [0, \infty)$, there is a unique Poisson probability contained in the interval $[k, k+1)$.

The final step of the method 3.1.1 has us compare (3.16) to (3.14): when does the former exceed the latter? For this to be true, we must have

$$\sqrt{\frac{\lambda}{c \cdot k}} \exp\left(-\lambda \cdot h(k/\lambda)\right) - \sqrt{\frac{\lambda}{c(k+1)}} \exp\left(-\lambda \cdot h([k+1]/\lambda)\right) \geq \frac{1}{\sqrt{2\pi k}} \exp\left(-\lambda \cdot h(k/\lambda)\right) \ .$$

Dividing through by the right-hand side, this occurs when

$$\sqrt{\frac{2\pi\lambda}{c}} - \sqrt{\frac{2\pi k\lambda}{c(k+1)}} \exp\left(-\lambda \left[h\left(\frac{k+1}{\lambda}\right) - h\left(\frac{k}{\lambda}\right)\right]\right) \geq \ 1 \ .$$

Re-arranging further, taking the natural logarithm, and subtracting, we find (3.16) is larger than (3.14) when

$$0 \geq -\log\left(1 - \sqrt{\frac{c}{2\pi\lambda}}\right) + \frac{1}{2}\log\left(\frac{k}{k+1}\right) - \lambda\left[h\left(\frac{k+1}{\lambda}\right) - h\left(\frac{k}{\lambda}\right)\right] =: d(k, \lambda, c) \ . \quad (3.17)$$

For the inequality at (3.17) to be well-defined, we must restrict $\lambda > c/(2\pi)$. This is because of the initial log-term which includes $c/(2\pi\lambda)$. However, this term does not depend on k. For the time being, we assume that $\lambda \geq 1/\sqrt{2}$ and $k \geq \lambda$. We analyze the behavior of (3.17) with respect to k. Differentiating twice, we find

$$\frac{\partial^2}{\partial k^2} d(k, \lambda, c) = \frac{2k^2 - 1}{2k^2(k+1)^2} > 0 \ .$$

The positivity of this expression follows by $\lambda \geq 1/\sqrt{2}$ and and $k \geq \lambda$. Thus the first partial derivative increases monotonically in k. At $k = \lambda$, we find

$$\left.\frac{\partial}{\partial k} d(k, \lambda, c)\right|_{k=\lambda} = \frac{1}{2\lambda(\lambda+1)} - \log\left(\frac{\lambda+1}{\lambda}\right) = \frac{1}{2\lambda(\lambda+1)}\left(1 - 2\lambda(\lambda+1)\log\left(\frac{\lambda+1}{\lambda}\right)\right) \ . \quad (3.18)$$

This is negative when

$$1 \leq 2\lambda(\lambda+1)\log\left(\frac{\lambda+1}{\lambda}\right) =: p_1(\lambda) \ .$$

The third derivative of $p_1(\lambda)$ is $2\lambda^{-2}(\lambda+1)^{-2} > 0$. This implies $p_1''(\lambda)$ increases in λ. At $\lambda = 1/\sqrt{2}$, we find $p_1''(\lambda) < -0.47$. Furthermore, as $\lambda \nearrow \infty$, $p_1''(\lambda) \nearrow 0$. This can be seen by writing

$$p_1''(\lambda) = -\frac{4}{\lambda+1} - \frac{2}{\lambda(\lambda+1)} + \frac{4}{1+\frac{1}{\lambda}} \log\left(1+\frac{1}{\lambda}\right) + \frac{4}{\lambda+1} \log\left(1+\frac{1}{\lambda}\right) .$$

A similar analysis shows $p_1'(\lambda) > 0$ for $\lambda \geq 1/\sqrt{2}$ (it decreases to 2 as $\lambda \nearrow \infty$). We conclude $p_1(\lambda)$ increases in λ. When $\lambda = 1/\sqrt{2}$, $p_1(1/\sqrt{2}) > 2.12$. Therefore, (3.18) is negative for $\lambda \geq 1/\sqrt{2}$. Returning to the function $d(k,\lambda,c)$ we also find we may write

$$\frac{\partial}{\partial k} d(k,\lambda) = \frac{1}{1+\frac{1}{k}} \log\left(\frac{1}{1+\frac{1}{k}}\right) + \frac{1}{k+1} \log\left(\frac{1}{1+\frac{1}{k}}\right) + \frac{1}{2k(k+1)} .$$

In this form, we see the first partial derivative increases to a limit of 0 as $k \nearrow \infty$. We conclude that when $\lambda \geq 1/\sqrt{2}$ that

$$\frac{\partial}{\partial k} d(k,\lambda,c) < 0 .$$

Thus, $d(k,\lambda,c)$ decreases in k for any choice of $\lambda \geq \left(1/\sqrt{2}\right) \vee (c/(2\pi))$.

Case: $c = 1$ **and** $1/\sqrt{2} \leq \lambda \leq 4$

We now analyze $d(k,\lambda,c)$ under the further restriction that $c = 1$ and $1/\sqrt{2} \leq \lambda \leq 4$. Evaluating $d(k,\lambda,1)$ at $k = \lambda$ we find

$$d(\lambda,\lambda,1) = 1 - \left(\lambda + \frac{3}{2}\right) \log\left(1+\frac{1}{\lambda}\right) - \log\left(1 - \frac{1}{\sqrt{2\pi\lambda}}\right) =: p_2(\lambda) . \tag{3.19}$$

Differentiating twice, we find

$$p_2''(\lambda) =: \frac{p_3(\lambda) + p_4(\lambda)}{2\left(2\sqrt{\pi\lambda} - \sqrt{2}\right)^2 \lambda^2(\lambda+1)^2} \tag{3.20}$$

where we define

$$p_3(\lambda) := -11\pi\lambda + 15\sqrt{2\pi\lambda} - 8$$

and

$$p_4(\lambda) := 22\sqrt{2\pi}\lambda^{3/2} + 3\sqrt{2\pi}\lambda^{5/2} - 16\pi\lambda^2 - 2\lambda^2 - \pi\lambda - 12\lambda .$$

We see the sign of (3.20) is determined by the sign of $p_3(\lambda)$ and $p_4(\lambda)$. We now show both are negative when $1/\sqrt{2} \le \lambda \le 4$.

We first consider $p_3(\lambda)$. Differentiating, we find

$$p_3'(\lambda) = 15\sqrt{\frac{\pi}{2\lambda}} - 11\pi < 12 \ .$$

The inequality follows by decrease in λ and evaluating at $\lambda = 1/\sqrt{2}$. Therefore, $p_3(\lambda)$ decreases when $\lambda \in [1/\sqrt{2}, \lambda]$.

The fourth derivative of $p_4(\lambda)$ may be written

$$\frac{9}{8\lambda^{5/2}}\sqrt{\frac{\pi}{2}}(22 - 5\lambda) \ .$$

For $1/\sqrt{2} \le \lambda \le (22/5)$, this implies $p_4''(\lambda)$ is convex. Since $4 < 22/5$, we see $p_4''(\lambda)$ is convex under our assumptions. At $\lambda = 1/\sqrt{2}$ we find $p_4''(\lambda) < -31$ and at $\lambda = 4$ that $p_4''(\lambda) < -27$. Convexity implies $p_4''(\lambda) < 0$ for $1/\sqrt{2} \le \lambda \le 4$. Therefore, $p_4'(\lambda)$ decreases in λ. At $\lambda = 1/\sqrt{2}$ we find $p_4'(\lambda) < 8$, which implies $p_4(\lambda)$ decreases in λ when $1/\sqrt{2} \le \lambda \le 4$. Evaluating at $\lambda = 1/\sqrt{4}$ we find $p_4(\lambda) < -0.88$. This implies $p_4(\lambda) < 0$ when $1/\sqrt{2} \le \lambda \le 4$.

We conclude $p_2''(\lambda) < 0$ when $1/\sqrt{2} \le \lambda \le 4$ by (3.20). This implies $p_2'(\lambda)$ decreases in λ for $\lambda \in [1/\sqrt{2}, 4]$. Evaluating $p_2'(\lambda)$ at $\lambda = 4$ we find it is greater than 0.02. Therefore, $p_2'(\lambda) > 0$ when $\lambda \in [1/\sqrt{2}, 4]$. Evaluating $p_2(\lambda)$ at $\lambda = 4$, we find it is smaller than -0.004.

We therefore have shown that $p_2(\lambda) < 0$ when $1/\sqrt{2} \le \lambda \le 4$. Combining (3.19) and (3.17), this implies (3.16) is larger than (3.14) for all $k \ge \lambda$ when $1/\sqrt{2} \le \lambda \le 4$. This proves the following Lemma:

Lemma 10. *Suppose* $1/\sqrt{2} \le \lambda \le 4$ *and* $X \sim$ *Poisson*(λ). *Then for all* $k \ge \lambda$ *we have*

$$P(X = k) \le \int_k^{k+1} \left[\frac{1}{2t}\sqrt{\frac{\lambda}{t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) + \sqrt{\frac{\lambda}{t}} \log\left(\frac{t}{\lambda}\right) \exp\left(-\lambda \cdot h(t/\lambda)\right)\right] \, dt$$

where

$$h(x) := x(\log(x) - 1) + 1 \ .$$

Case: $c = 2\pi$ **and** $\lambda \ge 4$ **and** $k \ge 2\lambda$

We now analyze $d(k,\lambda,c)$ under the further restrictions that $c = 2\pi$, $\lambda \geq 4$ and $k \geq 2\lambda$. Since $d(k,\lambda,2\pi)$ decreases in k, evaluating at $k = 2\lambda$ gives an upper bound of

$$\begin{aligned} p_5(\lambda) := d(2\lambda,\lambda,2\pi) = &\left[1 - \log\left(1 - \sqrt{\frac{1}{\lambda}}\right)\right] + \frac{1}{2}\log\left(\frac{2}{2+\frac{1}{\lambda}}\right) \\ &+ \frac{1}{\frac{1}{\lambda}}\left[\log(4) - 2\log\left(2 + \frac{1}{\lambda}\right)\right] - \log\left(2 + \frac{1}{\lambda}\right) \ . \end{aligned}$$

We show $p_5(\lambda) < 0$ when $\lambda \geq \pi$. Notice first that

$$1 - \log\left(1 - \sqrt{\frac{1}{\lambda}}\right)$$

decreases in λ for $\lambda \geq 4$. Hence, an upper bound on this expression is obtained simply by evaluating at $\lambda = 4$, where we see

$$1 - \log\left(1 - \sqrt{\frac{1}{\lambda}}\right) \leq 1 + \log(2) \ .$$

Therefore

$$p_5(\lambda) \leq 1 + \log(2) + \frac{1}{2}\log\left(\frac{2}{2+\frac{1}{\lambda}}\right) + \frac{1}{\frac{1}{\lambda}}\left[\log(4) - 2\log\left(2 + \frac{1}{\lambda}\right)\right] - \log\left(2 + \frac{1}{\lambda}\right) =: p_6(\lambda) \ .$$

Differentiating twice, we find

$$p_6''(\lambda) = -\frac{(8\lambda + 3)}{2\lambda^2(2\lambda+1)^2} < 0 \ .$$

Therefore, $p_6'(\lambda)$ decreases in λ. At $\lambda = 4$ we find $p_6'(\lambda)$ is larger than 0.0283. Writing

$$p_6'(\lambda) = \frac{2}{2\lambda+1} + \frac{3}{2\lambda(2\lambda+1)} + \frac{2\log(4)}{2+\frac{1}{\lambda}} + \frac{\log(4)}{2\lambda+1} - \frac{4}{2+\frac{1}{\lambda}}\log\left(2 + \frac{1}{\lambda}\right) - \frac{2}{2\lambda+1}\log\left(2 + \frac{1}{\lambda}\right) \ ,$$

we see that as $\lambda \nearrow \infty$ that $p_6'(\lambda) \searrow \log(4) - 2\log(2) = 0$. Therefore, $p_6'(\lambda) > 0$, and so $p_6(\lambda)$ increases in λ. At $\lambda = 4$, we find $p_6(\lambda) < -0.11$, and in the limit we see by L'Hôspital's that it tends to 0. Therefore, $p_6(\lambda) < 0$ for any $\lambda \geq 4$. This implies $p_5(\lambda) < 0$ too, which proves the following Lemma.

Lemma 11. *Suppose $\lambda \geq 4$ and $X \sim$ Poisson(λ). Then for all $k \geq 2\lambda$ we have*

$$P(X = k) \leq \int_k^{k+1} \left[\frac{1}{2t}\sqrt{\frac{\lambda}{2\pi t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) + \sqrt{\frac{\lambda}{2\pi t}} \log\left(\frac{t}{\lambda}\right) \exp\left(-\lambda \cdot h(t/\lambda)\right) \right] \, dt$$

where

$$h(x) := x(\log(x) - 1) + 1 \ .$$

Case: $\alpha > 1$**,** $c \geq 2\pi$**,** $k \geq \alpha\lambda$**, and** $\lambda \geq (c \cdot \alpha^2)/(2\pi(\alpha-1)^2)$

We now introduce an additional parameter α, where $\alpha > 1$, $\alpha \in \mathbb{R}$. We analyze $d(k, \lambda, c)$ under the further restrictions that $c > 2\pi$, $k \geq \alpha\lambda$ and

$$\lambda \geq \frac{c \cdot \alpha^2}{2\pi(\alpha - 1)^2} =: \lambda_{min} \ . \tag{3.21}$$

Since $d(k, \lambda, c)$ decreases in k, evaluating at $k = \alpha\lambda$ gives an upper bound of

$$p_7(\lambda) := d(\alpha\lambda, \lambda, c) = \left[1 - \log\left(1 - \sqrt{\frac{c}{2\pi\lambda}}\right)\right] + \frac{1}{\frac{1}{\alpha\lambda}}\left(\log(\alpha) - \log\left(\alpha + \frac{1}{\lambda}\right)\right)$$
$$- \log\left(\alpha + \frac{1}{\lambda}\right) - \frac{1}{2}\log\left(\frac{1}{\alpha\lambda} + 1\right) \ .$$

Similar to the previous case, we notice

$$1 - \log\left(1 - \sqrt{\frac{c}{2\pi\lambda}}\right) \leq 1 - \log\left(1 - \sqrt{\frac{c}{2\pi\left[\frac{c \cdot \alpha^2}{2\pi(\alpha-1)^2}\right]}}\right) = 1 - \log\left(1 - \frac{\alpha - 1}{\alpha}\right) = 1 + \log(\alpha) \ ,$$

with the inequality following by decrease in λ and our assumptions. Therefore,

$$p_7(\lambda) \leq 1 + \log(\alpha) + \frac{1}{\frac{1}{\alpha\lambda}}\left(\log(\alpha) - \log\left(\alpha + \frac{1}{\lambda}\right)\right) - \log\left(\alpha + \frac{1}{\lambda}\right) - \frac{1}{2}\log\left(1 + \frac{1}{\alpha\lambda}\right) =: p_8(\lambda) \ .$$

We show $p_8(\lambda) < 0$ for any finite λ obeying the assumptions. Differentiating twice we find

$$p_8''(\lambda) = -\frac{4\alpha\lambda + 3}{2\lambda^2(1 + \alpha\lambda)^2} < 0 \ .$$

Note this is negative for any $\alpha > 0$ and $\lambda > 0$. Thus, $p_8'(\lambda)$ decreases in λ. We find

$$p_8'(\lambda) = \frac{\alpha}{\alpha\lambda + 1} + \frac{3}{2\lambda(\alpha\lambda + 1)} - \alpha\log\left(1 + \frac{1}{\alpha\lambda}\right) \ . \tag{3.22}$$

Since we assume $c \geq 2\pi$, $\alpha > 1$, and $\lambda \geq \lambda_{min}$ (defined by (3.21)), we have that $\lambda > 1$. Evaluating $p_8'(\lambda)$ at $\lambda = 1$ we find it equals

$$p_9(\alpha) := \frac{\alpha}{\alpha+1} + \frac{3}{2(\alpha+1)} - \alpha \log\left(1 + \frac{1}{\alpha}\right) .$$

Differentiating we find

$$p_9''(\alpha) = \frac{2\alpha+1}{\alpha(\alpha+1)^3} > 0 .$$

Hence $p_9'(\alpha)$ increases in α. We may write

$$p_9'(\alpha) = \frac{\alpha}{(\alpha+1)^2} + \frac{1}{2(\alpha+1)^2} - \log\left(1 + \frac{1}{\alpha}\right) .$$

At $\alpha = 1$, this is smaller than -0.31. Since $p_9'(\alpha)$ increases in α, and since the way we have written it shows its limit is 0, we conclude $p_9'(\alpha) < 0$ for any finite α. Therefore, $p_9(\alpha)$ decreases in α. At $\alpha = 1$, we find $p_9(\alpha) > 0.55$. Its limit is 0 as $\alpha \nearrow \infty$. By monotone decrease, we conclude $p_9(\alpha) > 0$. Therefore, $p_8'(\lambda) > 0$ when $\lambda = 1$, From the form of $p_8'(\lambda)$ at (3.22), we see as $\lambda \nearrow \infty$ that $p_8'(\lambda) \searrow 0$. By monotone decrease, we conclude $p_8'(\lambda) > 0$. This implies $p_8(\lambda)$ increases in λ for $\lambda \in [1, \infty)$.

Evaluating $p_8(\lambda)$ at $\lambda = 1$, we find it equals

$$p_{10}(\alpha) :=\leq 1 + \log(\alpha) + \alpha\left(\log(\alpha) - \log(\alpha+1)\right) - \log(\alpha+1) - \frac{1}{2}\log\left(1 + \frac{1}{\alpha}\right) .$$

Differentiating, we find

$$p_{10}''(\alpha) = -\frac{4\alpha+3}{2\alpha^2(\alpha+1)^2} < 0 ,$$

implying $p_{10}'(\alpha)$ increases in α. Since at $\alpha = 1$, $p_{10}'(\alpha) > 0.55$, and

$$p_{10}'(\alpha) = \frac{1}{\alpha+1} + \frac{3}{2\alpha(\alpha+1)} - \log\left(1 + \frac{1}{\alpha}\right) \searrow 0 \text{ as } \alpha \nearrow \infty ,$$

we see by monotone decrease that $p_{10}'(\alpha) > 0$ under the assumptions. Thus $p_{10}(\alpha)$ increases in α. At $\alpha = 1$, we find $p_{10}(\alpha) < -0.73$. Furthermore as $\alpha \nearrow \infty$ we see $p_{10}(\alpha) \to 0$. By monotone increase, we conclude $p_{10}(\alpha) < 0$. This implies $p_8(\lambda) < 0$ when $\lambda = 1$. By L'Hôspital's, we see $p_8(\lambda) \to 0$ as $\lambda \nearrow \infty$ for any choice of α. By monotone increase for

$\lambda \in [1, \infty)$ we conclude $p_8(\lambda) < 0$ for any $\lambda \in [1, \infty)$. In particular, this implies $p_8(\lambda) < 0$ when $\lambda \in [\lambda_{min}, \infty)$, since $\lambda_{min} \geq 1$ under the assumptions.

This proves the following Lemma.

Lemma 12. *Suppose $c \geq 2\pi$, $\alpha > 1$, and*

$$\lambda \geq \frac{c \cdot \alpha^2}{2\pi(\alpha - 1)^2} =: \lambda_{min} \ .$$

Let $X \sim Poisson(\lambda)$. Then for all $k \geq \alpha\lambda$ we have

$$P(X = k) \leq \int_k^{k+1} \left[\frac{1}{2t}\sqrt{\frac{\lambda}{c \cdot t}} \exp\left(-\lambda \cdot h(t/\lambda)\right) + \sqrt{\frac{\lambda}{c \cdot t}} \log\left(\frac{t}{\lambda}\right) \exp\left(-\lambda \cdot h(t/\lambda)\right) \right] \, dt$$

where

$$h(x) := x(\log(x) - 1) + 1 \ .$$

3.4 Revisiting Upper Bounds on Hypergeometric Probabilities

The Poisson distribution provides an instructive example of the difficulties that arise when applying the method of section 3.1.1. Notice that we are able to take $k = \lambda$ only in Lemma 10; Lemma 11 only applies for $k \geq 2\lambda$ and Lemma 12 only for $k \geq \alpha\lambda$. There is a gap in the regions where the resulting bound applies that depends on the mean (and hence variance) of the Poisson distribution. A similar gap will appear in the analysis of the hypergeometric tail. We now turn to this problem.

Once again, we wish to apply the method 3.1.1. Now however, we wish to bound individual hypergeometric probabilities. These are determined by products of binomial coefficients. We have

$$p_{n,k} := P\left(H_{n,D,N} = k\right) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \ .$$

In this section, and throughout the rest of the thesis, we will analyze these probabilities in the special case that $n \leq D \leq N/2$. To obtain upper bounds on these hypergeometric probabilities, we again use Nanjundiah's refinement [35] of Stirling's formula. As noted by Nanjundiah, his formula follows from a refinement of Robbin's earlier analysis [42]. We include Nanjundiah's formula here for ease of reference.

Lemma 13. *(Nanjundiah, 1959) For $n \in \{1, 2, 3, \dots\}$ we have*

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \exp\left(\frac{1}{12n} - \frac{1}{360n^3}\right) \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \exp\left(\frac{1}{12n}\right) \ .$$

Lemma 14. *Let $N \geq 10,\ n,\ D \in \mathbb{N}$, $3 \leq n \leq D \leq N/2$. Let $H_{n,D,N} \sim$ Hypergeometric(n, D, N), and define*

$$\begin{aligned}
x \equiv x_k &:= \frac{Nk - nD}{n(N-D)} \ , \\
k^* &:= \left\lceil n\frac{D}{N} \right\rceil \ , \\
A_1(x) &:= \frac{1}{2}\log(1-x) \ , \\
A_2(x) &:= \frac{1}{2}\log\left(1 + \frac{N-D}{D}x\right) \ , \\
A_3(x) &:= \frac{1}{2}\log\left(1 - \frac{n(N-D)}{D(N-n)}x\right) \ , \\
A_4(x) &:= \frac{1}{2}\log\left(1 + \frac{n}{N-n}x\right) \ , \\
B_1(x) &:= \frac{n(N-D)}{N}(1-x)\log(1-x) \ , \\
B_2(x) &:= \frac{nD}{N}\left(1 + \frac{N-D}{D}x\right)\log\left(1 + \frac{N-D}{D}x\right) \ , \\
B_3(x) &:= \frac{(N-n)D}{N}\left(1 - \frac{n(N-D)}{D(N-n)}x\right)\log\left(1 - \frac{n(N-D)}{D(N-n)}x\right) \ , \\
B_4(x) &:= \frac{(N-n)(N-D)}{N}\left(1 + \frac{n}{N-n}x\right)\log\left(1 + \frac{n}{N-n}x\right) \ , \\
C_1(x) &:= \frac{N}{12n(N-D)(1-x)} - \frac{N^3}{360n^3(N-D)^3(1-x)^3} \ , \\
C_2(x) &:= \frac{N}{12n(D+(N-D)x)} - \frac{N^3}{360n^3(D+(N-D)x)^3} \ , \\
C_3(x) &:= \frac{N}{12(D(N-n)-n(N-D)x)} - \frac{N^3}{360(D(N-n)-n(N-D)x)^3} \ , \\
C_4(x) &:= \frac{N}{12(N-D)(N-n+nx)} - \frac{N^3}{360(N-D)^3(N-n+nx)^3} \ , \\
\textit{and } R_H &:= \frac{1}{12N} - \frac{1}{360N^3} - \frac{1}{12n} - \frac{1}{12D} - \frac{1}{12(N-D)} - \frac{1}{12(N-n)} \ .
\end{aligned} \tag{3.23}$$

Then for $k^* \leq k < n$*,*

$$P(H_{n,D,N} = k) \leq \sqrt{\frac{N^3}{2\pi n D(N-D)(N-n)}} \cdot \exp\left(-\left[\sum_{i=1}^{4} A_i(x) + B_i(x) + C_i(x)\right] - R_H\right) \tag{3.24}$$

$$\leq \sqrt{\frac{N^3}{2\pi n D(N-D)(N-n)}} \cdot \exp\left(-\left[\sum_{i=1}^{4} A_i(x) + B_i(x)\right] - C_1(x) - C_2(x) - R_H\right) . \tag{3.25}$$

Furthermore, if $N/(n(N-D)) \leq x$*, then*

$$P(H_{n,D,N} = k) \leq \sqrt{\frac{N^3}{2\pi n D(N-D)(N-n)}} \cdot \exp\left(-\left[\sum_{i=1}^{4} A_i(x) + B_i(x)\right] - R_H\right) . \tag{3.26}$$

Proof. By definition we have

$$\begin{aligned}
P(H_{n,D,N} = k) &= \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \\
&= \frac{\frac{D!}{(D-k)!k!}\frac{(N-D)!}{(N-D-n+k)!(n-k)!}}{\frac{N!}{(N-n)!n!}} \\
&= \left[\frac{D!}{(D-k)!k!}\right]\left[\frac{(N-D)!}{(N-D-n+k)!(n-k)!}\right]\left[\frac{(N-n)!n!}{N!}\right] \\
&=: [A]\,[B]\,[C] \ ,
\end{aligned}$$

with the final line defining A, B, and C. Using Lemma 13 we have

$$\begin{aligned}
A &\leq \frac{\sqrt{2\pi D}\left(\frac{D}{e}\right)^D \exp\left(\frac{1}{12D}\right)}{\left(\frac{(D-k)}{e}\right)^{(D-k)}\sqrt{2\pi(D-k)}\exp\left(\frac{1}{12(D-k)} - \frac{1}{360(D-k)^3}\right)\left(\frac{k}{e}\right)^k\sqrt{2\pi k}\exp\left(\frac{1}{12k} - \frac{1}{360k^3}\right)} \\
&= \left[\exp\left(\frac{1}{12D} - \frac{1}{12k} - \frac{1}{12(D-k)} + \frac{1}{360(D-k)^3} + \frac{1}{360k^3}\right)\right]\left[\sqrt{\frac{D}{2\pi k(D-k)}}\right]\left[\frac{D^D}{(D-k)^{(D-k)}k^k}\right] \\
&=: [D]\,[E]\,[F]
\end{aligned}$$

with the final line defining D, E, and F. Similar analysis gives

$$
\begin{aligned}
B \leq & \left[\exp\left(\frac{1}{12(N-D)} - \frac{1}{12(N-D-n+k)} - \frac{1}{12(n-k)} + \frac{1}{360(N-D-[n-k])^3} + \frac{1}{360[n-k]^3}\right)\right] \\
& \cdot \left[\sqrt{\frac{N-D}{2\pi[n-k](N-D-[n-k])}}\right] \left[\frac{(N-D)^{(N-D)}}{(N-D-[n-k])^{(N-D-[n-k])}(n-k)^{(n-k)}}\right] \\
& =: [G]\,[H]\,[I] \,.
\end{aligned}
$$

Here, the final line defines G, H, and I. Using the lemma on the final term yields

$$
\begin{aligned}
C \leq & \frac{\left(\frac{N-n}{e}\right)^{N-n}\sqrt{2\pi(N-n)}\exp\left(\frac{1}{12(N-n)}\right)\left(\frac{n}{e}\right)^n\sqrt{2\pi n}\exp\left(\frac{1}{12n}\right)}{\left(\frac{N}{e}\right)^N\sqrt{2\pi N}\exp\left(\frac{1}{12N} - \frac{1}{360N^3}\right)} \\
& = \exp\left(\frac{1}{12(N-n)} + \frac{1}{12n} - \frac{1}{12N} + \frac{1}{360N^3}\right)\sqrt{2\pi\frac{(N-n)n}{N}}\frac{[N-n]^{N-n}n^n}{(N)^{(N)}} \\
& = \left[\exp\left(\frac{1}{12(N-n)} + \frac{1}{12n} - \frac{1}{12N} + \frac{1}{360N^3}\right)\right]\left[\sqrt{2\pi N\left(\frac{N-n}{N}\right)\left(\frac{n}{N}\right)}\right]\left[\left(\frac{N-n}{N}\right)^{N-n}\left(\frac{n}{N}\right)^n\right] \\
& =: [J][K][L] \,.
\end{aligned}
$$

Here, the final line defines J, K, and L. Combining these three bounds we have

$$
P(H_{m,n} = k) \leq [DGJ][EHK][FIL] \,.
$$

Consider first FIL. We have

$$
\begin{aligned}
FIL = & \left[\frac{D^D}{(D-k)^{(D-k)}k^k}\right]\left[\frac{(N-D)^{(N-D)}}{(N-D-[n-k])^{(N-D-[n-k])}(n-k)^{(n-k)}}\right]\left[\left(\frac{N-n}{N}\right)^{N-n}\left(\frac{n}{N}\right)^n\right] \\
= & \left[\frac{D^D}{(D-k)^{(D-k)}(N-D-[n-k])^{(N-D-[n-k])}}\right]\left[\frac{(N-D)^{(N-D)}}{k^k(n-k)^{(n-k)}}\right]\left[\left(\frac{N-n}{N}\right)^{N-n}\left(\frac{n}{N}\right)^n\right] \\
= & \left[\left(\frac{D^{D-k}}{(D-k)^{(D-k)}}\right)\left(\frac{(N-D)^{N-D-n+k}}{(N-D-n+k)^{(N-D-n+k)}}\right)\left(\frac{N-n}{N}\right)^{N-D-n+k}\left(\frac{N-n}{N}\right)^{D-k}\right] \\
& \cdot \left[\left(\frac{D^k}{k^k}\right)\left(\frac{(N-D)^{n-k}}{(n-k)^{(n-k)}}\right)\left(\frac{n}{N}\right)^{n-k}\left(\frac{n}{N}\right)^k\right] \\
=: & [\alpha][\beta] \qquad (3.27)
\end{aligned}
$$

Recall that we are assuming $4 \leq n < D < N/2$. Recall the definition

$$x_i := \frac{Nk - nD}{n(N-D)} \ .$$

We now examine the α and β terms. Considering the α term first, we see

$$\begin{aligned}
&\alpha \\
&= \left[\frac{D(N-n)}{N(D-k)}\right]^{D-k} \left[\frac{(N-n)(N-D)}{N(N-D-n+k)}\right]^{N-D-n+k} \\
&= \exp\left[-(D-k)\log\left(\frac{N(D-k)}{D(N-n)}\right) - (N-D-n+k)\log\left(\frac{N(N-D-n+k)}{(N-n)(N-D)}\right)\right] \quad (3.28) \\
&= \exp\left[-(N-n)\frac{D}{N}\begin{pmatrix}\left(1-\frac{N-D}{D}\cdot\frac{n}{N-n}\cdot\frac{Nk-nD}{n(N-D)}\right)\log\left(1-\frac{N-D}{D}\cdot\frac{n}{N-n}\cdot\frac{Nk-nD}{n(N-D)}\right) \\ +\frac{(N-D)}{D}\left(1+\frac{n}{N-n}\cdot\frac{Nk-nD}{n(N-D)}\right)\log\left(1+\frac{n}{N-n}\cdot\frac{Nk-nD}{n(N-D)}\right)\end{pmatrix}\right] \\
&= \exp\left[-(N-n)\frac{D}{N}\begin{pmatrix}\left(1-\frac{(N-D)n}{D(N-n)}\cdot x_i\right)\log\left(1-\frac{(N-D)n}{D(N-n)}\cdot x_i\right) \\ +\frac{(N-D)}{D}\left(1+\frac{n}{N-n}\cdot x_i\right)\log\left(1+\frac{n}{N-n}\cdot x_i\right)\end{pmatrix}\right] \ .
\end{aligned}$$

We now perform a similar analysis for the β term, to find

$$\begin{aligned}
&\beta \\
&= \left[\left(\frac{N-D}{n-k}\right)\left(\frac{n}{N}\right)\right]^{n-k}\left[\left(\frac{D}{k}\right)\left(\frac{n}{N}\right)\right]^{k} \\
&= \exp\left[-(n-k)\log\left(\frac{(n-k)N}{(N-D)n}\right) - k\log\left(\frac{kN}{Dn}\right)\right] \\
&= \exp\left[-\frac{n}{N}\left((N-D)\frac{N(n-k)}{n(N-D)}\log\left(\frac{(n-k)N}{(N-D)n}\right) + D\frac{Nk}{nD}\log\left(\frac{kN}{nD}\right)\right)\right] \quad (3.29) \\
&= \exp\left[-n\frac{D}{N}\begin{pmatrix}\frac{(N-D)}{D}\left(1-\frac{Nk-nD)}{n(N-D)}\right)\log\left(1-\frac{Nk-nD)}{n(N-D)}\right) \\ +\left(1+\frac{N-D}{D}\cdot\frac{Nk-nD}{n(N-D)}\right)\log\left(1+\frac{(N-D)}{D}\cdot\frac{Nk-nD}{n(N-D)}\right)\end{pmatrix}\right] \\
&= \exp\left[-n\frac{D}{N}\left(\frac{(N-D)}{D}(1-x_i)\log(1-x_i) + \left(1+\frac{N-D}{D}\cdot x_i\right)\log\left(1+\frac{(N-D)}{D}\cdot x_i\right)\right)\right] \ .
\end{aligned}$$

Combining these two analyses, and using the definitions of $B_1(x) - B_4(x)$ in (3.23), we see

$$FIL = \exp\left[-B_1(x) - B_2(x) - B_3(x) - B_4(x)\right] \ . \quad (3.30)$$

Next, we examine EHK to find

$$EHK = \sqrt{\left[\frac{D}{2\pi k(D-k)}\right]\left[\frac{N-D}{2\pi[n-k](N-D-[n-k])}\right]\left[2\pi N\left(\frac{N-n}{N}\right)\left(\frac{n}{N}\right)\right]}$$
$$= \sqrt{\left[\frac{N^3}{2\pi nD(N-D)(N-n)}\right]\left[\frac{(N-D)n}{N(n-k)}\right]\left[\frac{nD}{Nk}\right]\left[\frac{D(N-n)}{N(D-k)}\right]\left[\frac{(N-D)(N-n)}{N(N-D-n+k)}\right]}$$
$$= \sqrt{\frac{N^3}{2\pi nD(N-D)(N-n)}}\exp\left(-\frac{1}{2}\left(\begin{array}{c}\log\left[\frac{N(n-k)}{(N-D)n}\right]+\log\left[\frac{Nk}{nD}\right]\\ +\log\left[\frac{N(D-k)}{D(N-n)}\right]+\log\left[\frac{N(N-D-n+k)}{(N-D)(N-n)}\right]\end{array}\right)\right) . \tag{3.31}$$

Now

$$\log\left[\frac{N(n-k)}{(N-D)n}\right]+\log\left[\frac{Nk}{Dn}\right] = \log\left[1-x_i\right]+\log\left[1+\frac{N-D}{D}x_i\right] , \tag{3.32}$$

with the equality following by the work at (3.29). Similarly,

$$\log\left[\frac{N[D-k]}{D(N-n)}\right]+\log\left[\frac{N(N-D-[n-k])}{(N-D)(N-n)}\right] = \log\left[1-\frac{(N-D)n}{D(N-n)}\cdot x_i\right]+\log\left[1+\frac{n}{N-n}\cdot x_i\right] ,$$

with the equality following by the work at (3.28). Substituting these expressions into the EHK identity, and using the definitions of $A_1(x) - A_4(x)$ in (3.23), yields

$$EHK = \sqrt{\frac{N^3}{2\pi nD(N-D)(N-n)}}\exp\left(-A_1(x)-A_2(x)-A_3(x)-A_4(x)\right) .$$

Note this defines the function $L(x_i)$. Combining this with (3.30) gives

$$[EHK][FIL] = \sqrt{\frac{N^3}{2\pi nD(N-D)(N-n)}}\cdot\exp\left(-\left[\sum_{i=1}^{4}A_i(x)+B_i(x)\right]\right) . \tag{3.33}$$

We finally consider DGJ. Recall

$$\begin{aligned} DGJ = &\exp\left(\frac{1}{12(N-D)}+\frac{1}{12(N-n)}+\left[\frac{1}{360k^3}-\frac{1}{12k}\right]\right)\\ &\cdot\exp\left(\left[\frac{1}{360(N-D-[n-k])^3}-\frac{1}{12(N-n-(D-k))}\right]\right)\\ &\cdot\exp\left(\frac{1}{12D}+\left[\frac{1}{360(D-k)^3}-\frac{1}{12(D-k)}\right]\right)\\ &\cdot\exp\left(\left[\frac{1}{360[n-k]^3}-\frac{1}{12(n-k)}\right]+\frac{1}{12n}-\frac{1}{12N}+\frac{1}{360N^3}\right) .\end{aligned}$$

We consider each of the bracketed expressions in turn. First, we see

$$\begin{aligned}\frac{1}{360k^3}-\frac{1}{12k} &= \frac{1}{360\left(\frac{nD}{N}\right)^3\left(\frac{Nk}{nD}\right)^3}-\frac{1}{12\left(\frac{nD}{N}\right)\left(\frac{Nk}{nD}\right)}\\ &= \frac{1}{360\left(\frac{nD}{N}\right)^3\left(1+\frac{N-D}{D}x_i\right)^3}-\frac{1}{12\left(\frac{nD}{N}\right)\left(1+\frac{N-D}{D}x_i\right)}\\ &= \frac{N^3}{360n^3(D+(N-D)x_i)^3}-\frac{N}{12n(D+(N-D)x_i)}\ .\end{aligned}$$

Next, we may write

$$\begin{aligned}&\frac{1}{360(N-D-[n-k])^3}-\frac{1}{12(N-n-(D-k))}\\ &= \frac{1}{360\left(\frac{(N-D)(N-n)}{N}\right)^3\left(\frac{N(N-D-[n-k])}{(N-D)(N-n)}\right)^3}-\frac{1}{12\left(\frac{(N-D)(N-n)}{N}\right)\left(\frac{N(N-n-(D-k))}{(N-D)(N-N)}\right)}\\ &= \frac{1}{360\left(\frac{(N-D)(N-n)}{N}\right)^3\left(1+\frac{n}{N-n}x_i\right)^3}-\frac{1}{12\left(\frac{(N-D)(N-n)}{N}\right)\left(1+\frac{n}{N-n}x_i\right)}\\ &= \frac{N^3}{360(N-D)^3(N-n+nx_i)^3}-\frac{N}{12(N-D)(N-n+nx_i)}\ .\end{aligned}$$

Similarly, we find

$$\begin{aligned}\frac{1}{360(D-k)^3}-\frac{1}{12(D-k)} &= \frac{1}{360\left(\frac{D(N-n)}{N}\right)^3\left(\frac{N(D-k)}{(N-n)D}\right)^3}-\frac{1}{12\left(\frac{D(N-n)}{N}\right)\left(\frac{N(D-k)}{(N-n)D}\right)}\\ &= \frac{1}{360\left(\frac{D(N-n)}{N}\right)^3\left(1-\frac{n(N-D)}{(N-n)D}x_i\right)^3}-\frac{1}{12\left(\frac{D(N-n)}{N}\right)\left(1-\frac{n(N-D)}{(N-n)D}x_i\right)}\\ &= \frac{N^3}{360\left(D(N-n)-n(N-D)x_i\right)^3}-\frac{N}{12\left(D(N-n)-n(N-D)x_i\right)}\ .\end{aligned}$$

Finally we may write

$$\begin{aligned}\frac{1}{360(n-k)^3}-\frac{1}{12(n-k)} &= \frac{1}{360\left(\frac{(N-D)n}{N}\right)^3\left(\frac{N(n-k)}{(N-D)n}\right)^3}-\frac{1}{12\left(\frac{(N-D)n}{N}\right)\left(\frac{N(n-k)}{(N-D)n}\right)}\\ &= \frac{1}{360\left(\frac{(N-D)n}{N}\right)^3(1-x_i)^3}-\frac{1}{12\left(\frac{(N-D)n}{N}\right)(1-x_i)}\\ &= \frac{N^3}{360(N-D)^3n^3\left(1-x_i\right)^3}-\frac{N}{12(N-D)n\left(1-x_i\right)}\ .\end{aligned}$$

Substituting these expressions into the DGJ term, and combining it with the expression at (3.33) proves the claim at (3.24).

The claim at (3.25) will follow if we can show $C_3(x) \geq 0$ and $C_4(x) \geq 0$ under the assumptions. Notice first that

$$\begin{aligned} C_3(x) &= \frac{N}{360(D(N-n) - n(N-D)x)^3}(30(D(N-n) - n(N-D)x)^2 - N^2) \\ &=: \frac{N}{360(D(N-n) - n(N-D)x)^3}(p_3(x)) \ . \end{aligned}$$

Since the assumptions imply $0 \leq x \leq 1 - N/(n(N-D))$, the sign of $C_3(x)$ will follow the sign of $p_3(x)$. We see $p_3(x)$ decreases in x, and so

$$p_3(x) \geq p_3(1 - N/(n(N-D))) = N^2(29 + 60(D-n) + 30(D-n)^2) > 0 \ .$$

Hence $C_3(x) \geq 0$. Similarly, we find

$$\begin{aligned} C_4(x) &= \frac{N}{360(N-D)^3(N-n+nx)^3}(30(N-D)^2(N-n+nx)^2 - N^2) \\ &=: \frac{N}{360(N-D)^3(N-n+nx)^3}(p_4(x)) \ . \end{aligned}$$

Here, the sign will follow $p_4(x)$, which increases in x. Hence

$$p_4(x) \geq p_4(0) = 30(N-D)^2(N-n)^2 - N^2 \geq 30(N/2)^2(N/2)^2 - N^2 = (30/16)N^4 - N^2 > 0 \ ,$$

and so $C_4(x) \geq 0$ too. This implies (3.25).

To show (3.26), we must show $C_1(x)$ and $C_2(x)$ are positive when $N/(n(N-D)) \leq x \leq 1 - N/(n(N-D))$. Here we begin with $C_1(x)$ and find

$$\begin{aligned} C_1(x) &= \frac{N}{360n^3(1-x)^3(N-D)^3}\left(30n^2(1-x)^2(N-D)^2 - N^2\right) \\ &=: \frac{N}{360n^3(1-x)^3(N-D)^3}(p_1(x)) \ . \end{aligned}$$

the positivity of $C_1(x)$ will follow if we show $p_1(x) \geq 0$. But since $p_1(x)$ decreases in x , we have $p_1(x) \geq p_1(1 - N/(n(N-D))) = 29N^2 > 0$, and the claim follows. Hence $C_1(x) \geq 0$.

Finally, we turn to $C_2(x)$ and find

$$
\begin{aligned}
C_2(x) &= \frac{N}{360n^3(D+(N-D)x)^3}\left(30n^2(D+(N-D)x)^2 - N^2\right) \\
&=: \frac{N}{360n^3(D+(N-D)x)^3}(p_2(x)) \ .
\end{aligned}
$$

Here the sign will follow $p_2(x)$, which increases in x. Hence

$$
p_2(x) \geq p_2(N/(n(N-D))) = 30D^2n^2 + 60DnN + 29N^2 > 0 \ ,
$$

and so $C_2(x) \geq 0$. Since $C_1(x) \geq 0$ too, (3.26) follows, and the proof is complete. □

The notation of the common scale used in this thesis is from Dudley [10]. The motivation for this parametrization comes from Bretagnolle and Massart's use of Tusnády's lemma [6] to obtain explicit constants in their analysis of the Komlós-Major-Tusnády inequality [30].

Lemma 15. *Let* $N \geq 10,\ n,\ D \in \mathbb{N}$, $3 \leq n < D < N/2$. *Let* $H_{n,D,N} \sim$ *Hypergeometric*(n, D, N). *Then*

$$
\begin{aligned}
&P(H_{n,D,N} = n) \leq \\
&\sqrt{\frac{D(N-n)}{N(D-n)}} \exp\left((D-n)\log\left(\frac{D}{D-n}\right) + n\log\left(\frac{D}{N}\right) + (N-n)\log\left(\frac{N-n}{N}\right)\right) \\
&\qquad \cdot \exp\left(\frac{1}{12D} + \frac{1}{12(N-n)} - \frac{1}{12(D-n)} + \frac{1}{360(D-n)^3} - \frac{1}{12N} + \frac{1}{360N^3}\right) \ .
\end{aligned}
\tag{3.34}
$$

Proof. By definition and Lemma 13, we have

$$\begin{aligned}
P(H_{n,D,N} = n) &= \frac{\binom{D}{n}}{\binom{N}{n}} \\
&= \frac{D!}{(D-n)!n!}\frac{(N-n)!n!}{N!} \\
&\le \frac{\sqrt{2\pi D}\left(\frac{D}{e}\right)^D \exp\left(\frac{1}{12D}\right)\sqrt{2\pi(N-n)}\left(\frac{(N-n)}{e}\right)^{(N-n)}\exp\left(\frac{1}{12(N-n)}\right)}{\sqrt{2\pi(D-n)}\left(\frac{(D-n)}{e}\right)^{(D-n)}\exp\left(\frac{1}{12(D-n)} - \frac{1}{360(D-n)^3}\right)\sqrt{2\pi N}\left(\frac{N}{e}\right)^N \exp\left(\frac{1}{12N} - \frac{1}{360N^3}\right)} \\
&= \sqrt{\frac{D(N-n)}{N(D-n)}}\frac{D^D(N-n)^{(N-n)}}{(D-n)^{(D-n)}N^N} \\
&\cdot \exp\left(\frac{1}{12D} + \frac{1}{12(N-n)} - \frac{1}{12(D-n)} + \frac{1}{360(D-n)^3} - \frac{1}{12N} + \frac{1}{360N^3}\right) .
\end{aligned}$$

Now

$$\begin{aligned}
\frac{D^D(N-n)^{(N-n)}}{(D-n)^{(D-n)}N^N} &= \left(\frac{D}{D-n}\right)^{(D-n)} \left(\frac{D}{N}\right)^n \left(\frac{N-n}{N}\right)^{(N-n)} \\
&= \exp\left((D-n)\log\left(\frac{D}{D-n}\right) + n\log\left(\frac{D}{N}\right) + (N-n)\log\left(\frac{N-n}{N}\right)\right) .
\end{aligned}$$

Substituting this into the previous display gives the claim. □

3.5 Lower Bounds on Gaussian Probabilities

Here, we derive several lower bounds on the Normal tail that will prove useful in the sequel. We make use of lower bounds due to Feller and Dudley.

Lemma 16. *(Dudley [10, P.22]) For any $0 \le a < b$ and a standard normal variable Y,*

$$P(Y \in [a,b]) \ge \sqrt{\frac{1}{2\pi}}(b-a)\exp\left(-\frac{a^2}{4} - \frac{b^2}{4}\right)\phi(a,b) , \tag{3.35}$$

where

$$\phi(a,b) := \left[\frac{4}{(b^2-a^2)}\right]\sinh\left[\frac{(b^2-a^2)}{4}\right] \ge 1 .$$

Lemma 17. *(Feller [13]) Let Φ denote the standard normal CDF, and ϕ the density function. Then for $x > 0$*

$$1 - \Phi(x) \ge \phi(x)\left(\frac{1}{x} - \frac{1}{x^3}\right) = \frac{\phi(x)}{x}\left(1 - \frac{1}{x^2}\right) . \tag{3.36}$$

Lemma 18. *Let $N \geq 150$, n, $D \in \mathbb{N}$, $10 \leq n < D < N/2$, and $N/(2n(N-D)) \leq x \leq 1 - N/(n(N-D))$. Let $Y \sim$ Normal$(0,1)$. Then*

$$\sqrt{2}P\left(2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x-\frac{N}{2n(N-D)}\right) \leq Y \leq 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x+\frac{N}{2n(N-D)}\right)\right)$$
$$\geq \sqrt{\frac{4N}{\pi n(N-n)}} \cdot \exp\left(G_1(x)+R_N\right) \tag{3.37}$$

where we define

$$G_1(x) := -\frac{2n(N-D)^2}{N(N-n)}x^2 \ ,$$

and

$$R_N := -\frac{N}{2n(N-n)} \ .$$

Proof. We use Lemma 16. In the notation of that Lemma, we identify

$$a = 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x-\frac{N}{2n(N-D)}\right)$$

and

$$b = 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x+\frac{N}{2n(N-D)}\right) \ .$$

Therefore

$$\sqrt{\frac{2}{2\pi}}(b-a) = \sqrt{\frac{4N}{\pi n(N-n)}} \ .$$

Moreover

$$-\frac{a^2}{4} - \frac{b^2}{4} = -\frac{2n(N-D)^2}{N(N-n)}x^2 - \frac{N}{2n(N-n)} \ .$$

Using these expressions, along with the bound $\gamma(a,b) \geq 1$, we apply Lemma 16 to obtain the claim. □

In the previous lemma, the quantity

$$\frac{N}{2n(N-D)} \tag{3.38}$$

is of central importance. In Theorem 6 of chapter 4, we will define points on a lattice where we need to verify that the bound holds. The distance between points on this lattice depends on the parameters: for (n, D, N) subject to our constraints, they are separated by $N/(n(N-D))$.

The quantity (3.38) is half the distance at which mass changes on the lattice in Theorem 6. By defining Gaussian intervals equal to length $N/(n(N-D))$, but by shifting back by half that distance, Lemma 18 eliminates a linear x-term in the lower bound. This elimination is crucial for the subsequent proof, because it dramatically simplifies the argument demonstrating monotone increase. In shifting the interval by $N/(2n(N-D))$, we see an echo of the continuity correction for the binomial distribution.

Lemma 19. *Let $N \geq 150$, n, $D \in \mathbb{N}$, $10 \leq n < D < N/2$, and $0 \leq x < N/(2n(N-D))$. Let $Y \sim Normal(0,1)$. Then*

$$\sqrt{2}P\left(2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x-\frac{N}{2n(N-D)}\right) \leq Y \leq 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x+\frac{N}{2n(N-D)}\right)\right)$$
$$\geq \sqrt{\frac{1}{\pi nN(N-n)}}w(x)\exp\left(G_1(x)-\frac{N}{4n(N-n)}\right), \tag{3.39}$$

where we define

$$G_1(x) := -\frac{2n(N-D)^2}{N(N-n)}x^2 ,$$

and

$$w(x) := (N+2n(N-D)x)\exp\left(-\frac{N-D}{N-n}x\right)+(N-2n(N-D)x)\exp\left(\frac{N-D}{N-n}x\right) .$$

Proof. We begin by observing that

$$P\left(2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x-\frac{N}{2n(N-D)}\right)\leq Y\leq 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x+\frac{N}{2n(N-D)}\right)\right)$$
$$=P\left(0\leq Y\leq 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x+\frac{N}{2n(N-D)}\right)\right)$$
$$+P\left(2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x-\frac{N}{2n(N-D)}\right)\leq Y\leq 0\right)$$
$$=P\left(0\leq Y\leq 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x+\frac{N}{2n(N-D)}\right)\right)$$
$$+P\left(0\leq Y\leq 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(\frac{N}{2n(N-D)}-x\right)\right)\ , \qquad (3.40)$$

with the final equality following by the symmetry of Y around 0. We now use Lemma 16 twice. In the notation of that Lemma, we now identify

$$a=0\ ,$$

$$b_1=2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(x+\frac{N}{2n(N-D)}\right)\ ,$$

and

$$b_2=2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(\frac{N}{2n(N-D)}-x\right)\ .$$

Therefore

$$\sqrt{\frac{2}{2\pi}}(b_1-a)=\sqrt{\frac{1}{\pi nN(N-n)}}(N+2n(N-D)x)$$

and

$$-\frac{a^2}{4}-\frac{b_1^2}{4}=-\frac{n(N-D)^2}{N(N-n)}x^2-\frac{N-D}{N-n}x-\frac{N}{4n(N-n)}\ .$$

Additionally,

$$\sqrt{\frac{2}{2\pi}}(b_2-a)=\sqrt{\frac{1}{\pi nN(N-n)}}(N-2n(N-D)x)$$

and

$$-\frac{a^2}{4}-\frac{b_2^2}{4}=-\frac{n(N-D)^2}{N(N-n)}x^2+\frac{N-D}{N-n}x-\frac{N}{4n(N-n)}\ .$$

Using these expressions, along with the bounds $\gamma(a, b_1) \geq 1$ and $\gamma(a, b_2) \geq 1$, we apply Lemma 16 to find (3.40) is bounded below by

$$\begin{aligned}
&\sqrt{\frac{1}{\pi n N(N-n)}}(N + 2n(N-D)x) \exp\left(-\frac{n(N-D)^2}{N(N-n)}x^2 - \frac{N-D}{N-n}x - \frac{N}{4n(N-n)}\right) \\
&+\sqrt{\frac{1}{\pi n N(N-n)}}(N - 2n(N-D)x) \exp\left(-\frac{n(N-D)^2}{N(N-n)}x^2 + \frac{N-D}{N-n}x - \frac{N}{4n(N-n)}\right) \\
&=\sqrt{\frac{1}{\pi n N(N-n)}} \exp\left(-\frac{n(N-D)^2}{N(N-n)}x^2 - \frac{N}{4n(N-n)}\right) \\
&\quad \cdot \left[(N + 2n(N-D)x) \exp\left(-\frac{N-D}{N-n}x\right) + (N - 2n(N-D)x) \exp\left(\frac{N-D}{N-n}x\right)\right] . \\
&=:\sqrt{\frac{1}{\pi n N(N-n)}} \exp\left(-\frac{n(N-D)^2}{N(N-n)}x^2 - \frac{N}{4n(N-n)}\right) \cdot [w(x)] \ . \qquad (3.41)
\end{aligned}$$

The final observation that

$$-\frac{n(N-D)^2}{N(N-n)}x^2 \geq -\frac{2n(N-D)^2}{N(N-n)}x^2$$

combines with (3.41) to give the claim. □

Lemma 20. *Let $N \geq 150$, n, $D \in \mathbb{N}$, $10 \leq n < D < N/2$, and $0 \leq x < N/(2n(N-D))$. Let $Y \sim \text{Normal}(0,1)$. Then*

$$w(x) = (N + 2n(N-D)x) \exp\left(-\frac{N-D}{N-n}x\right) + (N - 2n(N-D)x) \exp\left(\frac{N-D}{N-n}x\right)$$

decreases in x. For $x \in [N/(4n(N-D)), N/(2n(N-D))]$, we have

$$w(x) \geq 2N \exp\left(-\frac{N}{2n(N-n)}\right) \ . \qquad (3.42)$$

For $x \in [N/(8n(N-D)), N/(4n(N-D))]$, we have

$$w(x) \geq 2N \exp\left(-\frac{N}{4n(N-n)}\right) \ . \qquad (3.43)$$

For $x \in [N/(16n(N-D)), N/(8n(N-D))]$, we have

$$w(x) \geq 2N \exp\left(-\frac{N}{8n(N-n)}\right) \ . \qquad (3.44)$$

For $x \in [0, N/(16n(N-D))]$, we have

$$w(x) \geq 2N \exp\left(-\frac{N}{16n(N-n)}\right) . \tag{3.45}$$

Proof. Differentiating, we find $w'(x)$ equals

$$\begin{aligned}&\exp\left(-\frac{N-D}{N-n}x\right)\frac{N-D}{N-n}(2n(N-n)-N-2n(N-D)x)\\ -&\exp\left(\frac{N-D}{N-n}x\right)\frac{N-D}{N-n}(2n(N-n)-N+2n(N-D)x) .\end{aligned}$$

To show decrease, we must show this quantity is negative, which is equivalent to showing

$$\exp\left(-2\frac{N-D}{N-n}x\right)(2n(N-n)-N-2n(N-D)x) \leq (2n(N-n)-N+2n(N-D)x) ,$$

or on the log-scale to

$$0 \leq \log(2n(N-n)-N+2n(N-D)x)-\log(2n(N-n)-N-2n(N-D)x)+2\frac{N-D}{N-n}x =: p_1(x) .$$

Differentiating twice and combining terms, we find $p_1''(x)$ may be written

$$\frac{32n^3(N-D)^3(2n(N-n)-N)x}{[(2n(N-n)-N)^2-4n^2(N-D)^2x^2]^2} \geq 0 .$$

Therefore $p_1'(x)$ increases in x. Evaluating $p_1'(x)$ at $x=0$ we find it bounded below by

$$2\frac{N-D}{N-n}+\frac{4n(N-D)}{2n(N-n)-N} > 0 .$$

Therefore, $p_1(x)$ increases in x, and since at $x \geq 0$, decrease in x follows. Using this decrease, the claim at (3.42) follows by evaluating $w(x)$ at $x = N/(2n(N-D))$.

To show the second inequality (3.43), we evaluate $w(x)$ at $x = N/(4n(N-D))$, and find we must show

$$\frac{1}{2}N\exp\left(\frac{N}{4n(N-n)}\right)+\frac{3}{2}N\exp\left(-\frac{N}{4n(N-n)}\right) \geq 2N\exp\left(-\frac{N}{4n(N-n)}\right) .$$

Dividing through by the right-hand side and re-arranging, the claim becomes

$$\exp\left(\frac{N}{2n(N-n)}\right) \geq 1 ,$$

which follows since $N/(2n(N-n)) > 0$. Hence inequality (3.43) holds. The inequalities at (3.44) and (3.45) follow in the same fashion as (3.43), with the constants differing from the point of evaluation. $\square$

Lemma 21. *Let* $N \geq 500,\ n,\ D \in \mathbb{N}$, $30 \leq n \leq D \leq N/2$. *Let* $Y \sim$ *Normal*$(0,1)$. *Then*

$$\sqrt{2}\ P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right)\left(1-\frac{N}{2n(N-D)}\right)\right)$$

$$\geq \frac{811}{841}\sqrt{\frac{(N-n)N}{4\pi n(N-D)^2}\frac{2n(N-D)}{2n(N-D)-N}}\exp\left(-\frac{(2n(N-D)-N)^2}{2nN(N-n)}\right)\ .$$

Proof. Define

$$t := 2\sqrt{\frac{Nn}{N-n}}\left(\frac{N-D}{N}\right)\left(1-\frac{N}{2n(N-D)}\right)\ .$$

Then

$$1-\frac{1}{t^2} = 1-\frac{nN(N-n)}{(2n(N-D)-N)^2}\ .$$

Differentiating shows that this expression increases in n and decreases in D. So, under the assumptions it is bounded below by

$$1-\frac{nN(N-n)}{(2n(N-D)-N)^2} \geq \frac{900}{841N}+\frac{811}{841} > \frac{811}{841}\ .$$

Referring again to our definition of t, we find

$$\frac{\sqrt{2}}{t\sqrt{2\pi}} = \sqrt{\frac{(N-n)N}{4\pi n(N-D)^2}\frac{2n(N-D)}{2n(N-D)-N}}\ .$$

Applying Feller's normal bound (3.36) with the previous two observations yields the claim. $\square$

What guides us to seek Gaussian upper bounds for the hypergeometric tail? Hájek's Theorem 3.1 [20] tells us to expect Gaussian behavior in the hypergeometric case when the sample size $n \nearrow \infty$ and the population N grow in such a way that $N-n \nearrow \infty$ too. In particular, this occurs when $N = 2n$.

3.6 The ABCs

To carry out the comparison of the method of section 3.1.1, we will consider the function

$$\Gamma(x) := G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x)\right] + R(D) \tag{3.46}$$

where we define

$$R(D) := -\frac{1}{2}\log\left(\frac{N^2}{8D(N-D)}\right) + R_N + R_H \ .$$

In this definitions, R_H,$A_i(x)$,$B_i(x)$, $i \in \{1,2,3,4\}$, are as defined as in Lemma 14 and R_N and $G_1(x)$ are as defined in Lemma 18.

The bulk of the next chapter is dedicated to proving (3.46) is positive under various restrictions on x. The proof of positivity is technical, and relies on a careful partitioning of the values of x under consideration. The Lemmas are aided by understanding the behavior of certain sub-collections of the functions appearing in (3.46). We characterize that behavior in this section.

We introduce some additional notation. In the subsequent Lemmas we will denote the first partial derivative with respect to x using lower case letters rather than writing B' or $\partial B/\partial x$. For example, in the following $B_1'(x) \equiv b_1(x) \equiv (\partial/\partial x)b_1(x)$. Similarly, $A_1'(x) \equiv a_1(x)$, $B_1''(x) \equiv b_1'(x)$, and so on.

Lemma 22. *Let* $N \geq 500$, n, $D \in \mathbb{N}$, $20 \leq n < D < N/2$, *and* $0 \leq x \leq 1 - N/(n(N-D))$. *Using the functions in* (3.46), *we have that both*

$$\frac{N-n}{N}G_1(x) + B_1(x) + B_2(x)$$

and

$$\frac{n}{N}G_1(x) + B_3(x) + B_4(x)$$

increase in x. *Jointly these imply that*

$$G_1(x) + B_1(x) + B_2(x) + B_3(x) + B_4(x) \tag{3.47}$$

increases in x from 0. *Furthermore,*

$$A_2(x) + A_4(x) \tag{3.48}$$

increases in x and

$$A_1(x) + A_3(x) \tag{3.49}$$

decreases in x. Finally

$$R(D) \equiv R(D, n) \tag{3.50}$$

increases in D and increases in n.

Proof. To demonstrate the first claim, differentiate twice with respect to x to find

$$\frac{N-n}{N} g_1'(x) + b_1'(x) + b_2'(x) = \frac{n(N-D)(N-2D-2(N-D)x)^2}{N^2(1-x)(D+(N-D)x)}$$

and

$$\frac{n}{N} g_1'(x) + b_3'(x) + b_4'(x) = \frac{n^2(N-D)((N-2D)(N-n)+2(N-D)nx)^2}{N^2(N-n)(N-n+nx)(D(N-n)-n(N-D)x)} \ .$$

These imply $g_1'(x) + b_1'(x) + b_2'(x) + b_3'(x) + b_4'(x)$ is positive under the assumptions. The first claim (3.47) follows since at $x = 0$ we have that $((N-n)/N)g_1(x) + b_1(x) + b_2(x)$, $(n/N)g_1(x) + b_3(x) + b_4(x)$, and the sum $g_1(x) + b_1(x) + b_2(x) + b_3(x) + b_4(x)$ all equal 0 too.

The second claim (3.48) follows since

$$a_2(x) + a_4(x) = \frac{N-D}{2(D+(N-D)x)} + \frac{n}{2(N-n+nx)} \ ,$$

which is positive under the assumptions for $x \geq 0$.

The third claim (3.49) follows since

$$a_1(x) + a_3(x) = -\frac{1}{2(1-x)} - \frac{n(N-D)}{2(D(N-n)-n(N-D)x)} \ ,$$

which is negative under the assumptions for $0 \leq x < 1$.

The final claim (3.50) follows by differentiation. First,

$$\frac{\partial}{\partial n} R(D,n) = \frac{7N(N-2n)}{12n^2(N-n)^2} > 0 \ ,$$

giving increase in n independent of D. Similarly,

$$\frac{\partial}{\partial D} R(D,n) = \frac{12D^3 - 18D^2N + 2DN(3N-1) + N^2}{12D^2(N-D)^2} \ .$$

The sign of this depends on the sign of the numerator, which is independent of n. Analyzing the numerator, we find it concave in D since its second derivative with respect to D is $-36(N-2D)$. At $D = N/2$ the numerator equals 0, and at $D = 20$ it equals $121N^2 - 7240N + 96000$, which is positive for $N \geq 41$. The last claim follows by concavity.

□

Lemma 23. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $N/3 \leq D \leq N/2$, $31 \leq n \leq D$, and $1/4 \leq x \leq 1 - N/(n(N-D))$. Using the functions in* (3.46)*, we have that*

$$A_1(x) + A_2(x) + A_3(x) + A_4(x)$$

is concave in x. Furthermore, both

$$A_1(x) + A_2(x)$$

and

$$A_3(x) + A_4(x)$$

decrease in x.

Proof. To see the concavity claim holds, we differentiate the sum twice to find

$$-\frac{1}{2(1-x)^2} - \frac{(N-D)^2}{2(D+(N-D)x)^2} - \frac{n^2(N-D)^2}{2(D(N-n)-n(N-D)x)^2} - \frac{n^2}{2(N-n+nx)^2} < 0 \ .$$

To see the decrease claim holds, note first that $a_1(x) + a_2(x)$ has a zero at

$$\frac{N-2D}{2(N-D)},$$

and $a_3(x) + a_4(x)$ has a zero at

$$-\frac{(N-2D)(N-n)}{2n(N-D)} .$$

This second zero is negative under the assumptions, and since we have seen $a_3'(x)+a_4'(x) < 0$ one decrease follows. Under the assumptions,

$$\frac{N-2D}{2(N-D)} \leq \frac{1}{4} ,$$

with the upper bound achieved at $D = N/3$. Since we have also $a_1'(x) + a_2'(x) < 0$ the other sum decreases as well, which gives the claim. $\square$

Lemma 24. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $(15/100)N \leq D \leq N/2$, $31 \leq n \leq D$, and $0 \leq x \leq N/(n(N-D))$. Using the functions in (3.46), we have that*

$$C_i(x) \geq 0$$

for $i \in \{1, 2, 3, 4\}$. In addition $C_2(x)$ decreases in x under the assumptions.

Proof. We show this for each $C_i(x)$ in turn. First consider $C_1(x)$. Combining terms we write it as

$$\frac{N \cdot [p_1(x)]}{360n^3(N-D)^3(1-x)^3} \tag{3.51}$$

with

$$p_1(x) := 30n^2(N-D)^2(1-x)^2 - N^2$$

We see $p_1(x)$ decreases in x. Evaluating at $x = N/(n(N-D))$ shows it is bounded below by

$$p_{1a}(n) := 30n^2(N-D)^2 - 60n(N-D)N + 29N^2 .$$

Since $p_{1a}''(n) = 60(N-D)^2 > 0$, $p_{1a}'(n)$ is bounded below by $60(31D^2 - 61DN + 30N^2) > 0$. Therefore, $p_{1a}(n)$ increases in n and is bounded below by $28830D^2-55800DN+26999N^2 > 0$. Therefore $p_1(x) \geq 0$ and the claim follows for $C_1(x)$ by (3.51).

Turning next to $C_2(x)$, we combine terms to write it as

$$\frac{N \cdot p_2(x)}{360n^3(D+(N-D)x)^3} \tag{3.52}$$

where we define

$$p_2(x) := 30n^2(D+(N-D)x)^2 - N^2 \ .$$

By its definition, we see $p_2(x)$ increases in x. Hence at $x = 0$ it is bounded below by

$$30D^2n^2 - N^2 \geq \frac{25907N^2}{40} > 0$$

with the inequality following by the assumptions on D and n. Thus, $p_2(x) \geq 0$ which gives the claim for $C_2(x)$ by (3.52).

For $C_3(x)$, we combine terms to see it equals

$$\frac{N \cdot p_3(x)}{360(D(N-n) - nx(N-D))^3} \tag{3.53}$$

with

$$p_3(x) := 30(D(N-n) - nx(N-D))^2 - N^2$$

This decreases in x, and so is bounded below by

$$30(D(N-n) - N)^2 - N^2 \geq 30(D(N-D) - N)^2 - N^2 \geq 31(29791 - 1860N + 29N^2) > 0 \ ,$$

with the inequalities following by decrease in n and increase in D. Therefore $p_3(x) > 0$ under the assumptions, and so too is $C_3(x)$ by (3.53).

Finally for $C_4(x)$ we combine terms to find

$$\frac{N \cdot p_4(x)}{360(N-D)^3(N-n+nx)^3} \tag{3.54}$$

where we define

$$p_4(x) := 30(N-D)^2(N-n+nx)^2 - N^2 \ .$$

This increases in x, and so is bounded below by

$$30(N-D)^2(N-n)^2 - N^2 \geq 30(N-D)^4 - N^2 \geq \frac{30}{16}N^4 - N^2 > 0 \ .$$

Hence $p_4(x) \geq 0$, and the claim follows for $C_4(x)$ (3.54).

It remains to show $C_2(x)$ decreases in x. Differentiating, we find

$$c_2(x) = -\frac{N(N-D)}{120n^3(D+(N-D)x)^4} \cdot p_5(x) \ ,$$

with

$$p_5(x) := 10n^2(N-D)^2x^2 + 20Dn^2(N-D)x + 10D^2n^2 - N^2 \ .$$

The claim will follow if we show $p_5(x) \geq 0$. By inspection, we see $p_5(x)$ increases in x. At $x = 0$ it is therefore bounded below by

$$10D^2n^2 - N^2 \geq \frac{8609}{40}N^2 > 0 \ ,$$

and so decrease in x follows by the assumptions. □

Chapter 4

SHIFTED SERFLING GAUSSIAN BOUNDS FOR THE HYPERGEOMETRIC TAIL

> The traveler often has the choice between climbing a peak or using a cable car.
>
> William Feller [14]

In this chapter, we carry out the comparison between the upper bounds on hypergeometric probabilities and the lower bound on Gaussian probabilities prepared in chapter 3. The comparison is elaborate, and so we organize the chapter as follows: the main results, Theorem 6 and Theorem 7, are presented first. We call the bounds developed in these Theorems **shifted Serfling Gaussian probability bounds**. These theorems rely on proving the expression $\Gamma(x)$, defined at (3.46), is positive under two different conditions on x. The Propositions and Lemmas proving the positivity of $\Gamma(x)$ constitute the remainder of the chapter.

Recall in our analysis of hypergeometric probabilities we assumed that $n \leq D \leq N/2$. In the proof that $\Gamma(x)$ is positive, we will further constrain the hypergeometric parameters (n, D, N). A visual depiction of the constraints are presented in Figure 4.2. Section headers in this chapter refer to the regions depicted in Figure 4.2.

The regions $\alpha, \beta, \gamma 1$, and $\gamma 2$ represent constraints we place on the expression $\Gamma(x)$. The regions δ and ϵ represent constraints we place on the first partial derivative with respect to x of $\Gamma(x)$. We denote this first partial derivative by $\gamma(x)$. The regions ζ and η represent constraints we place on the second partial derivative with respect to x of $\Gamma(x)$. We denote this second partial derivative by $\gamma'(x)$.

The proof relies on the monotone increase of $\Gamma(x)$ in x. After proving that this occurs, different arguments are needed when $D \leq (15/100)N$ (the α region), when $(15/100)N \leq D \leq N/3$ (the β region), and when $N/3 \leq D \leq N/2$ (the regions $\gamma 1$ and $\gamma 2$).

For the α region, $\Gamma(x)$ is not positive when x is small. In that region, we demonstrate the positivity of $\Gamma(x)$ when $x \geq (N-5D)/(5(N-D))$.

In the definition of $\Gamma(x)$ there is a remainder term $R(D)$. This remainder does not depend on x. In the β region, the remainder $R(D)$ decreases as D decreases to $(15/100)N$. However in the regions $\gamma 1$ and $\gamma 2$, the remainder term is large and positive. The positivity arguments are thus fundamentally different on the β region, where positivity holds in spite of this remainder term, and the $\gamma 1$ and $\gamma 2$ region, where positivity follows *because of* the remainder term.

Finally, special arguments are needed when $x < N/(n(N-D))$ (when x is small) and $x = 1$ (when x is big). For small x, we provide arguments which address the β, $\gamma 1$ and $\gamma 2$ regions simultaneously. When x is small the need for specialized arguments essentially occurs because the remainder term keeps changing the closer x gets to 0. For large x, we provide arguments which address the α, β, $\gamma 1$ and $\gamma 2$ regions simultaneously. The large x argument is provided in its entirety in Proposition 6. The need for a special large-x argument occurs since the hypergeometric and Gaussian probabilities involved are exceptional, and so different quantities are examined after applying Stirling's formula.

To motivate what might seem an excess of detail, we conclude the introduction by providing a comparison between the shifted Serfling Gaussian probability bounds and existing probability bounds which apply to the hypergeometric distribution.

4.1 Shifted Serfling Gaussian Probability Bound

Theorem 6. *Let* $Y \sim N(0,1)$*,* $N \geq 500$ *and* $(15/100)N \leq D \leq N/2$*, and* $31 \leq n \leq D$*. Let* $H_{n,D,N} \sim$ *Hypergeometric*(n, D, N)*. Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

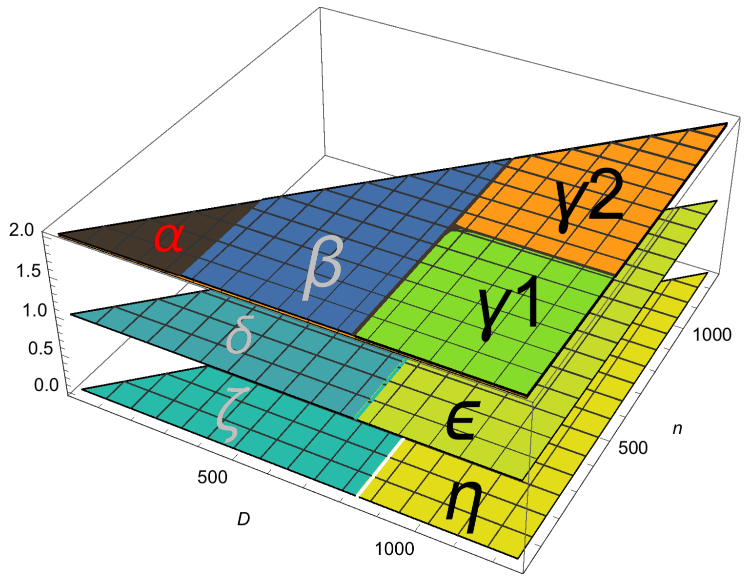

Figure 4.1: This picture provides an overview of the proof strategy. The argument proceeds sequentially. For a given population size N, we let the sample size n and population of successes D vary. We take $N = 2500$ in this figure. The top layer in the plot represents the main function $\Gamma(x)$: we wish to demonstrate positivity for $x \in [0, 1]$. The middle layer represents the first partial derivative of $\Gamma(x)$ with respect to x. The bottom layer represents the second partial derivative. For the bottom and middle layer, the different regions are for $D \leq N/3$ and $D > N/3$. For the top layer, four regions are considered in the proof: $D < (15/100)N$ (the α region), $(15/100)N \leq D \leq N/3$ (the β region), $N/3 < D \leq N/2$ and $31 \leq n \leq N/5$ (the $\gamma 1$ region), and $N/3 < D \leq N/2$ and $N/5 \leq n \leq D$ (the $\gamma 2$ region).

Then for

$$j \in \left\{ \frac{(k^* + p)N - nD}{N - D} \;\middle|\; 0 \leq p \leq n - k^*,\ p \in \mathbb{N} \right\} \ , \tag{4.1}$$

we have

$$P\left(H_{n,D,N} - n\frac{D}{N} \geq \frac{(N-D)}{N} j\right) \leq \sqrt{2} P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}} \left(\frac{N-D}{N}\right)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right)\right) \ . \tag{4.2}$$

Proof. On the j-lattice (5.10), we find that the hypergeometric quantity (5.11) is bounded

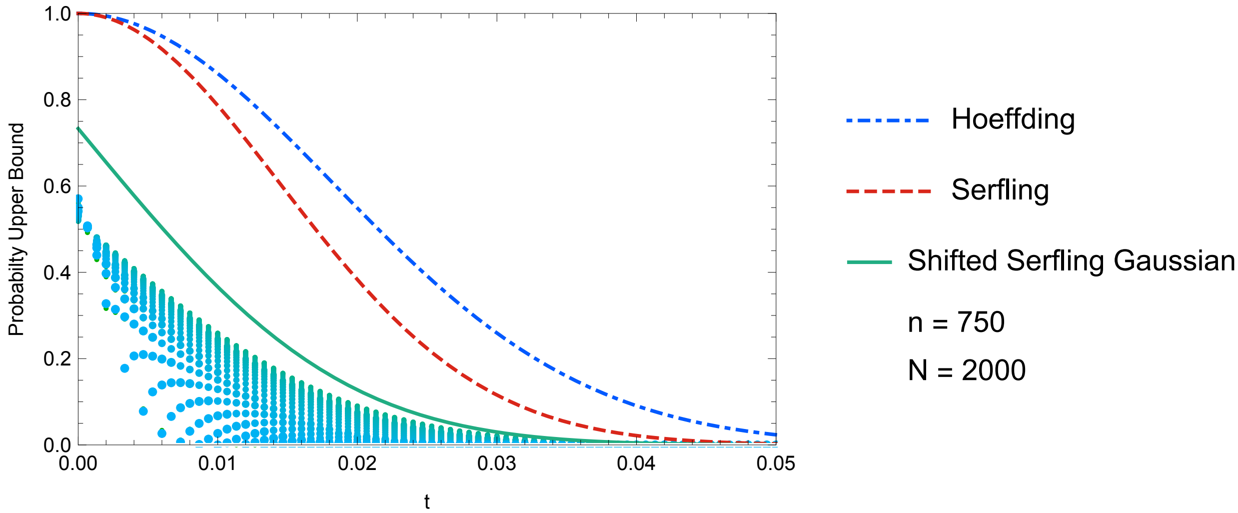

Figure 4.2: The shifted Serfling Gaussian bound compared to Serfling's inequality and Hoeffding's inequality.

above by

$$\begin{aligned}
&P\left(H_{n,D,N} - n\frac{D}{N} \geq \frac{(N-D)}{N}j\right) \\
=&P\left(H_{n,D,N} \geq \frac{nD+(N-D)j}{N}\right) \\
=&P(H_{n,D,N} = n) + \sum_{k=j}^{n-1} P\left(H_{n,D,N} = \frac{nD+(N-D)k}{N}\right) \\
\leq&\sqrt{2}\ P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right)\left(1 - \frac{N}{2n(N-D)}\right)\right) + \\
&\sum_{k=j}^{n-1}\sqrt{2}\cdot P\begin{pmatrix} 2\sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right) \leq \\ Y \\ \leq 2\sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right)\left(\frac{j}{n} + \frac{N}{2n(N-D)}\right)\end{pmatrix} \\
=&\sqrt{2}P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right)\right)
\end{aligned}$$

with the inequality following by Proposition 2, Proposition 4, Proposition 5, and Proposition 6. The particular proposition that applies depends on the choice of parameters and the magnitude of j. The claim follows. □

Theorem 7. *Let* $Y \sim N(0,1)$, $N \geq 500$ *and* $31 \leq D \leq (15/100)N$, *and* $31 \leq n \leq D$. *Let* $H_{n,D,N} \sim$ *Hypergeometric*(n,D,N). *Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Then for

$$j \in \left\{ \frac{(k^*+p)N - nD}{N-D} \;\middle|\; q^* \leq p \leq n-k^*-1, q^* = \frac{n}{5} - k^*,\ p \in \mathbb{N} \right\} , \tag{4.3}$$

we have

$$\begin{aligned} &P\left(H_{n,D,N} - n\frac{D}{N} \geq \frac{(N-D)}{N} j\right) \leq \\ &\sqrt{2} P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}} \left(\frac{N-D}{N}\right)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right)\right) . \end{aligned} \tag{4.4}$$

Proof. The proof is nearly identical to the proof of Theorem 6. However, only Proposition 3 and Proposition 6 are needed for the inequality under the current assumptions. The claim follows. □

4.2 Shifted Serfling Gaussian Interval Propositions

We first use the preceding Lemmas to derive Gaussian bounds when $(15/100)N \leq D \leq N/3$ and $N/(n(N-D)) \leq x \leq 1 - N/(n(N-D))$.

Proposition 2. *Let* $Y \sim N(0,1)$, $N \geq 500$ *and* $(15/100)N \leq D \leq N/3$, *and* $31 \leq n \leq D$. *Let* $H_{n,D,N} \sim$ *Hypergeometric*(n,D,N). *Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Then for

$$j \in \left\{ \frac{(k^*+p)N - nD}{N-D} \;\middle|\; 1 \leq p \leq n-k^*-1,\ p \in \mathbb{N} \right\} , \tag{4.5}$$

we have

$$P\left(H_{n,D,N} = \frac{nD + (N-D)j}{N}\right) \leq$$
$$\sqrt{2}P\left(2\ \zeta(n,D,N)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right) \leq Y \leq 2\ \zeta(n,D,N)\left(\frac{j}{n} + \frac{N}{2n(N-D)}\right)\right)\ ,$$

where

$$\zeta(n,D,N) := \sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right)\ .$$

Proof. Write $x = j/n$. We consider the continuous relaxation of the problem, and prove the claim for $N/(n(N-D)) \leq x \leq 1 - N/(n(N-D))$. This will imply the claim holds on the j-lattice (4.5).

We first use Lemma 18 and Lemma 14. Under the current assumptions it is enough to show that (3.37) in Lemma 18 is larger than (3.26) in Lemma 14. Combining terms, taking the logarithm and then differences, this problem reduces to showing the function defined at (3.46) is positive under the current assumptions. We re-define this function here for convenience. The function is defined as

$$\Gamma(x) := G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x)\right] + R(D)$$

where we define

$$R(D) := -\frac{1}{2}\log\left(\frac{N^2}{8D(N-D)}\right) + R_N + R_H$$

In this definitions, R_H,$A_i(x)$,$B_i(x)$, $i \in \{1,2,3,4\}$, are as defined as in Lemma 14 and R_N and $G_1(x)$ are as defined in Lemma 18.

To show it is positive, we first show it increases in x under the current assumptions. By Lemma 53 and Lemma 54, we know the first partial derivative of $\Gamma(x)$, which we denote as $\gamma(x)$ in the subsequent discussion, increases in x.

When $x \in [N/(n(N-D)), (N-2D)/(2(N-D))]$, Lemma 38, Lemma 42, and Lemma 43 imply $\gamma(x) \geq 0$. Notice that under the restrictions on D, the lower bound of the interval is guaranteed to be smaller than the upper bound.

When $x \in [(N-2D)/(2(N-D)), 1/2]$, Lemma 44 and Lemma 48 jointly imply $\gamma(x) \geq 0$. Again, the current conditions on D guarantee the lower bound of this interval does not exceed the upper bound.

Note that Lemma 53 and Lemma 48 imply that

$$\frac{n}{N} g_1(x) + b_3(x) + b_4(x) + a_3(x) + a_4(x) \geq 0$$

for $x \geq (N-2D)/(2(N-D))$. This combined with Lemma 49 and Lemma 50 show $\gamma(x) \geq 0$ when $x \in [1/2, 1 - N/(n(N-D))]$. Hence we have demonstrated that $\gamma(x) \geq 0$ under the assumptions. Therefore $\Gamma(x)$ increases in x.

Evaluating $\Gamma(x)$ at $x = N/(n(N-D))$, Lemma 26 and then Lemma 27 imply that this quantity is positive. By monotone increase in x, the claim is proved. □

We next use the preceding Lemmas to derive Gaussian bounds when $31 \leq D \leq (15/100)N$ and $(N-5D)/(5(N-D)) \leq x \leq 1 - N/(n(N-D))$.

Proposition 3. *Let $Y \sim N(0,1)$, $N \geq 500$ and $31 \leq D \leq (15/100)N$, and $31 \leq n \leq D$. Let $H_{n,D,N} \sim$ Hypergeometric(n, D, N). Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Then for

$$j \in \left\{ \frac{(k^*+p)N - nD}{N-D} \;\middle|\; q^* \leq p \leq n - k^* - 1, q^* = \frac{n}{5} - k^*,\ p \in \mathbb{N} \right\} , \tag{4.6}$$

we have

$$P\left(H_{n,D,N} = \frac{nD + (N-D)j}{N}\right) \leq$$
$$\sqrt{2} P\left(2\ \zeta(n,D,N)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right) \leq Y \leq 2\ \zeta(n,D,N)\left(\frac{j}{n} + \frac{N}{2n(N-D)}\right)\right) ,$$

where

$$\zeta(n,D,N) := \sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right) .$$

Proof. Write $x = j/n$. Once again we consider the continuous relaxation of the problem. Now however the definition of the lattice (4.6) means we will prove the claim for $(N - 5D)/(5(N-D)) \leq x \leq 1 - N/(n(N-D))$.

The proof follows the same structure as Proposition 2. By the same argument as the one given in Proposition 2, the claim will follow by showing $\Gamma(x) \geq 0$ under the current assumptions. The same collection of Lemmas used in Proposition 2 also show that $\gamma(x) \geq 0$ under the current assumptions. Therefore, $\Gamma(x)$ increases in x under the current assumptions for $x \in [N/(n(N-D)), 1 - N/(n(N-D))]$.

We therefore evaluate $\Gamma(x)$ at $x = (N-5D)/(5(N-D))$ to obtain a lower bound on $\Gamma(x)$. This follows since

$$\frac{N}{n(N-D)} \leq \frac{N-5D}{5(N-D)} \Longleftrightarrow 1 \leq n\left[\frac{1}{5} - \frac{D}{N}\right] ,$$

and we are assuming $D/N \leq 3/20$ and $n \geq 31$. Lemma 25 implies this lower bound is positive, and so monotone increase proves the claim.

□

Next, we use the preceding Lemmas to derive Gaussian bounds when $N/3 \leq D \leq N/2$ and $N/(n(N-D)) \leq x \leq 1 - N/(n(N-D))$.

Proposition 4. *Let $Y \sim N(0,1)$, $N \geq 500$ and $N/3 \leq D \leq N/2$, and $31 \leq n \leq D$. Let $H_{n,D,N} \sim$ Hypergeometric(n, D, N). Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Then for

$$j \in \left\{ \frac{(k^*+p)N - nD}{N-D} \;\middle|\; 1 \leq p \leq n - k^* - 1,\ p \in \mathbb{N} \right\} , \tag{4.7}$$

we have

$$P\left(H_{n,D,N} = \frac{nD + (N-D)j}{N}\right) \leq$$
$$\sqrt{2}P\left(2\,\zeta(n,D,N)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right) \leq Y \leq 2\,\zeta(n,D,N)\left(\frac{j}{n} + \frac{N}{2n(N-D)}\right)\right) ,$$

where

$$\zeta(n,D,N) := \sqrt{\frac{Nn}{N-n}} \left(\frac{N-D}{N}\right) .$$

Proof. Once again, write $x = j/n$. Again we consider the continuous relaxation of the problem. Now however the definition of the lattice (4.7) means we will prove the claim for $(N)/(n(N-D)) \leq x \leq 1 - N/(n(N-D))$.

The proof follows the same structure as Proposition 2. By the same argument as the one given in Proposition 2, the claim will follow by showing $\Gamma(x) \geq 0$ under the current assumptions.

First suppose that $x \in [N/(n(N-D)), 1/2]$. Then Lemma 28 and Lemma 29 imply the claim.

Next suppose $x \in [1/2, 1-(N/(n(N-D)))]$. We split $\Gamma(x)$ into

$$\Gamma_1(x) := \left(\frac{N-n}{N}\right) G_1(x) + B_1(x) + B_2(x) + A_1(x) + A_2(x)$$

and

$$\Gamma_2(x) := \left(\frac{n}{N}\right) G_1(x) + B_3(x) + B_4(x) + A_3(x) + A_4(x) .$$

We show both are positive under the assumptions. First, by Lemma 52, we see $\Gamma_1(x)$ increases in x. By Lemma 33, it is positive under the current assumptions.

Turning to $\Gamma_2(x)$, we see that Lemma 55 and Lemma 51 imply it is increasing in x. If we assume $n < N/5$, Lemma 30 and Lemma 31 imply $\Gamma_2(x)$ is positive. Finally, if we assume $n \geq N/5$, Lemma 32 implies $\Gamma_2(x)$.

We infer $\Gamma_1(x) \geq 0$ and $\Gamma_2(x) \geq 0$ when $x \in [1/2, 1-(N/(n(N-D)))]$. Hence, $\Gamma(x) = \Gamma_1(x) + \Gamma_2(x) \geq 0$, proving the claim. □

We next use the preceding Lemmas to derive Gaussian bounds when $(15/100)N \leq D \leq N/2$ and $0 \leq x \leq N/(n(N-D))$.

Proposition 5. *Let* $Y \sim N(0,1)$*,* $N \geq 500$ *and* $N/3 \leq D \leq N/2$*, and* $31 \leq n \leq D$*. Let* $H_{n,D,N} \sim$ *Hypergeometric*(n,D,N)*. Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Then for

$$j \in \left\{ \frac{(k^* + p)N - nD}{N - D} \;\middle|\; 0 \le p \le 1,\ p \in \mathbb{N} \right\} , \tag{4.8}$$

we have

$$P\left(H_{n,D,N} = \frac{nD + (N-D)j}{N}\right) \le$$
$$\sqrt{2}P\left(2\ \zeta(n,D,N)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right) \le Y \le 2\ \zeta(n,D,N)\left(\frac{j}{n} + \frac{N}{2n(N-D)}\right)\right) .$$

where

$$\zeta(n,D,N) := \sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right) .$$

Proof. Again we write $x = j/n$, and again we consider the continuous relaxation of the problem. Now however the definition of the lattice (4.8) means we will prove the claim for $0 \le x \le N/(n(N-D))$.

We first restrict $x \in [N/(2n(N-D)), N/(n(N-D))]$, and use Lemma 18 and Lemma 14. The same logic as Lemma 5 again yields the function $\Gamma(x)$. Here, Lemma 34 gives positivity.

We next restrict $x \in [N/(4n(N-D)), N/(2n(N-D))]$ and use Lemma 19, Lemma 20 with the bound (3.42), and Lemma 14. Taking logarithms and the differences yields the quantity analyzed in Lemma 35. Lemma 35 shows the quantity is positive, proving the claim for $x \in [N/(4n(N-D)), N/(2n(N-D))]$.

Next, we restrict $x \in [N/(8n(N-D)), N/(4n(N-D))]$ and use Lemma 19, Lemma 20 with the bound (3.43), and Lemma 14. Taking logarithms and the differences yields the quantity analyzed in Lemma 36. Lemma 36 shows the quantity is positive, proving the claim for $x \in [N/(8n(N-D)), N/(4n(N-D))]$.

Finally, we restrict $x \in [0, N/(8n(N-D))]$ and use Lemma 19, Lemma 20 with the bound (3.44), and Lemma 14. Taking logarithms and the differences yields the quantity analyzed in Lemma 37. Lemma 37 shows the quantity is positive, proving the claim for $x \in [0, N/(8n(N-D))]$.

Combined, the preceding arguments prove the claim.

□

We conclude this section by obtaining a proof for the far tail: we let $31 \le D \le N/2$, $30 \le n \le D$ and $x = 1$.

Proposition 6. *Let $Y \sim N(0,1)$, $N \ge 500$, $31 \le D \le N/2$, and $30 \le n \le D$. Let $H_{n,D,N} \sim$ Hypergeometric(n,D,N). Then we have*

$$P\left(H_{n,D,N} = n\right) \le \sqrt{2}\ P\left(Y \ge 2\ \sqrt{\frac{Nn}{N-n}}\ \left(\frac{N-D}{N}\right)\left(1 - \frac{N}{2n(N-D)}\right)\right) \ .$$

Proof. Use Lemma 21 to get a lower bound on this Normal quantity, and Lemma 15 to get an upper bound on the hypergeometric quantity. Then, combine terms under the radicals, take logarithms and then the difference, and we find the claim will follow if we show

$$p_1(n) + p_2(n) \ge 0 \ , \tag{4.9}$$

where

$$\begin{aligned} p_1(n) := \log\left(\frac{2n(N-D)}{2n(N-D)-N}\right) + \frac{1}{2}\log\left(\frac{N^2(D-n)}{4\pi Dn(N-D)^2}\right) - \frac{(2n(N-D)-N)^2}{2nN(N-n)} \\ -n\log\left(\frac{D}{N}\right) - (D-n)\log\left(\frac{D}{D-n}\right) - (N-n)\log\left(\frac{N-n}{N}\right) + \log\left(\frac{811}{841}\right) \ , \end{aligned}$$

and

$$p_2(n) := \frac{1}{12(D-n)} - \frac{1}{360(D-n)^3} - \frac{1}{12D} - \frac{1}{12(N-n)} - \frac{1}{360N^3} + \frac{1}{12N} \ .$$

We will demonstrate that both $p_1(n) \ge 0$ and $p_2(n) \ge 0$, which in turn will imply (4.9). Consider $p_1(n)$ first. Differentiate once, to define

$$\begin{aligned} p_1'(n) :=& \frac{4(D-n)+2}{N-n} - \frac{(2D-2n+1)^2}{2(N-n)^2} - \frac{2D}{(2n-1)(2n(N-D)-N)} \\ &- \frac{1}{2(D-n)} + \frac{1}{2n^2} + \frac{1}{2n} - \frac{2}{2n-1} + \log\left(\frac{D}{D-n}\right) - \log\left(\frac{D}{N}\right) + \log\left(\frac{N-n}{N}\right) - 2 \\ =&: p_3(D) \ . \end{aligned} \tag{4.10}$$

Now differentiating $p_3(D)$ with respect to D we find

$$p_3'(D) = \frac{p_4(n)}{2(D-n)^2(N-n)^2} - \frac{2N}{(2n(N-D)-N)^2} \ , \tag{4.11}$$

where we define

$$p_4(n) := 2n^3+(4N-10D-3)n^2+2(N-2D-1)(N-4D)n-(N-2D)(2D(N-2D-1)-N) \ .$$

We now show $p_4(n) \leq 0$. The third derivative of $p_4(n)$ is 12, implying the first derivative is convex in n. The second derivative of $p_4(n)$ is $8N-20D+12n-6$, which for $D \in [31,(2/5)N]$ is positive by inspection. So, we know that $p_4(n)$ is convex for $D \in [31,(2/5)N]$. Now evaluate $p_4'(n)$ at $n = 30$ to define

$$p_5(D) := 16D^2 - 12DN - 592D + 2N^2 + 238N + 5220 \ .$$

This is a quadratic in D, and since the coefficient of D^2 is positive, it is convex in D. Restricting it to the interval $D \in [(2/5)N, N/2]$, we evaluate at the boundary points to find it equals $5220 + (6/5)N - (6/25)N^2$ and $-58(N-90)$. Since both are negative, convexity implies $p_5(D)$ is negative when $D \in [(2/5)N, N/2]$. This implies $p_4'(n) < 0$ when $n = 30$ and since we have seen that $p_4'(n)$ is convex in n, this implies $p_4(n)$ is maximized either at $n = 30$ or $n = D-1$ when $D \in [(2/5)N, N/2]$. Since we have also seen that $p_4(n)$ is convex when $D \in [31,(2/5)N]$, we see that $p_4(n)$ is thus maximized either at $n = 30$ or $n = D-1$ for $D \in [31, N/2]$. We now consider these cases. First evaluate at $n = 30$ to define

$$p_6(D) := -8D^3 + D^2(8N+476) - 2D\left(N^2+180N+4380\right) + 61N^2 + 3540N + 51300 \ .$$

First notice that $p_6''(D) = 952 - 48D + 16N$, and so $p_6(D)$ is convex for $D \in [31, N/3]$. Evaluating at these points, we find it equals $-1152 + 68N - N^2$ and $51300 + 620N - (55/9)N^2 - (2/27)N^3$ which are both negative under the assumptions. Hence, $p_6(D) < 0$ for $D \in [31, N/3]$ For $D \in [N/3, N/2]$ we notice the third derivative is -48 and so $p_6'(D)$ is concave on this interval. At $D = N/3$ $p_6'(D)$ equals $(2/3)(-13140 - 64N + N^2)$ and at $D = N/2$ it equals $116N - 8760$. Thus, $p_6(D)$ increases for $D \in [N/3, N/2]$ and since at $N/2$

we find $p_6(D)$ equals $51300 - 840N$, we conclude $p_6(D) < 0$ for $D \in [N/3, N/2]$. Therefore $p_6(D) < 0$ for $D \in [31, N/2]$.

We next evaluate $p_4(n)$ at $n = D - 1$ and obtain

$$p_7(D) := -D^2 + 2D(N-3) - N^2 + 6N - 5 \ .$$

Since $p_7'(D) = 2(N-D) - 6 > 0$, we see it increases for $D \in [31, N/2]$. At $D = N/2$ we find $p_7(D)$ equals $-5 + 3N - (1/4)N^2 < 0$, and so we see $p_7(D) < 0$. Therefore $p_4(n) < 0$ when $n = D - 1$, and by the analysis of $p_4(n)$ when $n = 30$ and the analysis of the derivatives of $p_4(n)$ we conclude $p_4(n) < 0$ under the assumptions. Therefore by (4.11), we see that $p_3'(D) < 0$. Thus the expression (4.10) is bounded below at $D = N/2$. When we evaluate the expression at (4.10) at $D = N/2$ we obtain

$$\begin{aligned} p_8(n) :=& \frac{1}{2n^2} - \frac{1}{2n^2 - 3n + 1} - \frac{(N-2n+1)^2}{2(N-n)^2} + \frac{2N - 4n + 2}{N - n} - \frac{1}{N - 2n} + \frac{1}{2n} - \frac{2}{2n-1} \\ &+ \log\left(\frac{N-n}{N}\right) + \log\left(\frac{N}{N-2n}\right) + \log(2) - 2 \ . \end{aligned} \tag{4.12}$$

Differentiating, we find

$$p_8'(n) = \frac{p_9(n)}{(N-n)^3(N-2n)^2} + \frac{n^3 + 3n - 2}{2(n-1)^2 n^3} \ , \tag{4.13}$$

where we define

$$p_9(n) := -2n^3(N-1) + n^2\left(N^2 + 2N - 4\right) - 2n(N-2)N - N^2 \ .$$

We now analyze $p_9(n)$. Differentiating, we find the third derivative to be $-12(N-1) < 0$. Therefore, $p_9'(n)$ is concave in n, and $p_9''(n)$ decreases in n. Evaluating $p_9''(n)$ at $n = N/4$, we find it equals $-8 + 7N - N^2 < 0$. Therefore, $p_9(n)$ is concave for $n \in [N/4, N/2 - 1]$. Considering $p_9'(n)$ now, we evaluate at $n = 30$, and $n = N/4$ to find it equals $5160 - 5276N + 58N^2$ and $(1/8)N(N^2 - 5N + 16)$ respectively. Since both are positive under the assumptions, concavity implies $p_9'(n) > 0$ for $n \in [30, N/4]$. Thus we have shown $p_9(n)$ increases for $n \in [30, N/4]$ and is concave for $n \in [N/4, N/2]$. Evaluating at $n = 30$ we find

$p_9(n)$ equals $50400 - 52080N + 839N^2$ and evaluating at $n = N/2 - 1$ we find $p_9(n)$ equals $-6 + 7N - (7/2)N^2 + (1/4)N^3$. Since both are positive under the assumptions, we conclude $p_9(n) > 0$ under them. By (4.13) this implies $p_8'(n) > 0$ Therefore $p_8(n)$ increases in n, and so evaluating at $n = 30$ yields a lower bound of

$$-\frac{(N-59)^2}{2(N-30)^2} + \frac{2(N-59)}{N-30} - \frac{1}{N-60} + \log\left(\frac{N}{N-60}\right) + \log\left(\frac{N-30}{N}\right) - \frac{105301}{52200} + \log(2) \; .$$

Differentiating, we find this expression is minimized at $N = 870\sqrt{195} + 12210$, where it attains a value of approximately 0.1758. This being positive, we see $p_8(n) > 0$, and so (4.12) and (4.10) imply $p_1'(n) > 0$. Thus, $p_1(n)$ increases in n, and evaluating at $n = 30$ provides a lower bound which we define to be $p_{10A}(D)$. We then further define the lower bound

$$p_{10}(D) := p_{10A}(D) - 10\log\left(1 + \frac{D}{N}\right) < p_{10A}(D) \; . \tag{4.14}$$

Differentiating twice, we find

$$\begin{aligned} p_{10}''(D) = &\frac{2\left(-2D^4 + 120D^3 - 1800D^2 + 225N - 6750\right)}{(D-30)^2 D^2 (N-30)} \\ &+ \frac{30(D-31)}{(D-30)^2 D} + \frac{10}{(N+D)^2} + \frac{3600}{(59N-60D)^2} + \frac{4}{N} \; . \end{aligned}$$

For $D \in [31, 36]$ this is positive since all terms but the first are positive by inspection, and

$$-2D^4 + 120D^3 - 1800D^2 + 225N - 6750 \geq -100062 + 225N \geq 12438$$

when we assume $N \geq 500$. Thus, $p_{10}''(D) \geq 0$ when $D \in [31, 36]$. For $D \in [36, N/7]$ we write

$$\begin{aligned} p_{10}''(D) := &\frac{3\left(-40D^2 + 2400D + N^2 - 30N - 36000\right)}{(D-30)^2(N-30)N} \\ &+ \frac{3(9D-310)}{(D-30)^2 D} + \frac{450}{(D-30)^2 D^2} + \frac{10}{(D+N)^2} + \frac{3600}{(60D-59N)^2} \; . \end{aligned}$$

Each term in this expression is positive since $-40D^2 + 2400D + N^2 - 30N - 36000$ is concave in D, and equals $-1440 - 30N + N^2$ at $D = 36$ and $-36000 + (2190/7)N + (9/49)N^2$ at $D = N/7$. Next, for $D \in [N/7, (9/25)N]$ we write

$$\begin{aligned} p_{10}''(D) = &\frac{8\left(-15D^2 + 900D + 2N^2 - 60N - 13500\right)}{(D-30)^2(N-30)N} + \frac{2(7D-465)}{(D-30)^2 D} \\ &+ \frac{450}{(D-30)^2 D^2} + \frac{10}{(D+N)^2} + \frac{3600}{(60D-59N)^2} \; . \end{aligned}$$

Again each term is positive, with the first term following by concavity of the quadratic in D and equality with $-13500+(480/7)N+(83/49)N^2$ and $-13500+264N+(7/125)N^2$ at the boundary points. Finally, for $D \in [(9/25)N, N/2]$ we write

$$p''_{10}(D) = -\frac{30\left(-4D^3+240D^2+DN^2-30DN-3600D-31N^2+930N\right)}{(D-30)^2D(N-30)N}$$
$$+\frac{450}{(D-30)^2D^2}+\frac{10}{(D+N)^2}+\frac{3600}{(60D-59N)^2}\ .$$

We observe that $-4D^3+240D^2+DN^2-30DN-3600D-31N^2+930N$ is concave in D, since its second derivative is $480-24D$. Evaluating this cubic at the D-boundary, we find that when $D=(9/25)N$ it equals

$$\frac{N(2709N^2-167125N-5718750)}{15625}$$

and when $D=N/2$ it equals $14N^2-870N$. Since both are positive, we conclude $p''_{10}(D)>0$ for $D\in[(9/25)N, N/2]$. Combining the four preceding expansions, we have shown $p''_{10}(D)>0$ for $D\in[31,N/2]$. Hence $p'_{10}(D)$ increases in D. Evaluating at $D=N/2$ we find $p'_{10}(D)$ is bounded above by

$$-\frac{2(N-59)}{N-30}-\frac{400}{87N}+\frac{60}{(N-60)N}-\log\left(\frac{N}{N-60}\right)+2\ .$$

The derivative of this expression is

$$\frac{2\left(287N^4-51660N^3+5472000N^2-219834000N+788940000\right)}{87(N-60)^2(N-30)^2N^2}$$

which is positive under the assumptions, and therefore the expression increases in N. Taking the limit, we find it $\nearrow 0$ as $N\nearrow\infty$ and at $N=100$ approximately equals -0.1186. Therefore, the expression is negative and we have thus demonstrated that $p'_{10}(D)<0$. Therefore $p_{10}(D)$ decreases in D and evaluating at $D=N/2$ yields the lower bound

$$-\frac{841N}{60(N-30)}-\frac{1}{2}(N-60)\log\left(\frac{N}{N-60}\right)-(N-30)\log\left(\frac{N-30}{N}\right)$$
$$+\frac{1}{2}\log\left(\frac{N-60}{30\pi N}\right)+30\log(2)-10\log\left(\frac{3}{2}\right)+\log\left(\frac{30}{29}\right)-\log\left(\frac{841}{811}\right)\ .$$

Numerical checking shows this quantity is positive for $N \in [500, 869]$. For $N \geq 869$ we find the second derivative

$$-\frac{N^4 - 120N^3 - 698400N^2 + 42120000N + 48600000}{(N-60)^2(N-30)^3N^2}$$

is negative, implying the first derivative of the expression decreases for $N \geq 869$. At $N = 869$ the first derivative approximately equals 3.5×10^{-7}, and in the limit the first derivative decreases from this value to 0. Finally the expression itself is approximately 0.4472 at $N = 869$, and as $N \nearrow \infty$ it increases to approximately 0.4477. Therefore, the expression is positive under the assumptions, and we see $p_{10}(N) \geq > 0$. Therefore (4.14) implies $p_{10A}(N) > 0$, which in turn implies $p_1(n) \geq 0$. This nearly gives the claim.

It remains to show $p_2(n) \geq 0$. We differentiate to find

$$p_2'(n) = \frac{-2D^2 + 4Dn - n^2 - 2nN + N^2}{24(D-n)^2(N-n)^2} + \frac{5(D-n)^2 - 1}{120(D-n)^4} \; .$$

Since

$$-2D^2 + 4Dn - n^2 - 2nN + N^2 \geq -n^2 + \frac{N^2}{2} > 0 \; ,$$

we see $p_2'(n) > 0$ (the first inequality following by decrease in D and evaluating at $D = N/2$). Hence, evaluating $p_2(n)$ at $n = 30$ provides the lower bound

$$-\frac{1}{360(D-30)^3} - \frac{1}{12D} + \frac{1}{12(D-30)} - \frac{1}{360N^3} - \frac{1}{12(N-30)} + \frac{1}{12N} \; . \tag{4.15}$$

Differentiating this lower bound with respect to D, we find it equals

$$\frac{-600D^3 + 45001D^2 - 1080000D + 8100000}{120(D-30)^4D^2} \; ,$$

which is negative since the cubic in the numerator decreases in D and equals -8639 at $D = 31$. Hence, by evaluating (4.15) at $D = N/2$ we find the lower bound of

$$\frac{300N^5 - 42001N^4 + 1800050N^3 - 21601800N^2 + 60000N - 720000}{40(N-60)^3(N-30)N^3} \; ,$$

which is positive under the assumptions (and tends to 0 in the limit). Therefore $p_2(n) > 0$ and the claim follows. □

4.3 α: Main Positivity when $31 \leq D \leq (15/100)N$ and $x = (N-5D)/(5(N-D))$

Lemma 25. *Let* $N \geq 500,\ n,\ D \in \mathbb{N},\ 31 \leq D \leq (15/100)N$, *and* $31 \leq n \leq D$. *Using the functions in* (3.46), *we have that the function*

$$G_1\left(D|x = \frac{N-5D}{5(N-D)}\right) + \left[\sum_{i=1}^{4} A_i\left(D|x = \frac{N-5D}{5(N-D)}\right) + B_i\left(D|x = \frac{N-5D}{5(N-D)}\right)\right] + R(D) \tag{4.16}$$

is positive.

Proof. Differentiating (4.16) twice, we may write the second derivative as

$$\begin{aligned} &\frac{p_1(n)}{100D^2N(N-n)} + \frac{n \cdot p_2(n)}{10D^2(5D-n)^2} + \frac{4n \cdot p_3(n)}{5(N-D)^2(5(N-D)-4n)^2} \\ &+\frac{24n(N-D)-5}{30(N-D)^3} + \frac{3(n-30)}{200D^2} + \frac{3Dn-100}{600D^3} \end{aligned} \tag{4.17}$$

where we define

$$\begin{aligned} p_1(n) &:= -18n^2N - 400D^2n - 45nN + 45N^2 + 18nN^2 \ , \\ p_2(n) &:= -2n^2 + 5n + 10Dn - 50D \ , \\ \text{and } p_3(n) &:= -16n^2 + 10n - 20Dn + 20nN - 25N + 25D \ . \end{aligned}$$

We claim (4.17) is positive. As written, this is clear if we show $p_i(n) \geq 0$ for $i \in \{1, 2, 3\}$. Notice first that each $p_i(n)$ is concave in n. Therefore, we check each function at $n = 31$ and $n = D$, and demonstrate positivity at each end-point. Recall that we currently assume $D \leq (15/100)N$, since we will use this fact to derive lower bounds in the following discussion. Starting with $p_1(n)$ at $n = 30$ it equals

$$-12000D^2 - 17550N + 585N^2 \geq -17550N + 315N^2 > 0 \ ,$$

and at $n = D$ it equals

$$-400D^3 - 18D^2N + 9DN(2N-5) + 45N^2 \ .$$

This upper bound is concave in D, so checking $D = 31$ this becomes

$$603N^2 - 18693N - 11916400$$

and at $D = (15/100)N$ it equals

$$\frac{9}{200}N^2(21N + 850) \ .$$

Since both are positive, we see that $p_1(n) \geq 0$ by concavity.

Turning to $p_2(n)$, we evaluate at $n = 31$ and find it equals $260D - 1767 > 0$ and at $n = D$ it equals $D(8D - 45)$. Once again, concavity carries the day.

Finally with $p_3(n)$, evaluating at $n = 31$ we find it is

$$595(ND) - 15066 \geq \frac{595N}{2} - 15066 > 0 \ ,$$

and at $n = D$ it equals

$$-36D^2 + 5D(4N + 7) - 25N \ .$$

Since this final expression is concave in D we check $D = 31$ to find it equals $-33511+595N > 0$ and at $D = (15/100)N$ it equals $(1/100)N(219N - 1975)$. These are positive under the assumptions and sowe have that $p_3(n) \geq 0$. Hence we have demonstrated that (4.17) is positive. This implies the first derivative of (4.16) increases in D.

Evaluate the first derivative of (4.16) at $D = (15/100)N$ and defined this upper bound to be $p_4(n)$. We differentiate to find

$$\begin{aligned} p_4'(n) = &-\frac{160}{(17N-16n)^2} - \frac{3N-4n-200}{5(3N-4n)^2} \\ &- \frac{N^2+5n^2-5nN}{5(3N-4n)(N-n)^2} - \frac{15N^2+82nN-96n^2}{5(17N-16n)(3N-4n)(N-n)} \ . \end{aligned} \tag{4.18}$$

This quantity is negative, since

$$N^2 + 5n^2 - 5nN \geq \frac{29}{80}N^2 > 0$$

and the remaining terms respect their leading sign by inspection under the assumptions. Therefore, $p_4(n)$ decreases in n and evaluating at $n = 31$ yields the upper bound

$$\frac{28000}{7803N^2}+\frac{124(N-38)}{N(3N-124)}+\frac{31}{5(N-31)}-\frac{2078}{51N}-\frac{10}{17N-496}-\log\left(1+\frac{31}{17(N-31)}\right)+\log\left(\frac{3N-124}{3N-93}\right) .$$

The derivative of this expression

$$\frac{\begin{bmatrix} 604490607N^7 - 99954723744N^6 + 6957305641068N^5 \\ -268170337019776N^4 + 6210862143182112N^3 \\ -84550549224064000N^2 + 574023913718169600N \\ -1017860221665280000 \end{bmatrix}}{39015(17N-496)^2(3N-124)^2(N-31)^2N^3}$$

is positive, and so we see it increases in N. Taking the limit we see as $N \nearrow \infty$ the expression $\nearrow 0$, and at $N = 100$ it approximately equals -0.077. Therefore, $p_4(n) < 0$ under the assumptions, and by its definition before (4.18) this implies the first derivative of (4.16) is negative under the assumptions.

Therefore, the expression (4.16) decreases in D, and evaluating the expression at $D = (15/100)N$ yields the lower bound which we define to be $p_5(n)$. Differentiating, we find

$$\begin{aligned} p_5'(n) = &-\frac{n^2}{200(N-n)^2}+\frac{7}{12n^2}-\frac{n}{100(N-n)}+\frac{1}{N-n}-\frac{7}{12(N-n)^2}-\frac{2}{3N-4n} \\ &-\frac{8}{17N-16n}-\frac{1}{200}-\frac{1}{5}\log\left(1-\frac{n}{3(N-n)}\right)-\frac{4}{5}\log\left(1+\frac{n}{17(N-n)}\right) \\ &+\frac{1}{5}\log\left(\frac{4}{3}\right)-\frac{4}{5}\log\left(\frac{17}{16}\right) . \end{aligned}$$

We define

$$p_6(n) := p_5'(n) - \frac{1}{8n} < p_5'(n) . \tag{4.19}$$

We differentiate to find

$$p_6'(n) = \frac{3n-28}{24n^3}+\frac{p_7(n)}{1500(3N-4n)^2(N-n)^3}+\frac{p_8(n)}{500(16n-17N)^2(4n-3N)^2(n-N)^2} \tag{4.20}$$

where we define

$$p_7(n) := \begin{bmatrix} -240n^4 + 12n^3(29N - 2000) + n^2\left(-99N^2 + 60000N - 28000\right) \\ -12nN\left(3N^2 + 4125N - 3500\right) + 6N^2\left(2N^2 + 2250N - 2625\right) \end{bmatrix}$$

and

$$p_8(n) := \begin{bmatrix} -20480n^5 + 512n^4(103N - 4000) + 16n^3N(488000 - 3341N) \\ +4n^2N^2(10677N - 2801000) + nN^3(7176000 - 32857N) + N^4(11339N - 1732000) \end{bmatrix} .$$

We now show $p_7(n) > 0$ and $p_8(n) > 0$, which will imply $p_6'(n) > 0$ by (4.20) . First consdier $p_7(n)$. The fourth derivative is -5760, implying the third derivative is bounded below by $1224N - 144000$ for $n \in [31, (15/100)N]$. Therefore, $p_7'(n)$ is convex in n. Evaluating $p_7'(n)$ at $n = 31$, we find it equals

$$-36N^3 - 55638N^2 + 4765284N - 99527360 ,$$

and at $n = (15/100)N$ we find $p_7'(n)$ equals

$$-\frac{909N^3}{20} - 33120N^2 + 33600N .$$

Since both are negative under the assumptions, convexity implies $p_7'(n) < 0$. Therefore $p_7(n)$ decreases in n, and evaluating at $n = (15/100)N$ yields the lower bound

$$\frac{10851N^4}{2000} + 7344N^3 - 10080N^2 .$$

This is positive under the assumptions, and so we see $p_7(n) > 0$ as claimed.

Turning to $p_8(n)$, the fifth derivative -2457600. Under the assumption $n \in [31, (15/100)N]$, the remaining derivatives either have monotone decrease or monotone decrease. Evaluating at $n = (15/100)N$, we find: the fourth is bounded below by $897024N - 49152000$; the third is bounded above by $39475200N - (792672/5)N^2$; the second bounded below by $-15933760N^2 + (1254048/25)N^3$; and the first above by $-(2874098/125)N^4 + 4314192N^3$. Therefore $p_8(n)$ decreases in n on the interval $[31, (15/100)N]$, and at $n = (15/100)N$ is bounded below by

$$\frac{18040271N^5}{2500} - \frac{4411874N^4}{5} ,$$

which is positive. Therefore $p_8(n) \geq 0$, and combined with the positivity of $p_7(n)$ we thus see $p_6'(n) > 0$ by (4.20). This implies that $p_6(n)$ increases in n, and evaluating $p_6(n)$ at $n = 31$ produces the lower bound

$$\frac{69}{100(N-31)} - \frac{3233}{600(N-31)^2} + \frac{2}{124-3N} + \frac{8}{496-17N} - \frac{4}{5}\log\left(1+\frac{31}{17(N-31)}\right)$$
$$-\frac{1}{5}\log\left(1-\frac{31}{3N-93}\right) - \frac{2429}{288300} + \frac{1}{5}\log\left(\frac{4}{3}\right) - \frac{4}{5}\log\left(\frac{17}{16}\right).$$

The derivative of this expression is

$$\frac{\begin{bmatrix} -125307N^5 + 20636394N^4 - 1341532068N^3 \\ +43499680376N^2 - 707334624512N \\ +4622037900800 \end{bmatrix}}{300(N-31)^3(3N-124)^2(17N-496)^2}$$

is negative under the assumptions, implying the expression decreases monotonically in N. As $N \nearrow \infty$, the expression decreases to a value of approximately 0.0006, and since at $N = 100$ it is approximately 0.003, we see that the expression is negative under the assumptions. Therefore $p_6(n) > 0$ under the assumptions and so (4.19) implies that $p_5'(n) > 0$. Therefore $p_5(n)$ increases in n, and evaluating $p_5(n)$ at $n = 31$ yields the lower bound

$$-\frac{1}{360N^3} - \frac{1061N}{6200(N-31)} - \frac{349}{612N} - \frac{1}{12N-372} - \frac{1}{372} + \frac{67}{10}\log\left(\frac{4}{3}\right) - \frac{253}{10}\log\left(\frac{17}{16}\right)$$
$$+\frac{1}{2}\log\left(\frac{51}{50}\right) + \frac{3(N-38)}{20}\log\left(\frac{31}{93-3N}+1\right) + \frac{17N-486}{20}\log\left(1+\frac{31}{17(N-31)}\right).$$

The second derivative of this expression is

$$\frac{\begin{bmatrix} -63030033N^9 + 12918360486N^8 - 1189360318052N^7 \\ +61991956477244N^6 - 1972215622704458N^5 \\ +38377351553527970N^4 - 420310302226280240N^3 \\ +1967591176306787360N^2 - 12282482942684160N \\ +57472750373314560 \end{bmatrix}}{15300(496-17N)^2(124-3N)^2(N-31)^3N^5},$$

which is negative under the assumptions. Therefore the first derivative of the expression decreases in N from a value of approximately 0.00011 at $N = 100$ to 0 in the limit. This implies the expression increases in N. A final evaluation at $N = 100$ yields an approximate value of 0.2205, and in the limit it increases to approximately 0.2297. Since the expression provides a lower bound on $p_5(n)$, we thus see $p_5(n) \geq 0$. Since by definition $p_5(n)$ provides a lower bound on (4.16) , the claim is proved. □

4.4 β: Main Positivity when $(15/100)N \leq D \leq N/3$ and $x = N/(n(N-D))$

Lemma 26. *Let $N \geq 500$, $n, D \in \mathbb{N}$, $(15/100)N \leq D \leq N/3$, and $31 \leq n \leq N/3$. Using the functions in (3.46), we have that the function*

$$G_1\left(D|x = \frac{N}{n(N-D)}\right) + \left[\sum_{i=1}^{4} A_i\left(D|x = \frac{N}{n(N-D)}\right) + B_i\left(D|x = \frac{N}{n(N-D)}\right)\right] + R(D) \tag{4.21}$$

increases in D.

Proof. To prove the claim, we begin by considering the second-order partial derivatives of (4.21) with respect to D. We first observe that

$$r'(D) = -\frac{1}{6D^3} - \frac{1}{2D^2} - \frac{3}{(N-D)^2} - \frac{1}{(N-D)^3} ,$$

which is negative under the assumptions. Next observe

$$a_1'\left(D|x = \frac{N}{n(N-D)}\right) + b_1'\left(D|x = \frac{N}{n(N-D)}\right) = -\frac{N^2}{2(N-D)^2(n(N-D)-N)^2}$$

and

$$a_3'\left(D|x = \frac{N}{n(N-D)}\right) + b_3'\left(D|x = \frac{N}{n(N-D)}\right) = -\frac{N^2}{2D^2(D(N-n)+N)^2}$$

are both negative by inspection. We next observe that

$$a_2'\left(D|x = \frac{N}{n(N-D)}\right) + b_2'\left(D|x = \frac{N}{n(N-D)}\right) - \frac{12}{25D^2} = \frac{-24D^2n^2 + 52DnN + 51N^2}{50D^2(Dn+N)^2} .$$

This is negative since the numerator decreases in D, and so is bounded above by

$$-24D^2n^2 + 52DnN + 51N^2 \leq \frac{3}{50}\left(-9n^2 + 130n + 850\right) N^2 \ ,$$

which is negative for $n \geq 20$. Finally we have

$$\begin{aligned} & a_4'\left(D|x = \frac{N}{n(N-D)}\right) + b_4'\left(D|x = \frac{N}{n(N-D)}\right) - \frac{3}{150D^2} \\ &= \left[\frac{1}{50D^2(N-D)^2((N-n)(N-D)+N)^2}\right] \cdot p_1(D) \end{aligned} \tag{4.22}$$

where we define

$$p_1(D) := \begin{bmatrix} -(N-n)^2D^4 + 2N(N-n)(2(N-n)-49)D^3 \\ +2N^2(37 - 3n^2 - (3N-47)N + n(6N-47))D^2 \\ +2N^3(N-n+1)(2(N-n)+1)D - N^4(N-n+1)^2 \end{bmatrix} .$$

We see that demonstrating $p_1(D) \leq 0$ will show (4.22) is negative too. We proceed by differentiation. The fourth derivative of $p_1(D)$ is $-24(N-n)^2$. Evaluating at $D = N/3$ we find the third derivative is bounded below by $4(4(N-n)-147)(N-n)N > 0$. This implies $p_1''(D)$ increases in D, and so evaluating at $D = N/3$ we find it is bounded above by

$$-\frac{4N^2}{3}\left(4n^2 - 2n(4N+3) + 4N^2 + 6N - 111\right) \leq -\frac{4N^2}{3}\left(\frac{16N^2}{9} + 4N - 111\right) < 0 \ .$$

The first inequality in the preceding follows since the parenthetical expression decreases in n, and so evaluating at $n = N/3$ yields an upper bound on the negative quantity. Hence, $p_1'(D)$ decreases in D. Hence, evaluating again at $D = N/3$ produces the lower bound for $p_1'(D)$ of

$$\frac{2N^3}{27}(8(N-n)+231)(2(N-n)+3) \ ,$$

which is positive by inspection. Hence $p_1(D)$ increases in D and we evaluate a final time at $D = N/3$ to obtain the upper bound of $(N^4/81) \cdot p_2(n)$, where we define

$$p_2(n) := -16n^2 + 8n(4N-69) + 8N(69-2N) + 639 \ .$$

Since $p_2'(n) = 8(4(N-n)-69) > 0$, we evaluate $p_2(n)$ at $n = N/3$ to obtain the upper bound of $639 + 368N - (64/9)N^2$, which is negative under the assumptions.

The preceding arguments taken together imply that the second derivative of (4.21) is negative. So, the first derivative of (4.21) decreases in D. Evaluate the first derivative of (4.21) at $D = N/3$. We define this to be $p_3(n)$. We then find

$$p_3''(n) = \frac{9}{(n+3)^3(2n-3)^3(2(N-n)+3)^3(N-n-3)^3 N} \cdot p_4(n) \tag{4.23}$$

where we define

$$p_4(n) := -8n^7 p_5(n) + \frac{3n^3}{5} p_6(n) + p_7(n) \tag{4.24}$$

and further define

$$p_5(n) := 20n^2 - 54(2n+3)N + 54n + 252N^2 - 81\ ,$$

$$p_6(n) := \begin{bmatrix} 180n^3(-39-81N-8N^2+24N^3) \\ -20n^2(486-972N-1539N^2+108N^3+160N^4) \\ +5n(20412+21141N-9558N^2-5184N^3+840N^4+256N^5) \\ -3(-2187+59049N+51759N^2-21870N^3-2862N^4+792N^5+64N^6) \end{bmatrix} ,$$

and

$$p_7(n) := \begin{bmatrix} -\frac{1}{5}n^3\left(64N^6+792N^5-2862N^4-21870N^3+51759N^2+59049N-2187\right) \\ +9n^2\left(40N^6-36N^5-3330N^4+8505N^3+9477N^2-15309N-21870\right) \\ -27n\left(8N^6-372N^5+1206N^4+1701N^3-6075N^2-8748N-4374\right) \\ -27\left(56N^6-252N^5-378N^4+2187N^3+1458N^2-1458\right) \end{bmatrix} .$$

We first suppose $n \in [30, N(6/25)]$, and show on this interval that $p_5(n) \geq 0$, $p_6(n) \leq 0$, and $p_7(n) \leq 0$. By (4.24) this will imply $p_4(n) \leq 0$ when $n \in [30, (6/25)N]$. First, since $p_5'(n) = 54 + 40n - 108N < 0$, by evaluating at $n = N/4$ we find $p_5(n)$ is bounded below by $-81 - (297/2)N + (905/4)N^2$, and so is positive.

Next, since the third derivative of $p_6(n)$ is $1080(-39-81N-8N^2+24N^3) > 0$, we find $p_6''(n)$ increases in n. Evaluating at $n = (6/25)N$, we find the second derivative is bounded above by

$$-\frac{8}{5}\left(112N^4+3996N^3-25353N^2-17982N+12150\right) < 0\ .$$

So, $p_6'(n)$ decreases in n. Again evaluating at $n = (6/25)N$, we find the first derivative is bounded above by

$$\frac{1}{125}\left(61312N^5 + 364296N^4 - 1708128N^3 - 4958982N^2 + 12629925N + 12757500\right) > 0\ .$$

Hence, $p_6(n)$ increases in n. A final evaluation at $n = (6/25)N$ yields the negative upper bound

$$-\frac{9\left(3264N^6 + 525112N^5 - 1366866N^4 - 19153854N^3 + 45301275N^2 + 53004375N - 2278125\right)}{3125}\ ,$$

and so we infer $p_6(n) \leq 0$ when $n \in [30, (6/25)N]$.

We next show $p_7(n) \leq 0$ when $n \in [30, (6/25)N]$. The third derivative of $p_7(n)$ is

$$-\frac{6}{5}\left(64N^6 + 792N^5 - 2862N^4 - 21870N^3 + 51759N^2 + 59049N - 2187\right)$$

Hence, we evaluate $p_7''(n)$ at $n = 30$ to find it bounded above by

$$-18\left(88N^6 + 1620N^5 - 2394N^4 - 52245N^3 + 94041N^2 + 133407N + 17496\right)\ ,$$

which is also negative. Similarly, we evaluate $p_7'(n)$ at $n = 30$ to find it bounded above by the negative quantity

$$-27\left(488N^6 + 16188N^5 + 10566N^4 - 605799N^3 + 839565N^2 + 1478412N + 389286\right)\ .$$

A final evaluation of $p_7(n)$ at $n = 30$ yields the negative upper bound of

$$-27\left(1096N^6 + 157788N^5 + 462402N^4 - 6872283N^3 + 7327908N^2 + 16140060N + 5990922\right)\ .$$

We thus have shown for $n \in [30, (6/25)N]$ that $p_5(n) \geq 0$, $p_6(n) \leq 0$, and $p_7(n) \leq 0$. Hence by (4.24) this implies $p_4(n) \leq 0$ when $n \in [30, (6/25)N]$. It remains to show $p_4(n) \leq 0$ when $n \in [(6/25)N, N/3]$. This follows by direct differentiation. The details are very similar to the arguments given for $p_5(n), p_6(n)$, and $p_7(n)$, so we simply sketch the remaining details.

The eighth derivative of $p_4(n)$ is $-58060800n + 17418240(2N - 1)$, which is positive. Hence the seventh derivative increases in n, and at $n = N/3$ is bounded above by a negative quantity. This implies the sixth derivative decreases in n, and at $n = N/3$ is bounded below by a positive quantity. This implies the fourth derivative is convex in n.

Evaluate the fourth derivative at $n = (6/25)N$ and $n = N/3$ to find it is negative at both end-points. By convexity, this implies the second derivative is concave in n. Next evaluate the second derivative at $n = (6/25)N$ and $n = N/3$ to find it is positive at both end-points. By concavity, this implies $p_4(n)$ is convex for $n \in [(6/25)N, N/3]$. Finally evaluate $p_4(n)$ at $n = (6/25)N$ and $n = N/3$ to find it is negative at both end-points. By convexity, we conclude $p_4(n) \leq 0$ for $n \in [(6/25)N, N/3]$. Hence $p_4(n) \leq 0$ for $n \in [30, N/3]$, and so by (4.23) we infer the first derivative of (4.21) is concave in n.

The argument concludes by evaluating the first derivative of (4.21) at $n = 30$ and $n = N/3$. At these points we obtain two univariate functions in N, which we define to be $p_8(N)$ and $p_9(N)$. Both $p_8(N)$ and $p_9(N)$ are positive (which can be verified analytically by combining terms and determining the sign of the numerator), monotone decreasing in N, and tend to 0 as $N \nearrow \infty$. Hence the first derivative of (4.21) is positive by concavity, and the claim follows. □

Lemma 27. *Let $N \geq 500$ and $31 \leq n \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$G_1\left(n \mid D = \frac{15}{100}N\ ,\ x = \frac{N}{n(N-D)}\right) + R\left(n \mid D = \frac{15}{100}N\right) + \left[\sum_{i=1}^{4} A_i\left(n \mid D = \frac{15}{100}N\ ,\ x = \frac{N}{n(N-D)}\right) + B_i\left(n \mid D = \frac{15}{100}N\ x = \frac{N}{n(N-D)}\right)\right] \tag{4.25}$$

is positive.

Proof. We begin by showing the second partial derivative of (4.25) with respect to n is positive. We observe that the second partial derivative may be written as

$$p_1(n) + p_2(n) + p_3(n) \tag{4.26}$$

where we define

$$p_1(n) := \frac{127449n^4 - 345360n^3 - 3469600n^2 + 7904000n - 4960000}{6(20-17n)^2 n^3 (3n+20)^2} \ ,$$

$$p_2(n) := -\frac{289}{2(17(N-n)+20)^2} + \frac{1}{(N-n)^2} - \frac{31}{6(N-n)^3} - \frac{9}{2(3(N-n)-20)^2} \ ,$$

and

$$p_3(n) := \frac{400}{(17(N-n)+20)(3(N-n)-20)(N-n)} \ .$$

We see that $p_3(n) \geq 0$ by definition. We next observe that $p_1(n)$ decreases in n since we assume $n \geq 31$ and

$$p_1'(n) = \frac{\begin{bmatrix} -6499899n^6 + 11589240n^5 + 342390000n^4 + 165280000n^3 \\ -2823200000n^2 + 4422400000n - 1984000000 \end{bmatrix}}{2n^4(3n+20)^3(17n-20)^3} \ .$$

Hence $p_1(n) \geq p_1(N/3)$. Examining the components of $p_2(n)$, we see

$$-\frac{289}{2(17(N-n)+20)^2} \geq -\frac{289}{2(17(N-31)+20)^2} \ ,$$

$$\frac{1}{(N-n)^2} - \frac{31}{6(N-n)^3} \geq \frac{1}{(N-31)^2} - \frac{31}{6(N-31)^3} \ ,$$

and

$$-\frac{9}{2(3(N-n)-20)^2} \geq -\frac{9}{2(3(N-31)-20)^2} \ .$$

Using these lower bounds, along with the earlier bound $p_1(n) \geq p_1(N/3)$, we find

$$p_1(n) + p_2(n) \geq \frac{p_4(N)}{6(N-31)^3 N^3 (N+20)^2 (3N-113)^2 (17N-507)^2 (17N-60)^2} \ , \qquad (4.27)$$

where we further define

$$p_4(N) := \begin{bmatrix} 946420668N^{11} - 230213147943N^{10} \\ +23614525238453N^9 - 1321582345545379N^8 \\ +42809381061292226N^7 - 737434886489376657N^6 \\ +2634203023700043909N^5 + 15001990658707301088 3N^4 - 30664979813872364133 60N^3 \\ +251141633222492708712 00N^2 - 881687081467440368640 00N \\ +11785440977577594288 0000 \end{bmatrix} \ .$$

It is readily confirmed that $p_4(N) \geq 0$ for $N \geq 500$. Additionally, the ratio in (4.27) decreases monotonically to 0 as $N \nearrow \infty$, and so we infer that $p_1(n) + p_2(n) \geq 0$. Hence the second derivative at (4.26) is positive, implying the first derivative increases in N. Hence, evaluating the first derivative of (4.25) at $n = N/3$, we find it bounded above by

$$\begin{aligned} p_5(N) := &-\frac{30}{N^2+20N} + \frac{45}{34N^2+60N} + \frac{15}{20N-2N^2} + \frac{279}{16N^2} + \frac{90}{N(17N-60)} \\ &-\frac{3}{20}\log\left(1-\frac{10}{N}\right) + \frac{3}{20}\log\left(1+\frac{20}{N}\right) - \frac{17}{20}\log\left(1+\frac{30}{17N}\right) + \frac{17}{20}\log\left(1-\frac{60}{17N}\right) . \end{aligned}$$

Differentiation shows $p_5'(N) > 0$, and so $p_5(N)$ increases monotonically. As $N \nearrow \infty$, $p_5(N) \nearrow 0$, and at $N = 100$, $p_5(N) \approx -0.0030$. Hence, $p_5(N) < 0$, which implies that (4.25) decreases monotonically in n. Evaluating (4.25) at $n = N/3$ a final time yields the lower bound

$$\begin{aligned} &-\frac{1}{360N^3} - \frac{14927}{1224N} - \frac{1}{2}\log\left(1-\frac{10}{N}\right) + \frac{1}{10}N\log\left(1-\frac{10}{N}\right) + \frac{3}{2}\log\left(1+\frac{20}{N}\right) \\ &+\frac{1}{20}N\log\left(1+\frac{20}{N}\right) + \frac{3}{2}\log\left(1+\frac{30}{17N}\right) + \frac{17}{30}N\log\left(1+\frac{30}{17N}\right) - \frac{1}{2}\log\left(1-\frac{60}{17N}\right) \\ &+\frac{17}{60}N\log\left(1-\frac{60}{17N}\right) + \frac{1}{2}\log\left(\frac{51}{50}\right) . \end{aligned}$$

This lower bound decreases monotonically to $(1/2)\log(51/50) > 0.009$, and at $N = 100$ approximately equals 0.0895. The claim follows. □

4.5 $\gamma 1$: *Main Positivity when* $N/3 \leq D \leq N/2$ *and* $x \in [N/(n(N-D)), 1/2]$

Lemma 28. *Let* $N \geq 500$, n, $D \in \mathbb{N}$, $N/3 \leq D \leq N/2$, $31 \leq n \leq D$, *and* $N/(n(N-D)) \leq x \leq 1/4$. *Under these conditions, the function defined at* (3.46)

$$G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x)\right] + R(D)$$

is positive.

Proof. Under the assumptions, Lemma 22 implies the quantity of interest is bounded below by

$$G_1(0) + \left[\sum_{i=1}^{4} B_i(0)\right] + A_1(1/4) + A_2\left(\frac{N}{n(N-D)}\right) + A_3(1/4) + A_4\left(\frac{N}{n(N-D)}\right) + R(n)$$
$$= \left[A_1(1/4) + A_3(1/4) - \frac{1}{2}\log\left(\frac{N^2}{8(D(N-D))}\right)\right]$$
$$+ \left[A_2\left(\frac{N}{n(N-D)}\right) + A_4\left(\frac{N}{n(N-D)}\right) + R(D) + \frac{1}{2}\log\left(\frac{N^2}{8(D(N-D))}\right)\right] =: [p_1(n)] + [p_2(D)] \ .$$

We see the claim follows if we show $p_1(n)$ and $p_2(D)$ are both positive. We begin by analyzing $p_1(n)$. Differentiating twice, we find

$$p_1''(n) = -\frac{N(N-D)\left(6D(N-n) + DN + N(N-2n)\right)}{2(N-n)^2(3D(N-n) + N(D-n))^2} \ ,$$

which is negative by inspection. This implies $p_1(n)$ is concave in n. We first evaluate $p_1(n)$ at $n = 31$, and define this to be $p_3(D)$. Then

$$p_3'(D) = \frac{2N(N-31) - D(4N-93)}{(N-D)(D(4N-93) - 31N)} \ .$$

This is positive under the assumptions since $2N(N-31) - D(4N-93) > (31/2)N > 0$ and $D(4N-93) - 31N > (2/3)N(2N-93) > 0$. Hence, $p_3(D)$ increases in D, and evaluating at $D = N/3$ gives the lower bound of

$$\frac{1}{2}\log\left(\frac{4N-186}{3N-93}\right) \ ,$$

which is positive under the assumptions. We next evaluate $p_1(n)$ at $n = D$, and define this to be $p_4(D)$. Then

$$p_4'(D) = \frac{N-2D}{2D(N-D)} > 0 \ ,$$

and so $p_4(D)$ increases in D. Since at $D = N/3$, we find that $p_4(N/3)$ is exactly 0, we conclude $p_1(n) \geq 0$ by concavity.

Turning to $p_2(D)$, we differentiate to find

$$p_2'(D) = \frac{p_5(n)}{12(D^2)(N-D)^2(Dn+N)((N-n)(N-D)+N)} \ , \tag{4.28}$$

where we further define

$$p_5(n) := \begin{bmatrix} -DN(2D^2-3DN+N^2)n^2 \\ +N\left(-12D^4+26D^3N-D^2N(21N+4)+DN^2(7N+4)-N^3\right)n \\ N^2\left(6D^4-6D^3(3N+2)+2D^2N(9N+10)-DN\left(6N^2+9N+2\right)+N^2(N+1)\right) \end{bmatrix} .$$

We now show $p_5(n) \leq 0$. Differentiating twice, we find $p_5''(n) = -2DN(2D^2-3DN+N^2) < 0$. Hence $p_5'(n)$ decreases in n. Evaluating $p_5'(n)$ at $n = D$, we find it is bounded below by

$$N(8D^3(N-2D) + DN(24D^2 - D(23N+4) + N(7N+4)) - N^3) .$$

Now, the quadratic in D appearing this lower bound, $24D^2 - D(23N+4) + N(7N+4)$, is minimized at $D = (1/48)(23N+4)$. At this point, the quadratic equals $(1/96)(-16 + 200N + 143N^2)$, which is positive. Now, the quadratic is multiplied by $DN > N^2/3$; using these lower bounds, we find

$$DN(24D^2 - D(23N+4) + N(7N+4)) - N^3 > \frac{N^2}{288}\left(143N^2 - 88N - 16\right) > 0 ,$$

with positivity following by the assumption on N. This implies $p_5'(n) > 0$. Hence $p_5(n)$ increases in n, and so bounded above by

$$-(N-2D)N(7D^3(N-D) + DN(7D^2 - D(13N+8) + 6N^2) + (8D-N-1)N^2) .$$

This quantity is negative, since $7D^2 - D(13N+8) + 6N^2$ decreases in D and is bounded below by $(1/4)N(5N-16) > 0$, and $8D - N - 1 > (5/3)N - 1 > 0$. Therefore, $p_5(n) \leq 0$. By (4.28), this implies $p_2'(D) < 0$, and so $p_2(D)$ decreases in D. Evaluate $p_2(D)$ at $D = N/2$, and define this quantity to be $p_6(n)$. Differentiating to find

$$p_6'(n) = -\frac{p_7(n)}{12n^2(n+2)(N-n+2)(N-n)^2} , \tag{4.29}$$

where we further define

$$p_7(n) := 2n^3(5N+24) - 3n^2N(5N+24) + nN\left(5N^2+52N+56\right) - 14N^2(N+2) .$$

Differentiating twice, we find $p_7''(n) = -6(5N+24)(N-2n) < 0$, which implies $p_7(n)$ is concave in n. At $n = 31$, we find $p_7(n)$ equals $3\left(47N^3 - 4277N^2 + 76818N + 476656\right)$, which is positive. At $n = N/2$, $p_7(n)$ equals 0, and so concavity implies $p_7(n) \geq 0$. Hence by (4.29), we see $p_6'(n) \leq 0$. Hence, $p_6(n)$ decreases in n, and so evaluating at $n = N/2$ we find it bounded below by

$$\log\left(1 + \frac{4}{N}\right) - \frac{1}{360N^3} - \frac{31}{12N} \ .$$

Since the derivative of this quantity with respect to N is

$$\frac{-170N^3 + 1240N^2 + N + 4}{120N^4(N+4)} \ ,$$

we see it has monotone decrease in N. In the limit, it goes to 0, and at $N = 100$ it is approximately equal to 0.0134. This implies $p_2(n) \geq 0$, and so the claim follows. □

Lemma 29. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $N/3 \leq D \leq N/2$, $31 \leq n \leq D$, and $1/4 \leq x \leq 1/2$. Under these conditions, the function defined at* (3.46)

$$G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x)\right] + R(D)$$

is positive.

Proof. Under the assumptions, Lemma 22 implies that

$$G_1(x) + \sum_{i=1}^{4} B_i(x) \geq G_1(0) + \sum_{i=1}^{4} B_i(0) = 0 \ .$$

Since we assume $1/4 \leq x \leq 1/2$, we have by Lemma 23 that

$$\sum_{i=1}^{4} A_i(x) \geq \sum_{i=1}^{4} A_i\left(n \middle|\ x = \frac{1}{2}\right) \ .$$

Therefore the claim will follow by showing

$$p_1(n) := R(n) + \sum_{i=1}^{4} A_i\left(n \middle|\ x = \frac{1}{2}\right) \geq 0 \tag{4.30}$$

where we write $R(n)$ in place of $R(D)$. We begin by showing $p_1(n)$ is concave in n. Differentiating twice, we find

$$p_1''(n) = -\frac{7}{6n^3} - \frac{7}{6(N-n)^3} - \frac{1}{2(2N-n)^2} - \frac{p_2(n)}{2(N-n)^2(D(N-n)+N(D-n))^2} \quad (4.31)$$

where we define

$$p_2(n) := -(N+D)^2n^2 - 2N(-3D^2-2DN+N^2)n + N^2(-7D^2+2DN+N^2)\ .$$

Once again, we differentiate twice to find $p_2''(n) = -2(N+D)^2 < 0$, which implies $p_2(n)$ is concave in n. At $n = 31$, $p_2(n)$ equals to

$$D^2\left(-7N^2+186N-961\right) + 2DN\left(N^2+62N-961\right) + N^2\left(N^2-62N-961\right)\ .$$

Differentiating this expression twice, we see it is concave in D. Further evaluating at the D-boundary, we find when $D = N/3$ it equals $(8/9)N^2(N^2-1922)$ and when $D = N/2$ it equals $(1/4)N^2(N^2+186N-8649)$. Since both are positive, concavity implies $p_2(n)$ is positive when $n = 31$. Turning the upper n-boundary, we evaluate $p_2(n)$ at $n = D$ to find it equals

$$-D^4 + 4D^3N - 4D^2N^2 + N^4\ .$$

The third derivative of this expression is $24(N-D) > 0$, which implies the first derivative of it is convex in D. Evaluating the first derivative of this expression at $D = N/3$, we find it equals $-(40/27)N^3$, and when $D = N/2$ it equals $-(3/2)N^3$. Hence convexity implies the expression decreases in D. Evaluating the expression at $D = N/2$, we find it equals $(7/16)N^4 > 0$. Hence, when $n = D$, we see $p_2(n) \geq 0$. This combined with the positivity of $p_2(n) \geq 0$ when $n = 31$ and the concavity of $p_2(n)$ imply that $p_2(n) \geq 0$. By (4.31), we therefore see that $p_1(n)$ is concave in n.

We next evaluate $p_1(n)$ at $n = 31$ and define this to be $p_3(D)$. Differentiate once with respect to D to find

$$p_3'(D) = -\frac{1}{12(N-D)^2} + \frac{N-3D}{4(N-D)(N+D)} + \frac{p_4(D)}{12D^2(N-D)(2DN-31(N+D))}\ , \quad (4.32)$$

where we define

$$p_4(D) := -D^3(6N-93) - D^2(95N-31) + 188DN^2 - 31N^2 \ .$$

First note that $N-3D \le 0$ for $N/3 \le D \le N/2$, and so the second term in (4.32) is negative. Hence if we show $p_4(D) \le 0$, then we can infer the entire expression (4.32) is negative. We do so now. Since $p_4''(D) = -6D(6N-93) - 2(95N-31) < 0$, we see $p_4'(D)$ decreases in D. Hence we evaluate $p_4'(D)$ at $D = N/3$ to find it bounded above by $(1/3)N(62+467N-6N^2)$, which is negative. Hence, $p_4(D)$ decreases in D, and again evaluating at $D = N/3$ we find it bounded above by $-(2/9)N^2(124-250N+N^2)$, which is also negative. So, $p_4(D) \le 0$, and thus $p_3'(D) \le 0$ by (4.32).

We thus have shown $p_3(D)$ decreases in D. Evaluating at $D = N/2$, we find it equals

$$-\frac{1}{360N^3} - \frac{N}{62(N-31)} - \frac{1}{4N} + \frac{1}{372-12N} - \frac{1}{372} + \frac{1}{2}\log\left(1 + \frac{31}{2(N-31)}\right) + \frac{1}{2}\log\left(1 - \frac{31}{2N-62}\right) + \frac{1}{2}\log\left(\frac{3}{2}\right) \ .$$

The derivative of this expression is

$$\frac{1}{120N^4} + \frac{1}{4N^2} + \frac{7}{12(N-31)^2} + \frac{961}{(N-31)(2N-93)(2N-31)} \ ,$$

and so we see it increases in N. Since it approximately equals 0.005 at $N = 61$, and increases to a limiting value of about 0.1839 as $N \nearrow \infty$, we thus conclude $p_3(D) \ge 0$ (recalling $p_3(D)$ is obtained by evaluating $p_1(n)$ at $n = 31$).

Turning to the upper n-boundary, we evaluate $p_1(n)$ at $n = D$, and obtain $p_5(D)$. Differentiating twice, we find

$$p_5''(D) = -\frac{4}{3D^3} - \frac{4}{3(N-D)^3} - \frac{1}{2(N+D)^2} - \frac{1}{2(2N-D)^2} \ ,$$

which implies $p_5(D)$ is concave in D. Evaluating at $D = N/3$, we find $p_5(D)$ equals to

$$\frac{1}{2}\log\left(\frac{10}{9}\right) - \frac{1}{360N^3} - \frac{35}{12N}$$

and when $D = N/2$, $p_5(D)$ equals to

$$\frac{1}{2}\log\left(\frac{9}{8}\right) - \frac{1}{360N^3} - \frac{31}{12N} \ .$$

Both expressions are positive for $N \geq 61$, and so concavity implies $p_5(D) \geq 0$. Hence $p_1(n) \geq 0$ when $n = D$, and the concavity of $p_1(n)$ implies the claim. □

4.6 $\gamma 2$*: Main Positivity when* $N/3 \leq D \leq N/2$ *and* $x \in [1/2, 1 - N/(n(N-D))]$

Lemma 30. *Let* $N \geq 500$*,* $N/3 \leq D \leq N/2$*,* $10 \leq n \leq N/5$*, and* $1/2 \leq x \leq 1 - N/(n(N-D))$*. Using the functions in* (3.46)*, we have that*

$$A_3\left(n|x = 1 - \frac{N}{n(N-D)}\right) + A_4\left(n|x = \frac{N}{n(N-D)}\right) + R(n)$$

is positive.

Proof. The second derivative of this expression may be written

$$-\frac{7}{6(N-n)^3} - \frac{7}{6n^3} + \frac{p_1(n)}{2(D-n+1)^2(n-N)^2} \ ,$$

where we define

$$p_1(n) := n^2 + 2n(N - 2D - 2) + 2 + 4D + 2D^2 - N^2$$

Since $p_1'(n) = 2n + 2(N - 2D - 2) > 0$, we see

$$p_1(n) \leq -D^2 + 2DN - N^2 + 2 \leq 2 - \frac{1}{4}N^2 < 0 \ ,$$

and so we see the second derivative of the expression of interest is negative. Thus the expression of interest is concave in n. We first evaluate the expression of interest at $n = 10$, and define this to be $p_2(D)$. We differentiate to find

$$p_2'(D) = \frac{N - 2D + 8}{2(D-9)(N-D-1)} + \frac{N^2 - 2DN}{12D^2(D-N)^2} \ ,$$

which is positive. Hence $p_2(D)$ increases in D, and so a lower bound for $p_2(D)$ is obtained by evaluating at $D = N/3$. Doing so we obtain the function

$$p_3(N) := -\frac{1}{360N^3} - \frac{N}{20(N-10)} - \frac{1}{12(N-10)} - \frac{7}{24N} + \frac{1}{2}\log\left(1 - \frac{17}{N-10}\right) + \frac{1}{2}\log\left(\frac{17}{2(N-10)} + 1\right) - \frac{1}{120} + \frac{1}{2}\log\left(\frac{16}{9}\right) .$$

Differentiating, we find

$$p_3'(N) = \frac{35N^2+1}{120N^4} + \frac{7}{12(N-10)^2} + \frac{17(N+24)}{2(N-27)(N-10)(2N-3)} > 0 .$$

At $N = 50$ we find $p_3(N)$ is approximately 0.028, so monotone increase implies $p_3(N) > 0$ under the assumptions. Thus, the expression of interest is positive as well when $n = 10$.

We conclude by evaluating the the expression of interest at $n = N/5$, and defining this to be $p_4(D)$. We differentiate twice to find

$$p_4''(D) = -\frac{1}{6D^3} - \frac{1}{6(N-D)^3} - \frac{1}{2(N-D-1)^2} - \frac{25}{2(N-5D-5)^2} ,$$

which implies $p_4(D)$ is concave in D. Evaluating at $D = N/3$ we find it equals

$$-\frac{1}{360N^3} - \frac{63}{16N} + \frac{1}{2}\log\left(\frac{10N-15}{8N}\right) + \frac{1}{2}\log\left(\frac{1}{2} + \frac{15}{4N}\right) + \log\left(\frac{4}{3}\right)$$

and when $D = N/2$ it equals

$$-\frac{1}{360N^3} - \frac{187}{48N} + \frac{1}{2}\log\left(\frac{5(N-2)}{4N}\right) + \frac{1}{2}\log\left(\frac{5}{2N} + \frac{3}{4}\right) + \frac{\log(2)}{2} .$$

Both expressions are positive for $N \geq 50$, as can be verified by examining their first derivatives with respect to N and then evaluating at $N = 50$. Hence, concavity implies $p_4(D) > 0$. The claim follows by the expression of interests concavity in n. □

Lemma 31. *Let $N \geq 500$, $N/3 \leq D \leq N/2$, $10 \leq n \leq N/5$, and $1/2 \leq x \leq 1 - N/(n(N-D))$. Using the functions in (3.46), we have that*

$$\frac{n}{N}G_1(D|x=1/2) + B_3\left(D|x=1/2\right) + B_4\left(D|x=1/2\right)$$

is positive.

Proof. The second derivative of the expression of interest is

$$\frac{n^2}{2D^2N^2(N-n)(D(N-n)+N(D-n))} \cdot p_1(n) \tag{4.33}$$

where we define

$$p_1(n) := (2D^3 + 2D^2N - N^3)n + N^4 - 4D^3N \ .$$

Since $2D^3 + 2D^2N - N^3 < -(1/4)N^3 < 0$, we see $p_1(n)$ decreases in n. Evaluating at $n = N/5$, we thus find it bounded below by

$$\frac{2}{5}N\left(-9D^3 + D^2N + 2N^3\right) \geq \frac{9}{20}N^4 > 0 \ ,$$

from which we infer (4.33) is positive. Therefore, the first derivative of the expression of interest increases in D. We evaluate the first derivative of the expression of interest at $D = N/2$, and define this to be $p_2(n)$. Differentiating twice, we find

$$p_2''(n) = -\frac{nN\left(n^2 - 6nN + 4N^2\right)}{(2N-n)(2N-3n)^2(N-n)^3} < 0 \ ,$$

implying $p_2'(n)$ decreases in n. Evaluating $p_2'(n)$ at $n = 10$, we find it yields the upper bound

$$-\frac{5\left(N^2 - 20N + 150\right)}{(N-15)(N-10)^2N} + \frac{1}{2N}\log\left(1 + \frac{5}{N-10}\right) - \frac{1}{2N}\log\left(1 - \frac{5}{N-10}\right) \ ,$$

which is negative and monotone increasing to 0 for $N \geq 50$ (this may be verified, for example, by combining terms and demonstrating the resulting numerator is negative by analyzing its derivatives). Hence $p_2'(n) < 0$ and so $p_2(n)$ decreases in n as well. Evaluating at $n = 10$ we obtain an upper bound on $p_2(n)$ of

$$\frac{N-5}{(N-10)N} \cdot p_3(N)$$

where we define

$$p_3(N) := -(N-10)\log\left(1 + \frac{5}{N-10}\right) + (N-10)\log\left(1 - \frac{5}{N-10}\right) + 10 \ .$$

The second derivative of $p_3(N)$ is negative, and the first decreases to 0 as $N \nearrow \infty$. Thus, $p_3(N)$ increases monotonically in N to 0 from about -0.0525 when $N = 50$. We infer $p_2(n) < 0$, which implies the expression of interest decreases in D.

We thus obtain a lower bound on the expression of interest by evaluating at $D = N/2$. We define this lower bound to be $p_4(n)$. Differentiating twice, we find

$$p_4''(n) = \frac{n^2 N^2}{4(N-n)^3\,(3n^2 - 8nN + 4N^2)} > 0$$

implying the first derivative increases in n. Evaluating $p_4'(n)$ at $n = 10$ we find it equals

$$\frac{p_5(N)}{4(N-10)^2}$$

where we define

$$p_5(N) := -10(N-5) - (N-10)^2 \log\left(1 + \frac{5}{N-10}\right) - 3(N-10)^2 \log\left(1 - \frac{5}{N-10}\right) .$$

The third derivative of $p_5(N)$ is

$$-\frac{500\,(N^3 - 225N + 1500)}{(N-15)^3(N-10)(N-5)^3} ,$$

which we can use to determine $p_5'(N) < 0$, and consequently that $p_5(N) \searrow 0$ as $N \nearrow \infty$ from a starting value of about 2.49 at $N = 50$. This implies $p_4'(n) > 0$ and so $p_4(n)$ increases in n. Evaluating $p_4(n)$ at $n = 10$ thus yields a lower bound,

$$p_6(N) = \frac{(N^2 - 15N + 50)\log\left(1 + \frac{5}{N-10}\right) + (N^2 - 25N + 150)\log\left(1 - \frac{5}{N-10}\right) - 25}{2(N-10)} .$$

Differentiating twice, we find

$$p_6''(N) = \frac{625}{(N-10)^3\,(N^2 - 20N + 75)} ,$$

from which we can determine $p_6(N)$ is monotone decreasing to 0 from a value of about 0.0008 at $N = 50$. Hence, $p_4(n) > 0$, and the claim follows. □

Lemma 32. *Let $N \geq 500$, $N/3 \leq D \leq N/2$, $N/5 \leq n \leq D$, and $1/2 \leq x \leq 1 - N/(n(N-D))$. Using the functions in* (3.46)*, we have that*

$$\frac{n}{N} G_1(D|x = 1/2) + B_3(D|x = 1/2) + B_4(D|x = 1/2) + A_3(D|x = 1/2) + A_4(D|x = 1/2)$$

is positive.

Proof. The second derivative of the expression of interest may be written

$$\frac{n}{2D^2N^2(N-n)(D(n-2N)+nN)^2}\cdot p_1(n) \tag{4.34}$$

where we define

$$p_1(n) := \begin{bmatrix} -(N+D)\left(2D^3+2D^2N-N^3\right)n^3 \\ -N(N+2D)\left(-4D^3-2D^2N+(D+1)N^2+N^3\right)n^2 \\ N^2\left(-8D^4+2DN^2(N+3)+N^3\right)-4DN^5 \end{bmatrix} .$$

We begin by showing $p_1(n) > 0$ under the assumptions. The third derivative of the expression of interest equals $-6(D+N)\left(2D^3+2D^2N-N^3\right) < 0$, implying the second derivative decreases in n. Evaluating $p_1''(n)$ at $n = N/5$, we find it bounded above by

$$-\frac{2N}{5}(-34D^4-28D^3N+(12D+5)N^3+2(3D+5)DN^2+2N^4) =: -\frac{2N}{5}\cdot(p_2(D)) . \tag{4.35}$$

Since $p_2''(D) = -408D^2-168DN+12N^2 \leq -(268/3)N^2 < 0$, we see $p_2(D)$ is concave in D. We evaluate at the D-boundary to find when $D = N/3$ that $p_2(D)$ equals $(1/81)N^3(422N+675)$ and when $D = N/2$ it equals $(1/8)N^3(31N+80)$. Since both are positive, concavity implies $p_2(D) > 0$. This $p_1''(n) < 0$ by (4.35), which implies $p_1(n)$ is concave in n.

We will thus show $p_1(n) > 0$ by using concavity and demonstrating $p_1(n) > 0$ when $n = N/5$ and $n = D$. First considering $n = N/5$ we evaluate to find

$$\frac{2}{125}N^3\left(-81D^4+18D^3N+2(9D+5)N^3-(D+180)DN^2-2N^4\right) =: \frac{2}{125}N^3(p_3(D)) . \tag{4.36}$$

Since $p_2''(D) = -972D^2+108DN-2N^2 \leq -74N^2 < 0$, we see here that $p_3(D)$ is concave in D. We again evaluate at the D-boundary to find here that when $D = N/3$ that $p_3(D)$ equals $(2/9)N^3(16N-225)$ and when $D = N/2$ it equals $(1/16)(N^3)(63N-1280)$. Since both are positive under the assumptions, concavity implies $p_3(D) > 0$. This implies $p_1(n) > 0$ when $n = N/5$ by (4.36).

Turning to the upper boundary, we evaluate $p_1(n)$ at $n = D$ to find it equals

$$D(N-D)(2D^5 - 2D^4N + (D-3)N^4 - (D-2)DN^3) =: D(N-D)(p_4(D)) \ .$$

The fourth derivative of $p_4(D)$ is $240D - 48N > 0$, implying the second derivative of $p_4(D)$ is convex in D. Checking at the D-boundary, we find when $D = N/3$ that $p_4''(D)$ equals $-(86/27)N^3$ and when $D = N/2$ that $p_4''(D)$ equals $-3N^3$. Convexity therefore implies $p_4''(D) < 0$, which in turn implies $p_4(D)$ is concave in D. We again check the D-boundary to find when $D = N/3$ that $p_4(D)$ equals $(1/243)N^4(50N - 567)$ and when $D = N/2$ it equals $(1/16)N^4(3N - 32)$.

Concavity implies $p_4(D) > 0$. This implies $p_1(n) > 0$ when $n = D$. Combined with (4.36) and the concavity of $p_1(n)$, we infer $p_1(n) > 0$. By (4.34), we infer the second derivative of the expression of interest is positive. Hence, we may evaluate the first derivative of the expression of interest at $D = N/2$ to obtain an upper bound on the first derivative. We do so, and define this as $p_5(n)$. We differentiate twice to find

$$p_5''(n) = \frac{p_6(n)}{(2N-n)(2N-3n)^3(N-n)^3} \tag{4.37}$$

where we define

$$p_6(n) := 3n^4(N+8) - 20n^3N(N+6) + 24n^2N^2(N+9) - 8nN^3(N+21) + 48N^4 \ .$$

The fourth derivative of $p_6(n)$ is $72(N+8) > 0$, implying the third derivative increases in n and at $n = N/2$ is bounded above by $-12N(7N+36)$. Hence the third derivative of $p_3(n)$ is negative, implying $p_6''(n)$ decreases in n and $p_6'(n)$ is concave. Suppose now that $n \in [N/5, N/3]$. Then by monotone decreases, $p_6''(n) \geq 4N^2(56+3N) > 0$ for $n \in [N/5, N/3]$. Thus, $p_6(n)$ is convex for $n \in [N/5, N/3]$, At the boundary of this interval $p_6(n)$ equals $-(1/625)(497N - 13824)N^4$ and $-(1/27)(19N - 320)N^4$. Since both are negative, convexity implies $p_6(n) < 0$ when $n \in [N/5, N/3]$. Now suppose $n \in [N/3, N/2]$. We evaluate $p_6'(n)$ at the end-points to find it equals $(16/9)(N-34)N^3$ and $(5/2)(N-12)N^3$ at them. Since both are positive, concavity implies $p_6(n)$ increases for $n \in [N/3, N/2]$. Evaluating $p_6(n)$ at

$n = N/2$, we find it bounded above by $-(1/16)(5N-72)N^4 < 0$ for $n \in [N/3, N/2]$. This is negative, and so we infer $p_6(n) < 0$ for $n \in [N/3, N/2]$.

We have thus demonstrated that $p_6(n) < 0$ for $n \in [N/5, N/2]$. Hence by (4.37) we see $p_5''(n) < 0$. This implies $p_5'(n)$ decreases in n. Evaluating at $n = N/5$ we thus find $p_5'(n)$ is bounded above by

$$\frac{100}{49N^2} - \frac{33}{224N} + \frac{1}{2N}\log\left(\frac{8}{7}\right) + \frac{1}{2N}\log\left(\frac{9}{8}\right) \ ,$$

which is negative for $N \geq 100$. Hence $p_5'(n) < 0$ and so $p_5(n)$ decreases in n. We evaluate $p_5(n)$ at $n = N/5$ to find it bounded above by

$$\frac{2}{7N} + \frac{9}{40} - \frac{1}{40}9\log\left(\frac{6561}{2401}\right)$$

which is negative for $N \geq 242$. Hence $p_5(n) < 0$ under the assumptions, and by the discussion preceding (4.37) we see that the expression of interest decreases in D. We therefore evaluate the expression of interest at $D = N/2$ to obtain a lower bound, which we define here as $p_7(n)$. Differentiating twice we find

$$p_7''(n) = \frac{N}{4(2N-n)^2(2N-3n)^2(N-n)^3}p_8(n) \tag{4.38}$$

where we define

$$p_8(n) := 3n^4(N+8) - 4n^3N(2N+17) + 4n^2N^2(N+11) + 16nN^3 - 16N^4 \ .$$

The fourth derivative of $p_8(n)$ is $72(N+8)$, which implies the third derivative is bounded above at $n = N/2$ by $-12N(10+N)$. Hence $p_8''(n)$ decreases in n, and at $n = N/5$ is bounded above by $-(4/25)(N-112)N^2 < 0$. Therefore $p_8(n)$ is concave in n. Evaluating at the n-boundary, we find when $n = N/5$ that $p_8(n)$ equals $(1/625)N^4(63N - 7216)$ and when $n = N/2$ it equals $(1/16)N^4(3N-64)$. Since both are positive, concavity implies $p_8(n) > 0$. Thus (4.38) implies $p_7''(n) > 0$, which in turn implies $p_7'(n)$ increases in n. Evaluating $p_7'(n)$ at $n = N/5$ therefore yields the lower bound

$$-\frac{25}{252N} - \frac{9}{128} + \frac{3}{4}\log\left(\frac{8}{7}\right) - \frac{1}{4}\log\left(\frac{9}{8}\right) \ .$$

This is positive for $N \geq 255$, and so $p_7(n)$ increases in n under the assumptions. Therefore we may evaluate $p_7(n)$ at $n = N/5$ to find it bounded below by

$$-\frac{N}{160} - \frac{1}{20}7N\log\left(\frac{8}{7}\right) + \frac{9}{20}N\log\left(\frac{9}{8}\right) - \frac{1}{2}\log\left(\frac{64}{63}\right) .$$

This is positive for $N \geq 481$, and so under the assumptions, $p_7(n) \geq 0$. This implies the expression of interest is positive under the assumptions by the discussion preceding (4.38), and the claim is proved. □

Lemma 33. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $N/3 \leq D \leq N/2$, and $31 \leq n \leq D$. Using the functions in* (3.46)*, we have that*

$$\frac{N-n}{N}G_1(n|x=1/2) + B_1(n|x=1/2) + B_2(n|x=1/2) + A_1(n|x=1/2) + A_2(n|x=1/2)$$

is positive.

Proof. We define the first derivative of the expression of interest as $p_1(D)$ where

$$p_1(D) := -\frac{D^2}{2N^2} + \frac{D}{N} - \frac{1}{2} - \frac{1}{2}\log(4) + \frac{1}{2}\log\left(\frac{D+N}{D}\right) + \frac{D}{2N}\log\left(\frac{D+N}{D}\right) \tag{4.39}$$

We further define a lower bound for $p_1(D)$ through the relation

$$p_2(D) := p_1(D) - \log\left(1 + \left(\frac{D}{8N}\right)^2\right) . \tag{4.40}$$

Differentiating, we find

$$p_2''(D) = \frac{-2D^5 - 2D^4N - 132D^3N^2 - 131D^2N^3 + 64N^5}{2D^2N^2(D+N)\left(D^2+64N^2\right)} + \frac{4D^2}{\left(D^2+64N^2\right)^2} .$$

This is positive since for $D \in [N/3, N/2]$ we have

$$-2D^5 - 2D^4N - 132D^3N^2 - 131D^2N^3 + 64N^5 \geq \frac{233}{16}N^5 .$$

Thus $p_2'(D)$ increases in D. We consider the behavior of $p_2(D)$ on the sub-interval $[N/3, (109/250)N]$. Evaluating $p_2'(D)$ at $D = (109/250)N$, we find it equals

$$\frac{65197807161}{109323757250N} + \frac{\log\left(\frac{359}{109}\right)}{2N} < -\frac{0.00038}{N} < 0 ,$$

which implies $p_2(D)$ decreases for $D \in [N/3, (109/250)N]$. Evaluating $p_2(D)$ at $D = (109/250)N$, we find it equals

$$-\frac{19881}{125000} + \frac{359}{500}\log\left(\frac{359}{109}\right) - \log\left(\frac{4011881}{2000000}\right) > 0.00067 > 0 \ ,$$

and so we see $p_2(D) > 0$ for $D \in [N/3, (109/250)N]$.

Now consider $p_1(D)$ on the interval $[(109/250)N, N/2]$. The second derivative of $p_1(D)$ is

$$\frac{-2D^3 - 2D^2N + N^3}{2D^2N^2(D+N)}$$

which is positive since

$$-2D^3 - 2D^2N + N^3 \geq \frac{1}{4}N^3 > 0 \ .$$

Therefore $p_1'(D)$ increases in D when $D \in [(109/250)N, N/2]$, and so evaluating at $D = (109/250)N$ yields the lower bound

$$\frac{1}{2N}\log\left(\frac{359}{109}\right) - \frac{15881}{27250N} > \frac{0.013}{N} > 0 \ .$$

Thus $p_1(D)$ increases for $n \in [(109/250)N, N/2]$. But since $p_1(D) > p_2(D)$ and we have already seen $p_2(D) > 0$, we infer $p_1(D) > 0$ for $D \in [(109/250)N, N/2]$. Combined with the analysis of $p_2(D)$ for $D \in [N/3, (109/250)N]$, we conclude $p_1(D) > 0$ for $D \in [N/3, N/2]$.

Hence by (4.39) we see that the expression of interest increases in n. We evaluate the expression of interest at $n = 31$ to obtain a lower bound, which we define as $p_3(D)$. Additionally, we define a further lower bound through the relation

$$p_4(D) := p_3(D) - \log\left(1 + \frac{D}{64N}\right) \ . \tag{4.41}$$

We now consider the behavior of $p_4(D)$ on the interval $[N/3, (2823/6250)N]$. We find the second derivative of $p_4(D)$ to be

$$p_4''(D) = \frac{1}{(64N+D)^2} + \frac{-62D^4 - 124D^3N - 62D^2N^2 + 33DN^3 + 32N^4}{2D^2N^2(N+D)^2}$$

This is positive, since

$$-62D^4 - 124D^3N - 62D^2N^2 + 33DN^3 + 32N^4$$

is concave in D, and equals $(2491/81)N^4$ when $D = N/3$ and $(109/8)N^4$ when $D = N/2$. Therefore $p_4'(D)$ increases in D, and at $D = (2823/6250)N$, is bounded above by

$$-\frac{0.00018}{N} < 0 \ .$$

Thus, $p_4(D)$ decreases for $D \in [N/3, (2823/6250)N]$, and so is bounded below by 0.0058 at $D = (2823/6250)N$. This implies $p_4(D) \geq 0$ for $D \in [N/3, (2823/6250)N]$.

We conclude by examining $p_3(D)$ for $D \in [(2823/6250)N, N/2]$. Differentiating twice, we find

$$p_3''(D) = \frac{-62D^4 - 124D^3N - 62D^2N^2 + 33DN^3 + 32N^4}{2D^2N^2(D+N)^2} \ ,$$

which we already know to be positive by the analysis of $p_4''(D)$. Thus, $p_3'(D)$ increases in D for $D \in [(2823/6250)N, N/2]$ Evaluating $p_3'(D)$ at $D = (2823/6250)N$, we find it is bounded below by

$$\frac{0.0153}{N} > 0 \ .$$

Hence $p_3(D)$ increases for $D \in [(2823/6250)N, N/2]$. But since $p_3(D) > p_4(D)$ by (4.41), this implies $p_3(D) > 0$ for $D \in [(2823/6250)N, N/2]$. Combined with the analysis of $p_4(D)$, this implies $p_3(D) > 0$ for $D \in [N/3, N/2]$. Since $p_3(D)$ provides a lower bound for the expression of interest, the claim is proved. □

4.7 $\beta \cup \gamma 1 \cup \gamma 2$: *Main Positivity when* $(15/100)N \leq D \leq N/2$ *and* $x \in [0, N/(n(N-D))]$

Lemma 34. *Let* $N \geq 500$, n, $D \in \mathbb{N}$, $(15/100)N \leq D \leq N/2$, $31 \leq n \leq D$, *and* $N/(2n(N-D)) \leq x \leq N/(n(N-D))$. *Using the functions in* (3.46), *and the* $C_i(x)$ *functions in Lemma* (14), *we have that the function*

$$G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x) + C_i(x)\right] + R(D)$$

is positive.

Proof. By Lemma 22 we know

$$G_1(x) + B_1(x) + B_2(x) + B_3(x) + B_4(x) \geq 0$$

and by Lemma 24 we know that under the current assumptions

$$C_1(x) + C_2(x) + C_3(x) + C_4(X) \geq 0 \ .$$

Therefore it is enough to show

$$A_1(x) + A_2(x) + A_3(x) + A_4(x) + R(D) > 0$$

under the current assumptions. Again by Lemma 22 we know

$$\begin{aligned}\sum_{i=1}^{4} A_i(x) \geq\ & A_1\left(D|x = \frac{N}{n(N-D)}\right) + A_2\left(D|x = \frac{N}{2n(N-D)}\right) \\ & + A_3\left(D|x = \frac{N}{n(N-D)}\right) + A_4\left(D|x = \frac{N}{2n(N-D)}\right) \\ & =: p_{1a}(D) \ .\end{aligned}$$

So the claim will follow if we show

$$p_1(D) := p_{1a}(D) + R(D) \tag{4.42}$$

is positive. Differentiating twice, we find $p_1''(D)$ may be written

$$\begin{aligned}&-\frac{1}{6D^3} - \frac{1}{6(N-D)^3} - \frac{N(2n(N-D)-N)}{2(N-D)^2(n(N-D)-N)^2} - \frac{2n^2}{(N+2Dn)^2} \\ &-\frac{N(2D(N-n)-N)}{2D^2(D(N-n)-N)^2} - \frac{2(N-n)^2}{(2(N-D)(N-n)+N)^2} \ .\end{aligned}$$

This is negative by inspection, and so $p_1(D)$ is concave in D. Thus, showing positivity at the D-boundary will demonstrate positivity. We first evaluate $p_1(D)$ at $D = N/2$ to obtain

$$\begin{aligned}p_2(n) :=& -\frac{N}{2n(N-n)} - \frac{1}{12(N-n)} - \frac{1}{12n} - \frac{1}{360N^3} - \frac{1}{4N} + \frac{1}{2}\log(2) \\ &+\frac{1}{2}\log\left(1 + \frac{1}{N-n}\right) + \frac{1}{2}\log\left(1 - \frac{2}{N-n}\right) + \frac{1}{2}\log\left(1 + \frac{1}{n}\right) + \frac{1}{2}\log\left(1 - \frac{2}{n}\right) \ .\end{aligned} \tag{4.43}$$

Differentiating twice, we find

$$p_2''(n) = -\frac{13n^4+19n^3-45n^2+4n+28}{6(n-2)^2n^3(n+1)^2} - \frac{p_3(n)}{2(N-n+1)^2(N-n)^2(N-n-2)^2} - \frac{7}{6(N-n)^3} \tag{4.44}$$

where we define

$$p_3(n) := -2n^3 + n^2(6N+11) + n\left(-6N^2-22N+8\right) + 2N^3 + 11N^2 - 8N - 8 \ .$$

Consider $p_3(n)$. Its second derivative is $22+12(N-n) > 0$, and so its first derivative increases in n. Evaluating $p_3'(n)$ at $n = N/2$, we thus find it bounded above by $8-11N-(3/2)N^2 < 0$, and so $p_3(n)$ decreases in n. Evaluating again at $n = N/2$ we thus find $p_3(n)$ bounded below by $-8-4N+(11/4)N^2+(1/4)N^3 > 0$. Therefore $p_3(n) \geq 0$ and so (4.44) is negative. Therefore $p_2'(n)$ decreases in n, and so evaluating at $p_2'(n)$ at $n = N/2$ we find it bounded below by 0. This implies $p_2'(n) \geq 0$ and so $p_2(n)$ increases in n. Thus we evaluate $p_2(n)$ at $n = 31$ to find a lower bound for $p_2(n)$ of

$$-\frac{1}{360N^3} - \frac{N}{62(N-31)} - \frac{1}{12(N-31)} - \frac{1}{4N} + \frac{1}{2}\log\left(1+\frac{1}{N-31}\right) + \frac{1}{2}\log\left(1-\frac{2}{N-31}\right)$$
$$-\frac{1}{372} + \frac{\log(2)}{2} - \frac{1}{2}\log\left(\frac{31}{29}\right) + \frac{1}{2}\log\left(\frac{32}{31}\right) \ .$$

This increases in N since its derivative is

$$\frac{160N^6-11640N^5+295231N^4-3657815N^3+28547557N^2-121923N+951390}{120(N-33)(N-31)^2(N-30)N^4} > 0 \ ,$$

which is positive for $N \geq 100$. Evaluating the expression at $N = 100$ we find it approximately equals 0.2918 and in the limit it approaches a value close to 0.3102. Therefore, the lower bound of $p_2(n)$ is positive under the assumptions, implying $p_2(n)$ is positive, and hence $p_1(D)$ is positive at $D = N/2$.

Next, we evaluate $p_1(D)$ at $D = (15/100)N$ to obtain

$$p_4(n) := -\frac{N}{2n(N-n)} - \frac{1}{12(N-n)} + \frac{1}{2}\log\left(1-\frac{20}{3(N-n)}\right) + \frac{1}{2}\log\left(\frac{10}{17(N-n)}+1\right) - \frac{1}{12n}$$
$$+\frac{1}{2}\log\left(1+\frac{10}{3n}\right) + \frac{1}{2}\log\left(1-\frac{20}{17n}\right) - \frac{1}{360N^3} - \frac{349}{612N} + \frac{1}{2}\log\left(\frac{51}{50}\right) \ .$$

Since

$$p_4'(n) = \frac{-303n^2 + 3170n - 1400}{12n^2(3n+10)(17n-20)} - \frac{5(31(N-n)+40)}{(17(N-n)+10)(3(N-n)-20)(N-n)} - \frac{7}{12(N-n)^2}$$

is negative, we see $p_4(n)$ decreases in n under the assumptions. Evaluating $p_4(n)$ at $n = (15/100)N$ we obtain the lower bound

$$-\frac{1}{360N^3} - \frac{3149}{612N} + \frac{1}{2}\log\left(\frac{200}{9N}+1\right) + \log\left(1 - \frac{400}{51N}\right) + \frac{1}{2}\log\left(\frac{200}{289N}+1\right) + \frac{1}{2}\log\left(\frac{51}{50}\right) .$$

This increases in N for $N \geq 500$, and at $N = 500$ approximately equals 0.00623. Hence, $p_4(n) \geq 0$ for $31 \leq D \leq (15/100)N$. This implies $p_1(D) \geq 0$ on the rectangle $31 \leq n \leq (15/100)N$ and $(15/100)N \leq D \leq N/2$. For $n \geq (15/100)N$, we evaluate $p_1(D)$ at $D = n$ to obtain

$$\begin{aligned} p_5(n) :=& \frac{1}{2}\log\left(1+\frac{N}{2n^2}\right) - \frac{1}{2}\log\left(\frac{N^2}{8n(N-n)}\right) + \log\left(1 - \frac{N}{n(N-n)}\right) + \frac{1}{2}\log\left(1 + \frac{N}{2(N-n)^2}\right) \\ &- \frac{1}{6n} - \frac{1}{360N^3} + \frac{1}{12N} - \frac{N}{2n(N-n)} - \frac{1}{6(N-n)} . \end{aligned} \tag{4.45}$$

Differentiating twice, we find $p_5''(n)$ may be written

$$\begin{aligned} &-\frac{1}{3n^3} - \frac{n^3 + 9n^2N - 9nN^2 + 3N^3}{3n^3(N-n)^3} - \frac{4n^4 - 8n^2N - N^2}{2n^2\left(2n^2+N\right)^2} - \frac{p_6(n)}{2(N-n)^2\left(2(n-N)^2+N\right)^2} \\ &-\frac{N \cdot p_7(n)}{n^2(n-N)^2\left(n(N-n)-N\right)^2} \end{aligned} \tag{4.46}$$

where we define

$$p_6(n) := 4n^4 - 16n^3N + 8n^2N(3N-1) - 16n(N-1)N^2 + N^2\left(4N^2 - 8N - 1\right)$$

and

$$p_7(n) := -6n^4 + 12n^3N - 2n^2N(4N+1) + 2nN^2(N+1) - N^3 .$$

We see showing $p_6(n) \geq 0$ and $p_7(n) \geq 0$ will imply (4.46) is negative. Consider $p_6(n)$ first. Since the third derivative of $p_6(n)$ is $-96(N-n) < 0$, we see $p_6''(n)$ is decreasing in n. Evaluating at $n = N/2$ shows $p_6''(n)$ is bounded below by $4N(3N-4) > 0$. The same logic shows $p_6'(n)$ is bounded above by $-2(N-4)N^2 < 0$ and so $p_6(n)$ decreases in N. At $n = N/2$ we find it equals $(1/4)N^2(N^2 - 8N - 4) > 0$. Hence $p_6(n) \geq 0$.

Turning to $p_7(n)$, we find the third derivative is $72(N - 2n)$, and so $p_7'(n)$ is convex in n. At $n = N/2$, $p_7'(n)$ equals 0, and so convexity implies $p_7(n)$ is minimized either at $n = (15/100)N$ or $n = N/2$. Evaluating $p_7(n)$ at these points yields the values $(12597/80000)N^4 - (149/200)N^3$ and $(1/8)(N-4)N^3$, both of which are positive.

Therefore, $p_7(n) \geq 0$. Combined with the positivity of $p_6(n)$, we infer (4.46) is negative. Therefore $p_5(n)$ is concave in n. We conclude by evaluating $p_5(n)$ at $n = (15/100)N$ and $n = N/2$ to obtain

$$-\frac{1}{360N^3} - \frac{3149}{612N} + \frac{1}{2}\log\left(\frac{200}{9N} + 1\right) + \log\left(1 - \frac{400}{51N}\right) + \frac{1}{2}\log\left(\frac{200}{289N} + 1\right) + \frac{1}{2}\log\left(\frac{51}{50}\right)$$

and

$$-\frac{1}{360N^3} - \frac{31}{12N} + \log\left(\frac{2}{N} + 1\right) + \log\left(1 - \frac{4}{N}\right) + \frac{\log(2)}{2}$$

respectively. Both increase in N (which can be verified by differenitation), and at $N = 500$ both are positive. Therefore, $p_5(n) \geq 0$, and the claim follows by the concavity of $p_1(D)$. □

Lemma 35. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $(15/100)N \leq D \leq N/2$, $31 \leq n \leq N/2$, and $N/(4n(N-D)) \leq x \leq N/(2n(N-D))$. Using the functions in functions in Lemma 14 and Lemma 19, we have that the function*

$$G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x) + C_i(x)\right] + R_1(D)$$

is positive, where we define

$$R_1(D) := -\frac{1}{2}\log\left(\frac{N^2}{8D(N-D)}\right) + R_H - \frac{3N}{4n(N-n)} .$$

Proof. Using the same logic as that used at the beginning of Lemma 34, we define

$$\sum_{i=1}^{4} A_i(x) \geq A_1\left(D|x = \frac{N}{2n(N-D)}\right) + A_2\left(D|x = \frac{N}{4n(N-D)}\right)$$
$$+A_3\left(D|x = \frac{N}{2n(N-D)}\right) + A_4\left(D|x = \frac{N}{4n(N-D)}\right) =: p_{1a}(D) .$$

Here we also use Lemma 24 to observe $C_2(x)$ decreases in x, and so the claim will follow by showing

$$p_1(D) := p_{1a}(D) + R_1(D) + C_2\left(D|x = \frac{N}{2n(N-D)}\right) \tag{4.47}$$

is positive. Differentiating twice, we find $p_1''(D)$ may be written

$$-\frac{1}{6D^3} - \frac{1}{6(N-D)^3} - \frac{4nN(N-D)-N^2}{2(N-D)^2(2n(N-D)n-N)^2}$$
$$-\frac{N(4D(N-n)-N)}{2D^2(2D(N-n)-N)^2} - \frac{8\left(n^2-2nN+N^2\right)}{(4(N-D)(N-n)+N)^2}$$
$$-\frac{4n^2\left(960D^5n^5+2080D^4n^4N+1920D^3n^3N^2+1004D^2n^2N^3+272DnN^4+29N^5\right)}{15(N+2Dn)^5(N+4Dn)^2},$$

which is negative by inspection under the assumptions. Therefore, $p_1(D)$ is concave in D. We now define $p_2(n)$ by evaluating $p_1(D)$ at $D = N/2$. We then differentiate twice to find

$$p_2''(n) = -\frac{4n^3+15n^2+15n+5}{3n^3(n+1)^3} - \frac{5}{3(N-n)^3} - \frac{4}{15(n+1)^5} - \frac{4n^3+11n^2-4n-2}{2(n-1)^2n^2(2n+1)^2}$$
$$-\frac{p_3(n)}{2(2(N-n)+1)^2(N-n)^2(-n+N-1)^2} \tag{4.48}$$

where we define

$$p_3(n) := -4n^3 + n^2(12N+11) - 2n\left(6N^2+11N-2\right) + 4N^3 + 11N^2 - 4N - 2\ .$$

We examine $p_3(n)$: it's third derivative is -24, and so its second derivative is bounded below by $22 + 12N > 0$. Evaluating at $n = N/2$ we find $p_3'(n)$ is bounded above by $4 - 11N - 3N^2 < 0$. Therefore, $p_3(n)$ decreases in n, and so is bounded below by $-2 - 2N + (11/4)N^2 + (1/2)N^3 > 0$. Therefore, $p_3(n) \geq 0$, and so inspecting (4.48) we see that $p_2''(n) < 0$. Thus $p_2(n)$ is concave in n. Evaluating $p_2(n)$ at $n = 31$ we find it equals

$$-\frac{1}{360N^3} - \frac{1}{4N} - \frac{5}{6(N-31)} - \frac{990751}{45711360} + \frac{1}{2}\log\left(1-\frac{1}{N-31}\right) + \frac{1}{2}\log\left(\frac{1}{2(N-31)}+1\right) + \frac{\log(2)}{2}$$
$$-\frac{1}{2}\log\left(\frac{31}{30}\right) + \frac{1}{2}\log\left(\frac{63}{62}\right)\ ,$$

and at $n = N/2$ we find $p_2(n)$ equals

$$-\frac{1}{360N^3}+\frac{N^2}{3(N+2)^3}+\frac{4N}{3(N+2)^3}-\frac{43}{12N}+\frac{52}{45(N+2)^3}+\log\left(1+\frac{1}{N}\right)+\log\left(1-\frac{2}{N}\right)+\frac{\log(2)}{2}\ .$$

Both expressions increase in N (as can be verified by differentiation), and at $N = 100$ are greater than 0.29. Therefore, they are positive, and $p_2(n) \geq 0$ by concavity. We conclude by evaluating $p_1(D)$ at $D = (15/100)N$ and define this to be $p_4(n)$. We differentiate to find

$$p_4'(n) = -\frac{5(31(N-n)+20)}{2(17(N-n)+5)(3(N-n)-10)(N-n)} - \frac{2\left(27n^4 - 540n^3 - 2700n^2 - 6000n - 5000\right)}{3n^2(3n+10)^4} - \frac{50\left(225n^4 - 444n^3 - 5370n^2 - 12300n - 10000\right)}{n(3n+10)^4\left(51n^2+55n-50\right)} - \frac{p_5(n)}{6n^2\left(51n^2+55n-50\right)(N-n)^2}\,, \tag{4.49}$$

where we define

$$p_5(n) := 267n^4 + n^3(110 - 24N) + 2n^2\left(6N^2 + 165N - 50\right) - 15nN(11N+20) + 150N^2\ .$$

We see (4.49) is negative since for $n \geq 25$ the quartic

$$27n^4 - 540n^3 - 2700n^2 - 6000n - 5000$$

is positive, and for $n \geq 7$ the quartic

$$225n^4 - 444n^3 - 5370n^2 - 12300n - 10000$$

is positive. Additionally, $p_5(n) > 0$ through analysis of its derivatives: we find its second derivative is minimized at $n^* = (2/89)N - (55/534)$, where it attains the value $-(20825/89) + (60060/89)N + (1992/89)N^2$. Therefore, the first derivative increases from $n = 31$, and is bounded below by $32127718 - 49032N + 579N^2$. This is positive under the assumptions, and so $p_5(n)$ increases in n. Furthermore, $p_5(n)$ is bounded below at $n = 31$ where it has the value $249761017 - 407154N + 6567N^2$ which is also positive under the assumptions. Hence we see $p_5(n) \geq 0$, which in turn implies $p_4'(n) < 0$ by (4.49).

Therefore, $p_4(n)$ is bounded below at $n = N/2$, where it has the value

$$-\frac{1}{360N^3} - \frac{2389}{612N} - \frac{1600}{9(3N+20)^3} + \frac{10}{9N+60} + \frac{1}{2}\log\left(1+\frac{10}{3N}\right) + \frac{1}{2}\log\left(1-\frac{20}{3N}\right) + \frac{1}{2}\log\left(1+\frac{10}{17N}\right) + \frac{1}{2}\log\left(1-\frac{20}{17N}\right) + \frac{1}{2}\log\left(\frac{51}{50}\right)\ .$$

This quantity increases in N, as can be verified by differentiation. At $N = 500$ it is approximately equal to 0.0003, and so is positive under the assumptions. Therefore, $p_4(n) > 0$, which implies $p_1(D) > 0$ by concavity. This implies the claim. □

Lemma 36. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $(15/100)N \leq D \leq N/2$, $31 \leq n \leq N/2$, and $N/(8n(N-D)) \leq x \leq N/(4n(N-D))$. Using the functions in functions in Lemma 14 and Lemma 19, we have that the function*

$$G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x) + C_i(x)\right] + R_2(D)$$

is positive, where we define

$$R_2(D) := -\frac{1}{2}\log\left(\frac{N^2}{8D(N-D)}\right) + R_H - \frac{N}{2n(N-n)} \; .$$

Proof. We begin by defining

$$p_1(D) := p_{1a}(D) + R_2(D) + C_2\left(D|x = \frac{N}{4n(N-D)}\right) \tag{4.50}$$

where

$$\begin{aligned} p_{1a}(D) :=& A_1\left(D|x = \frac{N}{4n(N-D)}\right) + A_2\left(D|x = \frac{N}{8n(N-D)}\right) \\ &+ A_3\left(D|x = \frac{N}{4n(N-D)}\right) + A_4\left(D|x = \frac{N}{8n(N-D)}\right) \; . \end{aligned} \tag{4.51}$$

Using the same logic as that used at the beginning of Lemma 35, we see it is enough to show $p_1(D) \geq 0$ under the assumptions. Differentiating twice, we find $p_1''(D)$ may be written as

$$\begin{aligned} &-\frac{1}{6D^3} - \frac{1}{6(N-D)^3} - \frac{8nN(N-D) - N^2}{2(N-D)^2(4n(N-D)-N)^2} \\ &-\frac{N(8D(N-n)-N)}{2D^2(4D(N-n)-N)^2} - \frac{32\left(n^2 - 2nN + N^2\right)}{(8(N-D)(N-n)+N)^2} \\ &-\frac{64n^2\left(7680D^5n^5 + 7040D^4n^4N + 2880D^3n^3N^2 + 1192D^2n^2N^3 + 218DnN^4 + 13N^5\right)}{15(4Dn+N)^5(8Dn+N)^2} \; . \end{aligned}$$

This is negative by inspection, and so we see $p_1(D)$ is concave in D. We next evaluate $p_1(D)$ at $D = N/2$ to define $p_2(n)$. We differentiate twice to find

$$\begin{aligned} p_2''(n) = &-\frac{7}{6(N-n)^3} - \frac{800n^5 + 2480n^4 + 2976n^3 + 1400n^2 + 350n + 35}{30n^3(2n+1)^5} \\ &-\frac{16n^3 + 22n^2 - 4n - 1}{n^2(2n-1)^2(4n+1)^2} - \frac{p_3(n)}{(2(N-n)-1)^2(N-n)^2(4(N-n)+1)^2} \end{aligned} \quad (4.52)$$

where we define

$$p_3(n) := -16n^3 + n^2(48N + 22) + n\left(-48N^2 - 44N + 4\right) + 16N^3 + 22N^2 - 4N - 1\ .$$

The second derivative of $p_3(n)$ is $44 + 96(N - n) > 0$ and so $p_3'(n)$ is bounded above by $4 - 22N - 12N^2 < 0$. Therefore $p_3(n)$ decreases in n, and is bounded below by $-1 - 2N + (11/2)N^2 + 2N^3 > 0$ at $n = N/2$. Thus, $p_3(n) \geq 0$, and so (4.52) implies $p_2''(n) < 0$. Thus $p_2(n)$ is concave in n, and at $n = 31$ equals

$$\begin{aligned} &-\frac{1}{360N^3} - \frac{N}{62(N-31)} - \frac{1}{12(N-31)} - \frac{1}{4N} + \frac{1}{2}\log\left(1 - \frac{1}{2(N-31)}\right) + \frac{1}{2}\log\left(1 + \frac{1}{4(N-31)}\right) \\ &+\frac{3630643}{1395262260} + \frac{\log(2)}{2} - \frac{1}{2}\log\left(\frac{62}{61}\right) + \frac{1}{2}\log\left(\frac{125}{124}\right) \end{aligned} \quad (4.53)$$

while at $n = N/2$ it equals

$$-\frac{1}{360N^3} - \frac{31}{12N} - \frac{8}{45(N+1)^3} + \frac{1}{3N+3} + \log\left(\frac{N-1}{N}\right) + \log\left(\frac{1}{2N} + 1\right) + \frac{\log(2)}{2}$$

Differentiation show both increase in N for $N \geq 100$, and at $N = 100$ approximately equal 0.3161 and 0.3189 respectively. Hence, both are positive under the assumptions, and so concavity implies $p_2(n) \geq 0$. Turning to the lower boundary, we evaluate $p_1(D)$ at $D = (15/100)N$ to define the sum $p_4(n) + p_5(n)$, with

$$p_4(n) := -\frac{N}{2n(N-n)} - \frac{1}{12(N-n)} - \frac{1}{12n} - \frac{200}{9(3n+5)^3} + \frac{5}{9n+15} - \frac{1}{360N^3} - \frac{349}{612N} + \frac{1}{4}\log\left(\frac{51}{50}\right) \quad (4.54)$$

and

$$\begin{aligned} p_5(n) := &\frac{1}{2}\log\left(1 + \frac{5}{34(N-n)}\right) + \frac{1}{2}\log\left(1 - \frac{5}{3(N-n)}\right) + \frac{1}{2}\log\left(1 + \frac{5}{6n}\right) \\ &+\frac{1}{2}\log\left(1 - \frac{5}{17n}\right) + \frac{1}{4}\log\left(\frac{51}{50}\right)\ . \end{aligned} \quad (4.55)$$

We conclude by showing both $p_4(n) \geq 0$ and $p_5(n) \geq 0$. This will imply $p_1(D) > 0$ when $D = (15/100)N$, which implies the claim by concavity. Consider $p_4(n)$ first. Differentiating twice, we find

$$p_4''(n) = -\frac{9n^3 + 945n^2 + 1575n + 875}{6n^3(3n+5)^3} - \frac{7}{6(N-n)^3} - \frac{2400}{(3n+5)^5} < 0 \ .$$

Therefore $p_4(n)$ is concave in n. Evaluating at $n = 31$ we find it equals

$$-\frac{1}{360N^3} + \frac{N}{1922 - 62N} - \frac{349}{612N} + \frac{1}{372 - 12N} + \frac{1876883}{131296284} + \frac{1}{4}\log\left(\frac{51}{50}\right)$$

and at $n = N/2$ it $p_4(n)$ equals

$$-\frac{1}{360N^3} + \frac{10\left(27N^2 + 180N + 140\right)}{9(3N+10)^3} - \frac{1777}{612N} + \frac{1}{4}\log\left(\frac{51}{50}\right) \ .$$

Both increase in N (which can be seen through differentiating and combining terms), and at $N = 500$ approximately equal 0.0007 and 0.0013 respectively. Therefore $p_4(n) > 0$ by concavity. Turning to $p_5(n)$ we differentiate to find

$$-\frac{5(11n - 10)}{2n(6n+5)(17n-5)} - \frac{5(31(N-n) + 10)}{2(34(N-n)+5)(3(N-n)-5)(N-n)} < 0 \ ,$$

and so $p_5(n)$ decreases in n under the assumptions. Evaluating at $n = N/2$, we find it bounded below by

$$\frac{1}{2}\log\left(1 + \frac{5}{3N}\right) + \frac{1}{2}\log\left(1 - \frac{10}{3N}\right) + \frac{1}{2}\log\left(1 + \frac{5}{17N}\right) + \frac{1}{2}\log\left(1 - \frac{10}{17N}\right) + \frac{1}{4}\log\left(\frac{51}{50}\right) \ .$$

This increases in N (again seen by differentiation), and at $N = 500$ approximately equals 0.0029. Therefore, $p_5(n) > 0$. Since we have shown (4.54) and (4.55) are positive, $p_1(D) > 0$ and the claim follows. □

Lemma 37. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $(15/100)N \leq D \leq N/2$, $31 \leq n \leq N/2$, and $0 \leq x \leq N/(8n(N-D))$. Using the functions in functions in Lemma 14 and Lemma 19, we have that the function*

$$G_1(x) + \left[\sum_{i=1}^{4} A_i(x) + B_i(x) + C_i(x)\right] + R_3(D)$$

is positive, where we define

$$R_3(D) := -\frac{1}{2}\log\left(\frac{N^2}{8D(N-D)}\right) + R_H - \frac{3N}{8n(N-n)} \ .$$

Proof. We begin by defining

$$p_1(D) := p_{1a}(D) + R_2(D) + C_2\left(D|x = \frac{N}{4n(N-D)}\right) \tag{4.56}$$

where

$$p_{1a}(D) := A_1\left(D|x = \frac{N}{8n(N-D)}\right) + A_2\left(D|x=0\right) + A_3\left(D|x = \frac{N}{8n(N-D)}\right) + A_4\left(D|x=0\right) \ .$$

Using the same logic as that used at the beginning of Lemma 36, we see it is enough to show $p_1(D) \geq 0$ under the assumptions. Differentiating twice, we find $p_1''(D)$ may be written as

$$-\frac{1}{6D^3} - \frac{1}{2(N-D)^2} - \frac{1}{6(N-D)^3} - \frac{N(16n(N-D)-N)}{2(D-N)^2(8Dn-8nN+N)^2} - \frac{N(16D(N-n)-N)}{2D^2(8D(n-N)+N)^2}$$
$$-\frac{491520D^5n^5 + 143360D^4n^4N + 35840D^3n^3N^2 + 39808D^2n^2N^3 + 600DnN^4 + 15N^5}{30D^2(8Dn+N)^5} \ .$$

This is negative by inspection, and so we see $p_1(D)$ is concave in D. We first evaluate $p_1(D)$ at $D = N/2$ to define $p_2(n)$. Differentiating twice, we find

$$\begin{aligned} p_2''(n) = &-\frac{35840n^5 + 60160n^4 + 50304n^3 + 8800n^2 + 1100n + 55}{60n^3(4n+1)^5} - \frac{11}{12(N-n)^3} \\ &-\frac{(4N-1)\cdot p_{3a}(n)}{2(4n-1)^2n^2(4(N-n)-1)^2(N-n)^2} \end{aligned} \tag{4.57}$$

where

$$p_{3a}(n) := -96n^4 + 192n^3N - 2n^2\left(64N^2 + 4N - 1\right) + 2nN\left(16N^2 + 4N - 1\right) - 4N^3 + N^2 \ .$$

We further observe

$$p_3(n) := p_{3a}(n) + 128n^2N^2 - 135n^2N^2 < p_{3a}(n) \ .$$

The third derivative of $p_3(n)$ is $-2304n + 1152N$, which is positive for $n \in [31, N/2]$. This implies $p_3'(n)$ is convex in n. Evaluating $p_3'(n)$ at $n = N/2$ we find it equals $-7N^3 < 0$.

Convexity thus implies that $p_3(n)$ is minimized at either $n = 31$ or $n = N/2$. At these points, $p_3(n)$ equals $-88656094+5712122N-129486N^2+988N^3$ and $(1/4)N^2(2-8N+N^2)$. Both are positive under the assumptions, and so $p_3(n) > 0$. By its definition, this implies $p_{3a}(n) > 0$, which in turn implies $p_2''(n) < 0$ by (4.57). Therefore, $p_2(n)$ is concave in n. Evaluating at $n = 31$, we find $p_2(n)$ equals

$$-\frac{1}{360N^3}-\frac{3N}{248(N-31)}-\frac{1}{4N}+\frac{1}{372-12N}+\frac{1}{2}\log\left(\frac{1}{124-4N}+1\right)+\frac{28820189}{10898437500}+\frac{\log(2)}{2}-\frac{1}{2}\log\left(\frac{124}{123}\right)$$

and at $n = N/2$ we find $p_2(n)$ equals

$$-\frac{1}{360N^3}-\frac{25}{12N}-\frac{64}{45(2N+1)^3}+\frac{2}{6N+3}+\log\left(1-\frac{1}{2N}\right)+\frac{\log(2)}{2} .$$

Both increase in N for $N \geq 100$, which can be verified by differentiation. Furthermore, at $N = 100$ they equal 0.3221 and 0.3240 respectively. This implies $p_2(n) > 0$ by concavity. We next evaluate $p_1(D)$ at $D = (15/100)N$ which we define as the sum $p_4(n)+p_5(n)$, where

$$p_4(n) := \frac{3N}{8n^2-8nN}+\frac{1}{12n-12N}-\frac{1}{12n}+\frac{10}{3(6n+5)}-\frac{1600}{9(6n+5)^3}-\frac{1}{360N^3}-\frac{349}{612N}+\frac{1}{4}\log\left(\frac{51}{50}\right)$$

and

$$p_5(n) := \frac{1}{2}\log\left(\frac{5}{6n-6N}+1\right)+\frac{1}{2}\log\left(1-\frac{5}{34n}\right)+\frac{1}{4}\log\left(\frac{51}{50}\right) .$$

We conclude by showing $p_4(n) \geq 0$ and $p_5(n) \geq 0$. Beginning with $p_4(n)$, we differentiate to find

$$p_4'(n) = -\frac{36n^2+60n-3175}{(6n+5)^4}-\frac{456n^4-120n^3N+60n^2N(N+22)-110nN(6N-5)-275N^2}{24n^2(6n+5)^2(N-n)^2} .$$

This is negative since both numerators are positive for $n \in [31, N/2]$ (the quartic in n has positive second and first derivatives and so is bounded below at $n = 31$). Therefore, $p_4(n)$ decreases in n and is bounded below by the quantity

$$-\frac{1}{360N^3}-\frac{1471}{612N}+\frac{10}{3(3N+5)}-\frac{1600}{9(3N+5)^3}+\frac{1}{4}\log\left(\frac{51}{50}\right) .$$

This is increasing for $N \geq 500$ and at $N = 500$ approximately equal to 0.00235. Therefore, $p_4(n) > 0$. Turning to $p_5(n)$, we differentiate twice to find

$$p_5''(n) = -\frac{5(68n-5)}{2n^2(34n-5)^2}-\frac{5(12(N-n)-5)}{2(6(N-n)-5)^2(N-n)^2} < 0 .$$

Therefore, $p_5(n)$ is concave in n. Checking the n-boundary, at $n = 31$ we find $p_5(n)$ equals

$$\frac{1}{2}\log\left(\frac{5}{186-6N}+1\right)+\frac{1}{4}\log\left(\frac{51}{50}\right)-\frac{1}{2}\log\left(\frac{1054}{1049}\right)$$

and at $n = N/2$ it equals

$$\frac{1}{2}\log\left(1-\frac{5}{3N}\right)+\frac{1}{2}\log\left(1-\frac{5}{17N}\right)+\frac{1}{4}\log\left(\frac{51}{50}\right) .$$

Both expressions increase in N for $N \geq 500$, and approximately equal 0.0016 and 0.0029 at $N = 500$ respectively. Thus both are positive and $p_5(n) > 0$ under the assumptions. Since $p_4(n) > 0$ too, we see $p_1(D) > 0$ at $D = (15/100)N$. By concavity $p_1(D) > 0$, which proves the claim. ☐

4.8 δ: First Partial Derivatives for $D \leq N/3$

4.8.1 Restricting $x \in [N/(n(N-D)), (N-2D)/(2(N-D))]$

Lemma 38. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $25 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{N-n}{N}g_1\left(D|x=\frac{N}{n(N-D)}\right)+a_1\left(D|x=\frac{N-2D}{2(N-D)}\right)$$
$$+a_2\left(D|x=\frac{N-2D}{2(N-D)}\right)+\sum_{i=1}^{2}b_i\left(D|x=\frac{N}{n(N-D)}\right)$$

decreases in D.

Proof. Differentiating twice, we may write the expression of interest as

$$\frac{n\left(nD(2N-D)+N^2\right)}{D^2(N+nD)^2}+\frac{n\cdot p_1(D)}{D(N-D)(nD+N)(n(N-D)-N)^2} \tag{4.58}$$

where we define

$$p_1(n) := \left[(N-D)^3n^2 - N(D^2-4DN+2N^2)n + N^3\right] .$$

We see if $p_1(n) \geq 0$, then (4.58) will be positive. Since $p_1''(n) = 2(N-D)^3 > 0$, we find $p_1'(n)$ is bounded below by evaluating at $n = 25$, which yields the lower bound

$$25D^2(N-2D)+2N(62D^2-73DN+24N^2) .$$

Since the quadratic in D is minimized at $D = (73/124)N$, and attains the positive value of $(623/248)N^2$ there, we infer that $p_1'(n) \geq 0$. Hence, we may evaluate $p_1(n)$ at $n = 25$ to find it is bounded below by

$$D^2(314N - 625D) + N(1536D^2 - 1775DN + 576N^2) \ .$$

Again since the quadratic in D is minimized at $(1775/3072)N$, and attains the positive value of $(388319/6144)N^2$ there, we conclude that $p_1(n) \geq 0$. Hence, (4.58) is positive, implying the first derivative of the expression of interest is bounded above when evaluated at $D = N/3$. Doing this, we find the upper bound may be written

$$p_2(n) := -\frac{(3n-4)}{N} + \frac{6n^3}{(n+3)(2n-3)N} + \frac{n}{N}\left(\log\left(1+\frac{3}{2n}\right) + \log\left(1+\frac{3}{n}\right)\right) \ .$$

Since

$$p_2''(n) = \frac{27\left(14n^3 - 45n^2 + 81n + 54\right)}{(n+3)^3(2n-3)^3N} > 0 \ ,$$

we see $p_2'(n)$ increases in n, and so is bounded above at $n = N/3$. Evaluating there, we obtain the expression

$$\frac{p_3(N)}{(2N-9)^2N(9+N)^2} \ ,$$

where we define

$$p_3(N) := 27(N-9)\left(2N^2+81\right) + (2N-9)^2(N+9)^2\left(\log\left(1-\frac{9}{2N}\right) - \log\left(1+\frac{9}{N}\right)\right) \ .$$

The fifth derivative of $p_3(N) > 0$, the fourth derivative is negative and increases to 0, the third positive and decreases to 0, the second negative an increases to -1701, and the first negative and decreases to $-\infty$. Hence $p_3(N)$ decreases in N, and at $N = 100$ is approximately -7.9×10^{-6}. Thus $p_2'(n)$ is negative, and so we infer $p_2(n)$ decreases in n. Evaluating $p_2(n)$ at $n = 25$ gives the upper bound of

$$\frac{157 - 16450\log\left(\frac{56}{47}\right)}{658N} \ ,$$

which is negative since the numerator is approximately -2725.11. Hence, $p_2(n) < 0$, and the claim follows. □

Lemma 39. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $25 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{1}{6} b_3'' \left(D | x = \frac{N}{n(N-D)} \right) + b_4'' \left(D | x = \frac{N}{n(N-D)} \right)$$

is positive.

Proof. We may write the expression of interest as

$$\frac{nN}{6D^2(N-D)(D(N-n)-N)^2((N-n)(N-D)+n)^2} \cdot p_1(n) \tag{4.59}$$

where we define $p_1(n)$ as

$$\left[\begin{array}{c} -2D(N-D)^3 n^3 \\ -(N^4 - 6DN^3(1+N) + D^4(5+6N) - 2D^3N(1+9N) + 2D^2N^2(4+9N))n^2 \\ +2N(N^3(1+N) + D^4(5+3N) + D^2N^2(8+9N) - DN^2(2+3N(2+N)) - D^3(5+N(2+9N)))n \\ +N^2\left(D^4(-(2N+5)) + 2D^3\left(3N^2+N+5\right) - D^2\left(6N^3+8N^2+5\right)\right) \\ +N^2\left(2DN^2(N+1)(N+2) - N^2(N+1)^2\right) \end{array} \right] .$$

The sign of (4.59) is determined by the sign of $p_1(n)$, and so we prove the claim by showing $p_1(n) \geq 0$. Notice first that $p_1'''(n) = -12D(N-D)^3 < 0$. Evaluating $p_1''(n)$ at $n = D$, we find it is bounded below by $2 \cdot p_2(D)$, where we define

$$p_2(D) := 6D^5 - D^4(24N+5) + 2D^3N(18N+1) - 8D^2N^2(3N+1) + 6DN^3(N+1) - N^4 \; .$$

Now, the fourth derivative of $p_2(D)$ is $-120 + 720D - 576N$, which is negative under the assumptions. Evaluating the third derivative of $p_2(D)$ at $D = N/3$, we find it is bounded below by $64N^2 - 28N > 0$. Evaluating $p_2''(D)$ at $D = N/3$, we find it is bounded above by $-(8/9)N^2(4N+21) < 0$, implying $p_2(D)$ is concave in D. At $D = 26$, $p_2(D)$ equals $155N^4 - 16068N^3 + 627328N^2 - 10932272N + 69003376$ and at $D = N/3$ $p_2(D)$ equals $(2/81)N^4(16N+5)$. Both are positive under the assumptions, and so concavity implies $p_2(D) \geq 0$. Hence, $p_1''(n) \geq 0$. This implies $p_1'(n)$ increases in n. Evaluating $p_1'(n)$ at $n = D$, we obtain the upper bound $2 \cdot p_3(D)$, where we define

$$p_3(D) := \begin{bmatrix} 3D^6 - 5D^5(3N+1) + D^4N(30N+7) - 5D^3N\left(6N^2+2N+1\right) \\ +D^2N^3(15N+14) - DN^3\left(3N^2+7N+2\right) + N^4(N+1) \end{bmatrix} .$$

The fifth derivative of $p_3(D)$ is $-600 + 2160D - 1800N < 0$. Hence the fourth derivative of $p_3(D)$ decreases, and evaluating at $D = N/3$ gives a positive lower bound of $240N^2 - 32N$. This implies the third derivative of $p_3(D)$ increases, and again evaluating at $D = N/3$ gives the negative upper bound of $-(2/3)N(45+56N+40N^2)$. So, $p_3''(D)$ decreases and is bounded below by $(2/27)N^2(184N - 135)$ from which we see $p_3(D)$ is convex. At $D = 26$ we find $p_3(D)$ equals $867340448 - 175109688N + 13533520N^2 - 517868N^3 + 9959N^4 - 77N^5$ and at $D = N/3$ it equals $-(4/243)N^4(8N^2 + 5N - 9)$. By convexity in infer $p_3(D) \leq 0$, and so $p_1'(D)$ is negative as well. This implies $p_1(n)$ decreases in n. Evaluating $p_1(n)$ at $n = D$, we find it is bounded below by

$$p_4(D) := \begin{bmatrix} 2D^7 - D^6(12N+5) + 6D^5N(5N+2) - D^4N\left(40N^2+17N+10\right) \\ +2D^3N^2\left(15N^2+12N+5\right) - D^2N^2\left(12N^3+21N^2+4N+5\right) \\ +2DN^4\left(N^2+4N+3\right) - N^4(N+1)^2 \end{bmatrix} .$$

The sixth derivative of $p_4(D)$ equals $-3600 + 10080D - 8640N < 0$. By monotonicity and evaluating at $D = N/3$, we find under the assumptions the fifth derivative is positive, the fourth is negative, and the third positive. This implies $p_4'(D)$ is convex. At $D = N/3$, we find $p_4(D)$ equals

$$-\frac{2N^3}{729}\left(128N^3 - 36N^2 - 1890N + 1215\right) ,$$

which is negative under the assumptions. This combined with convexity implies $p_4(D)$ is minimized at either $D = 26$ or $D = N/3$. Evaluating at these points yields

$$51N^6 - 7906N^5 + 513239N^4 - 17859920N^3 + 348845068N^2 - 3568982560N + 14519041472$$

and

$$\frac{2N^4}{2187}\left(64N^3 + 60N^2 - 216N - 1701\right) ,$$

both of which are positive. We conclude $p_4(D) \geq 0$, giving the claim. □

Lemma 40. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $25 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{2}{3}b_3''\left(D|x = \frac{N}{n(N-D)}\right) + a_3''\left(D|x = \frac{N}{n(N-D)}\right)$$

is positive.

Proof. Combining terms, we may write the expression of interest as

$$\frac{n}{3D^2(D(N-n)-N)^3} \cdot p_1(n) \tag{4.60}$$

where we define

$$p_1(n) := ND^2n^2 - DN(D - 6N + 2DN)n + N^2(2N + D^2(N+1) - D(6N-2)) \ .$$

Since $p_1''(n) = 2ND^2 > 0$, we find $p_1'(n)$ is bounded above by $DN \cdot p_2(D)$, where we define

$$p_2(D) := 2D^2 - D(2N+1) + 6N$$

Since $p_2'(D) = -1 + 4D - 2N < 0$, we see $p_2(D) \leq p_2(26) = 1326 - 46N < 0$, and hence $p_1'(n) < 0$. So, $p_1(n)$ decreases in n and evaluating at $n = D$ gives the lower bound $N \cdot [D\ p_3(D) + p_4(D)]$, where we define

$$p_3(D) := D^3 + \frac{1}{4}DN(N+7) + 2N(1-3N)$$

and

$$p_4(D) := -(2N+1)D^3 + \frac{3}{4}D^2N(N+7) + 2N^2 \ .$$

Since $p_3'(D) = 3D^2 + (1/4)N(N+7) > 0$, we see $p_3(D)$ increases in D and is bounded below by $17576 + (95/2)N + (1/2)N^2 > 0$. Turning to $p_4(D)$, we see its third derivative is negative, implying the first derivative is concave. At $D = 26$, the first derivative equals $-2028 - 3783N + 39N^2$ and at $D = N/4$ it equals $(39/16)N^2$. Concavity implies $p_4(D)$ increases for $D \in [26, N/4]$, and consequently is bounded below by the positive quantity $-17576 - 31603N + 509N^2$. For $D \in [N/4, N/3]$, we find the second derivative is bounded above by $-(3/2)(N-6)N < 0$, implying $p_4(D)$ is concave on this interval. Evaluating at the boundary points, we find $p_4(D)$ equals $(1/64)N^2(128+20N+N^2)$ and $(1/108)N^2(216+59N+N^2)$ at them, and so is positive by concavity. The positivity of $p_3(D)$ and $p_4(D)$ implies $p_1(n)$ is positive at $n = D$, which proves the claim. □

Lemma 41. *Let* $N \geq 500$, n, $D \in \mathbb{N}$, *and* $25 \leq n \leq D \leq N/3$. *Using the functions in* (3.46)*, we have that the function*

$$\frac{1}{6}b_3''\left(D|x = \frac{N}{n(N-D)}\right) + a_4''\left(D|x = \frac{N}{n(N-D)}\right)$$

is positive.

Proof. Combining terms, we may write the expression of interest as

$$\frac{nN}{6D^2(D(N-n)-N)^2((N-n)(N-D)+N)^3} \cdot p_1(n)$$

where we define $p_1(n)$ to be

$$\left[\begin{array}{c} 2D(N-D)^3n^4 \\ +(N^4-8DN^3(1+N)+12D^2N^2(1+2N)-4D^3N(1+6N)+D^4(5+8N))n^3 \\ -3N(N^3(N+1)+D^4(4N+5)+D^2N(1+12N(N+1)))n^2 \\ +3N(DN^2(3+4N(2+N))+D^3(3+4N(1+3N)))n^2 \\ +N^2(3(D-1)D^2(5D-1)+2(D-1)D(1-2D+4D^2)N-3(2D-1)^3N^2)n \\ +N^2(6(2D-1)^2N^3+(3-8D)N^4)n \\ +N^3(-N(N+1)^3-D^4(2N+5)+D^3(9+4N+6N^2)) \\ +N^3(D(1+N)^2(2N(N+2)-1)-3D^2(1+N(1+2N(2+N)))) \end{array}\right].$$

We see the claim will follow if we show $p_1(n) \geq 0$. The fourth derivative of $p_1(n)$ is $48D(N-D)^3 > 0$. This implies the third derivative of $p_1(n)$ increases in n. Evaluating the third derivative at $n = D$, we obtain the upper bound of $-6 \cdot p_2(D)$, where we define

$$p_2(D) := 8D^5 - D^4(32N+5) + 4D^3N(12N+1) - 4D^2N^2(8N+3) + 8DN^3(N+1) - N^4 .$$

The fourth derivative of $p_2(D)$ is $-120 + 960D - 768N < 0$, implying the third derivative decreases in D. Evaluating the third derivative at $D = N/3$, we find it is bounded below by $(16/3)N(16N - 3) > 0$, implying the second derivative increases in D. Evaluating $p_2''(D)$ at $D = N/3$, we find it is bounded above by $-(4/27)N^2(32N + 153) < 0$, implying $p_2(D)$ is concave. At $D = 26$, $p_2(D)$ equals $207N^4 - 21424N^3 + 835536N^2 - 14552928N + 92766128$ and at $D = N/3$ it equals $(2/243)N^4(64N + 51)$. Both are positive, so concavity implies $p_2(D) \geq 0$. This in turn implies $p_1''(n)$ decreases in n. Evaluating $p_1''(n)$ at $n = D$, we find the lower bound $-6 \cdot p_3(D)$, where we define

$$p_3(D) := \begin{bmatrix} 4D^6 - 5D^5(4N+1) + D^4N(40N+9) - D^3N(8N(5N+2)+3) \\ +D^2N^2(20N(N+1)+1) - DN^3(N(4N+9)+3) + N^4(N+1) \end{bmatrix}.$$

The fifth derivative of $p_3(D)$ is $-600 + 2880D - 2400N < 0$. Monotonicity and evaluating at $D = N/3$ shows that the fourth derivative of $p_3(D)$ is positive and the third derivative is negative. The same holds for $p_3''(D)$, where evaluating at $D = N/3$ gives the positive lower bound $(4/27)N^2(110N - 27)$. This implies $p_3(D)$ is convex; evaluating $p_3(D)$ at $D = 26$ yields $-103N^5 + 13287N^4 - 689598N^3 + 17998500N^2 - 233567464N + 1176256224$ and at $D = N/3$ yields $-(4/729)N^5(32N + 51)$. Both are negative, and we conclude $p_3(D) \leq 0$ by convexity. Hence $p_1''(n) \geq 0$. Evaluating $p_1'(n)$ at $n = D$ produces the upper bound

$$p_4(D) := \begin{bmatrix} -8D^7 + 3D^6(16N+5) - 6D^5N(20N+7) + D^4N\left(160N^2+75N+18\right) \\ -12D^3N^2\left(10N^2+9N+2\right) + 3D^2N^2\left(16N^3+29N^2+8N+1\right) \\ -2DN^3\left(4N^3+15N^2+12N+1\right) + 3N^4(N+1)^2 \end{bmatrix}.$$

The sixth derivative of $p_4(D)$ is $10800 - 40320D + 34560N > 0$. Monotonicity and evaluating at $D = N/3$ then shows: that the fifth derivative of $p_4(D)$ is negative; the fourth derivative of $p_4(D)$ is positive; the third derivative of $p_4(D)$ is negative (with the upper bound of $-(16/27)N^3(441 + 80N)$). This implies that $p_4'(D)$ is concave. Evaluating $p_4'(D)$ at $D = N/3$ yields the expression $(8/729)N^4(-1215 + 81N + 128N^2)$, which is positive. Positivity combined with concavity implies that $p_4(D)$ is maximized either at $D = 26$ or $D = N/3$.

Evaluating at these points yields the expressions

$$-205N^6+31674N^5-2050929N^4+71234124N^3-1391911716N^2+14337165024N-59620744768$$

and $-(8/2187)N^4(64N^3+153N^2-729)$, respectively. Since both are negative, we infer $p_4(D) \leq 0$. Hence, $p_1'(n) \leq 0$, and so $p_1(n)$ decreases in n. Evaluating $p_1(n)$ at $n = D$ thus gives the lower bound

$$p_5(D) := \begin{bmatrix} -2D^8 + D^7(14N+5) - D^6N(42N+19) + D^5N\left(70N^2+39N+9\right) \\ -D^4N^2\left(70N^2+61N+21\right) + D^3N^2\left(42N^3+65N^2+24N+3\right) \\ -D^2N^3\left(14N^3+39N^2+24N+5\right) + DN^3(N+1)^2\left(2N^2+7N-1\right) - N^4(N+1)^3 \end{bmatrix}.$$

The seventh derivative of $p_5(D)$ is $25200 - 80640D + 70560N > 0$. Monotonicity and evaluating at $D = N/3$ gives: the sixth derivative of $p_5(D)$ is negative; the fifth derivative of $p_5(D)$ is positive; the fourth derivative of $p_5(D)$ is negative; the third derivative of $p_5(D)$ is positive (with lower bound of $(2/81)N^2(729+1458N+3666N^2+224N^3)$). This implies $p_5'(D)$ is convex in D. Evaluating $p_5'(D)$ at $D = N/3$ yields $-(N^3/2187)(640N^4+840N^3-9720N^2-5832N+2187)$, which is negative. Negativity combined with convexity implies $p_5(D)$ is minimized either at $D = 26$ or $D = N/3$. Evaluating at these values yields the expressions

$$\begin{bmatrix} 51N^7 - 9181N^6 + 712215N^5 - 30861975N^4 + 804239202N^3 \\ -12520632696N^2 + 106682875104N - 377495078272 \end{bmatrix}$$

and

$$\frac{4N^4}{6561}\left(64N^4+204N^3-2916N-2187\right).$$

As both are positive, we infer $p_1(n) \geq 0$, giving the claim. ☐

Lemma 42. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $25 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{n}{N}g_1\left(D|x = \frac{N}{n(N-D)}\right) + \sum_{i=3}^{4} b_i\left(D|x = \frac{N}{n(N-D)}\right) + a_i\left(D|x = \frac{N}{n(N-D)}\right)$$

decreases in D.

Proof. Differentiate the expression of interest twice, and combine Lemma 39, Lemma 40, and Lemma 41, to see the second derivative is positive under the assumptions (notice that at $x = N/(n(N-D))$, the term $(n/N)g_1''(D|x = N/(n(N-D))) = 0$). Hence, the first derivative increases in D, and so evaluating the first derivative of the expression of interest at $D = N/3$ gives the following upper bound:

$$p_1(n) := \frac{4n}{N(N-n)} - \frac{6n}{N(N-n-3)} + \frac{9n(N-n-1)}{2N(N-n-3)^2} + \frac{3n}{N(2(N-n)+3)} - \frac{9n}{2N(2(N-n)+3)^2} + \frac{n}{N}\left(\log\left(1-\frac{3}{N-n}\right) - \log\left(1+\frac{3}{2(N-n)}\right)\right) .$$

Differentiating twice, we find $p_1''(n)$ may be written as

$$\frac{p_2(n)}{(N-n)^3(N-n-3)^4N(2(N-n)+3)^4}$$

where we define

$$p_2(n) := \left[\begin{array}{c} -8(2N-9)n^8 \\ +4\left(32N^2-114N+999\right)n^7 \\ -\left(448N^3-1176N^2+19476N+11745\right)n^6 \\ +\left(896N^4-1512N^3+32940N^2+58671N+45684\right)n^5 \\ -\left(1120N^5-840N^4+12420N^3+117180N^2+208089N+10206\right)n^4 \\ +\left(896N^6+168N^5-30060N^4+116910N^3+375516N^2+75330N-2187\right)n^3 \\ -N\left(448N^6+504N^5-43524N^4+58185N^3+334854N^2+164754N+5103\right)n^2 \\ +N\left(128N^7+264N^6-23004N^5+11475N^4+147096N^3+144342N^2+16767N-69984\right)n \\ +N\left(-16N^8-48N^7+4500N^6+54N^5-25353N^4-44712N^3-9477N^2+69984N+52488\right) \end{array}\right]$$

While $p_2(n)$ is complicated at first glance, it is amenable to analysis via repeated differentiation. In particular, the eighth derivative is $-322560(2N-9) < 0$. Monotonicity and evaluating at $D = N/3$ then gives: the seventh derivative is positive; the sixth derivative is negative; the fifth, positive; the fourth, negative; the third, positive; the second, negative; the first, positive. This implies that $p_2(n)$ increases in n. Evaluating $p_2(n)$ at $n = N/3$, we find it bounded above by

$$-\frac{4096N^9}{6561} - \frac{1024N^8}{243} + 512N^7 + \frac{1568N^6}{3} - 2000N^5 - 12240N^4 - 4536N^3 + 46656N^2 + 52488N ,$$

which is negative under the assumptions. This implies $p_1'(n)$ decreases in n. Evaluating $p_1'(n)$ at $n = 25$, we find it is bounded above by $(1/(2N)) \cdot p_3(N)$, where we define

$$\begin{aligned} p_3(N) := &-\frac{53}{N-28} - \frac{57}{(N-28)^2} + \frac{900}{(N-28)^3} + \frac{8}{N-25} + \frac{200}{(N-25)^2} + \frac{106}{2N-47} \\ &- \frac{900}{(2N-47)^3} + \frac{291}{(47-2N)^2} + 2\log\left(1 - \frac{3}{N-25}\right) - 2\log\left(1 + \frac{3}{2(N-25)}\right) . \end{aligned}$$

Differentiation reveals $p_3(N)$ has monotone increase in N; taking the limit, we find it tends to 0 and $N \nearrow \infty$. At $N = 100$, we find it is approximately -0.018, and so infer that $p_3(N) \leq 0$. This implies $p_1'(n) \leq 0$. Hence $p_1(n)$ decreases in n. Evaluating $p_1(n)$ at $n = 25$, we find it is bounded above by $(25/(2N)) \cdot p_4(N)$, where we define $p_4(N)$ to be

$$-\frac{3}{N-28} + \frac{18}{(N-28)^2} + \frac{8}{N-25} + \frac{6}{2N-47} - \frac{9}{(47-2N)^2} + 2\log\left(1 - \frac{3}{N-25}\right) - 2\log\left(1 + \frac{3}{2(N-25)}\right) .$$

Differentiating $p_4(N)$ shows it has monotone increase in N as well, and in the limit approaches 0. Now at $N = 100$ we find it is approximately -0.0139, and so infer $p_4(N) \leq 0$. Hence $p_1(n) \leq 0$ and so we conclude the expression of interest decreases in D, as claimed. □

Lemma 43. *Let $N \geq 500$, $n, \in \mathbb{N}$, and $25 \leq n \leq N/3$. Using the functions in* (3.46), *we have that the function*

$$\begin{aligned} &g_1\left(n|\ D = \frac{N}{3}, x = \frac{N}{n(N-D)}\right) + \sum_{i=1}^{2} a_i\left(n|D = \frac{N}{3}, x = \frac{N-2D}{2(N-D)}\right) \\ &+ \left[\sum_{j=3}^{4} a_j\left(n|D = \frac{N}{3}, x = \frac{N}{n(N-D)}\right)\right] + \left[\sum_{i=1}^{4} b_i\left(n|D = \frac{N}{3}\ , x = \frac{N}{n(N-D)}\right)\right] \end{aligned}$$

is positive

Proof. Differentiating the expression of interest twice, we find we may write it as

$$-\frac{6}{(n+3)^3} - \frac{18(n-6)}{(2n^2+3n-9)^2} - \frac{2}{3(2(N-n)+3)^3(N-n)^3(N-n-3)^3} \cdot p_1(n) \ , \qquad (4.61)$$

where we define

$$p_1(n) := \left[\begin{array}{c} 4(N-9)n^6 \\ -3(27-42N+8N^2)n^5 \\ +3(-567+18N-30N^2+20N^3)n^4 \\ +(1215+4266N+594N^2-180N^3-80N^4)n^3 \\ +3N(-567-864N-432N^2+120N^3+20N^4)n^2 \\ -3N(-1944+81N+270N^2-333N^3+78N^4+8N^5)n \\ +N(-5832-5832N+729N^2+837N^3-270N^4+54N^5+4N^6) \end{array}\right]$$

We see that if we show $p_1(n) \geq 0$, it will imply (4.61) is negative, and hence the expression of interest is concave. We proceed by differentiation. The sixth derivative equals $2880(N-9) > 0$. Monotonicity and evaluating at $D = N/3$ gives the following: the fifth derivative is negative; the fourth derivative, positive; the third, negative; the second, positive; the first, negative. Hence $p_1(n)$ decreases in n and at $n = N/3$ equals

$$\frac{256N^7}{729} + \frac{704N^6}{81} - \frac{176N^5}{3} + 416N^4 + 504N^3 - 3888N^2 - 5832N \ ,$$

which is positive. Hence (4.61) is negative, and so the expression of interest is concave. At $n = 25$, the expression equals

$$\frac{-16N^3+749N^2-7228N-35625}{3(N-28)(N-25)(2N-47)} + \frac{50}{3}\left(\log\left(1+\frac{3}{2(N-25)}\right) - \log\left(1-\frac{3}{N-25}\right) + \log\left(\frac{56}{47}\right)\right) \ ,$$

and at $n = N/3$ it equals

$$\frac{-34N^2+54N+324}{(2N-9)(4N+9)} + \frac{2N}{9}\left(\log\left(1+\frac{9}{N}\right) - 2\log\left(1-\frac{9}{2N}\right) + \log\left(1+\frac{9}{4N}\right)\right) \ .$$

Standard differentiation arguments show both expressions increase monotonically in N, the first to a limiting value of approximately 0.2534 and the second to a value of 1/4. At $N = 100$, the first expression approximately equals 0.191 and the second approximately equals 0.177. Concavity then implies the expression of interest is positive, giving the claim. $\square$

4.8.2 Restricting $x \in [(N-2D)/(2(N-D)), 1/2]$

Lemma 44. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $29 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{N-n}{N} g_1 \left(D | x = \frac{N-2D}{2(N-D)} \right) + a_1 \left(D | x = \frac{1}{2} \right) + a_2 \left(D | x = \frac{1}{2} \right) + \sum_{i=1}^{2} b_i \left(D | x = \frac{N-2D}{2(N-D)} \right)$$

is positive.

Proof. Differentiating twice, we may write the expression of interest as

$$\frac{n\left(8D^3 - 8D^2N + N^3\right)}{D^2N^2(N-D)} + \frac{4N}{(N+D)^3} \ ,$$

which is positive under the assumptions. Evaluating the first derivative at $D = N/3$, we find it bounded above by

$$-\frac{8n(\log(8)-1)+27}{24N} \ .$$

Since $\log(8) > 1$, this is negative, and so the expression of interest decreases in D. Evaluating the expression at $D = N/3$, we find it equals

$$\frac{1}{18}(4n(\log(8)-2)-9) \ ,$$

which increases in n since $\log(8) > 2$. For $n \geq 29$, this is positive, and the claim follows. □

Lemma 45. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $25 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{1}{6} b_3'' \left(D | x = \frac{N-2D}{2(N-D)} \right) + b_4'' \left(D | x = \frac{N-2D}{2(N-D)} \right)$$

is positive.

Proof. Combining terms, we may write the expression of interest as

$$\frac{n^2}{6D^2(2D-n)^{(}2(N-D)-n)^2(N-D)N} \cdot p_1(n) \tag{4.62}$$

where we define

$$p_1(n) := \begin{bmatrix} -(N^2+5D^2)n^3 + 4(N^3 - 2D^2N + 7D^3)n^2 \\ -4(5D^4 + 6D^3N - 8D^2N^2 + 2DN^3 + N^4)n \\ +16D(N-D)^3n \end{bmatrix} .$$

We see the claim follows if we show $p_1(n) \geq 0$. The third derivative of $p_1(n)$ is $-6(N^2 + 5D^2) < 0$, and so the second derivative decreases in n and is bounded below by

$$26D^3 - 16D^2N - 6DN^2 + 8N^3 .$$

Differentiation shows this expression decreases in D, and so evaluating at $D = N/3$ we find it is further bounded below by $(140/27)N^3 > 0$. Hence, $p_1'(n)$ increases in n, and evaluating $p_1'(n)$ at $n = D$ yields the upper bound of

$$21D^4 - 40D^3N + 29D^2N^2 - 4N^4 .$$

Differentiation show this expression increases in D, and we now evaluate at $D = N/3$ to find it is bounded above by $-2N^4 < 0$. Therefore, $p_1(n)$ decreases in n, and so we evaluate at $n = D$ to find that $p_1(n)$ is bounded below by

$$D(3D^4 - 48D^3N + 79D^2N^2 - 52DN^3 + 12N^4) =: D(p_2(D)) .$$

Again differentiation shows $p_2(D)$ decreases in D, and so we evaluate a final time at $D = N/3$ to find that $p_2(D)$ is bounded below by $(46/27)N^4 > 0$. Hence, $p_2(D) \geq 0$, and so $D \cdot p_2(D) \geq 0$. Hence $p_1(n) \geq 0$ and the claim follows. □

Lemma 46. *Let $N \geq 500$, $n,\ D \in \mathbb{N}$, and $25 \leq n \leq D \leq N/3$. Using the functions in* (3.46), *we have that the function*

$$\frac{n}{N} g_1'' \left(D | x = \frac{N-2D}{2(N-D)} \right) + \frac{1}{3} b_3'' \left(D | x = \frac{N-2D}{2(N-D)} \right) + a_3'' \left(D | x = \frac{N-2D}{2(N-D)} \right)$$

is positive.

Proof. Combining terms, we may write the expression of interest as

$$\frac{2n}{3D^2(2D-n)^3(N-n)N^2} \cdot p_1(n) \tag{4.63}$$

where we define

$$p_1(n) := \begin{bmatrix} -(N-3D)(N+4D)n^4 \\ +(-72D^3+2D^2N+7DN^2+N^3)n^3 \\ +2D(72D^3-3N^3-DN(3+5N))n^2 \\ +2D^2(-48D^3+N^2(9+4N))n \\ -12D^2N^3 \end{bmatrix} .$$

We see the claim will follow if we show $p_1(n) \geq 0$. The fourth derivative of $p_1(n)$ is $-24(N-3D)(N+4D) < 0$, which implies the third derivative of $p_1(n)$ decreases in n. Evaluating at $n = D$, we find that the third derivative of $p_1(n)$ is bounded below by

$$-144D^3 - 12D^2N + 18DN^2 + 6N^3 .$$

This expression is concave in D, at $D = 25$ equals $-2250000 - 7500N + 450N^2 + 6N^3$, and at $D = N/3$ equals $(16/3)N^3$. Hence it is positive, and so $p_1''(n)$ increases in n. When we evaluate $p_1''(n)$ at $n = D$, we thus find it is bounded above by $2DN(-3N^2 + D(5N-6))$, which is negative. Therefore, $p_1(n)$ is concave in n. Evaluating at $n = 25$ we find the expression

$$p_2(D) := \begin{bmatrix} -2400D^5 + 90000D^4 - 1125000D^3 \\ +4D^2(1171875 + 6875N - 1450N^2 + 47N^3) \\ -625DN(625 - 175N + 6N^2) \\ +15625(N-25)N^2 \end{bmatrix} , \tag{4.64}$$

and evaluating at $n = D$ we obtain the expression $D^2 p_3(D)$, where we define

$$p_3(D) := -12D^4 + D^3N - 2D^2N(2N+3) + 3DN^2(N+6) - 12N^3 .$$

The claim will follow if we show $p_2(D) \geq 0$ and $p_3(D) \geq 0$. We first consider $p_2(D)$. We observe we may further decompose it into the expression

$$7500D^2 \cdot p_4(D) + 25D \cdot p_5(D) + N^2 \cdot p_6(D) \ , \tag{4.65}$$

where we define

$$\begin{aligned} p_4(D) &:= 625 - 150D + 12D^2 \ , \\ p_5(D) &:= -96D^4 + 2DN(550 + N^2) - 15625N \ , \\ \text{and } p_6(D) &:= 2D^2(69N - 2900) - 625D(6N - 175) + 15625(N - 25) \ . \end{aligned}$$

Thus if we show $p_4(D) \geq 0$, $p_5(D) \geq 0$, and $p_6(D) \geq 0$, (4.65) will imply $p_2(D) \geq 0$. We first show such positivity for $D \in [25, N/4]$. On this interval, we first see that $p_4(D)$ increases in D, and at $D = 25$ equals 4375. Next we see $p_5(D)$ is concave in D, equals $-37500000 + 11875N + 50N^3$ when $D = 25$, and equals $(N/8)(-125000 + 2200N + N^3)$ when $D = N/4$. Both expressions are positive for $N \geq 100$ and so concavity implies $p_5(D) \geq 0$. Finally, since $4 \cdot 25 \cdot 69 > 625 \cdot 6$, we see $p_6(D)$ increases in D on the interval, and so is bounded below by $625(13N - 2050)$. This is positive under the assumptions, and so (4.65) implies $p_2(D) \geq 0$ for $D \in [25, N/4]$.

For $D \in [N/4, N/3]$ we demonstrate positivity by the direct analysis of $p_2(D)$. The fourth derivative of $p_2(D)$ is $-144000(2D - 15) < 0$, implying the third derivative of$p_2(D)$ decreases in D. Evaluating the third derivative at $D = N/4$ we find it bounded above by $-9000(750 - 60N + N^2)$, which is negative. So we see $p_2''(D)$ decreases in D, and evaluating again at $D = N/4$ gives the upper bound of $9375000 - 1632500N + 55900N^2 - 374N^3$, which is again negative under the assumptions. Hence, $p_2(D)$ is concave for $D \in [N/4, N/3]$. Evaluating at the D-boundary, we find it equals

$$\frac{N^2}{32}(301N^3 - 30350N^2 + 867500N - 6250000)$$

when $D = N/4$, and

$$\frac{2N^2}{81}(446N^2 - 31725N + 545625)$$

when $D = N/3$. Since both are positive for $N \geq 500$, concavity implies $p_2(D) \geq 0$ for $D \in [N/4, N/3]$. Combined with the preceding argument, $p_2(D) \geq 0$ for $D \in [25, N/3]$.

We conclude by showing $p_3(D) \geq 0$. Differentiating twice, we find $p_3''(D) = -144D^2 - 12N - N(8N - 6D) < 0$, and so $p_3(D)$ is concave in D. When $D = 25$, $p_3(D)$ equals

$$63N^3 - 2050N^2 + 11875N - 4687500$$

and when $D = N/3$ it equals

$$\frac{4N^3}{9}(N - 15) \ .$$

Both expressions are positive under the assumptions. The claim follows. $\square$

Lemma 47. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $25 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{1}{6}b_3''\left(D|x = \frac{N-2D}{2(N-D)}\right) + a_4''\left(D|x = \frac{N-2D}{2(N-D)}\right)$$

is positive.

Proof. Combining terms

$$\frac{n^2}{6D^2(2D-n)^2(2(N-D)-n)^3N} \cdot p_1(n) \tag{4.66}$$

where we define

$$p_1(n) := \begin{bmatrix} (N+D)n^4 - 2(3(N^2 - D^2) + 2DN)n^3 \\ +12(D^3 + D^2(2 - 3N) + DN^2 + N^3)n^2 \\ +8(D^4 + 12D^2N^2 - 4DN^3 - N^4 - 4D^3(3 + 2N))n \\ -32D(D^3(-3 + N) - 3D^2N^2 + 3DN^3 - N^4) \end{bmatrix} .$$

We see the claim will follow if we show $p_1(n) \geq 0$. Since the fourth derivative of $p_1(n)$ is $24(N + D) > 0$, we see the third derivative of $p_1(n)$ increases in n. Evaluating the third derivative of $p_1(n)$ at $n = D$, we find it is bounded above by $60D^2 - 36N^2$, which is negative.

Hence, the second derivative is decreasing in n, and evaluating at $n = D$ we find $p_2''(n)$ is bounded below by

$$12(6D^3 - D^2(7N - 4) - DN^2 + 2N^3) \ .$$

Differentiating this expression once shows it decreases in D under the assumptions, and so evaluating at $D = N/3$ we find that the $p_1''(n)$ is bounded below by $(8/3)N^2(5N + 2) > 0$. Therefore $p_2'(n)$ increases in n, and evaluating $p_1'(n)$ at $n = D$ we find it is bounded above by

$$2 \cdot (27D^4 - 24D^3(3N + 1) + 51D^2N^2 - 4DN^3 - 4N^4)$$

Differentiating this expression twice, we find it convex in D. When $D = 25$ it equals

$$-8N^4 - 200N^3 + 63750N^2 - 2250000N + 20343750$$

and when $D = N/3$ it equals

$$-\frac{4}{9}\left(9N^4 + 4N^3\right) \ .$$

Both are negative under the assumptions, and so by convexity we infer $p_1'(n) < 0$. Hence $p_1(n)$ decreases in n, and so evaluating $p_1(n)$ at $n = D$, we find it is bounded below by $D \cdot p_2(D)$, where we define

$$p_2(D) := 27D^4 - 3D^3(45N - 8) + 198D^2N^2 - 116DN^3 + 24N^4 \ .$$

The third derivative of $p_2(D)$ equals $144 + 648D - 810N < 0$, and so $p_2''(D)$ decreases in D. Evaluating at $D = N/3$ we find $p_2''(D)$ bounded below by $48N + 162N^2 > 0$, implying $p_2'(D)$ increases in D. At $D = N/3$, we find $p_2'(D)$ equals $8N^2 - 25N^3 < 0$, and so $p_2(D)$ decreases in D. Finally, at $D = N/3$, $p_2(D)$ equals $(8/9)N^3(3N + 1) > 0$, implying $p_2(D) \geq 0$. Hence $p_1(n) \geq 0$, and the claim follows. □

Lemma 48. *Let* $N \geq 500$, n, $D \in \mathbb{N}$, *and* $31 \leq n \leq D \leq N/3$. *Using the functions in* (3.46), *we have that the function*

$$\frac{n}{N} g_1\left(D|x = \frac{N-2D}{2(N-D)}\right) + \sum_{i=3}^{4} b_i\left(D|x = \frac{N-2D}{2(N-D)}\right) + a_i\left(D|x = \frac{N-2D}{2(N-D)}\right)$$

is positive.

Proof. Differentiate the expression of interest twice, and combine Lemma 45, Lemma 46, and Lemma 47, to see the second derivative is positive under the assumptions. Evaluating the first derivative of this expression at $D = N/3$, we find it bounded above by

$$\begin{aligned} p_1(n) := &-\frac{3n^2(4N-3n-3)}{N(4N-3n)^2} + \frac{10n^2}{3N(N-n)} + \frac{9n(2N-n)}{N(2N-3n)^2} - \frac{6n^2}{N(2N-3n)} \\ &+\frac{n}{N}\left(\log\left(1-\frac{n}{2(N-n)}\right) + \log\left(1+\frac{n}{4(N-n)}\right)\right) \ . \end{aligned} \tag{4.67}$$

Differentiating twice, we find that $p_1''(n)$ may be written

$$\frac{2\cdot p_2(n)}{3(4N-3n)^2(2N-3n)^2(N-n)^3} - \frac{6}{(4N-3n)(2N-3n)} + \frac{72\cdot p_3(n)}{(4N-3n)^4} + \frac{36\cdot p_4(n)}{(2N-3n)^4} \ , \tag{4.68}$$

where we define

$$\begin{aligned} p_2(n) &:= -81n^4(27N-7n) - 1047n^2(N-3n)N^2 - 8N^3(138n^2-99nN+19N^2) \\ p_3(n) &:= 3n^2-(4N-6)n+4N \ , \\ \text{and } p_4(n) &:= 6n^2-(4N-3)n+10N \ . \end{aligned}$$

Inspecting the terms of (4.68), we see if we show $p_2(n) \le 0$, $p_3(n) \le 0$, and $p_4(n) \le 0$, then we may conclude the entire expression is negative. We consider $p_2(n)$ first. As defined, the first two quantities are positive by assumption. The third quadratic in n is also positive, since $138n^2 - 99nN + 19N^2 \ge (4/3)N^2$ by monotone decrease when $n \in [29, N/3]$. Hence, each component respects its leading sign, and we infer $p_2(n) \le 0$. Turning to $p_3(n)$, we see it decreases in n, and so $p_3(n) \le p_3(29) = 2697 - 112N < 0$. Similarly, $p_4(n)$ decreases in n under the assumptions, and so $p_4(n) \le p_4(29) = 5133 - 106N < 0$. We thus see that $p_1''(n) \le 0$ by (4.68).

This implies $p_1'(n)$ decreases in n, and so evaluating at $n = 29$ we obtain a univariate function in N. We call this function $p_5(N)$. Analyzing it, we find it $p_5'(N) \ge 0$, that as $N \nearrow \infty$, $p_5(N) \nearrow 0$, and at $N = 100$ $p_5(N) \approx -0.0214$. The details of this analysis may be obtained by combining the terms of the expression, and demonstrating the numerator is negative for $N \geq 100$. Hence, we see that $p_1'(n) < 0$, and so $p_1(n)$ decreases in n. Evaluating $p_1(n)$ at $n = 29$, we find it equals

$$-\frac{29\left(1456N^4 - 18792N^3 - 5555646N^2 + 191624373N - 1661403069\right)}{3(4N-87)^2(2N-87)^2(N-29)N} + \frac{29}{N}\log\left(\frac{4N-174}{4N-87}\right) .$$

This quantity is negative, since the quartic in the first term is positive for $N \geq 100$ and the the argument of the logarithm is less than 1 for $N \geq 100$. We conclude $p_1(n) \leq 0$. This implies that the expression of interest decreases in D. Therefore evaluating at $D = N/3$ yields the lower bound

$$p_5(n) := -\frac{4n^2}{9(N-n)} + \frac{2n}{4N-3n} - \frac{2n}{2N-3n} - \frac{2n}{3}\log\left(\frac{4N-6n}{4N-3n}\right) . \tag{4.69}$$

Differentiating twice, we find we may write $p_5''(n)$ as

$$\left[\frac{15N^2(8N-9n)}{2\left(9n^2-18nN+8N^2\right)^2} - \frac{8N^2}{9(N-n)^3}\right] + \left[\frac{N^2(8N-9n)}{2\left(9n^2-18nN+8N^2\right)^2} - \frac{24N}{(2N-3n)^3}\right] + \frac{48N}{(4N-3n)^3} .$$

This is positive, since

$$\frac{15N^2(8N-9n)}{2\left(9n^2-18nN+8N^2\right)^2} - \frac{8N^2}{9(N-n)^3} = \frac{N^2\left(-81n^4 + 459n^3N - 603n^2N^2 + 153nN^3 + 56N^4\right)}{18(N-n)^3\left(9n^2-18nN+8N^2\right)^2}$$

and

$$\frac{N^2(8N-9n)}{2\left(9n^2-18nN+8N^2\right)^2} - \frac{24N}{(2N-3n)^3} = \frac{27n^2N^2 - 432n^2N - 42nN^3 + 1152nN^2 + 16N^4 - 768N^3}{2(2N-3n)^3(4N-3n)^2} .$$

Analyzing the numerators of these two expressions gives positivity for $n \in [29, N/3]$, (the first decreases in n, the second is concave in n), and so we infer $p_5''(n) \geq 0$. This implies $p_5'(n)$ increases in n, and so evaluating at $n = 29$ yields the lower bound

$$-\frac{174}{(2N-87)^2} + \frac{174}{(4N-87)^2} - \frac{116(2N-29)}{9(N-29)^2} + \frac{112N}{8N^2-522N+7569} - \frac{2}{3}\log\left(\frac{4N-174}{4N-87}\right) .$$

The derivative of this quantity may be written

$$-\frac{4\cdot\begin{pmatrix}3136N^7 - 350784N^6 + 12988404N^5 - 184380840N^4 + 1775982591N^3 \\ -28243852173N^2 - 433626201009N + 12575159829261\end{pmatrix}}{9(N-29)^3(2N-87)^3(4N-87)^3} ,$$

which implies monotone decrease in N. Taking the limit, we find it tends to 0 as $N \nearrow \infty$, and at $N = 100$ approximately equals 0.0847. Hence $p_5'(n) \geq 0$, and so $p_5(n)$ increases in n. Evaluating at $n = 29$, we thus find the lower bound

$$p_6(N) := -\frac{116\,(241N^2 - 15399N + 219501)}{9(N-29)(2N-87)(4N-87)} - \frac{58}{3}\log\left(\frac{4N-174}{4N-87}\right) .$$

Differentiating a final time with respect to N, we find the derivative may be written

$$p_6'(N) = -\frac{116\,(160N^4 - 10962N^3 + 227070N^2 - 3951018N + 57289761)}{9(87-4N)^2(87-2N)^2(N-29)^2} < 0 ,$$

implying that $p_6(N)$ decreases in N. Taking the limit, we find $p_6(N) \searrow 0$ as $N \nearrow \infty$, and that at $N = 100$, $p_6(N) \approx 0.738$. We conclude $p_6(N) \geq 0$. This implies $p_5(n) \geq 0$. This implies the claim. □

4.8.3 Restricting $x \in [1/2, 2/3]$

Lemma 49. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $31 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{N-n}{N} g_1\left(D|x = \frac{1}{2}\right) + a_1\left(D|x = \frac{2}{3}\right) + a_2\left(D|x = \frac{2}{3}\right) + \sum_{i=1}^{2} b_i\left(D|x = \frac{1}{2}\right)$$

is positive.

Proof. Differentiating twice, we may write the expression of interest as

$$\frac{9N}{(2N+D)^3} + \frac{n}{N^2D^2(N+D)^2} \cdot p_1(D) \tag{4.70}$$

where we define

$$p_1(D) := -4D^4 - 8D^3N - 4D^2N^2 + 3DN^3 + N^4 .$$

Since $p_1''(D) = -48D^2 - 48DN - 8N^2 < 0$, we see $p_1(D)$ is concave in D. At $D = 20$, $p_1(D)$ equals $N^4 + 60N^3 - 1600N^2 - 64000N - 640000$ and at $D = N/3$ it equals $(98/81)N^4$. Both are positive under the assumption, and so $p_1(D) \geq 0$ by concavity. Hence (4.70) is positive too, and so the first derivative of the expression of interest increases in D. Evaluating the first derivative at $D = N/3$, we find it bounded above by

$$-\frac{49(6\log(4)-7)n+243}{294N} .$$

Since $49(6\log(4)-7) > 64$, we see this upper bound is negative, and so the expression of interest decreases in D. Evaluating the expression at $D = N/3$, we find it equals

$$\frac{4(\log(8)-2)}{9}n - \frac{15}{14} ,$$

which is positive for $n \geq 31$. Hence the claim. □

4.8.4 Restricting $x \in [2/3, 1-N/(n(N-D))]$

Lemma 50. *Let $N \geq 500$, n, $D \in \mathbb{N}$, and $31 \leq n \leq D \leq N/3$. Using the functions in* (3.46)*, we have that the function*

$$\frac{N-n}{N}g_1\left(D|x=\frac{2}{3}\right)+a_1\left(D|x=\frac{2}{3}\right)+a_2\left(D|x=\frac{2}{3}\right)+\sum_{i=1}^{2}b_i\left(D|x=\frac{2}{3}\right)$$

is positive.

Proof. Differentiating twice, we may write the expression of interest as

$$\frac{n\left(3N^2-16D^2\right)}{3D^2N^2}+\frac{Dn(2N-D)+9DN+8nN^2}{D(2N+D)^3} ,$$

which is positive under the assumptions. Evaluating the first derivative at $D = N/3$, we find it bounded above by

$$-\frac{(882\log(7)-1624)n+729}{882N} .$$

This is negative since $882\log(7) - 1624 > 92$, and so the expression decreases in D. At $D = N/3$ the expression of interest equals

$$\frac{2}{27}n(9\log(7)-16) - \frac{15}{14} ,$$

which is positive for $n \geq 10$. The claim follows. □

4.9 ϵ: First Partial Derivatives for $N/3 \le D \le N/2$

Lemma 51. *Let $N \ge 500$, $N/3 \le D \le N/2$, and $N/5 \le n \le D$. Using the functions in* (3.46)*, we have that*

$$\frac{n}{N} g_1(D \mid x = 1/2) + b_3(D \mid x = 1/2) + b_4(D \mid x = 1/2) + a_3(D \mid x = 1/2) + a_4(D \mid x = 1/2)$$

is positive.

Proof. We begin by differentiating the expression of interest twice, and find we may write it as

$$\frac{n}{D^2(N-n)N^2(D(N-n)+N(D-n))^3} \cdot p_1(n) \tag{4.71}$$

where we define

$$p_1(n) := \begin{bmatrix} (N+D)\left(p_2(D)\right)n^4 \\ +N\left(-24D^5 - 48D^4N - 24D^3N^2 + D^2N^2(13N+4) + 10DN^4 + N^5\right)n^3 \\ +2DN^2\left(24D^4 + 24D^3N - DN^2(9N+8) - 3N^4\right)n^2 \\ 4D^2N^3\left(N^2(2N+5) - 8D^3\right)n - 8D^2N^6 \end{bmatrix},$$

and

$$p_2(D) := 4D^4 + 8D^3N + 4D^2N^2 - 3DN^3 - N^4 .$$

We will show $p_1(n) \ge 0$, which in turn will imply (4.71) is positive too. The sign of the fourth derivative of $p_1(n)$ is determined by the sign of $p_2(D)$. Since $p_2''(D) = 48D^2 + 48DN + 8N^2 > 0$, we see $p_2(D)$ is convex in D. Evaluating at $D = N/3$ we find $p_2(D)$ equals $-(98/81)N^4$ and at $D = N/2$ it equals $-(1/4)N^4$. Convexity thus implies $p_2(D) < 0$ and so the fourth derivative of $p_1(n)$ is negative. This implies the third derivative decreases in n. Evaluating the third derivative of $p_1(n)$ at $n = D$ we thus find it bounded below by the function $6 \cdot p_3(D)$, where we define

$$p_3(D) := 16D^6 + 24D^5N - 20D^3N^3 + D^2N^3(4-3N) + 6DN^5 + N^6 .$$

The fourth derivative of $p_3(D)$ is $5760D^2 + 2880DN > 0$, implying the third derivative of $p_3(D)$ increases in D and so is bounded below by $(1000/9)N^3$. This implies $p_3'(D)$ is convex in D. Evaluating at $D = N/3$, we find $p_3'(D)$ equals $-(8/81)(8N-27)N^4$ and when $D = N/2$ we find $p_3'(D)$ equals $-(1/2)(3N-8)N^4$. Since both are negative, convexity implies $p_3'(D) < 0$. Hence $p_3(D)$ decreases in D, and evaluating at $D = N/2$ yields the lower bound $(7/4)N^6$. Hence, $p_3(D) \geq 0$, implying the third derivative of $p_1(n)$ is positive.

We thus see $p_1'(n)$ is convex in n. We next evaluate $p_1'(n)$ at $n = D$ to obtain the function $D^2 \cdot p_4(D)$ where we define

$$p_4(D) := 16D^6 - 24D^5N - 4(D-3)D^2N^3 - 10(D-2)N^5 + (23D-32)DN^4 - N^6 \ .$$

The fourth derivative of $p_4(D)$ is $5760D^2 - 2880DN$, which increases in D and bounded above by 0 at $D = N/2$. So, $p_4''(D)$ is concave under the assumptions. At $D = N/3$, $p_4''(D)$ equals $(2/27)N^3(353N + 324)$ and at $D = N/2$ it equals $4N^3(N+6)$. Since both are positive, concavity implies $p_4''(D) > 0$ which in turn implies that $p_4(D)$ is convex in D. Again evaluating at the D-boundary, we find when $D = N/3$ that $p_4(D)$ equals $-(4/729)(365N - 1944)N^5$ and when $D = N/2$ it equals $-(1/4)N^5(5N-28)$. Since both values are negative, convexity implies $p_4(D) < 0$.

The preceding argument implies $p_1'(n) < 0$ when evaluated $n = D$. Since $p_1'(n)$ is also convex in n we infer $p_1(n)$ is minimized either at $n = N/5$ or $n = D$. We conclude the second derivative argument by showing positivity in both cases. Evaluating first at $n = N/5$ we find $p_1(n)$ equals $(4/625)N^4p_5(D)$, where we define

$$p_5(D) := -729D^5 + 243D^4N + 154D^2N^3 - 9(3D+80)D^2N^2 - 26DN^4 + N^5 \ .$$

The third derivative of $p_5(D)$ decreases in D and so is bounded above by $-3078N^2$ when $D = N/3$. Hence, $p_5'(D)$ is concave in D, and at the D-boundary we find when $D = N/3$ that $p_5'(D)$ equals $(16/3)N^3(11N-90)$. Since this is positive, concavity implies $p_5(D)$ is minimized either at $D = N/3$ or $D = N/2$. At these points we find $p_5(D)$ equals $(4/9)(N^4)(19N-180)$ and $(1/32)(N^4)(497N - 5760)$ respectively. Both begin positive, we infer $p_5(D) \geq 0$ and so $p_1(n)$ is positive when $n = N/5$.

Turning to the upper boundary, we evaluate $p_1(n)$ at $n = D$ and obtain the function $D^2(N-D)^2 \cdot p_6(D)$, where we define

$$p_6(D) := 4D^5 - 4D^4N + (3D-8)N^4 + (4-3D)DN^3 \ .$$

The fourth derivative of $p_6(D)$ is $480D - 96N$, which is positive under the assumptions. Hence, $p_6''(D)$ is convex in D. Checking the D-boundary, we find $p_6''(D)$ equals $-(226/27)N^3$ when $D = N/3$ and $-8N^3$ when $D = N/2$. Thus, $p_6(D)$ is concave in D, and we may now check these points to find $p_6(D)$ equals $(2/243)(N^4)(77N-810)$ when $D = N/3$ and $(1/8)N^4(5N-48)$ when $D = N/2$. Since both are positive, concavity implies $p_6(D) \geq 0$. This in turn implies $p_1(n) \geq 0$ when $n = D$.

Thus $p_1(n) \geq 0$, and so the second derivative of the expression of interest is positive by (4.71). Therefore the first derivative of the expression of interest increases in D. Thus we may evaluate the first derivative of the expression of interest at $D = N/2$ to obtain an upper bound, which we define to be $p_7(n)$. We next differentiate twice to find

$$p_7''(n) = \frac{4}{(2N-n)^2(2N-3n)^4(N-n)^3} \cdot p_8(n) \tag{4.72}$$

where we define

$$p_8(n) := \begin{bmatrix} -3n^6(N-4) - 4n^5N(16N+29) + 4n^4N^2(68N+113) \\ -4n^3N^3(95N+227) + 32n^2N^4(7N+31) - 16nN^5(3N+35) + 128N^6 \end{bmatrix} .$$

The sixth derivative of $p_8(n)$ is $-2160(N-4)$, implying the fifth derivative decreases in n. Evaluating the fifth derivative at $n = N/5$, we thus obtain the upper bound of $-48N(169N+254)$. This implies the fourth derivative decreases in n, and so evaluating the fourth derivative at $n = N/2$ yields the lower bound of $6N^2(403N+828)$. Hence the third derivative increases in n, and at $n = N/2$ is bounded above by $-3N^3(7N+528)$. This implies $p_8'(n)$ is concave in n. Evaluating at the n-boundary, we find when $n = N/5$ that $p_8'(n)$ equals $(2/3125)(N^5)(6541N - 404064)$. and when $n = N/5$ it equals $(1/16)N^5(103N-912)$. Since both are positive, concavity implies $p_8(n)$ increases in n. A final evaluation at $n = N/2$

produces the negative upper bound for $p_8(n)$ of $-(1/64)N^6(35N-468)$. We infer $p_8(n) < 0$ and so $p_7''(n) < 0$ by (4.72).

We have thus shown $p_7'(n)$ decreases in n. Evaluating $p_7'(n)$ at $n = N/5$ we thus obtain the upper bound

$$\frac{1800}{343N^2} + \frac{401}{3528N} - \frac{1}{N}\log\left(\frac{8}{7}\right) - \frac{1}{N}\log\left(\frac{9}{8}\right) .$$

This is bounded above by

$$\frac{5.25}{N^2} - \frac{0.137}{N} < 0 ,$$

with the negativity claim following since we have assumed $N \geq 45$. Thus $p_7'(n) < 0$ under the assumptions, and so $p_7(n)$ decreases in n. We thus evaluate $p_7(n)$ at $n = N/5$ to obtain the upper bound

$$\frac{32}{49N} + \frac{3}{70} - \frac{1}{5}\log\left(\frac{8}{7}\right) - \frac{1}{5}\log\left(\frac{9}{8}\right) ,$$

which is negative for $N \geq 89$. Therefore, $p_7(n) < 0$ and so the first derivative of the expression of interest is negative by the discussion preceding (4.72).

We have thus demonstrated the expression of interest decreases in D. We may thus evaluate it at $D = N/2$ to obtain a lower bound, which we define as $p_9(n)$. We differentiate twice to find

$$p_9''(n) = \frac{N}{(2N-n)^3(2N-3n)^3(N-n)^3} \cdot p_{10}(n) , \tag{4.73}$$

where we define

$$p_{10}(n) := \begin{bmatrix} 3n^6(7N+32) - 8n^5N(13N+54) + 180n^4N^2(N+4) \\ -16n^3N^3(8N+29) + 16n^2N^4(2N-3) + 192nN^5 - 64N^6 \end{bmatrix} .$$

We will now show $p_{10}(n) \geq 0$. First note that the fifth derivative increases in n, and is bounded above at $n = N/2$ by $-120N(41N+144)$. Therefore, the fourth derivative decreases in n, and the third derivative is concave in n.

We now argue in two parts: first we suppose $n \in [N/5, N/3]$ and demonstrate positivity on this set, and then the same for $n \in [N/3, N/2]$. Assuming $n \in [N/5, N/3]$, we see the fourth derivative is bounded below by $40N^2(96+25N) > 0$. Thus the second derivative is convex for $n \in [N/5, N/3]$. Evaluating at the boundary of this interval, we find the second derivative equals $-(22/125)N^4(107N + 2112)$ and $-(14/27)N^4(N41N + 672)$. Convexity thus implies $p''_{10}(n) < 0$ when $n \in [N/5, N/3]$, which in turn implies $p_{10}(n)$ is concave for $n \in [N/5, N/3]$. At the boundary we find $p_{10}(n)$ equals $(N^6/15625)(8001N - 472064)$ and $(1/243)N^6(155N - 3712)$. Both are positive under the assumptions, and so concavity implies $p_{10}(n) > 0$ when $n \in [N/5, N/3]$

Next assuming $n \in [N/3, N/2]$ we recall the third derivative of $p_{10}(n)$ is concave in n. Evaluating at the new boundary points, we find the third derivative equals $(8/3)N^3(196 + 27N)$ and $3N^3(272 + 49N)$ at these points. Concavity implies the third derivative is positive for $n \in [N/3, N/2]$. This implies $p'_{10}(n)$ is convex on this interval. At the boundary, $p'_{10}(n)$ equals $-(2/81)(23N - 3552)N^5$ and $(N^5/16)(41N - 624)$, and so convexity implies $p_{10}(n)$ decreases on this interval. Hence, for $n \in [N/3, N/2]$ we find $p_{10}(n)$ is bounded below by $(N^6/64)(21N - 320)$, which is positive.

We have thus demonstrated that $p_{10}(n) > 0$. By (4.73), we see $p''_9(n) > 0$. Therefore, we may evaluate $p'_9(n)$ at $n = N/5$ to obtain the lower bound

$$-\frac{1600}{3969N} - \frac{247}{2016} + \frac{1}{2}\log\left(\frac{8}{7}\right) + \frac{1}{2}\log\left(\frac{9}{8}\right) .$$

This is positive for $N \geq 128$, implying $p'_9(n) > 0$. Thus we may evaluate $p_9(n)$ at $n = N/5$ and find it bounded below by

$$-\frac{N}{40} + \frac{N}{10}\log\left(\frac{8}{7}\right) + \frac{N}{10}\log\left(\frac{9}{8}\right) - \frac{2}{63} .$$

This is positive for $N \geq 242$, and so $p_9(n) > 0$ under the assumptions. Since $p_9(n)$ was defined as a lower bound of the expression of interest at (4.73), the claim is proved. □

Lemma 52. *Let* $N \geq 500$, n, $D \in \mathbb{N}$, $N/3 \leq D \leq N/2$, $31 \leq n \leq D$, *and* $1/2 \leq x \leq 1 - N/(n(N-D))$. *Using the functions in* (3.46), *we have that*

$$\frac{N-n}{N} g_1(x) + b_1(x) + b_2(x) + a_1(x) + a_2(x)$$

is positive.

Proof. We first differentiate by x to obtain

$$\frac{N-n}{N} g_1'(x) + b_1'(x) + b_2'(x) + a_1'(x) + a_2'(x) \ . \tag{4.74}$$

Combining terms, we may write the expression of interest as

$$\frac{1}{2N^2(1-x)^2(D+(N-D)x)^2} \cdot p_1(x) \tag{4.75}$$

where we define

$$p_1(x) := \begin{bmatrix} -8n(N-D)^4x^4 + 16n(N-D)^3(N-2D)x^3 \\ -2(N-D)^2((5n+1)N^2 - 24Dn(N-D))x^2 \\ +2(2D^2-3DN+N^2)((n+1)N^2 - 8Dn(N-D))x \\ +2D(N-D)((n+1)N^2 - 4Dn(N-D)) - N^4 \end{bmatrix} .$$

We see if we can show $p_1(x) \geq 0$, it will imply (4.74) is positive. Since the fourth derivative of $p_1(x)$ equals $-192n(N-D)^4 < 0$, we see the third derivative decreases in x. Thus we evaluate the third derivative of $p_1(x)$ at $x = 1/2$ to find it bounded above at $-96Dn(N-D)^3$, which is also negative. This implies the first derivative of $p_1(x)$ is concave in x. Evaluating $p_1'(x)$ at $x = 1/2$, we find it equals

$$2D(N-D)(n(N^2-2D^2) - N^2) \ .$$

This is positive since $n(N^2-2D^2) - N^2 \geq (1/2)(n-2)N^2 > 0$. Since $p_1'(x)$ is also concave, we infer that $p_1(x)$ is minimized either at $x = 1/2$ or $x = 1 - (N/(n(N-D))$. Evaluating $p_1(x)$ first at $x = 1 - (N/(n(N-D)))$, we find it equals

$$\frac{n^3 - 8n^2 + 14n - 8}{n^3} N^4 \ ,$$

which is positive under the assumptions. Evaluating $p_1(x)$ at $x = 1/2$, we find it equals

$$\frac{1}{2}\left(D^2(n-1)N^2 - D^4 n - N^4\right) \ .$$

This is positive as well, since the assumptions imply

$$\frac{1}{2}\left(D^2(n-1)N^2 - D^4 n - N^4\right) \geq \frac{1}{2}(30D^2N^2 - 31D^4 - N^4) \geq \frac{158}{81} N^4 \ .$$

By the analysis of the first derivative, we conclude $p_1(x) \geq 0$, which implies (4.74) is positive by (4.75).

Evaluating the expression of interest at $x = 1/2$, we find it equals

$$p_2(n) := \frac{n(N-D)}{N} \log\left(1 + \frac{N}{D}\right) - \frac{2n(N-D)^2}{N^2} - \frac{2D}{N+D} \ . \tag{4.76}$$

Differentiating with respect to n, we find

$$\begin{aligned} p_2'(n) &= \frac{(N-D)}{N} \log\left(1 + \frac{N}{D}\right) - \frac{2(N-D)^2}{N^2} & (4.77) \\ &\geq \frac{(N-D)}{N} \log\left(1 + \frac{N}{D}\right) - \frac{2(N-D)^2}{N^2} - \log\left(1 + \left(\frac{D}{4N}\right)^2\right) \\ &=: p_{3b}(D) \ . & (4.78) \end{aligned}$$

Define the expression (4.77) to be $p_{3a}(D)$. The reason for defining these two functions is that $p_{3a}(D)$ is convex for $D \in [N/3, N/2]$, and in particular is minimized near $D \approx (37/100)N$. However, the root cannot be determined exactly, so we instead show that $p_{3b}(D)$ decreases for $D \in [N/3, (38/100)N]$, that $p_{3a}(D)$ increases for $D \in [(38/100)N, N/2]$, and finally that $p_{3b}(D)$ is positive at $D = (38/100)N$. These three arguments will jointly implies that $p_2'(n) \geq 0$, which will provide a lower bound for (4.76) in n.

Proceeding with the arguments, we first analyze $p_{3b}(D)$ for $D \in [N/3, (38/100)N]$. Differentiating $p_{3b}(D)$ twice we find

$$p_{3b}''(D) = \frac{p_4(D)}{D^2N^2(D+N)^2\left(D^2+16N^2\right)^2} \tag{4.79}$$

where we define

$$p_4(D) := -4D^8 - 8D^7N - 130D^6N^2 - 249D^5N^3 - 1181D^4N^4 - 2016D^3N^5 - 1024D^2N^6 + 768DN^7 + 256N^8 \ .$$

Differentiating twice we see

$$p_4''(D) = -224D^6 - 336D^5N - 3900D^4N^2 - 4980D^3N^3 - 14172D^2N^4 - 12096DN^5 - 2048N^6 < 0 \,,$$

implying $p_4(D)$ is concave in D. Since at $D = N/3$, $p_4(D)$ equals $(2019266/6561)N^8$ and at $D = (38/100)N$, $p_4(D)$ is larger than $262N^8$, we infer $p_4(D) > 0$ by concavity. Hence (4.79) implies $p'_{3b}(D)$ increases in D. Evaluating $p'_{3b}(D)$ at $D = (38/100)N$, we find it is bounded above by

$$\frac{1654361552}{1322831775N} - \frac{1}{N}\log\left(\frac{69}{19}\right) < \frac{1.251}{N} - \frac{1.289}{N} < 0 \ .$$

We thus infer $p_{3b}(D)$ decreases in D for $D \in [N/3, (38/100)N]$.

Next, we consider $p_{3a}(D)$ for $D \in [(38/100)N, N/2]$ Differentiating twice we find

$$p''_{3a}(D) = \frac{N-D}{D(N+D)^2} + \frac{N^2 - 4D^2}{D^2N^2} > 0 \ ,$$

which implies $p'_{3a}(D)$ increases in D. Evaluating $p'_{3a}(D)$ at $D = (38/100)N$, we find it equals

$$\frac{42532}{32775N} - \frac{1}{N}\log\left(\frac{69}{19}\right) > \frac{0.008}{N} > 0 \ .$$

This implies $p_{3a}(D)$ increases in D for $D \in [(38/100)N, N/2]$. We conclude by showing $p_{3b}(D)$ is positive at $D = (38/100)N$. Evaluating there, we find it approximately equals 0.0218, which is positive. Since this holds for any choice on N, we infer the preceding argument implies $p'_2(n) \geq 0$. Hence, $p_2(n)$ increases in n, and so evaluating (4.76) at $n = 31$ we find it equals

$$p_5(D) := \frac{31(N-D)}{N}\log\left(1 + \frac{N}{D}\right) - \frac{62(N-D)^2}{N^2} - \frac{2D}{N+D} \ .$$

Similar to the preceding analysis of $p'_2(n)$, $p_5(D)$ is also convex in D. As one might expect, a similar analysis demonstrates positivity. We define a lower bound of

$$p_6(D) := p_5(D) - \log\left(1 + \frac{D}{N}\right)$$

and show the following: for $D \in [N/3, (39/100)N]$, $p_6(D)$ decreases in D; for $D \in [(39/100)N, N/2]$, $p_5(D)$ increases in D; and at $D = (39/100)N$, $p_6(D)$ is positive. To see the first claim, we differentiate twice to find

$$p''_6(D) = \frac{-124D^5 - 372D^4N - 371D^3N^2 - 26D^2N^3 + 124DN^4 + 31N^5}{D^2N^2(D+N)^3} \ .$$

Differentiating the numerator twice shows it is concave in D. Evaluating the numerator at $D = N/3$ and $D = (39/100)N$, we find it larger than $50N^5$ and $43N^5$. Hence concavity shows $p_6''(D) > 0$, implying $p_6'(D)$ is bounded above at $D = (39/100)N$. Evaluating there we obtain the expression

$$\frac{734729429}{18837975N} - \frac{31}{N}\log\left(\frac{139}{39}\right) < -\frac{0.394}{N} < 0 ,$$

and so infer $p_6(D)$ decreases for $D \in [N/3, (39/100)N]$. Turning to $p_5(D)$, we differentiate twice to find

$$p_5''(D) = \frac{-124D^5 - 372D^4N - 372D^3N^2 - 27D^2N^3 + 124DN^4 + 31N^5}{D^2N^2(N+D)^3} ,$$

which is nearly identical to $p_6''(D)$. The numerator of $p_5''(D)$ is similarly concave and similarly positive at the endpoints (it is larger than $12N^5$ at $D = N/2$), implying $p_5'(D)$ increases in D. Evaluating $p_5'(D)$ at $D = (39/100)N$, we find it equals

$$\frac{748281929}{18837975N} - \frac{31}{N}\log\left(\frac{139}{39}\right) > \frac{0.32}{N} > 0 .$$

From this we see $p_5(D)$ increases for $D \in [(39/100)N, N/2]$. We conclude the argument by evaluating $p_6(D)$ at $D = (39/100)N$ and find it larger than 0.072. This being positive and independent of N implies $p_5(D) \geq 0$. This implies $p_2(n) \geq 0$. By (4.76), this implies the claim. □

4.10 ζ: Second Partial Derivatives for $D \leq N/3$

Lemma 53. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $20 \leq n < D < N/2$, and $0 \leq x \leq 1 - N/(n(N-D))$. Using the functions in* (3.46)*, we have that if $D \leq (9/20)N$, then*

$$\frac{n}{N}g_1(x) + b_3(x) + b_4(x) + a_3(x) + a_4(x)$$

increases in x.

Proof. We begin by differentiating to find

$$\frac{n}{N}g_1''(x) + b_3''(x) + b_4''(x) + a_3''(x) + a_4''(x) = \frac{n^3}{(N-n+nx)^3(D(N-n)-n(N-D)x)^3}p_1(x) \tag{4.80}$$

where we define

$$p_1(x) := \begin{bmatrix} -2n^3(N-n+1)(N-D)^3x^3 \\ -3n^2(N-D)^2(N-n+1)(N-n)(N-2D)x^2 \\ -n(N-D)(N-n)^2(-6D(N-n+1)N+6D^2(N-n+1)+N^2(N-n+3))x \\ +(N-2D)(N-n)^3(D(N-n+1)N-N^2-D^2(N-n+1)) \end{bmatrix} .$$

We see the sign of (4.80) is determined by the sign of $p_1(x)$. The third derivative $-12n^3(N-n+1)(N-D)^3 < 0$, implying the second derivative decreases in x. At $x = 0$, the second derivative equals $-6n^2(N-D)^2(N-n+1)(N-n)(N-2D) < 0$, implying $p_1(x)$ is concave in x. At $x = 0$, we find $p_1(x)$ equals $(N-2D)(N-n)^3(D(N-n+1)N-N^2-D^2(N-n+1))$. This is positive since, by monotone decrease in n, we have

$$D(N-n+1)N - N^2 - D^2(N-n+1) \geq D^3 - D^2(2N+1) + DN(N+1) - N^2 .$$

The lower bound is concave in D, at $D = 2$ equals $4 - 6N + N^2$ and at $D = N/2$ equals $(N^2/8)(N-6)$. Hence, at $x = 0$, $p_1(x) \geq 0$. At $x = 1 - N/(n(N-D))$, we find $p_1(x)$ may be written

$$N^3(N-2D+n-2)((N-n+1)(D-n)(N-D-2)-1) ,$$

which is positive. Hence $p_1(x) \geq 0$ by concavity, which implies (4.80) is positive, and therefore

$$\frac{n}{N}g_1'(x) + b_3'(x) + b_4'(x) + a_3'(x) + a_4'(x) \tag{4.81}$$

increases in x. Evaluating (4.81) at $x = 0$ we find it equals

$$\frac{n^2}{2D^2(N-n)^2N^2}p_2(n) \tag{4.82}$$

where we define

$$p_2(n) := -2D(N-D)(N-2D)^2n + N(-8D^4 + 16ND^3 - 2N(5N+1)D^2 + 2N^2(N-1)D - N^3) .$$

Since $p_2'(n) = -2D(N-D)(N-2D)^2 < 0$, we find a lower bound by evaluating at $n = D$, obtaining the function

$$p_3(D) := 8D^5 - 24D^4N + 26D^3N^2 - 2D^2N^2(6N+1) + 2DN^3(N+1) - N^4 .$$

We next show $p_3(D) \geq 0$ for $D \in [4, (9/20)N]$. The fourth derivative of $p_3(D)$ is $960D - 576N < 0$. Evaluating the third derivative at $D = N/4$, we find it equals $42N^2$, and so $p_3''(D)$ increases for $D \in [4, N/4]$. At $D = N/4$, $p_3''(D)$ equals $-4N^2 - (1/2)N^3 < 0$, implying $p_3(D)$ is concave for $D \in [4, N/4]$. At $D = 4$, $p_3(D)$ equals $7N^4 - 184N^3 + 1632N^2 - 6144N + 8192$ and at $D = N/4$ $p_3(D)$ equals $(N^4/128)(9N - 80)$. Both are positive under the assumptions, so concavity implies $p_3(D) \geq 0$ when $D \in [4, N/4]$.

Next suppose $D \in [N/4, N/3]$. Evaluating the third derivative at $D = N/3$, we find a new lower bound of $(52/3)N^3 > 0$. This implies $p_3'(D)$ is convex when $D \in [N/4, N/3]$. Evaluating $p_3'(D)$ at the end-points of this interval, we find it equals $-(N^3/32)(15N - 32)$ and $-(2/81)(16N - 27)N^3$. Both are negative under the assumptions, which implies $p_3(D)$ decreases for $D \in [N/4, N/3]$. At $D = N/3$, $p_3(D)$ equals $(N^4/243)(8N - 135)$, and so we infer $p_3(D) \geq 0$ when $D \in [N/4, N/3]$.

Finally, suppose $D \in [N/3, (9/20)N]$. Since the fourth derivative is negative, we see $p_3''(D)$ is concave in D. At $D = N/3$, $p_3''(D)$ equals $(4/27)N^2(13N - 27)$ and at $D = N/2$ it equals $2(N-2)N^2$. Hence the first derivative increases under this assumption. At $D = N/2$, the $p_1'(D)$ equals 0. and so $p_3(D)$ decreases for $D \in [N/3, N/2]$. At $D = (9/20)N$, we find $p_3(D)$ equals $(N^4/400000)(1089N - 202000)$, which is positive for $N \geq 500$. Taken together we have demonstrated $p_3(D) \geq 0$ for $D \in [4, (9/20)N]$. Hence $p_2(n) \geq 0$, which implies (4.82) is positive. This implies (4.81), which proves the final claim. □

Lemma 54. *Let $N \geq 500$, n, $D \in \mathbb{N}$, $20 \leq n < D < N/2$, and $0 \leq x \leq 1 - N/(n(N - D))$. Using the functions in* (3.46)*, we have that for $2/3 \leq x \leq 1 - N/(n(N - D))$,*

$$\frac{N - n}{N} g_1(x) + b_1(x) + b_2(x) + a_1(x) + a_2(x)$$

increases in x.

Proof. Differentiating, we find that

$$\frac{N - n}{N} g_1'(x) + b_1'(x) + b_2'(x) + a_1'(x) + a_2'(x)$$

may be written as

$$\frac{p_1(x)}{2N^2(1-x)^2(D+(N-D)x)^2} \tag{4.83}$$

where we define

$$p_1(x) := \begin{bmatrix} -8n(N-D)^4x^4 \\ +16n(N-D)^3(N-2D)x^3 \\ -2(N-D)^2((5n+1)N^2-24Dn(N-D))x^2 \\ +2(2D^2-3DN+N^2)((n+1)N^2-8Dn(N-D))x \\ +2D(N-D)((n+1)N^2-4Dn(N-D))-N^4 \end{bmatrix} .$$

We see it sufficient to show $p_1(x) \geq 0$. The fourth derivative of $p_1(x)$ is $-192n(N-D)^4 < 0$. This implies the third derivative of $p_1(x)$ decreases in x. At $x = 2/3$, the third derivative of $p_1(x)$ equals $-32n(N-D)^3(N+2D) < 0$. By monotone decrease, this implies $p_1'(x)$ is concave in x. First evaluating $p_1'(x)$ at $x = 2/3$, we find it equals

$$\frac{2(N-D)(N+2D)}{27} \cdot [n(3N^2-8D^2+4N(N-2D))-9N^2] =: \frac{2(N-D)(N+2D)}{27} \cdot [p_2(n)] . \tag{4.84}$$

Since $3N^2 - 8D^2 > N^2$ under the assumptions, we see $p_2(n)$ increases in n as defined. Evaluating at $n = 10$ gives the lower bound $-80D^2 - 80DN + 61N^2$ which decreases in D. Evaluating at $D = N/2$ gives the further lower bound of $N^2 > 0$, and so $p_2(n) \geq 0$. Hence at $x = 2/3$, $p_1'(x) \geq 0$. Since $p_1'(x)$ is concave, this implies $p_1(x)$ is minimized either at $x = 2/3$ or $x = 1 - N/(n(N-D))$. Evaluating $p_1(x)$ at $x = 2/3$, we find it may be written as $(1/81)p_3(n)$, where we define

$$p_3(n) := 2(N+2D)^2(N(N-D)+N^2-D^2)n - 9N^2(2D^2+2DN+5N^2) .$$

Inspecting this expression, we see it increases in n under the assumptions. Evaluating at $n = 20$ gives the lower bound

$$p_3(20) = [35N^4 - 160D^4] + DN[262N^2 - 320D^2] + 102D^2N^2 .$$

Using the assumptions $D \leq N/2$, we see each bracketed expression in the display is positive, implying $p_3(20) \geq 0$. Hence, $p_1(x) \geq 0$ when $x = 2/3$. Finally, when $x = 1 - N/(n(N-D))$, $p_1(x)$ may be written as

$$\left(1 - \frac{8}{n^3} + \frac{14}{n^2} - \frac{8}{n}\right) N^4 \ .$$

This is positive when $n \geq 20$, and so $p_1(x) \geq 0$ when $x = 1 - N/(n(N-D))$. By the preceding first derivative argument, we conclude $p_1(x) \geq 0$, which implies the claim. □

4.11 η: Second Partial Derivatives for $N/3 \leq D \leq N/2$

Lemma 55. *Let $N \geq 500$, $N/3 \leq D \leq N/2$, $N/5 \leq n \leq D$, and $1/2 \leq x \leq 1 - N/(n(N-D))$. Using the functions in (3.46), we have that*

$$\frac{n}{N} g_1'(x) + b_3'(x) + b_4'(x) + a_3'(x) + a_4'(x)$$

is positive.

Proof. Combining terms, we may write the expression of interest as

$$\frac{n^2}{2(N-n)N^2(N-n+nx)^2(D(N-n)-n(N-D)x)^2} \cdot p_1(x) \tag{4.85}$$

where we define

$$p_1(x) := \left[\begin{array}{c} -8n^4(N-D)^4x^4 - 16n^3(N-n)(N-2D)(N-D)^3x^3 \\ +2n^2(N-D)^2(N-n)(24D^2n - 24D(D+n)N + (24D+5n-1)N^2 - 5N^3)x^2 \\ +2n(N-n)^2(2D^2 - 3DN + N^2)(8D^2n - 8D(D+n)N + (8D+n-1)N^2 - N^3)x \\ -(N-n)^3(-16D^3(N-n)N - 2D(N-n+1)N^3) \\ -(N-n)^3(8D^4(N-n) + 2D^2N^2(5(N-n)+1) + N^4) \end{array}\right] .$$

We see the claim will follow if we can show $p_1(x) \geq 0$. We will first analyze the derivatives of $p_1(x)$ relaxing the conditions listed in the claim, and supposing $N/3 \leq D \leq N/2$ and $31 \leq n \leq D$. The third derivative of $p_1(x)$ is

$$-192n^4(N-D)^4x - 96n^3(N-n)(N-2D)(N-D)^3 \ ,$$

which is negative. Hence $p_1'(x)$ is concave in x. Evaluating $p_1'(x)$ at $x = 1/2$, we may write it as

$$2n(N-D)(N(N-2D)+Dn)\cdot p_2(n)\ , \tag{4.86}$$

where we define

$$p_2(n) := (N^2-2D^2)n^2 + N(8D^2+N-4DN-2N^2)n + N^2(8D(N-D)-N) - N^4\ .$$

We see the sign of $p_1'(x)$ at $x = 1/2$ is determined by the sign of $p_2(n)$. Since $p_2''(n) = 2(N^2-2D^2) > 0$, we evaluate $p_2'(n)$ at $n = D$ to find it bounded above by

$$-2N(N^2-4D^2) - N^2(2D-1) - 4D^3\ .$$

This is negative since we assume $D \leq N/2$. This implies $p_2(n)$ decreases in n, and so evaluating at $n = D$ we find $p_2(n)$ is bounded below by

$$(N-D)(2D^3-6D^2N+5DN^2-N^2(N+1)) =: (N-D)(p_3(D))\ . \tag{4.87}$$

Since $p_3''(D) = -12(N-D) < 0$, we see $p_3(D)$ is concave in D. Evaluating at the D-boundary, we find when $D = N/3$ that $p_3(D)$ equals $(1/27)N^2(2N-27)$ and when $D = N/2$, $p_3(D)$ equals $(1/4)(N-4)N^2$. Since both expressions are positive, concavity implies $p_3(D) \geq 0$. By (4.87), this implies $p_2(n) \geq 0$. By (4.86), this implies $p_1'(x) \geq 0$ at $x = 1/2$. Since $p_1'(x)$ is also concave, this implies $p_1(x)$ is minimized either at $x = 1/2$ or $x = 1 - N/(n(N-D))$.

We now evaluate $p_1(x)$ at $x = 1/2$ and we find we may write it as $(1/2)p_4(D)$, where we define

$$p_4(D) := \begin{bmatrix} -D^4(2N-n)^4 + 4D^3N^2(2N-n)^3 \\ +D^2N^2(2N-n)^2\left(n^2-2nN+n-N(5N+1)\right) \\ -2DN^4(2N-n)\left(n^2-2nN+n-N(N+1)\right) \\ +N^4\left(n^3-3n^2N-2(n+1)N^3+(n+4)nN^2\right) \end{bmatrix}\ . \tag{4.88}$$

The fourth derivative of $p_4(D)$ is $-24(2N-n)^4 < 0$, which implies the third derivative decreases in D. Evaluating the third derivative at $D = N/2$, we find it is bounded below by $12n(2N-n)^3N > 0$, which implies $p_4'(D)$ is convex in D. Next evaluating $p_4'(D)$ at the D-boundary, we find when $D = N/3$ it equals

$$\frac{2}{27}(2N-n)N^3(N+n)(-7n^2+n(22N-9)+N(9-7N)) \ ,$$

when $D = N/2$ it equals

$$\frac{1}{2}nN^3(2N-n)(-n^2+n(4N-2)-2(N-1)N) \ ,$$

and when $D = n$ it equals

$$2(2N-n)(N-n)^3(-2n^3+6n^2N+N^2-5nN^2+N^3) \ .$$

Since

$$-7n^2+n(22N-9)+N(9-7N) \leq -\frac{2N}{9}(2N-27) < 0 \ ,$$

(with the first inequality following by evaluating at $n = N/3$),

$$-n^2+n(4N-2)-2(N-1)N \leq -\frac{N}{4}(N-4) < 0 \ ,$$

(with the first inequality following by evaluating at $n = N/2$), and

$$-2n^3+6n^2N+N^2-5nN^2+N^3 \leq -\frac{N^2}{27}(2N-27) < 0$$

(with the first inequality following by convexity and evaluating at $n = N/3$), we conclude $p_4'(D) < 0$ by convexity. Hence, $p_4(D)$ decreases in D. Therefore we evaluate $p_4(D)$ at $D = N/2$ to obtain the lower bound $(N^4/16) \cdot p_5(n)$, where we define

$$p_5(n) := 3n^4 - 4(2N-5)n^3 + 4(N-13)Nn^2 + 48nN^2 - 16N^3 \ . \tag{4.89}$$

Since the third derivative of $p_5(n)$ is negative, we find that $p_5''(n)$ is bounded above at $n = N/5$. Evaluating there, we find it equals $-(4/25N(N+500) < 0$, which implies that $p_5(n)$ is concave in n. Evaluating at the n-boundary, we find $p_5(n)$ equals $(1/625)N^3(63N-5200)$ when $n = N/5$, and $p_5(n)$ equals $(1/16)N^3(3N-40)$ when $n = N/2$. Since both expressions are positive, we conclude $p_5(n) \geq 0$, and so $p_4(D) \geq 0$ too.

We finally evaluate $p_1(x)$ at $x = 1 - (N/(n(N-D)))$, where we obtain the function $N^4 \cdot p_6(n)$, and further define

$$p_6(n) := \begin{bmatrix} (-2(N-D)+3)n^3 \\ -(10D^2 + D(22-14N) + N(4N-13) + 12)n^2 \\ +((14D+15)N^2 - 28(D+1)^2 N + 2(D+1)^2(8D+9) - 2N^3)n \\ (2D+1)N^3 - 2(5D+4)(D+1)N^2 + 2(8D+7)(D+1)^2 N - 8(D+1)^4 \end{bmatrix} . \tag{4.90}$$

We conclude by showing $p_6(n) \geq 0$. Combined with the positivity of $p_4(D)$, and the analysis of $p_1'(x)$, this will imply $p_1(x) \geq 0$ and hence the claim. The third derivative of $p_7(n)$ is $-12(N-D)+18 < 0$, implying the second derivative decreases in n. Evaluating $p_7''(n)$ at $n = N/5$, we thus find it bounded above by

$$-\frac{4}{5}\left(25D^2 - D(38N-55) + N(13N-37) + 30\right) =: -\frac{4}{5}(p_7(D)) .$$

Differentiating, we see $p_7'(D) = 55 + 50D - 38N < 0$, and so $p_7(D)$ decreases in D. Evaluating at $D = N/2$ we thus find it bounded below by $(1/4)(120 - 38N + N^2) > 0$. This implies $p_7''(n) < 0$, which in turn implies $p_7(n)$ is concave in n. We first evaluate $p_7(n)$ at $n = N/5$ to find it equals $(8/25) \cdot p_8(D)$ where we further define

$$p_8(D) := (84D+71)N^3 - 10(25D+22)(D+1)N^2 + 25(12D+11)(D+1)^2 N - 125(D+1)^4 - 9N^4 .$$

The third derivative of $p_8(D)$ is $1800N - 3000D - 3000 > 0$, which implies $p_8'(D)$ is convex in D. At $D = N/3$ we find $p_8'(D)$ equals

$$-\frac{2}{27}(4N-15)(4N^2+195N-450)$$

and when $D = N/2$ we find $p_8'(D)$ equals

$$-\frac{1}{2}(N-10)(7N^2+10N-100) .$$

Both are negative under the assumptions, and so $p_8'(D) < 0$ by convexity. Hence, $p_8(D)$ decreases in D, and evaluating at $D = N/2$ produces the lower bound

$$\frac{3}{16}N^4 - \frac{31}{4}N^3 + \frac{35}{2}N^2 + 25N - 125 ,$$

which is positive. Hence, $p_8(D) \geq 0$, implying $p_7(n) \geq 0$ at $n = N/5$. We conclude by evaluating $p_7(n)$ at $n = D$ to obtain the function

$$p_9(D) := -D^3 + D^2(3N - 8) + D\left(-3N^2 + 16N - 14\right) + N^3 - 8N^2 + 14N - 8 \ .$$

Since $p_9'(D) \leq -14 + 8N - (3/4)N^2 < 0$, we find that evaluating $p_9(D)$ at $D = N/2$ yields a lower bound of

$$\frac{N^3}{8} - 2N^2 + 7N - 8 \ .$$

This is positive under the assumptions. Hence, concavity implies $p_7(n) \geq 0$, and the claim follows. $\square$

Chapter 5

CONCLUSION

> I have tried to encourage the reader to think of the computer as a physicist would his laboratory – it may be used to check existing ideas about the construction of the world, or as a tool for discovering new phenomena which then demand new ideas for their explanation.
>
> Tristan Needham [36]

In this chapter, we conclude by considering an implication of the shifted Serfling Gaussian exponential bound, stated in Theorem 6, and proved in chapter 4. When it is combined with the Bernstein type bound of Corollary 1, stated in chapter 1, we are able to demonstrate an improvement to Serfling's inequality when $31 \leq n \leq D \leq N/2$ and $N \geq 500$.

In the first section, we prove that this improvement holds. Then we show how computation demonstrates that the shifted Serfling Gaussian inequality holds for $50 \leq N \leq 500$ and $20 \leq n \leq D \leq N/2$. This in turn implies that Serfling's inequality may be improved for all $N \geq 62$ and $31 \leq n \leq D \leq N/2$. We then discuss the plausibility of a shifted Bernstein Gaussian bound. We conclude with a summary of contributions made in this dissertation.

5.1 Improving Serfling's Inequality

Proposition 7. *Let $N \geq 10$ and $1 \leq D \leq (15/100)N$, and $1 \leq n \leq D$. Suppose $0 \leq t \leq (N-5D)/(5N)$. Then*

$$\exp\left(-\frac{nt^2}{2\left(\frac{D}{N}\left(1-\frac{D}{N}\right)\left(1-\frac{n-1}{N-1}\right)+\frac{t}{3}\right)}\right) \leq \exp\left(-\frac{2nt^2}{1-\frac{n}{N}}\right) \ .$$

Proof. Taking logarithms and performing algebra, we find the claim equivalent to showing

$$\frac{\left((N-2D)^2-N\right)(N-n)}{(N-1)N} - \frac{4N}{3}t \geq 0 \ .$$

This decreases in t so it is enough to show this inequality holds when $t = ((N-5D)/(5N))$. That is,

$$\frac{\left((N-2D)^2-N\right)(N-n)}{(N-1)N} - \frac{4N(N-5D)}{15N} \geq 0 \ .$$

We similarly see this expression decreases in n, and so is bounded below when $n = D$. Hence it is enough to show

$$\frac{\left((N-2D)^2-N\right)(N-D)}{(N-1)N} - \frac{4N(N-5D)}{15N} \geq 0 \ .$$

Combining terms, we see this equivalent to showing

$$\frac{p_1(D)}{15(N-1)N(N-D)} \geq 0 \ ,$$

where we define

$$p_1(D) := 60D^4 - 180D^3N + 195D^2N^2 - 15D^2N - 70DN^3 + 10DN^2 + 11N^4 - 11N^3 \ .$$

We see the sign of this expression is determined by $p_1(D)$. The third derivative of $p_1(D)$ is bounded above by $-864N$ under the assumptions, and so the second derivative decreases in D. At $D = (15/100)N$, the second derivative equals $(3/5)N(407N-50)$, and so we see the first derivative increases in D under the assumptions. At $D = (15/100)N$, the first derivative is bounded above by $-(1/50)(1142N-275)N^2$, and so we see $p_1(D)$ decreases in D under the assumptions. A final evaluation at $D = (15/100)N$ shows that $p_1(D)$ is bounded below by

$$\frac{N^3(34483N - 78700)}{8000} \ ,$$

which is positive. Hence, the claim follows. □

The following bound is due to Pollak [41].

Lemma 56. *(Pollak 1956) Let Φ denote the standard normal CDF, and ϕ the density function. Then for $x > 0$*

$$1 - \Phi(x) < \frac{2\phi(x)}{\sqrt{8/\pi + x^2} + x} \ . \tag{5.1}$$

Proposition 8. *Let $N \geq 500$ and $5 \leq n \leq N/2$. Suppose $Y \sim N(0,1)$. Then for $1/(2n) \leq t \leq 1$,*

$$\sqrt{2}P\left(Y \geq 2\sqrt{\frac{nN}{N-n}}\left(t - \frac{1}{2n}\right)\right) \leq \exp\left(-\frac{2nNt^2}{N-n}\right) \ .$$

Proof. Using Pollak's bound (5.1) on the Gaussian tail, we find that the left-hand side is bounded above by

$$\frac{2}{\sqrt{8 + \frac{4\pi nN}{N-n}\left(t - \frac{1}{2n}\right)^2} + 2\sqrt{\frac{\pi nN}{N-n}}\left(t - \frac{1}{2n}\right)} \exp\left(-\frac{2nNt^2}{N-n} - \frac{2nN}{N-n}\left(\frac{1}{4n^2} - \frac{t}{n}\right)\right) \ .$$

Comparing this to the right-hand side of the claim, we see it is enough to show

$$\frac{2}{\sqrt{8 + \frac{\pi N(2nt-1)^2}{n(N-n)}} + \sqrt{\frac{\pi N(2nt-1)^2}{n(N-n)}}} \exp\left(-\frac{2nN}{N-n}\left(\frac{1}{4n^2} - \frac{t}{n}\right)\right) \leq 1 \ .$$

Taking the logarithm, the claim will follow if

$$p_1(t) := \log(2) - \log\left(\sqrt{8 + \frac{\pi N(2nt-1)^2}{n(N-n)}} + \sqrt{\frac{\pi N(2nt-1)^2}{n(N-n)}}\right) - \frac{2nN}{N-n}\left(\frac{1}{4n^2} - \frac{t}{n}\right) \leq 0 \ .$$

Differentiating twice, we find

$$p_1''(t) = \frac{4\pi^{3/2}nN^2(2nt-1)^2}{(N-n)\sqrt{\frac{N(2nt-1)^2}{n(N-n)}}\sqrt{8 + \frac{\pi N(2nt-1)^2}{n(N-n)}}} \cdot \frac{1}{p_2(t)}$$

where we define

$$p_2(t) := 4\pi n^2 N t^2 - 8n^2 - 4\pi n N t + 8nN + \pi N \ .$$

Inspecting $p_1''(t)$, we see it sign will follow that of $p_2(t)$. Since $p_2''(t) = 8n^2 N\pi$, we see $p_2'(t)$ increases in t. Evaluating at $t = 1/(2n)$ we find $p_2'(t)$ equals 0, and so $p_2(t)$ increases in t under the assumptions. Now evaluating $p_2(t)$ at $t = 0$ we find it equals $8n(N-n) > 0$ and so we conclude $p_2(t) > 0$ under the assumptions. Therefore $p_1''(t) > 0$, which implies $p_1(t)$ is convex in t.

We conclude by examining $p_1(t)$ at the boundary of $[1/(2n), 1]$. First when $t = 1/(2n)$ we find that $p_1(t)$ may be written

$$\frac{N}{2n(N-n)} - \frac{1}{2}\log(2) \ . \tag{5.2}$$

Since under the assumptions we have

$$\frac{N}{2n(N-n)} \leq \frac{N}{4(N-2)} < 0.32 < \frac{1}{2}\log(2) \ ,$$

we see $p_1(t) < 0$ at $t = 1/(2n)$. Finally at $t = 1$ we find that $p_1(t)$ may be written

$$p_3(n) := \log(2) - \log\left(\sqrt{8 + \frac{\pi(2n-1)^2 N}{n(N-n)}} + \sqrt{\frac{\pi(2n-1)^2 N}{n(N-n)}}\right) + \frac{(4n-1)N}{2n(N-n)} \tag{5.3}$$

We first consider the behavior of $p_3(n)$ for $n \in [5, N/8]$. Differentiating we find

$$p_3'(n) = p_4(n) + p_5(n) \tag{5.4}$$

where we define

$$p_4(n) := -\frac{\sqrt{\pi}N\left(n^2(N-4) + 2n - N\right)}{2n^2\sqrt{\frac{(2n-1)^2 N}{n(N-n)}}(N-n)^2\sqrt{\frac{\pi(2n-1)^2 N}{n(N-n)} + 8}}$$

and

$$p_5(n) := \frac{N\left(4n^2 - 2n + N\right)}{2n^2(N-n)^2} - \frac{3\sqrt{\pi}nN\sqrt{\frac{(2n-1)^2 N}{n(N-n)}}}{2(2n-1)^2(N-n)\sqrt{\frac{\pi(2n-1)^2 N}{n(N-n)} + 8}} \ .$$

Inspecting these definitions, we see $p_4(n) \leq 0$ under the assumptions, since the quantities in its fraction are all positive under the assumption and so the leading sign is respected. We now show for $n \in [5, N/8]$ that $p_5(n) \leq 0$. Re-arranging terms and squaring both sides, this is equivalent to showing

$$\frac{N^2\left(4n^2-2n+N\right)^2}{4n^4(n-N)^4} \leq \frac{9\pi(2n-1)^2 nN^3}{4(2n-1)^4(N-n)^3\left(8+\frac{\pi(2n-1)^2 N}{n(N-n)}\right)}.$$

After cross-multiplying and simplification, this is equivalent to showing

$$4N^2(N-n)^2 \cdot p_6(n) \geq 0 \tag{5.5}$$

where

$$p_6(n) := 9\pi n^6 N(N-n)^2 - (2n-1)^2\left(4n^2-2n+N\right)^2\left(\pi(1-2n)^2 N + 8n(N-n)\right)$$

We see the sign of $p_6(n)$ will determine the sign of (5.5). We now show it is positive under the assumptions. We first differentiate to find

$$p_6'(n) = 6\pi n^5 N\left(4n^2-5nN+N^2\right) + 4Np_{7a}(n) + n^3 p_{7b}(n) + n^5 p_{7c}(n) \tag{5.6}$$

where we define

$$p_{7a} := \begin{bmatrix} 12n^2(2(\pi-1)N^2+2(5\pi-7)N+3\pi-4) \\ -2n(2N(3\pi N-4N+5\pi-5)+\pi)+N(2(\pi-1)N+\pi) \end{bmatrix},$$

$$p_{7b} := \begin{bmatrix} 3n^2\left(\pi N^3-16\pi N^2-40(3\pi-4)N+96\right) \\ +320n\left((5\pi-4)N^2+2(5\pi-9)N-4\right) \\ -64\left(\pi N^3+(20\pi-26)N^2+(15\pi-28)N-2\right) \end{bmatrix},$$

and

$$p_{7c} := \begin{bmatrix} n^2(4096-2000\pi N)-32n\left(3\pi N^2-56(3\pi-2)N+224\right) \\ +45\left(\pi N^3-16\pi N^2-40(3\pi-4)N+96\right) \end{bmatrix}.$$

We will show that $p_{7a}(n)$,$p_{7b}(n)$, and $p_{7c}(n)$ are each positive, which will imply $p_6'(n) \geq 0$ since the remaining terms in (5.6) are positive by inspection. We observe that as defined, each p_7 function is a quadratic in n, and so the following analysis is similar for each function.

Beginning with $p_{7a}(n)$ we see the second derivative is positive under the assumptions, implying the first derivative is bounded below by

$$2(57\pi - 56)N^2 + 10(59\pi - 83)N + 179\pi - 240 > 0 \ .$$

Hence $p_{7a}(n)$ increases in n, and evaluating at $n = 5$ gives a positive lower bound of

$$(542\pi - 522)N^2 + (2901\pi - 4100)N + 890\pi - 1200$$

under the assumptions. Therefore, $p_{7a}(n) \geq 0$ under the assumptions.

Turning to $p_{7b}(n)$, the same basic analysis holds: its second derivative is positive under the assumptions, which leads to a lower bound of

$$3\pi N^3 + 16(7\pi - 8)N^2 - 8(12 + 5\pi)N + 160$$

on the first derivative. This is positive, so evaluating at $n = 5$ yields a lower bound of

$$11\pi N^3 + 16(345\pi - 296)N^2 + 8(755\pi - 1876)N + 928$$

on $p_{7b}(n)$, which implies it too is positive.

On the other hand, the second derivative of $p_{7c}(n)$ is

$$-2(2000\pi N - 4096) < 0 \ ,$$

implying it is concave in n. Evaluating at $n = 5$ and $n = N/8$, we find $p_{7c}(n)$ equals

$$45\pi N^3 - 1200\pi N^2 - 28520\pi N - 10720N + 70880$$

and

$$\frac{7\pi N^3}{4} - 48\pi N^2 - 384N^2 - 5400\pi N + 6304N + 4320 \ .$$

Since both are positive for $N \geq 125$, we conclude $p_{7c}(n) \geq 0$ by concavity. Taken together, these arguments imply (5.6) is positive. Hence, $p_6'(n) > 0$ under the assumptions, and so $p_6(n)$ increases in n. Evaluating at $n = 5$, we thus find $p_6(n)$ is bounded below by

$$134064\pi N^3 - 3240N^3 - 2587230\pi N^2 - 567000N^2 - 49628475\pi N - 23328000N + 131220000 \ ,$$

which is positive for $N \geq 50$ and hence under the assumptions. Therefore, $p_6(n) \geq 0$ and so (5.5) implies $p_5(n) \leq 0$ under the assumptions. Since we have also seen that $p_4(n) \leq 0$ too, by (5.4) we see $p_3'(n) < 0$ for $n \in [5, N/8]$. Therefore, $p_3(n)$ decreases in n for $n \in [5, N/8]$, and so evaluating $p_3(n)$ at $n = 5$ yields the upper bound of

$$\frac{19N}{10(N-5)} - \log\left(\sqrt{\frac{81\pi N}{5(N-5)} + 8} + 9\sqrt{\frac{\pi N}{5(N-5)}}\right) + \log(2)\ . \tag{5.7}$$

Differentiating this bound with respect to N, we find it equals

$$\frac{1}{2(N-5)\sqrt{(81\pi N + 40N - 200)N}} \cdot \left[45\sqrt{\pi} - 19\sqrt{\frac{N(81\pi N + 40N - 200)}{(N-5)^2}}\right]\ .$$

The initial fraction is positive for all $N \geq 25$. Analyzing the bracketed term, we find it is monotone increasing to a value of $45\sqrt{\pi} - 19\sqrt{40 + 81\pi}$, which is bounded above -246.2. Hence, this derivative is negative, and we see the bound (5.7) decreases in N. Hence evaluating at $N = 100$ yields an upper bound of -0.0259 and so the assumptions imply $p_3(n) \leq 0$ when $n \in [5, N/8]$.

We now assume $n \in [N/8, N/2]$ and consider how $p_3(n)$ behaves on this interval. By the preceding argument we know at $n = N/8$, that $p_3(n) < 0$. We will now show it is convex on this new interval. We now differentiate twice to find

$$\begin{aligned} p_3''(n) = &\frac{N\left(4n^3 - 3n^2 + 3nN - N^2\right)}{n^3(N-n)^3} \\ &+ \frac{\sqrt{\pi}(2n-1)N \cdot p_8(n)}{2n^3(N-n)^3\sqrt{\frac{(2n-1)^2N}{n(N-n)}}\left(\pi(2n-1)^2N + 8n(N-n)\right)\sqrt{8 + \frac{\pi(2n-1)^2N}{n(N-n)}}} \end{aligned} \tag{5.8}$$

where we define

$$p_8(n) := \left[\begin{array}{c} -16n^4(N-1)(\pi N - 2) + 4n^3N\left(2\pi N^2 - 2(5+\pi)N - 3\pi + 16\right) \\ +2n^2N\left(2(2+\pi)N^2 + (6\pi - 22)N + \pi\right) - 2nN^2(3(\pi-2)N + \pi) + \pi N^3 \end{array}\right]\ .$$

Consider the numerator of the first term of (5.8).

Differentiating shows $4n^3 - 3n^2 + 3nN - N^2$ increases in n, and hence is bounded below when $n = N/8$. At this value the cubic becomes $(1/128)(N - 86)N^2$, and so is positive under the assumptions. Hence the first term in $p_3''(n)$ is positive under the assumptions. It remains to analyze the second term in (5.8).

Examining this fraction, we see its sign is determined by the sign of $p_8(n)$. We now show $p_8(n) \geq 0$ under the assumptions. The fourth derivative of $p_8(n)$ is $-384(N-1)(\pi N - 2)$, implying the third derivative decreases in n. Evaluating the third derivative of $p_8(n)$ at $n = N/8$, we find it bounded above by the negative value $-72N(2N + \pi - 4)$, which implies $p_8'(n)$ is concave in n. Evaluating $p_8'(n)$ at $n = N/8$, we find it equals

$$\frac{N^2}{16}\left(4\pi N^3 + 6(1+2\pi)N^2 + (60-57\pi)N - 24\pi\right) ,$$

which is positive under the assumptions. This positivity combined with the fact that $p_8'(n)$ is concave in n implies that $p_8(n)$ is minimized either at $n = N/8$ or $n = N/2$ under the current assumptions. When $n = N/8$ we find $p_8(n)$ equals

$$\frac{N^3}{256}\left(3\pi N^3 + (14+13\pi)N^2 + (238-150\pi)N + 200\pi\right)$$

and when $n = N/2$ it equals

$$\frac{1}{2}(N-1)N^3(2(\pi-1)N - \pi) .$$

Both values are positive, and so we see $p_8(n) > 0$ under the current assumptions. Therefore by (5.8) we have shown that $p_3''(n) > 0$. This implies $p_3(n)$ is convex for $n \in [N/8, N/2]$. We have already demonstrated that $p_3(n)$ is negative at $n = N/8$. At $n = N/2$, we find $p_3(n)$ equals

$$\frac{2(2N-1)}{N} - \log\left(\sqrt{\frac{4\pi(N-1)^2}{N} + 8} + 2\sqrt{\pi}\sqrt{\frac{(N-1)^2}{N}}\right) + \log(2) . \tag{5.9}$$

Differentiating this expression with respect to N, we find it equals

$$\frac{4\sqrt{N+\frac{1}{N}-2}\sqrt{\pi\left(N+\frac{1}{N}-2\right)+2} - \sqrt{\pi}\left(N^2-1\right)}{2N^2\sqrt{N+\frac{1}{N}-2}\sqrt{\pi\left(N+\frac{1}{N}-2\right)+2}} .$$

The denominator is negative by inspection. After some algebra, the problem of showing the numerator is negative under the assumption is equivalent to showing that

$$\frac{(N-1)^2}{N^2}\left(\pi N^4 + 2\pi N^3 - 15\pi N^2 + 32(\pi-1)N - 16\pi\right)$$

is positive under the assumptions. Positivity here follows by analyzing the quartic in N. Hence, the expression at (5.9) decreases in N. For $N \geq 245$ it is found to be negative, and hence is negative under the assumptions. Therefore $p_3(n) < 0$ for $n \in [N/8, N/2]$. Combined with (5.7), we have demonstrated $p_3(n) < 0$ for $n \in [5, N/2]$. Therefore, (5.2), (5.3), and the convexity of $p_1(t)$ imply $p_1(t) < 0$ under the assumptions, proving the claim. □

Theorem 8. *Suppose* $H \equiv H_{n,D,N} \sim$ *Hypergeometric*(n, D, N) *with* $N \geq 500$, $31 \leq n \leq D \leq N/2$. *Let* $\bar{H}_n = H/n$ *and* $\mu = D/N$. *Then for all* $t \geq 0$ *we have*

$$P(\bar{H}_n - \mu \geq t) \leq \exp\left(-\frac{2nt^2}{1-\frac{n}{N}}\right)$$

Proof. Suppose first that $t = 0$. Then claim then follows since the right-hand side equals 1.

Next suppose $t > 0$. Define

$$t = \frac{\lambda}{n}$$

with $j > 0$. The left-hand side of the claim becomes

$$P(\bar{H}_n - \mu \geq t) = P\left(H_{n,D,N} \geq \frac{nD}{N} + nt\right) = P\left(H_{n,D,N} \geq \frac{nD}{N} + \lambda\right) \ .$$

Since the assumptions imply $(nD)/N \in \mathbb{N}$, it is enough to show the bound holds for $\lambda \in \mathbb{N}$. That is, we must control

$$\sum_{\lambda=p}^{n-\frac{nD}{N}} P\left(H_{n,D,N} = \frac{nD}{N} + \lambda\right)$$

for $p \in \{1, 2, \ldots, n-(nD)/N\}$. For notational convenience define

$$j = \frac{N}{N-D}\lambda \ .$$

We then further re-write the quantity as

$$\sum_{\frac{(N-D)}{N}j=p}^{n-\frac{nD}{N}} P\left(H_{n,D,N} = \frac{nD+(N-D)j}{N}\right)$$

First suppose that $(15/100)N \leq D \leq N/2$. Under these assumptions, we may use Theorem 6 to find

$$\sum_{\frac{(N-D)}{N}j=p}^{n-\frac{nD}{N}} P\left(H_{n,D,N} = \frac{nD+(N-D)j}{N}\right) \leq \sqrt{2}P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}}\left(\frac{p}{n}-\frac{1}{2n}\right)\right) .$$

On the original t-scale, this may be summarized as follows: for $1/(2n) \leq t$,

$$P\left(\bar{H}_n - \mu \geq t\right) \leq \sqrt{2}P\left(Y \geq 2\sqrt{\frac{nN}{N-n}}\left(t-\frac{1}{2n}\right)\right) .$$

But by Proposition 8,

$$\sqrt{2}P\left(Y \geq 2\sqrt{\frac{nN}{N-n}}\left(t-\frac{1}{2n}\right)\right) \leq \exp\left(-\frac{2nNt^2}{N-n}\right) ,$$

and so the claim holds when $(15/100)N \leq D \leq N/2$. For $31 \leq D \leq (15/100)N$, the claim follows by first noting the Bernstein-type corollary for the hypergeometric (Corollary 1 in chapter 1) and Proposition 7 imply the claim for $0 \leq t \leq (N-5D)/(5N)$. Then for $t \geq (N-5D)/(5N)$ we may use a similar argument to the one presented, but with Theorem 7 justifying the Gaussian inequality. This gives the claim. □

5.2 *Implications of Computation*

The restriction that $N \geq 500$ may be improved by computation. The j-lattice (5.10) of Theorem 6 contains a finite number of points for any fixed $N \in \mathbb{N}$. Hence, for $50 \leq N \leq 500$, we may check that the bound (5.11) holds for $21 \leq n \leq D \leq N/2$ by computer.

The program to make this check is recorded in the appendix. Combining this check with Theorems 6 and 10 gives the following theorems.

Theorem 9. *Let $Y \sim N(0,1)$, $N \geq 62$ and $(15/100)N \leq D \leq N/2$, and $31 \leq n \leq D$. Let $H_{n,D,N} \sim$ Hypergeometric(n,D,N). Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Then for

$$j \in \left\{ \frac{(k^* + p)N - nD}{N - D} \;\middle|\; 0 \leq p \leq n - k^*,\ p \in \mathbb{N} \right\} , \tag{5.10}$$

we have

$$\begin{aligned} &P\left(H_{n,D,N} - n\frac{D}{N} \geq \frac{(N-D)}{N} j\right) \\ &\leq \sqrt{2} \cdot P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}} \left(\frac{N-D}{N}\right)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right)\right) . \end{aligned} \tag{5.11}$$

Theorem 10. *Let* $Y \sim N(0,1)$*,* $N \geq 207$ *and* $31 \leq D \leq (15/100)N$*, and* $31 \leq n \leq D$*. Let* $H_{n,D,N} \sim$ *Hypergeometric*(n, D, N)*. Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Then for

$$j \in \left\{ \frac{(k^* + p)N - nD}{N - D} \;\middle|\; q^* \leq p \leq n - k^* - 1, q^* = \frac{n}{5} - k^*,\ p \in \mathbb{N} \right\} , \tag{5.12}$$

we have

$$\begin{aligned} &P\left(H_{n,D,N} - n\frac{D}{N} \geq \frac{(N-D)}{N} j\right) \\ &\leq \sqrt{2} \cdot P\left(Y \geq 2\sqrt{\frac{Nn}{N-n}} \left(\frac{N-D}{N}\right)\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right)\right) . \end{aligned} \tag{5.13}$$

We may similarly extend the improvement to Serfling's inequality to $N = 62$ through computation.

5.3 Bernstein Type Gaussian Bounds

The shifted Serfling Gaussian bound is designed to hold for deviations from the mean of moderate sizes over all choice of parameter values. If we wish to incorporate information about variance into the bounds, we might use the hypergeometric bounds of chapter 1 as a starting point. In particular, we may design Gaussian bounds that recover the exponential term through bounds on the Gaussian tail, such as Pollak's bound of (5.1). Using Corollary 1 from chapter 1 as inspiration, we conjecture that the following bound holds.

Conjecture 1. *Let* $Y \sim N(0,1)$*,* $N \geq 500$ *and* $(15/100)N \leq D \leq N/2$*, and* $31 \leq n \leq D$*. Let* $H_{n,D,N} \sim$ *Hypergeometric*(n, D, N)*. Define*

$$k^* := \left\lceil \frac{nD}{N} \right\rceil .$$

Additionally, let $f_{n,D,N}(x)$ *be as defined ahead in Lemma 57. Then for*

$$j \in \left\{ \frac{(k^*+p)N - nD}{N-D} \;\middle|\; 0 \leq p \leq n-k^*,\ p \in \mathbb{N} \right\} , \tag{5.14}$$

we have

$$P\left(H_{n,D,N} - n\frac{D}{N} \geq \frac{(N-D)}{N} j\right) \leq \left(1 + \frac{N}{2D}\right) \cdot P\left(Y \geq f_{n,D,N}\left(\frac{j}{n} - \frac{N}{2n(N-D)}\right)\right) . \tag{5.15}$$

This conjecture is designed to address the coverage deficiency that emerged in the α region of the proof of the shifted Serfling Gaussian bound. Here, we explicitly include the term $1 + N/(2D)$, which is large when $D \lll N$ (the α region). We may use this inflation factor to control the variance term that emerges when $x = j/n$ is small.

This conjecture features the function $f_{n,D,N}(x)$. We will show how we believe this function should be handled in trying to prove this conjecture the next few Lemmas.

Lemma 57. *Let* $N \geq 10$*,* n*,* $D \in \mathbb{N}$*,* $2 \leq n \leq D \leq N/2$*, and* $x \geq 0$*. Then the function*

$$f_{n,D,N}(x) \equiv f(x) := \sqrt{\frac{3nN^3}{3D(N-D)(N-n) + N^2(N-D)x}} \frac{(N-D)}{N} x \equiv \sqrt{\frac{3nN(N-D)x^2}{3D(N-n) + N^2 x}}$$

is positive, increasing in x*, and concave.*

Proof. By definition we see $f(0) = 0$. Differentiating once, we find

$$f'(x) = \sqrt{\frac{3}{4}} \sqrt{\frac{nN(N-D)}{(3D(N-n) + N^2 x)}} \left[\frac{6D(N-n) + N^2 x}{3D(N-n) + N^2 x}\right] ,$$

which is positive for $x \geq 0$ under the assumptions. Differentiating twice, we see

$$f''(x) = -N^2 \sqrt{\frac{3}{16}} \left[\frac{12D(N-n) + N^2 x}{(3D(N-n) + N^2 x)^2}\right] \sqrt{\frac{nN(N-D)}{3D(N-n) + N^2 x}} ,$$

which is negative by inspection under the assumptions. The claims follow. $\square$

The next Lemmas will use the properties of the function presented in Lemma 57.

Lemma 58. *Let* $N \geq 20$, n, $D \in \mathbb{N}$, $4 \leq n \leq D \leq N/2$, *and* $N/(2n(N-D)) \leq x \leq 1-N/(n(N-D))$. *Let* $f_{n,D,N}(x) \equiv f(x)$ *be as defined in Lemma 57. Under these conditions, we have*

$$f\left(x+\frac{N}{2n(N-D)}\right)-f\left(x-\frac{N}{2n(N-D)}\right) \geq \frac{N}{n(N-D)}f'\left(x+\frac{N}{2n(N-D)}\right) .$$

Proof. By the mean value theorem, we have for some $\zeta \in (x-N/[2n(N-D)], x+N/[2n(N-D)])$ that

$$f\left(x+\frac{N}{2n(N-D)}\right) = f\left(x-\frac{N}{2n(N-D)}\right)+\frac{N}{n(N-D)}f'(\zeta) .$$

By Lemma 57, we have $f'(\zeta) \geq f'(x+N/[2n(N-D)])$, which gives the claim. □

Lemma 59. *Let* $N \geq 500$, n, $D \in \mathbb{N}$, $4 \leq n \leq D \leq N/2$. *Let* $N/(n(N-D)) \leq x \leq 1-N/(n(N-D))$. *Let* $Y \sim$ *Normal*$(0,1)$. *For any* (n,D,N) *under these assumptions, let* $f_{n,D,N}(x)$ *be as in Lemma 57. Then*

$$\begin{aligned}
&\left(1+\frac{N}{2D}\right)\cdot P\left(f_{n,D,N}\left(x-\frac{N}{2n(N-D)}\right) \leq Y \leq f_{n,D,N}\left(x+\frac{N}{2n(N-D)}\right)\right)\\
\geq\ &\left(1+\frac{N}{2D}\right)\sqrt{\frac{3}{2}}\sqrt{\frac{N^3}{2\pi}}\frac{V_1(x)}{(V_2(x))^{3/2}}\exp\left(H_1(x)+H_2(x)\right)
\end{aligned}$$

where we define

$$\begin{aligned}
V_1(x) &:= 12nD(N-D)(N-n)+N^3+2N^2n(N-D)x\ ,\\
V_2(x) &:= 6nD(N-D)(N-n)+N^3+2N^2n(N-D)x\ ,\\
H_1(x) &:= -\frac{3N(N-2n(N-D)x)^2}{8\left(6Dn(N-D)(N-n)-N^3+2nN^2(N-D)x\right)}\ ,\\
\text{and}\quad H_2(x) &:= -\frac{3N(N+2n(N-D)x)^2}{8\left(6Dn(N-D)(N-n)+N^3+2nN^2(N-D)x\right)}\ .
\end{aligned}$$

Proof. Identifying

$$a = f_{n,D,N}\left(x-\frac{N}{2n(N-D)}\right)$$

and

$$b = f_{n,D,N}\left(x + \frac{N}{2n(N-D)}\right) ,$$

we see by Lemma 57 that $b > a$. Furthermore under these conditions Lemma 58 gives

$$\frac{1}{\sqrt{2\pi}}(b-a) \geq \sqrt{\frac{3}{2}}\sqrt{\frac{N^3}{2\pi}}\left(\frac{12nD(N-D)(N-n) + N^3 + 2nN^2(N-D)x}{(6nD(N-D)(N-n) + N^3 + 2nN^2(N-D)x)^{3/2}}\right) .$$

Additionally, we may write

$$\begin{aligned} -\frac{b^2}{4} - \frac{a^2}{4} &= -\frac{3N(N+2n(N-D)x)^2}{8\left(6Dn(N-D)(N-n) + N^3 + 2nN^2(N-D)x\right)} \\ &\quad - \frac{3N(N-2n(N-D)x)^2}{8\left(6Dn(N-D)(N-n) - N^3 + 2nN^2(N-D)x\right)} . \end{aligned}$$

Using these expressions, along with the bound $\gamma(a,b) \geq 1$, we apply Lemma 16 to prove the claim. □

Lemma 59 contains the lower bound on Gaussian intervals that need to be compared to the hypergeometric quantities that appear in Lemma 14 in chapter 3. At first appearance, this may seem to be intractable, since the “continuity correction” here does not eliminate the linear term: there are two exponential quantities $H_1(x)$ and $H_2(x)$. However, after some initial exploration, these quantities seem to be tractable at the level of the second partial derivative with respect to x, when weighted according to the coefficients of the $B_i(x)$ functions (with $i \in \{1,2,3,4\}$). Moreover, the “continuity correction” seems to introduce monotonicity and convexity for $x \in [N/(n(N-D)), 1 - N/(n(N-D))]$. The main difficulties in the argument emerge at the level of the first partial derivative with respect to x, since obtaining control of the n and D parameters appears to require complicated analysis.

Note that Lemma 59 contains two functions $V_1(x)$ and $V_2(x)$. These expressions are analogous to the term

$$\sqrt{\frac{4N}{\pi n(N-n)}}$$

that appears in Lemma 18. While $V_1(x)$ and $V_2(x)$ are more complicated than this constant term, they (along with the hypergeometric constant term) are made positive by the term

$1 + N/(2D)$ on the log-scale when $x \in [N/(n(N-D)), 1 - N/(n(N-D))]$. Therefore, preliminary analysis supports the validity of Conjecture 1; smoothing out the attendant technical details remains.

5.4 Concluding Remarks

This dissertation began by discussing the general problem of sampling without replacement from a bounded finite population. It argued that the hypergeometric distribution deserves special consideration in this problem because of its dominant position with respect to the convex order of Theorem 1. After reviewing existing bounds for the hypergeometric distribution, in chapter 2 we derived hypergeometric analogues of Talagrand's and León and Perron's exponential bounds for the binomial distribution.

In chapter 3, we discussed an alternative approach to deriving bounds for discrete distributions. We called this approach the de Moivre-Tusnády method. Using this method, we derived bounds for the Poisson tail, and then laid the ground work for obtaining bounds for the hypergeometric distribution. In chapter 4, we carried out the comparison prepared in chapter 3. This comparison yields the shifted Serfling Gaussian bounds, summarized in theorems 6 and 7.

Finally in this chapter, we showed how the shifted Serfling Gaussian bounds imply an improvement to Serfling's inequality in the case of the hypergeometric distribution. We also showed how improvements to the bounds can be extended via computer checking. We conjectured that Gaussian bounds which incorporate the variance are also feasible. In the future, it is natural to extend the shifted Serfling Gaussian bound to the remaining combinations of parameters for the hypergeometric distribution. The de Moivre-Tusnády method also seems useful in analyzing special cases of multivariate distributions, such as the multinomial and multivariate hypergeometric.

BIBLIOGRAPHY

[1] Micah Adler, Eran Halperin, Richard M. Karp, and Vijay V. Vazirani. A stochastic process on the hypercube with applications to peer-to-peer networks. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 575–584. ACM, 2003.

[2] Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

[3] George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.

[4] Alexandra Boldyreva, Nathan Chenette, Younho Lee, and Adam Oneill. Order-preserving symmetric encryption. In *Advances in Cryptology-EUROCRYPT 2009*, pages 224–241. Springer, 2009.

[5] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[6] Jean Bretagnolle and Pascal Massart. Hungarian constructions from the nonasymptotic viewpoint. *The Annals of Probability*, 17(1):239–256, 1989.

[7] Nathalie Castelle and Françoise Laurent-Bonvalot. Strong approximations of bivariate uniform empirical processes. In *Annales de l'IHP Probabilités et statistiques*, volume 34, pages 425–480, 1998.

[8] Sourav Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007.

[9] Richard M. Dudley. The Gaussian process and how to approach it. In *Proceedings of the International Congress of Mathematicians*, pages 143–146, 1974.

[10] Richard M. Dudley. *Uniform Central Limit Theorems*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2nd edition, 2014.

[11] Werner Ehm. Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.*, 11(1):7–16, 1991.

[12] Paul Erdös and Alfréd Rényi. On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci*, 4:49–61, 1959.

[13] William Feller. *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley and Sons, Inc., third edition, 1968.

[14] William Feller. *An Introduction to Probability Theory and Its Applications. Volume II.* John Wiley & Sons London-New York-Sydney-Toronto, 1971.

[15] David A. Freedman. *Statistical models: theory and practice.* Cambridge University Press, 2009.

[16] Peter W. Glynn. Upper bounds on Poisson tail probabilities. *Operations Research Letters*, 6(1):9–14, 1987.

[17] Larry Goldstein and Ümit Işlak. Concentration inequalities via zero bias couplings. *Statist. Probab. Lett.*, 86:17–23, 2014.

[18] Evan Greene and Jon A. Wellner. Finite sampling inequalities: an application to two-sample Kolmogorov-Smirnov statistics. *arXiv preprint arXiv:1502.00342*, 2015.

[19] Evan Greene and Jon A. Wellner. Exponential bounds for the hypergeometric distribution. *arXiv preprint arXiv:1507.08298*, 2015b.

[20] Jaroslav Hájek. Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematics Institute of the Hungarian Academy of Science*, 5(361):74, 1960.

[21] Jaroslav Hájek, Zbyněk Šidák, and Pranab K. Sen. *Theory of Rank Tests.* Probability and Mathematical Statistics. Academic Press, Inc., San Diego, CA, second edition, 1999.

[22] A. Bruce Hoadley. On the Probability of Large Deviations of Functions of Several Empirical CDFs. *The Annals of Mathematical Statitics*, 38(2), April 1967.

[23] Wassily Hoeffding. A combinatorial central limit theorem. *Ann. Math. Statistics*, 22:558–566, 1951.

[24] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):pp. 13–30, March 1963.

[25] Susan Holmes. Stein's method for birth and death chains. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 45–67. Inst. Math. Statist., Beachwood, OH, 2004.

[26] Don Hush and Clint Scovel. Concentration of the hypergeometric distribution. *Statist. Probab. Lett.*, 75(2):127–132, 2005.

[27] Samuel Karlin. Inequalities for symmetric sampling plans I. *The Annals of Statistics*, 2(6):1065–1094, 1974.

[28] J. H. B. Kemperman. Moment problems for sampling without replacement. I, II, III. *Indagationes Mathematicae (Proceedings)*, 76(3):149–188, 1973.

[29] J. Kiefer. On large deviations of the empiric D. F. of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.*, 11:649–660, 1961.

[30] János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent rv'-s, and the sample df. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1-2):111–131, 1975.

[31] Carlos A. León and François Perron. Extremal properties of sums of Bernoulli random variables. *Statistics & probability letters*, 62(4):345–354, 2003.

[32] J. E. Littlewood. On the probability in the tail of a binomial distribution. *Advances in Applied Probability*, 1(1):43–72, 1969.

[33] Albert W. Marshall, Ingram Olkin, and Barry Arnold. *Inequalities: Theory of Majorization and its Applications*. Springer Science & Business Media, 2010.

[34] Brendan D. McKay. On Littlewood's estimate for the binomial distribution. *Advances in Applied Probability*, pages 475–478, 1989.

[35] T. S. Nanjundiah. Note on Stirling's formula. *The American Mathematical Monthly*, 66(8):701–703, 1959.

[36] Tristan Needham. *Visual Complex Analysis*. Oxford University Press, 1998.

[37] Thomas K. Philips and Randolph Nelson. The moment bound is tighter than Chernoff's bound for positive tail probabilities. *The American Statistician*, 49(2):175–178, 1995.

[38] I. S. Pinelis and S. A. Utev. Exact exponential bounds for sums of independent random variables. *Theory of Probability and Its Applications*, 34(2):pp. 340–346, March 1990.

[39] Iosef Pinelis. On the Bennett-Hoeffding inequality. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(1), 2014.

[40] Jim Pitman. Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Combin. Theory Ser. A*, 77(2):279–303, 1997.

[41] H. O. Pollak. A remark on "Elementary inequalities for Mills' ratio" by Yûsaku Komatu. *Rep. Stat. Appl. Res. UJSE*, 4:110, 1956.

[42] Herbert Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 62:26–29, 1955.

[43] Angelika Rohde. Optimal calibration for multiple testing against local inhomogeneity in higher dimension. *Probability Theory and Related Fields*, 149(3-4):515–559, 2011.

[44] Jason Schweinsberg. Coalescent processes obtained from supercritical Galton–Watson processes. *Stochastic Processes and their Applications*, 106(1):107–139, 2003.

[45] R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *Ann. Statist.*, 2:39–48, 1974.

[46] Galen R. Shorack and Jon A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[47] George F. Simmons. *Differential Equations: with Applications and Historical Notes*, volume 452. McGraw-Hill New York, 1972.

[48] M. Stone. Extreme tail probabilities for the null distribution of the two-sample Wilcoxon statistic. *Biometrika*, 54(3/4), 1967.

[49] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1):28–76, 1994.

[50] Michel Talagrand. The missing factor in Hoeffding's inequalities. In *Annales de l'IHP Probabilités et statistiques*, volume 31, pages 689–702, 1995.

[51] Henry Teicher. An inequality on Poisson probabilities. *The Annals of Mathematical Statistics*, 26(1):147–149, 1955.

[52] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[53] Vladimir N. Vapnik. *Statistical Learning Theory*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, Inc., New York, 1998. A Wiley-Interscience Publication.

[54] V. A. Vatutin and V. G. Mikhaĭlov. Limit theorems for the number of empty cells in an equiprobable scheme for the distribution of particles by groups. *Teor. Veroyatnost. i Primenen.*, 27(4):684–692, 1982.

Appendix A
COMPUTER CHECKS

The following Mathematica 10.4.1.0 program is used to verify the shifted Serfling Gaussian inequality holds for $N \in \{50, 51, 52, \ldots, 500\}$ and $21 \leq n \leq D \leq N/2$. The platform specification: Mac OS X x86 (32-bit, 64-bit Kernel).

```
ks[n_, DD_, NN_] := Ceiling[(n DD)/NN]


HGCDF[n_, DD_, NN_, j_] :=
 1 - CDF[HypergeometricDistribution[n, DD, NN], j]


HGTail[n_, DD_, NN_] :=
 Table[HGCDF[n, DD,
   NN, (((n DD) + (NN -
         DD) (((ks[n, DD, NN] + p) NN - n DD)/(NN - DD)))/NN) -
    1], {p, 0, n - ks[n, DD, NN]}]


UpperBound[n_, DD_, NN_, j_] :=
 Sqrt[2] (1 -
    CDF[NormalDistribution[0, 1],
     2 Sqrt[(n NN)/(NN - n)] ((NN - DD)/
        NN) ((j/n) - (NN/(2 n (NN - DD))))])


UBTail[n_, DD_, NN_] :=
 Table[UpperBound[n, DD,
```

```
   NN, (((ks[n, DD, NN] + p) NN - n DD)/(NN - DD))], {p, 0,
   n - ks[n, DD, NN]}]

(*
search from N = NMIN to N = NMAX for violations of the conjectured \
inequality.
$MaxExtraPrecision = \[Infinity] allocates the maximum available \
resources to the bound
comparison. In the event of a violation, the population parameters \
are printed
in the list {"Violation at:" {n , D , N}}. To keep track of the \
search, as N passes
multiples of 10 the list {"New N Block" NN} is printed.
cmin is comparison value and so is exactly zero.
*)

cmin := 0
NMIN := 50
NMAX := 500

Block[{$MaxExtraPrecision = \[Infinity]},
 Do[If[Min[UBTail[n, DD, NN] - HGTail[n, DD, NN]] < cmin,
   Print[{"Violation at:", n, DD, NN}],
   If[And[Mod[DD, 10] == 0, n == 20], Print[{"New D Block", DD}],
    If[And[DD == 21, n == 20], Print[{"New N", NN}]]]], {NN, NMIN,
   NMAX}, {DD, 21, Floor[NN/2]}, {n, 20, DD}]]
```

VITA

Evan Greene grew up in Salinas, California. As an undergraduate, he attended the University of California, Berkeley, where he obtained the degree of Bachelor of Arts with majors in Applied Mathematics and English. In 2016, he obtained a Ph.D. in Statistics from the University of Washington.