APPROXIMATION TO THE MEAN CURVE IN THE LCS PROBLEM

CLEMENT DURRINGER[§,*], RAPHAEL HAUSER[†*] AND HEINRICH MATZINGER [‡*]

Abstract. The problem of sequence comparison via optimal alignments occurs naturally in many areas of applications. The simplest such technique is based on evaluating a score given by the length of a longest common subsequence divided by the average length of the original sequences. In this paper we investigate the expected value of this score when the input sequences are random and their length tends to infinity. The corresponding limit exists but is not known precisely. We derive a large-deviation, convex analysis and Montecarlo based method to compute a consistent sequence of upper bounds on the unknown limit.

AMS subject classifications. Primary 05A16, 62F10. Secondary 92E10.

Key words. Longest common subsequence problem, Chvátal-Sankoff constant, Steele conjecture, mean curve, large deviation theory, Montecarlo simulation, convex analysis.

1. Introduction.

A naive way of quantifying the similarity of two finite strings is to compare them character by character and count the number of matching symbols. For example, the strings T_1 = 'nebbiolo' and T_2 = 'nbbiolo' have only two common characters under this method of comparison and would not be judged very similar. The design of more meaningful methods of comparing strings often depends on specific applications such genetic sequence analysis, speech recognition or file comparison.

A family of sophisticated methods is based on looking at subsequences of the original strings and their likely descendants under a mutation process based on a hidden Markov model. The simplest method of this kind is to measure the similarity between strings by the relative length of a longest common subsequence (LCS) with

[§]Laboratoire de Statistique et Probabilités; Université Paul Sabatier; 118, Route de Narbonne; 31062 Toulouse Cedex 4 France; (durringe@cict.fr).

[*]All three authors would like to thank for support through the Sonderforschungsbereich Grant SFB 701 A3 from the German Research Foundation.

[†]Oxford University Computing Laboratory; Wolfson Building; Parks Road; Oxford, OX1 3QD; United Kingdom; (hauser@comlab.ox.ac.uk). This author was supported through grant GR/S34472 from the Engineering and Physical Sciences Research Council of the UK and through grant NAL/00720/G from the Nuffield Foundation.

[‡]Fakultät für Mathematik; Universität Bielefeld; 33501 Bielefeld; Germany; (matzing@mathematik.uni-bielefeld.de). Also: School of Mathematics; Georgia Institute of Technology; 686 Cherry Street; Atlanta GA 30332-0160; USA; (matzi@math.gatech.edu).

respect to the mean of the lengths of the original sequences. A subsequence of a string is defined as any string obtained by deleting some of the characters from the original string and keeping the remaining characters in the same order. For example, the strings T_1 and T_2 defined above contain 'nbbiolo' as a longest common subsequence and have a similarity score of 93.33% under this method of comparison.

In order to use string comparison algorithms for the detection of sequences that are closely related to one another, one must be able to tell when a similarity score is significantly higher than scores that are likely to occur when comparing random strings. The study of the distribution of scores under random input strings is therefore of great interest.

In this paper we study the asymptotic expectation of the LCS method. To fix ideas, let $(X_n^s)_{\mathbb{N}}$ $(s = 1, 2)$ be two independent sequences of i.i.d. standard Bernoulli random variables. For all $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}_+^2$ such that $\lambda_1 + \lambda_2 = 1$ let $L_n(\lambda)$ denote the length of a longest common subsequence of the two strings $(X_1^s, X_2^s, \dots, X_{\lfloor 2\lambda_s n \rfloor}^s)$ $(s = 1, 2)$. Note that $L_n(\lambda)$ is well-defined, but there may be more than one common subsequence of this length. A simple subadditivity argument shows that the limit $\gamma(\lambda) := \lim_{n\to\infty} \mathrm{E}[L_n(\lambda)/n]$ exists. The function $\lambda \to \gamma(\lambda)$ is called the *mean curve* of the LCS-problem. This definition has an immediate extension to the case where r sequences $(X_n^s)_{\mathbb{N}}$ $(s = 1, \dots, r)$ of i.i.d. random variables X_n^s with distribution μ in a finite alphabet $\mathbb{A}$ are considered. The mean curve is concave and symmetric in λ. Montecarlo simulation methods for the computation of probabilistic lower bounds on $\gamma(\lambda)$ are readily available through the exploitation of subadditivity. Furthermore, in the case where $r = 2$ methods for the computation of both probabilistic and deterministic upper and lower bounds on the value $\gamma(1/2, 1/2))$ have been derived in a number of papers [6, 9, 7, 13, 1, 4, 11, 8].

In this paper we discuss a method for the computation of upper bounds on the mean curve based on large deviations technology. This extends the upper bound method derived in [11] for the case $\gamma(1/2, 1/2))$, $r = 2$. Our method depends on a finite measure $\nu^{m,r}$ that encapsulates information about the probability that random sequences of given lengths are so-called m-matches, defined by the properties of having a LCS of length exactly m and by requiring that the last characters of each of the sequences must be aligned to achieve this score. Our method then boosts any known information about the score distribution of shorter strings (through partial knowledge of the measure $\nu^{m,r}$) to derive information about the expected LCS scores of longer r-tuples of random sequences by decomposing the latter into a concatenation of m-matches. In practice the measure $\nu^{m,r}$ is not known exactly, but it can be approximated arbitrarily well via simulations. Simulated partial knowledge about $\nu^{m,r}$ leads to probabilistic upper bounds $q_{m,\ell}(\lambda)$ which depend on the number ℓ of simulations as well as the parameter m. For any given $\alpha, \varepsilon > 0$ one can choose m_ε and ℓ_α such that for all $m \geq m_\varepsilon$ and $\ell \geq \ell_\alpha$,

$$\mathrm{P}[\gamma(\lambda) \leq q_{m,\ell}(\lambda) \leq \gamma(\lambda) + \varepsilon] \geq 1 - \alpha,$$

that is, $q_{m,\ell}(\lambda)$ approximates $\gamma(\lambda)$ to precision ε at the confidence level $1 - \alpha$.

The paper is structured as follows. In Section 2 we introduce the basic concepts of our analysis and prove some of their elementary properties. Section 3 serves to characterize upper bounds on $\gamma(\lambda)$ via a criterion that is amenable to numerical computations via optimization problems that depend on the parameter m. Furthermore, the criterion is used to show that the solutions $q_m(\lambda)$ of these optimization problems form a consistent sequence $(q_m(\lambda))_{m\in\mathbb{N}}$ of upper bounds on $\gamma(\lambda)$, in other words, $\lim_{m\to\infty} q_m(\lambda) = \gamma(\lambda)$. Practical computations and numerical results are discussed in Section 4. Our results yield strong numerical evidence that the so-called Steele-conjecture [15] is likely not true.

2. Basic Concepts and Notation. This section is devoted to building the main tools we need in the analysis of Sections 3 and 4.

2.1. The General Setup. For convenience we use the shorthand notation $\mathbb{N}_n := \{1, \dots, n\}$. We denote the canonical basis vectors of $\mathbb{R}^r$ by $e_1, \dots, e_r$ and write $e := \sum_{s=1}^r e_s$ for the r-vector of ones. Let $\mathbb{A}$ be a finite alphabet and μ a probability measure on $\mathbb{A}$ with $\mu_{\min} := \min_{a \in \mathbb{A}} \mu(a) > 0$. Throughout this paper $X = \big((X_n^1)_{\mathbb{N}}, \dots, (X_n^r)_{\mathbb{N}}\big)$ denotes an independent r-tuple of infinite sequences of i.i.d. random variables X_n^s with law μ.

2.2. Strings and Subsequences. We write $\mathbb{A}^* := \bigcup_{n \in \mathbb{N}} \mathbb{A}^n$ for the set of finite strings with characters in $\mathbb{A}$ and $\#x := n$ for the length of a finite string $x = (x_1, \dots, x_n) \in \mathbb{A}^*$. If $x \in \mathbb{A}^* \cup \mathbb{A}^{\mathbb{N}}$ and $i, j \leq \#x$ we write $x[i,j] := (x_i, \dots, x_j)$ for the piece of x between indices i and j. We take this to be the empty string if $j < i$. Furthermore, if $x \in \mathbb{A}^*$ we use the shorthand notation $x^{[-]} := x[1, \#x - 1]$ for the string that is left after truncating the last character. The set of subsequences of x,

$$Sub(x) := \big\{ \big(x_{\pi(1)}, \dots, x_{\pi(k)}\big) : \; k \in \mathbb{N}, \, \pi : \mathbb{N}_k \to \mathbb{N}_{\#x} \text{ strictly increasing} \big\}$$

consists of the strings obtained by deleting some of the characters of x and keeping the remaining ones in the original order.

2.3. Vectorized Notation. Extending the introduced notation to r-tuples $x = (x^1, \dots, x^r)$ of strings, we write $\#x$ for the vector $(\#x^1, \dots, \#x^r)$, x^- for the set of r-tuples $\big\{\big(x^{1[-]}, x^2, \dots, x^r\big), \dots, \big(x^1, \dots, x^{r-1}, x^{r[-]}\big)\big\}$, and $x[i,j]$ for the r-tuple $\big(x^1[i_1, j_1], \dots, x^r[i_r, j_r]\big)$ when $i, j \in \mathbb{N}^r$ are multi-indices. Finally, we write $|i| := \sum_s i_s$, and $i < j$ if $i_r < j_r$ for all r.

2.4. Longest Common Subsequences. The set of common subsequences of $x \in \mathbb{A}^{*r}$ is defined by $ComSub(x) := \bigcap_{s \in \mathbb{N}_r} Sub(x^s)$, and the length of a longest common subsequence of x by $LCS(x) := \max\big\{\#y : \; y \in ComSub(x)\big\}$. The function $x \mapsto LCS(x)$ can easily be evaluated: define a function $incr : \; \mathbb{N}_0^r \times \mathbb{A}^{*r} \to \{0, 1\}$

by $incr(i,x) := \delta\big(x^1_{i_1},\dots,x^r_{i_r}\big)$, where δ is the Kronecker delta, that is, $\delta = 1$ if all arguments are defined and equal, and $\delta = 0$ otherwise. Let $Score : \mathbb{N}_0^r \times \mathbb{A}^{*r} \to \mathbb{N}$ be defined by setting $Score(i,x) = 0$ if $i \not> 0$ and by requiring that the recursive relationship

$$Score(i,x) = \max\big\{Score(i-e,x) + incr(i,x), Score(i-e_1,x),\dots,Score(i-e_r,x)\big\}$$

be satisfied for all $i > 0$. Then it is easily verified that $Score(\#x,x) = LCS(x)$. The computation of $LCS(x)$ via recursive evaluation of $Score$ is referred to as *Wagner-Fischer algorithm* [16]. Since each recursion step involves finding a maximum, this is actually a dynamic programming algorithm.

2.5. The Mean Curve. Each multi-index $i \in \mathbb{N}^r$ defines a random variable $L(i) := LCS\big(X[e,i]\big)$. For $i,j \in \mathbb{N}^r$ the r-tuples of random strings $X[e,j]$ and $X[i+e,i+j]$ are identically distributed and the inequality $L(i) + LCS\big(X[i+e,i+j]\big) \leq L(i+j)$ holds trivially true. Therefore, the following subadditivity property holds:

$$\mathrm{E}[L(i)] + \mathrm{E}[L(j)] \leq \mathrm{E}[L(i+j)]. \tag{2.1}$$

Let $\Delta_r := \mathrm{conv}\{e_1,\dots,e_r\}$ be the standard r-simplex, that is, the set of vectors in $\mathbb{R}^r$ whose components form weights in a convex combination of r objects. For $\gamma \in \Delta_r$ we denote the multi-index $(\lfloor rn\gamma_1\rfloor,\dots,\lfloor rn\gamma_r\rfloor)$ by the short-hand $\lfloor rn\gamma\rfloor$. Each $n \in \mathbb{N}$ defines a random function $L_n : \Delta_r \to \mathbb{N}$ via $\lambda \mapsto L(\lfloor rn\lambda\rfloor)$. The central object of investigation of this paper is the *mean curve*

$$\begin{aligned}\gamma : \Delta_r &\to \mathbb{N}\\ \lambda &\mapsto \lim_{n\to\infty} \mathrm{E}[L_n(\lambda)/n].\end{aligned}$$

This function is well-defined because $(\inf_{k\geq n} \mathrm{E}[L_k/k])_{n\in\mathbb{N}}$ is an increasing sequence bounded by 1 so that $\gamma(\lambda) := \lim_{n\to\infty} \inf_{k\geq n} \mathrm{E}[L_k/k]$ exists, and furthermore, if

$(n_i)_{i\in\mathbb{N}}$ is an increasing subsequence of $\mathbb{N}$ such that $\lim_{i\to\infty} \mathrm{E}[L_{n_i}/n_i] = \gamma(\lambda)$, then for all $n \in \mathbb{N}$,

$$\frac{\mathrm{E}[L_{n_i}(\lambda)]}{n_i} \overset{(2.1)}{\geq} \frac{\lfloor n_i/n \rfloor \, \mathrm{E}[L_n(\lambda)]}{n_i} = \frac{\mathrm{E}[L_n(\lambda)]}{n} \cdot \frac{n}{n + \frac{n_i - \lfloor n_i/n \rfloor n}{\lfloor n_i/n \rfloor}},$$

so that

$$\gamma(\lambda) \geq \lim_{i\to\infty} \frac{\mathrm{E}[L_n(\lambda)]}{n} \cdot \frac{n}{n + \frac{n_i - \lfloor n_i/n \rfloor n}{\lfloor n_i/n \rfloor}} = \frac{\mathrm{E}[L_n(\lambda)]}{n}. \tag{2.2}$$

Subadditivity also implies that γ is concave as a direct consequence of

$$\begin{aligned}\gamma(\lambda)/2 + \gamma(\eta)/2 &= \lim_{n\to\infty} \frac{1}{2n} \mathrm{E}\big[L(\lfloor rn\lambda \rfloor) + L(\lfloor rn\eta \rfloor)\big] \\ &\overset{(2.1)}{\leq} \lim_{n\to\infty} \frac{1}{2n} \mathrm{E}\big[L(\lfloor rn\lambda + rn\eta \rfloor)\big] = \gamma(\lambda/2 + \eta/2).\end{aligned}$$

2.6. The Notion of m-Matches and an Associated Measure. Let $m \in \mathbb{N}$. A r-tuple $x \in \mathbb{A}^{*r}$ of finite strings is called *m-match* if $LCS(x) = m$ and $LCS(y) = m-1$ for all $y \in x^{[-]}$. The second condition simply says that the final character of each string must be part of any longest common subsequence of x. We write $M^{m,r}$ for the set of *$m-matches$* in $\mathbb{A}^{*r}$ and $\chi_{M^{m,r}} : \mathbb{A}^{*r} \to \{0,1\}$ for the indicator function of $M^{m,r}$, that is, $\chi(x) = 1$ if $x \in M^{m,r}$ and $\chi(x) = 0$ otherwise. For all $i \in \mathbb{N}^r$ let

$$\nu^{m,r}(i) := \mathrm{E}\big[\chi_{M^{m,r}}(X[e,i])\big].$$

Then $\nu^{m,r}(B) := \sum_{i\in B} \nu^{m,r}(i)$ defines a measure on $\mathbb{N}^r$ with support $(\mathbb{N} \setminus \mathbb{N}_{m-1})^r$. By embedding $\mathbb{N}^r$ in $\mathbb{R}^r$ we can also interpret $\nu^{m,r}$ as a Borel measure on $\mathbb{R}^r$.

Lemma 2.1. *For all $m \in \mathbb{N}$,*

i) $\nu^{m,r}$ is a finite nonnegative measure, *ii)* $\tilde{\nu}^{m,r} := \nu^{m,r}/\nu^{m,r}(\mathbb{R}^2)$ *is a well-defined probability measure on* $\mathbb{R}^r$*,*

iii) *the Laplace transform* $\int_{\mathbb{R}^r} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy$ *is finite for all* x *in an open domain* D *containing the negative orthant* $\mathbb{R}^r_- := \{x \in \mathbb{R}^r :\, x_s \le 0\, \forall\, s \in \mathbb{N}_r\}$*.*

Proof. The nonnegativity of $\nu^{m,r}$ is clear from the definition. Parts i) and ii) are straighforward consequences of iii). For all $k \in \mathbb{N}$ let

$$\mu^{m,r}(k) := \sum_{i \in \mathbb{N}^r : |i| = k} \nu^{m,r}(i).$$

To prove iii), it suffices to establish that there exist constants $\epsilon > 0$, $k_\epsilon \in \mathbb{N}$ such that

$$\mu^{m,r}(k) \le \exp(-2\epsilon k) \tag{2.3}$$

for all $k \ge k_\epsilon$, since it then follows that for all $x < \epsilon e$ we have

$$\begin{aligned}\int_{\mathbb{R}^r} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy < \int_{\mathbb{R}^r_+} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,\epsilon e\rangle}\, dy &= \sum_{k \in \mathbb{N}} \mu^{m,r}(k) \exp(\epsilon k) \\ &\le \sum_{k < k_\epsilon} \mu^{m,r}(k) \exp(\epsilon k) + \sum_{k \ge k_\epsilon} \exp(-\epsilon k) < \infty.\end{aligned}$$

It is easy to see that for all $x \in M^{m,r}$ there exists

$$(i_1, \ldots, i_m) \in \mathcal{CS}_{\#x} := \left\{ (i_1, \ldots, i_m) \mathbb{N}^{r \times m} :\, \sum_{l=1}^{m} i_l = \#x \right\}$$

such that

$$x\Big[e + \sum_{u<l} i_u \;,\; \sum_{u \le l} i_u\Big] \in M^{1,r}, \quad (l = 1, \ldots, m). \tag{2.4}$$

We define the cardinality signature $cs(x)$ of x to be the (unique) minimal tuple with respect to lexicographic order amongst the tuples $(i_1, \ldots, i_m) \in \mathcal{CS}_{\#x}$ of multi-indices for which (2.4) holds. We now have

$$\begin{aligned}
\mu^{m,r}(k) &= \sum_{i\in\mathbb{N}^r:|i|=k} \mathrm{P}\left[X[e,i]\in M^{m,r}\right] \\
&= \sum_{i\in\mathbb{N}^r:|i|=k}\ \sum_{j\in\mathcal{CS}_i} \mathrm{P}\big[X[e,i]\in M^{m,r}\|cs(X[e,i])=j\big]\cdot\mathrm{P}\left[cs(X[e,i])=j\right] \\
&\le \sum_{i\in\mathbb{N}^r:|i|=k}\ \sum_{j=(i_1,\dots,i_m)\in\mathcal{CS}_i}\ \prod_{l=1}^{m}\left(\sum_{a\in\mathbb{A}}\xi(a)^r(1-\xi(a))^{|i_r|-r}\right)\cdot\mathrm{P}\left[cs(X[e,i])=j\right] \\
&\le |\mathbb{A}|^m\cdot(1-\xi_{\min})^{k-mr},
\end{aligned}$$

which implies the required property (2.3). □

The measure $\nu^{m,r}$ is not known explicitly. However, given $x\in\mathbb{A}^{*r}$, we have $x[e,i]\in M^{m,r}$ if and only if $Score(i,x)=m$ and $Score(i-e_s,x)=m-1$ for all $s\in\mathbb{N}_r$. Hence, $\nu^{m,r}$ can be simulated using the Wagner-Fischer algorithm.

2.7. Parsing Strings into m-Matches. Let $x\in\mathbb{A}^{*r}$ be a r-tuple of finite strings with length $i=\#x$, and let $k,m\in\mathbb{N}$. If $LCS(x)\ge km$ then there exists at least one r-tuple $\varpi=(\varpi^1,\dots,\varpi^r)$ of increasing functions $\varpi^s:\mathbb{N}_{km}\to\mathbb{N}_{i_s}$, $(s\in\mathbb{N}_r)$ such that $\varpi^1(l)=\dots=\varpi^r(l)$ for all $l\in\mathbb{N}_{km}$. If we choose ϖ minimal among all such r-tuples of functions under the partial ordering defined by

$$\varpi\preceq\varsigma\Leftrightarrow\varpi^s(l)\le\varsigma^s(l)\quad\forall\, s\in\mathbb{N}_r,\, l\in\mathbb{N}_{km},$$

then $x\big[e,\varpi(m)\big]$, $x\big[\varpi(m)+e,\varpi(2m)\big]$,..., $x\big[\varpi((k-1)m)+e,\varpi(km)\big]$, $x\big[\varpi(km)+e,i\big]$ is a parsing of x into k collated m-matches and a remainder in $\mathbb{A}^{*r}$. Consider the index set

$$I_i^{k,m}:=\Big\{(u^0,\dots,u^k)\in\mathbb{N}^{r\times(k+1)}:\,u^0=0,\,u^j\ge me\quad\forall j\in\mathbb{N}_k,\,\sum_{j=1}^{k}u^j\le i\Big\}.$$

Then our argument shows that the following equivalence holds,

$$LCS(x) \geq km \Leftrightarrow \sum_{u \in I^{k,m}_{\#x}} \prod_{l=1}^{k} \chi_{M^{m,r}}\Big(x\Big[\sum_{j=0}^{l-1} u^j + e, \sum_{j=0}^{l} u^j\Big]\Big) \geq 1. \qquad (2.5)$$

The righthand side says that if we sum over all possible ways to parse x into k collated r-tuples of length at least me and a remainder, then there must be at least one such parsing for which the k first r-tuples are all m-matches. We also note that

$$\#I^{k,m}_i = \begin{cases} \prod_{s=1}^{r} \binom{i_s - k(m-1)}{k} & \text{if } i_s \geq km \; \forall\, s \in \mathbb{N}_r, \\ 0 & \text{otherwise.} \end{cases} \qquad (2.6)$$

3. Upper Bounds on the Mean Curve. In this section we characterize upper bounds on the mean curve in terms of computable quantities that will form the basis of the algorithm of Section 4. The idea that drives our method is based on the observation made in Section 2.7 that whenever $LCS(x) \geq km$ holds, x can be parsed into k m-matches plus a remainder. In general there are many such parsings for the same r-tuple of strings. Therefore, the number of different concatenations of k m-matches and remainders that form r-tuples of strings of length $\lfloor rn\lambda \rfloor$ is an upper bound on the number of r-tuples of the same length for which $LCS(x) \geq km$ holds.

To characterize bounds on $\gamma(\lambda)$, we proceed via a sequel of reformulations – each to be analyzed in a separate section – that gradually approach a form that is amenable to numerical computations. Proposition 3.1, the main result of Section 3.1, shows in essence that q is an upper bound on $\gamma(\lambda)$ if and only if

$$\limsup_{n \to \infty} \mathrm{P}\left[L_n(\lambda) \geq nq\right]^{1/n} < 1.$$

In reality, the equivalence is slightly weaker in one direction. In Proposition 3.4 of

Section 3.2 this criterion is further reformulated in the form

$$\limsup_{k\to\infty}\big(\big(\nu^{m,r}\big)^{*k}\big(kB^{m,r}_{\lambda,q,\varepsilon}\big)\big)^{1/k} < 1,$$

where $B^{m,r}_{\lambda,q,\varepsilon}$ is a certain box in $\mathbb{R}^r$. The proof of this result is where the decomposition mechanism comes into play and where a connection with the measure $\nu^{m,r}$ is established. Finally, in Proposition 3.9 of Section 3.3 this last criterion is further reduced to

$$\inf\Big\{\Lambda^{m,r}(x) - \frac{rm}{q}\langle\lambda, x\rangle :\; x \in \mathbb{R}^r_-\Big\} < 0,$$

where

$$\Lambda^{m,r}(x) := \log \int_{\mathbb{R}^r} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy$$

is the log-Laplace transforms of $\nu^{m,r}$. In this form, deciding whether a given value of q is an upper bound on $\gamma(\lambda)$ becomes an optimization problem that is amenable to numerical computations. In Section 3.4, finally, we use this last criterion to derive a consistent sequence of upper bounds on $\gamma(\lambda)$, see Theorem 3.14 which consitutes the main result of Section 3.

3.1. Characterizing Upper Bounds by Exponential Rates. The first step in our approach is to characterize upper bounds on $\gamma(\lambda)$ in terms of the tail events of the form $\mathrm{P}\left[L_n(\lambda) \geq nq\right]$.

For $q > 0$ and $\lambda \in \Delta_r$ we define the exponential rate

$$c(\lambda, q) := \limsup_{n\to\infty}\ \mathrm{P}\left[L_n(\lambda) \geq nq\right]^{1/n}.$$

The following proposition is the main result of this section.

Proposition 3.1. *For all $q > 0$, $\lambda \in \Delta_r$, the following implications hold,*

i) $c(\lambda, q) < 1 \Rightarrow \gamma(\lambda) \leq q$

ii) $\gamma(\lambda) < q \Rightarrow c(\lambda, q) < 1$.

Before proving this result, let us introduce two lemmas. The first is a classical result from the theory of large deviations.

Lemma 3.2 (Azuma-Hoeffding Theorem). *Let $\mathscr{F} = \big(\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}_t\big)$ be a filtration of σ-algebras for some $t \in \mathbb{N}$, $V = \big(V_0, V_1, \ldots, V_t\big)$ a $\mathscr{F}$-adapted martingale with $V_0 \equiv 0$, and $a > 0$ a positive number. If $\mathrm{P}\big[|V_s - V_{s-1}| \leq a\big] = 1$ for all $s \in \mathbb{N}_t$ then*

$$\mathrm{P}[V_t \geq \delta t] \leq \exp\big(-t\delta^2/(2a^2)\big) \quad \forall\, \delta > 0.$$

For a proof of Lemma 3.2, see Azuma [3] and Hoeffding [12]. For a modern proof see e.g. [17]. Our next result is a close relative of the large deviation result of Arriata-Waterman [2] and yields some useful inequalities.

Lemma 3.3. *Let $\lambda \in \Delta_r$. Then*

i) $\mathrm{P}\big[L_n(\lambda) - n\gamma(\lambda) \geq n\delta\big] \leq \mathrm{e}^{-n\delta^2/(2r)}$*, for all $\delta \geq 0$,*

ii) $\mathrm{P}\big[L_n(\lambda) - n\gamma(\lambda) \leq -n\delta\big] \leq \mathrm{e}^{-n\delta^2/(8r)}$ *for all $\delta \in [0, 2\gamma(\lambda)]$ and $n \gg 1$.*

Proof. i) Equation (2.2) shows that

$$\mathrm{P}\big[L_n(\lambda) - n\gamma(\lambda) \geq n\delta\big] \leq \mathrm{P}\big[L_n(\lambda) \geq \mathrm{E}[L_n(\lambda)] + n\delta\big]. \tag{3.1}$$

Let $\Gamma : \mathbb{N}_{|\lfloor rn\lambda\rfloor|} \cup \{0\} \to \mathbb{Z}^r$ be such that $\Gamma(0) = 0$, $\Gamma(|\lfloor rn\lambda\rfloor|) = \lfloor rn\lambda\rfloor$ and $\Gamma(k) - \Gamma(k-1) \in \{e_1, \ldots, e_r\}$ for all $k \in \mathbb{N}_{|\lfloor rn\lambda\rfloor|}$. For $k \in \mathbb{N}_{|\lfloor rn\lambda\rfloor|} \cup \{0\}$ let $\mathscr{G}_k := \sigma\big(X^s_u : s \in \mathbb{N}_r, u \in \mathbb{N}_{\Gamma_s(k)}\big)$ and

$$W_k := \mathrm{E}\big[L_n(\lambda) - \mathrm{E}[L_n(\lambda)] \| \mathscr{G}_k\big].$$

Then $\mathscr{G}_0 = \{\mathbb{R}, \emptyset\}$ and $\mathscr{G} = \big(\mathscr{G}_0 \subseteq \cdots \subseteq \mathscr{G}_{|\lfloor rn\lambda\rfloor|}\big)$ is a filtration of σ-algebras. Furthermore, $W = \big(W_0, \ldots, W_{|\lfloor rn\lambda\rfloor|}\big)$ is a $\mathscr{G}$-adapted martingale with $W_0 \equiv 0$ and $\mathrm{P}\big[|W_k - W_{k-1}| \leq 1\big] = 1$ for all $k \in \mathbb{N}_{|\lfloor rn\lambda\rfloor|}$. Applying Lemma 3.2 to $\big(\mathscr{G}, W\big)$, we find

$$\begin{aligned}\mathrm{P}\Big[L_n(\lambda) - \mathrm{E}[L_n(\lambda)] \geq n\delta\Big] &= \mathrm{P}\left[W_{|\lfloor rn\lambda\rfloor|} \geq \frac{n\delta}{|\lfloor rn\lambda\rfloor|} \cdot |\lfloor rn\lambda\rfloor|\right] \\ &\leq \exp\left(-\frac{n^2\delta^2}{2|\lfloor rn\lambda\rfloor|}\right) \leq \mathrm{e}^{-n\delta^2/(2r)}\,.\end{aligned}$$

Together with (3.1) this implies i).

ii) For $n \gg 1$ we have $\mathrm{E}[L_n(\gamma)] \geq n\big(\gamma(\lambda) - \delta/2\big)$, and then

$$\mathrm{P}[L_n(\lambda) - n\gamma(\lambda) \leq -n\delta] \leq \mathrm{P}\Big[L_n(\lambda) - \mathrm{E}[L_n(\lambda)] \leq -n\delta/2\Big]. \tag{3.2}$$

Furthermore, applying Lemma 3.2 to the $\mathscr{G}$-adapted martingale $-W$ we find

$$\begin{aligned}\mathrm{P}\Big[L_n(\lambda) - \mathrm{E}[L_n(\lambda)] \leq -n\delta/2\Big] &= \mathrm{P}\left[-W_{|\lfloor rn\lambda\rfloor|} \geq \frac{n\delta}{2|\lfloor rn\lambda\rfloor|} \cdot |\lfloor rn\lambda\rfloor|\right] \\ &\leq \exp\left(-\frac{n^2\delta^2}{8|\lfloor rn\lambda\rfloor|}\right) \leq \mathrm{e}^{-n\delta^2/(8r)}\,.\end{aligned}$$

Together with (3.2) his establishes part ii). ◻

Finally, we are ready to prove Proposition 3.1:

Proof. i) Let $\varepsilon > 0$ be such that $c(\lambda, q) + \varepsilon < 1$. Since $\mathrm{P}[L_n(\lambda) \geq nq] \leq (c(\lambda, q) + \varepsilon)^n$ for all sufficiently large values of n, we have

$$\lim_{n\to\infty} \mathrm{P}[L_n(\lambda) \leq nq] = 1. \tag{3.3}$$

Suppose now that $\delta := \gamma(\lambda) - q > 0$ then Lemma 3.3 ii) implies

$$\lim_{n\to\infty} \mathrm{P}[L_n(\gamma) \leq nq] \leq \lim_{n\to\infty} \mathrm{e}^{-n\delta^2/(8r)} = 0.$$

Since this contradicts (3.3), it must be the case that $\gamma(\lambda) \leq q$.

ii) Since $\delta := q - \gamma(\lambda) > 0$, Lemma 3.3 i) implies

$$c(\lambda, q) = \limsup_{n\to\infty} \mathrm{P}\left[L_n(\lambda) \geq nq\right]^{1/n} \leq \lim_{n\to\infty} \mathrm{e}^{-\delta^2/(2r)} < 1.$$

□

3.2. An Explicit Handle on Exponential Rates. In order to put Proposition 3.1 to good use, we next need to develop bounds on $\mathrm{P}[L_n(\lambda) \geq nq]$. This is the purpose of the present section. Let $q > 0$, $\varepsilon \geq 0$, $\lambda \in \Delta_r$ and $m \in \mathbb{N}$, and let us define the box

$$B^{m,r}_{\lambda,q,\varepsilon} := \Big[m, \frac{rm\lambda_1}{q} + \varepsilon\Big] \times \cdots \times \Big[m, \frac{rm\lambda_r}{q} + \varepsilon\Big].$$

We denote the k-fold convolution of $\nu^{m,r}$ with itself by $\big(\nu^{m,r}\big)^{*k}$. The following constitutes the main result of this section.

Proposition 3.4.

i) $\limsup_{k\to\infty}\big(\big(\nu^{m,r}\big)^{*k}\big(kB^{m,r}_{\lambda,q,\varepsilon}\big)\big)^{1/k} < 1 \Rightarrow c(\lambda, q) < 1.$

ii) $c(\lambda, q - \varepsilon) < 1 \Rightarrow$ *there exists* m_0 *such that for all* $m \geq m_0$,

$$\limsup_{k\to\infty}\big(\big(\nu^{m,r}\big)^{*k}\big(kB^{m,r}_{\lambda,q,0}\big)\big)^{1/k} < 1.$$

We prepare the proof of Proposition 3.4 via four lemmas.

Lemma 3.5. *There exists* $n_1 \in \mathbb{N}$ *such that for all* $n \geq n_1$,

$$\mathrm{P}[L_n(\lambda) \geq nq] \leq \big(\nu^{m,r}\big)^{*\lfloor nq/m\rfloor}\big(\lfloor nq/m\rfloor B^{m,r}_{\lambda,q,\varepsilon}\big).$$

Proof. Using (2.5) and the fact that

$$\chi_{M^{m,r}}\Big(X\Big[\sum_{j=0}^{l-1}u^j+e,\sum_{j=0}^{l}u^j\Big]\Big)$$

are i.i.d. random variables for $l=1,\dots\lfloor nq/m\rfloor$, we find

$$\begin{aligned}
\mathrm{P}[L_n(\lambda)\geq nq] &\leq \mathrm{P}\big[L_n(\lambda)\geq \lfloor nq/m\rfloor m\big]\\
&=\mathrm{P}\Big[\sum_{u\in I^{\lfloor nq/m\rfloor,m}_{\lfloor rn\lambda\rfloor}}\prod_{l=1}^{\lfloor nq/m\rfloor}\chi_{M^{m,r}}\Big(X\Big[\sum_{j=0}^{l-1}u^j+e,\sum_{j=0}^{l}u^j\Big]\Big)\geq 1\Big]\\
&\leq \mathrm{E}\Big[\sum_{u\in I^{\lfloor nq/m\rfloor,m}_{\lfloor rn\lambda\rfloor}}\prod_{l=1}^{\lfloor nq/m\rfloor}\chi_{M^{m,r}}\Big(X\Big[\sum_{j=0}^{l-1}u^j+e,\sum_{j=0}^{l}u^j\Big]\Big)\Big]\\
&=\sum_{u\in I^{\lfloor nq/m\rfloor,m}_{\lfloor rn\lambda\rfloor}}\prod_{l=1}^{\lfloor nq/m\rfloor}\nu^{m,r}(u^l)\\
&=\big(\nu^{m,r}\big)^{*\lfloor nq/m\rfloor}\Big(\big[\lfloor nq/m\rfloor m, rn\lambda_1\big]\times\cdots\times\big[\lfloor nq/m\rfloor m, rn\lambda_r\big]\Big). \qquad (3.4)
\end{aligned}$$

Let n_1 be sufficiently large so that $rn\lambda_s \leq \lfloor nq/m\rfloor\cdot(rm\lambda_s/q+\varepsilon)$ holds for all $s\in\mathbb{N}_r$ and $n\geq n_1$. Then the inclusion $\big[\lfloor nq/m\rfloor m, rn\lambda_1\big]\times\cdots\times\big[\lfloor nq/m\rfloor m, rn\lambda_r\big]\subseteq \lfloor nq/m\rfloor B^{m,r}_{\lambda,q,\varepsilon}$ holds for all $n\geq n_1$, and the claim follows. □

LEMMA 3.6.

i) $r\lambda_s\geq q$ *if and only if* $\lfloor rn\lambda_s\rfloor\geq\lfloor nq/m\rfloor m$ *for all* $m,n\in\mathbb{N}$*,*

ii) if $q>r\lambda_s$ *for some* $s\in\mathbb{N}_r$ *then* $nq>\lfloor rn\lambda_s\rfloor$ *and* $\mathrm{P}[L_n(\lambda)\geq nq]=0$*.*

Proof. Both parts are immediate to verify. □

LEMMA 3.7. *If* $r\lambda_s\geq q$ *for all* $s\in\mathbb{N}_r$*, then there exists* $n_2\in\mathbb{N}$ *such that for all* $n\geq n_2$*,*

$$\mathrm{P}[L_n(\lambda)\geq n(q-\varepsilon)]\geq\beta^{m,r}_{\lambda,q,n}\cdot\big(\nu^{m,r}\big)^{*\lfloor nq/m\rfloor}\big(\lfloor nq/m\rfloor B^{m,r}_{\lambda,q,0}\big),$$

where

$$(\beta^{m,r}_{\lambda,q,n})^{-1} = \prod_{s=1}^{r} \binom{\lfloor rn\lambda_s\rfloor - \lfloor nq/m\rfloor(m-1)}{\lfloor nq/m\rfloor}.$$

Proof. It follows from Lemma 3.6 i) and equation (2.6) that $I^{\lfloor nq/m\rfloor,m}_{\lfloor rn\lambda\rfloor}$ and $[\lfloor nq/m\rfloor, rn\lambda_1]\times\cdots\times[\lfloor nq/m\rfloor, rn\lambda_r]$ are nonempty sets. For $n \geq n_1 := m/\varepsilon$ we have $\lfloor nq/m\rfloor m \geq n(q-\varepsilon)$ and hence,

$$\begin{aligned}
&\mathrm{P}\big[L_n(\lambda) \geq n(q-\varepsilon)\big] \geq \mathrm{P}\big[L_n(\lambda) \geq \lfloor nq/m\rfloor m\big] \\
&\quad\overset{(2.5)}{=} \mathrm{P}\Big[\sum_{u\in I^{\lfloor nq/m\rfloor,m}_{\lfloor rn\lambda\rfloor}} \prod_{l=1}^{\lfloor nq/m\rfloor} \chi_{M^{m,r}}\Big(X\big[\sum_{j=0}^{l-1} u^j + e, \sum_{j=0}^{l} u^j\big]\Big) \geq 1\Big] \\
&\quad\geq \big(\#I^{\lfloor nq/m\rfloor,m}_{\lfloor rn\lambda\rfloor}\big)^{-1}\,\mathrm{E}\Big[\sum_{u\in I^{\lfloor nq/m\rfloor,m}_{\lfloor rn\lambda\rfloor}} \prod_{l=1}^{\lfloor nq/m\rfloor} \chi_{M^{m,r}}\Big(X\big[\sum_{j=0}^{l-1} u^j + e, \sum_{j=0}^{l} u^j\big]\Big)\Big] \\
&\quad\overset{(3.4),(2.6)}{=} \beta^{m,r}_{\lambda,q,n}\cdot\big(\nu^{m,r}\big)^{\lfloor nq/m\rfloor}\Big(\big[\lfloor nq/m\rfloor, rn\lambda_1\big]\times\cdots\times\big[\lfloor nq/m\rfloor, rn\lambda_r\big]\Big) \\
&\quad\geq \beta^{m,r}_{\lambda,q,n}\cdot\big(\nu^{m,r}\big)^{\lfloor nq/m\rfloor}\big(\lfloor nq/m\rfloor B^{m,r}_{\lambda,q,0}\big).
\end{aligned}$$

□

Lemma 3.8. *If $r\lambda_s \geq q$ for all $s\in\mathbb{N}_r$, then*

$$\lim_{m\to\infty}\lim_{n\to\infty}\big(\beta^{m,r}_{\lambda,q,n}\big)^{-1/\lfloor nq/m\rfloor} = 0.$$

Proof. The proof of Stirling's formula establishes Robbins' inequality [14]: for all $n\geq 1$,

$$\sqrt{2\pi}\, n^{n+\frac{1}{2}}\, \mathrm{e}^{-n+\frac{1}{12n+1}} \leq n! \leq \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, \mathrm{e}^{-n+\frac{1}{12n}}\,.$$

Using this inequality and the shorthand notation $k = \lfloor nq/m\rfloor$ and $n_s = \lfloor rn\lambda_s\rfloor -$

$\lfloor nq/m \rfloor(m-1)$, we find

$$\prod_{s=1}^{r} \frac{(2\pi)^{\frac{1}{2k}}\, n_s^{-\frac{n_s}{k}+\frac{1}{2k}}\, \mathrm{e}^{-\frac{n_s}{k}+\frac{1}{(12n_s+1)k}}}{(2\pi)^{\frac{1}{2k}}\, k^{-1+\frac{1}{2k}}\, \mathrm{e}^{-1+\frac{1}{12k^2}}\, (2\pi)^{\frac{1}{2k}}\, (n_s-k)^{-\frac{n_s}{k}+1+\frac{1}{2k}}\, \mathrm{e}^{-\frac{n_s}{k}+1+\frac{1}{12(n_s-k)k}}} \leq \left(\beta^{m,r}_{\lambda,q,n}\right)^{-\frac{1}{k}}$$

$$\leq \prod_{s=1}^{r} \frac{(2\pi)^{\frac{1}{2k}}\, n_s^{-\frac{n_s}{k}+\frac{1}{2k}}\, \mathrm{e}^{-\frac{n_s}{k}+\frac{1}{12n_sk}}}{(2\pi)^{\frac{1}{2k}}\, k^{-1+\frac{1}{2k}}\, \mathrm{e}^{-1+\frac{1}{(12k+1)k}}\, (2\pi)^{\frac{1}{2k}}\, (n_s-k)^{-\frac{n_s}{k}+1+\frac{1}{2k}}\, \mathrm{e}^{-\frac{n_s}{k}+1+\frac{1}{(12(n_s-k)+1)k}}}.$$

Since $n_s/k \to r\lambda_s m/q - (m-1)$ when $n \to \infty$, both sides of the inequality converge to

$$\prod_{s=1}^{r} \frac{\left(1 - \frac{q}{r\lambda_s m - q(m-1)}\right)^{\frac{r\lambda_s m}{q}-(m-1)}}{\frac{r\lambda_s m}{q} - m},$$

and for $m \to \infty$ this tends to $\prod_{s=1}^{r}(\mathrm{e} \cdot \lim_{m\to\infty}(r\lambda_s/q - q)m)^{-1} = 0$. ∎

The stage is now set for a proof of Proposition 3.4:

Proof. i) Using Lemma 3.5, we find

$$c(\lambda, q) = \limsup_{n\to\infty} \mathrm{P}\big[L_n(\lambda) \geq nq\big]^{1/n} \leq \limsup_{n\to\infty}\big(\left(\nu^{m,r}\right)^{*\lfloor nq/m\rfloor}\big(\lfloor nq/m\rfloor B^{m,r}_{\lambda,q,\varepsilon}\big)\big)^{1/n}$$
$$= \Big(\limsup_{n\to\infty}\big(\left(\nu^{m,r}\right)^{*\lfloor nq/m\rfloor}\big(\lfloor nq/m\rfloor B^{m,r}_{\lambda,q,\varepsilon}\big)\big)^{1/\lfloor nq/m\rfloor}\Big)^{\lim_{n\to\infty}\frac{\lfloor nq/m\rfloor}{n}} < 1^{q/n}.$$

ii) If $q > r\lambda_s$ for some $s \in \mathbb{N}_r$ then Lemma 3.6 implies $B^{m,r}_{\lambda,q,0} = \emptyset$, and the claim holds trivially. On the other hand, if $q \leq r\lambda_s$ for all s then Lemma 3.8 shows that there exists m_0 such that for all $m \geq m_0$, $lim_{n\to\infty}(\beta^{m,r}_{\lambda,q,n})^{-1/\lfloor nq/m\rfloor} < 1$, and then

$$\limsup_{k\to\infty}\big(\left(\nu^{m,r}\right)^{*k}\big(kB^{m,r}_{\lambda,q,0}\big)\big)^{1/k}$$
$$= \limsup_{n\to\infty}\big(\beta^{m,r}_{\lambda,q,0} \cdot \left(\nu^{m,r}\right)^{*\lfloor nq/m\rfloor}\big(\lfloor nq/m\rfloor B^{m,r}_{\lambda,q,0}\big)\big)^{\frac{1}{n}\cdot\frac{n}{\lfloor nq/m\rfloor}} \cdot \lim_{n\to\infty}\big(\beta^{m,r}_{\lambda,q,n}\big)^{-1/\lfloor nq/m\rfloor}$$
$$\stackrel{Lem3.7}{\leq} \Big(\limsup_{n\to\infty}\mathrm{P}[L_n(\lambda) \geq n(q-\varepsilon)]^{\frac{1}{n}}\Big)^{\lim_{n\to\infty}\frac{n}{\lfloor nq/m\rfloor}} = (c(\lambda, q-\varepsilon))^{m/q} < 1.$$

∎

3.3. Reduction to a Decision Problem. Propositions 3.1 and 3.4 combined show the following implications: if there exists $\varepsilon > 0$ such that

$$\limsup_{k\to\infty}\big((\nu^{m,r})^{*k}(kB^{m,r}_{\lambda,q,\varepsilon})\big)^{1/k} < 1 \tag{3.5}$$

then $\gamma(\lambda) \le q$. On the other hand, if $\gamma(\lambda) < q$ then there exists m_0 such that for all $m \ge m_0$ (3.5) holds with $\varepsilon = 0$. Thus, all that is needed to determine whether q is an upper bound on $\gamma(\lambda)$ or not is to decide whether (3.5) holds. In this section we will show that this decision problem can be replaced by one that is more amenable to numerical computations. This reformulation uses large deviations theory and convex analysis.

For all $m \in \mathbb{N}$ we denote the log-Laplace transforms of $\nu^{m,r}$ and $\tilde{\nu}^{m,r}$ by

$$\begin{aligned}\Lambda^{m,r}(x) &:= \log \int_{\mathbb{R}^r} \nu^{m,r}(y)\,\mathrm{e}^{\langle y,x\rangle}\,dy,\\ \tilde{\Lambda}^{m,r}(x) &:= \log \int_{\mathbb{R}^r} \tilde{\nu}^{m,r}(y)\,\mathrm{e}^{\langle y,x\rangle}\,dy.\end{aligned}$$

It follows from Lemma 2.1 that both functions are well-defined and finite on an open domain $D \subset \mathbb{R}^r$ that contains the nonpositive orthant $\mathbb{R}^r_- := \{x \in \mathbb{R}^r : x_i \le 0\ \forall i \in \mathbb{N}_r\}$. In particular, D contains a neighborhood of the origin. Consider the relation

$$\inf\Big\{\Lambda^{m,r}(x) - \frac{rm}{q}\langle\lambda, x\rangle :\ x \in \mathbb{R}^r_-\Big\} < 0. \tag{3.6}$$

The main result of this section is the following.

Proposition 3.9. *For all $q > 0$, $\lambda \in \Delta_r$ and $m \in \mathbb{N}$,*

i) if (3.6) *holds then there exists $\varepsilon > 0$ such that* (3.5) *holds,*

ii) if (3.5) *holds for $\varepsilon = 0$ then* (3.6) *holds.*

The following is an interesting immediate consequence of Proposition 3.9.

Corollary 3.10. *For all $q > 0$, $\lambda \in \Delta_r$, $m \in \mathbb{N}$,*

i) (3.5) *holds for $\varepsilon = 0$ if and only if* (3.5) *holds for some $\varepsilon > 0$,*

ii) (3.5) *with $\varepsilon = 0$ is equivalent to* (3.6).

We prepare the proof of Proposition 3.9 through a string of preliminaries. Functions that appear throughout this section are *proper functions* in the sense commonly used in convex analysis: we allow such functions to take values in $\mathbb{R} \cup \{+\infty\}$. By default we set $f(x) = +\infty$ for all x outside the domain $\mathrm{dom}(f) := \{x : f(x) < +\infty\}$ of f. A proper function $f : \mathbb{R}^r \to \mathbb{R} \cup \{+\infty\}$ is convex if its epigraph $epi(f) := \{(x,t) \in \mathbb{R}^{r+1} : t \geq f(x)\}$ is a convex set. This is equivalent to $\mathrm{dom}(f)$ and $f|_{\mathrm{dom}(f)}$ being convex.

Lemma 3.11. *$\Lambda^{m,r}$ and $\tilde{\Lambda}^{m,r}$ are convex on $\mathbb{R}^r$.*

Proof. Let us first argue that D is convex. Let $x = \xi_1 x_1 + \xi_2 x_2$ be a convex combination of two elements of D. It suffices to show that $\int_{\mathbb{R}^r} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy$ is finite, but this follows immediately from the convexity of $x \mapsto \exp\langle y,x\rangle$ and the assumption that $x_1, x_2 \in D$. Next, to show that $\Lambda^{m,r}$ is convex on D it suffices to show that $\big(\partial^2/\partial x_j \partial x_k \Lambda^{m,r}(x)\big)$ is positive definite for all $x \in D$. Let $u \in \mathbb{R}^r$, and let us write $\varphi(x) = \int_{\mathbb{R}^r} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy$. Then

$$\mathrm{D}^2_{u,u}\, \Lambda^{m,r}(x) = \frac{1}{\varphi(x)} \int_{\mathbb{R}^r} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle} \langle y,u\rangle^2 dy - \frac{1}{\varphi^2(x)} \Big(\int_{\mathbb{R}^r} \nu^{m,r}\, \mathrm{e}^{\langle y,x\rangle} \langle y,u\rangle dy \Big)^2.$$

Setting $\xi(y) = \mathrm{e}^{\langle y,x\rangle/2}$ and $\psi(y) = \mathrm{e}^{\langle y,x\rangle/2}\langle y,u\rangle$, we find

$$\begin{aligned}\mathrm{D}^2_{u,u}\, \Lambda^{m,r}(x) = \frac{1}{\varphi^2(x)} \Big(\int_{\mathbb{R}^r} \nu^{m,r}(y)\xi^2(y)dy \cdot \int_{\mathbb{R}^r} \nu^{m,r}(y)\psi^2(y)dy \\ - \Big(\int_{\mathbb{R}^r} \nu^{m,r}(y)\xi(y)\psi(y)dy \Big)^2 \Big) > 0,\end{aligned}$$

where the last inequality follows from the Cauchy-Schwartz inequality and the fact that ξ and ψ are not proportional to one another. The convexity of $\tilde{\Lambda}^{m,r}$ follows immediately from that of $\Lambda^{m,r}$. ∎

Lemma 3.12 (Large Deviation Theorem in $\mathbb{R}^r$). *Let $S_k = Z_1 + \cdots + Z_k$ be a sum of k i.i.d. random vectors Z_j with values in $\mathbb{R}^r$ and law $\mathscr{L}(Z)$. Let $\Lambda_Z(x) = \log \mathrm{E}\big[\exp\langle Z, x\rangle\big]$ be the log-Laplace transform of $\mathscr{L}(Z)$. If Λ_Z is finite in a neighborhood of the origin and B is a convex subset of $\mathbb{R}^r$, then*

$$\lim_{k\to\infty} \frac{1}{k} \mathrm{P}[S_k/k \in B] = -\inf_{y\in B} \Lambda_Z^*(y),$$

where $\Lambda_Z^*(y) = sup_{x\in\mathbb{R}^r}\big\{\langle y, x\rangle - \Lambda_Z(x)\big\}$ is the Legendre transform of Λ_Z.

For a proof of Lemma 3.12, see e.g. [10].

Lemma 3.13 (Fenchel Duality Theorem). *For convex proper functions $f : \mathbb{R}^r \to \mathbb{R} \cup \{+\infty\}$, $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a linear map $A : \mathbb{R}^r \to \mathbb{R}^n$, if the origin lies in the interior of $\mathrm{dom}(g) - A(\mathrm{dom}(f))$, then*

$$\inf_{x\in\mathbb{R}^r} \big\{f(x) + g(Ax)\big\} = \sup_{y\in\mathbb{R}^n} \big\{-f^*(A^*y) - g^*(-y)\big\},$$

where f^* is the Legendre transform of f and A^* the adjoint of A.

For a proof of Lemma 3.13, see e.g. Theorem 3.3.5 [5].

We are ready to give a proof of Proposition 3.9:

Proof. i) Let $Z_1, \dots, Z_k$ be i.i.d. random vectors with distribution $\mathscr{L}(Z_j) = \tilde{\nu}^{m,r}$ on $\mathbb{R}^r$ and $S_k = Z_1 + \cdots + Z_k$. Then

$$(\nu^{m,r})^{*k}\big(kB^{m,r}_{\lambda,q,\varepsilon}\big) = \nu^{m,r}(\mathbb{R}^r)^k \cdot \mathrm{P}\big[S_k \in kB^{m,r}_{\lambda,q,\varepsilon}\big]. \tag{3.7}$$

For $x \in \mathbb{R}^r$ let $H^{m,r,x}_{\lambda,q,\varepsilon} := \{z \in \mathbb{R}^r : \langle z - rm\lambda/q - \varepsilon e, x\rangle \le 0\}$. Then $B^{m,r}_{\lambda,q,\varepsilon} \subset H^{m,r,x}_{\lambda,q,\varepsilon}$

for all $x \in \mathbb{R}^r_+ := -\mathbb{R}^r_-$. Let $W_j^{\varepsilon,x} := \langle Z_j - rm\lambda/q - \varepsilon e, x\rangle$. Then for all $x \in \mathbb{R}^r_-$,

$$\begin{aligned}
\mathrm{P}\big[S_k \in kB^{m,r}_{\lambda,q,\varepsilon}\big] &\le \mathrm{P}\left[S_k/k \in H^{m,r,-x}_{\lambda,q,\varepsilon}\right] = \mathrm{P}\left[\sum_{j=1}^k W_j^{\varepsilon,-x} \le 0\right] \\
&\le \mathrm{E}\left[\mathrm{e}^{-\sum_{j=1}^k W_j^{\varepsilon,-x}}\right] = \left(\mathrm{E}\left[\mathrm{e}^{-W_1^{\varepsilon,-x}}\right]\right)^k = \Big(\int_{\mathbb{R}^r} \tilde{\nu}^{m,r}(y)\,\mathrm{e}^{\langle y - rm\lambda/q - \varepsilon e, x\rangle}\,dy\Big)^k.
\end{aligned}$$

Together with (3.7) this shows that for all $k \in \mathbb{N}$, $x \in \mathbb{R}^r_-$,

$$\begin{aligned}
\Big(\big(\nu^{m,r}\big)^{*k}\big(kB^{m,r}_{\lambda,q,\varepsilon}\big)\Big)^{1/k} &\le \int_{\mathbb{R}^r} \nu^{m,r}(y)\,\mathrm{e}^{\langle y - rm\lambda/q - \varepsilon e, x\rangle}\,dy \\
&= \exp\left(-\varepsilon\langle e, x\rangle + \Lambda^{m,r}(x) - \frac{rm}{q}\langle \lambda, x\rangle\right).
\end{aligned}$$

The claim of part i) now follows immediately.

ii) Theorem 3.12 shows that

$$\lim_{k\to\infty} \frac{1}{k}\log \mathrm{P}\left[S_k \in kB^{m,r}_{\lambda,q,0}\right] = \sup_{y \in B^{m,r}_{\lambda,q,0}} -\tilde{\Lambda}^{m,r*}(y).$$

Therefore, (3.7) implies

$$\begin{aligned}
\lim_{k\to\infty} \frac{1}{k}\,\log\Big(\big(\nu^{m,r}\big)^{*k}\big(kB^{m,r}_{\lambda,q,0}\big)\Big) &= \nu^{m,r}(\mathbb{R}^r) + \lim_{k\to\infty}\frac{1}{k}\log \mathrm{P}\left[S_k \in kB^{m,r}_{\lambda,q,0}\right] \\
&= \nu^{m,r}(\mathbb{R}^r) + \sup_{y \in B^{m,r}_{\lambda,q,0}} -\tilde{\Lambda}^{m,r*}(y) \\
&= \sup_{y\in\mathbb{R}^r}\left\{-\Lambda^{m,r*}(y) - g^*(-y)\right\}, \qquad (3.8)
\end{aligned}$$

where

$$g^*(y) = \begin{cases} 0 & \text{if } y \in -B^{m,r}_{\lambda,q,0}, \\ +\infty & \text{otherwise.} \end{cases}$$

It is easy to check that g^* is a convex proper function with Legendre transform

$$g(x) = g^{**}(x) = \sup_{y\in\mathbb{R}^r}\left\{\langle x, y\rangle - g^*(y)\right\} = \sup_{y \in -B^{m,r}_{\lambda,q,0}} \langle x, y\rangle.$$

Since the origin lies in the interior of $\mathrm{dom}(g) - D$, Theorem 3.13 establishes that

$$\sup_{y\in\mathbb{R}^r} \{-\Lambda^{m,r*}(y) - g^*(-y)\} = \inf_{x\in\mathbb{R}^r} \left\{ \Lambda^{m,r}(x) + \sup_{y\in -B^{m,r}_{\lambda,q,0}} \langle x,y\rangle \right\}. \tag{3.9}$$

For $x \in \mathbb{R}^r$ let $J_x := \{j \in \mathbb{N}_r : x_j > 0\}$. Then – using LP duality or by inspection – it is easy to check that

$$\sup_{y\in -B^{m,r}_{\lambda,q,0}} \langle x,y\rangle = -m \sum_{j\in J_x} x_j - \frac{rm}{q} \sum_{j\in J_x^c} \lambda_j x_j .$$

We claim that if there exists $j_0 \in J_x$ then $-e_{j_0}$ is a descent direction for the function

$$\varphi(x) := \Lambda^{m,r}(x) + \sup_{y\in -B^{m,r}_{\lambda,q,0}} \langle x,y\rangle.$$

In fact,

$$\frac{d}{dt}|_{t=0}\varphi(x - te_{j_0}) = \frac{\int_{\mathbb{R}^r} \nu^{m,r}(y)\cdot \mathrm{e}^{\langle y,x\rangle} \cdot (-y_{j_0}) dy}{\int_{\mathbb{R}^r} \nu^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy} + m < 0,$$

where the last inequality follows from the fact that $\nu^{m,r}(y) > 0$ implies $y_{j_0} \geq m$. Therefore,

$$\inf_{x\in\mathbb{R}^r} \left\{ \Lambda^{m,r}(x) + \sup_{y\in -B^{m,r}_{\lambda,q,0}} \langle y,x\rangle \right\} = \inf_{x\in\mathbb{R}^r_-} \left\{ \Lambda^{m,r}(x) - \frac{rm}{q}\langle \lambda, x\rangle \right\}. \tag{3.10}$$

The claim of part ii) now follows from (3.8), (3.9) and (3.10). □

3.4. A Consistent Sequence of Upper Bounds. By now we have built up all the necessary tools to present and prove the main result of Section 3 which shows that

$$q_m(\lambda) = \inf\left\{q > 0 : \inf\left\{\Lambda^{m,r}(x) - (rm/q)\langle\lambda, x\rangle : x \in \mathbb{R}^r_-\right\} < 0\right\} \tag{3.11}$$

defines a consistent sequence $\big(q_m(\lambda)\big)_{m\in\mathbb{N}}$ of upper bounds on $\gamma(\lambda)$.

THEOREM 3.14. *For all* $\lambda \in \Delta_r$*,*

i) $\gamma(\lambda) \leq q_m(\lambda)$ *for all* $m \in \mathbb{N}$*,*

ii) $\lim_{m\to\infty} q_m(\lambda) = \gamma(\lambda)$*.*

Proof. i) It suffices to show that if $q > 0$ is such that (3.6) holds then $\gamma(\lambda) \leq q$. If (3.6) holds, then by Proposition 3.9 i) there exists $\varepsilon > 0$ such that (3.5) holds, and by Proposition 3.4 i) this implies $c(\lambda, q) < 1$. Finally, Proposition 3.1 establishes that $\gamma(\lambda) \leq q$.

ii) It suffices to prove that for all $q > \gamma(\lambda)$ there exists m_0 such that $q \geq q_m$ for all $m \geq m_0$. Let $\varepsilon = (q - \gamma(\lambda))/2 > 0$. Then $q - \varepsilon > \gamma(\lambda)$ and Proposition 3.1 ii) implies $c(\lambda, q - \varepsilon) < 1$. By Proposition 3.4 ii) there exists m_0 such that for all $m \geq m_0$,

$$\limsup_{k\to\infty} \Big(\big(\nu^{m,r}\big)^{*k}\big(kB^{m,r}_{\lambda,q,0}\big)\Big)^{1/k} < 1.$$

Finally, Proposition 3.9 ii) shows that (3.6) holds, implying that $q \geq q_m$ for all $m \geq m_0$, as claimed. ☐

In [11] it was established that in the special case $r = 2$ and $\lambda = (0.5, 0.5)$ the approximation error $q_m(\lambda) - \gamma(\lambda)$ converges to zero at the rate $\mathcal{O}(m^{-\frac{1-\varepsilon}{2}})$, for $\varepsilon > 0$ arbitrary. We believe that the same order holds true in the general case, but we will not pursue this issue further here because this bound is too conservative in practical computations, and furthermore it is a theoretical result that only holds for large m which are beyond the scope of our numerical experiments.

4. Monte Carlo Simulation and Numerical Results. In Section 3 we saw how to construct a consistent sequence $(q_m(\lambda))_{m\in\mathbb{N}}$ of upper bound functions on the mean curve $\gamma(\lambda)$. So far this is only a theoretical tool, as the measure $\nu^{m,r}$ is not known. To turn this into a practical method for the computation of upper bounds,

$\nu^{m,r}$ has to be replaced by an approximation $\widehat{\nu}^{m,r}$ obtained by Montecarlo simulation as follows: for $\ell = 1, \ldots, \ell_0$ let $X_\ell = \big((X^1_{\ell,n})_{n\in\mathbb{N}}, \ldots, (X^r_{\ell,n})_{n\in\mathbb{N}}\big)$ be independent copies of r-tuples of random sequences with the required distribution. For all multi-indices $i \in \mathbb{N}^r$ set

$$\widehat{\nu}^{m,r}(i) = \frac{1}{\ell_0} \cdot |\{\ell : X_\ell[e,i] \in M^{m,r}\}| .$$

Note that this is possible to compute because almost surely only finitely many terms of each X_ℓ have to be simulated to find all $i \in \mathbb{N}^r$ for which $X_\ell[e,i] \in M^{m,r}$. Furthermore, since $\nu^{m,r}(i)$ decreases exponentially in $|i|$ (see proof of Lemma 2.1), $\widehat{\nu}^{m,r}(i) > 0$ is not observed for $|i|$ very large. To illustrate this point, Figure 4.1 shows sparsity patterns of $\widehat{\nu}^{1000,2}$ after $s = 10,000$ simulation runs. For both patterns random sequences over the binary alphabet were used, in the first case using the law $\mu_1(0) = 0.5 = \mu_1(1)$ for the distribution of the random variables $X^s_{\ell,n}$, and $\mu_2(0) = 0.2 = 1 - \mu_2(1)$ for the second. In the first case there were only $66,751$ multi-indices i where nonzero entries $\widehat{\nu}^{1000,2}(i) > 0$ occurred, and in the second case there occurred $76,150$ nonzero entries. We remark that the cost for every simulation run ℓ is of order $\mathcal{O}(\prod_{s=1}^r n^s_{\max})$, where $n^s_{\max}$ are the numbers of entries of $(X^s_{\ell,n})_{n\in\mathbb{N}}$ that have to be generated to determine all multi-indices $i \in \mathbb{N}^r$ for which $X_\ell[e,i] \in M^{m,r}$. In particular, in the two examples of Figure 4.1 we have $n^s_{\max} < 2500$ in both cases.

Figure 4.2 shows the curves $\widehat{q}_{1000}(\lambda)$ computed by numerically solving the bilevel optimization problem

$$\widehat{q}_m(\lambda) = \inf\Big\{q > 0 : \inf\big\{\widehat{\Lambda}^{m,r}(x) - (rm/q)\langle\lambda, x\rangle : x \in \mathbb{R}^r_-\big\} < 0\Big\}, \qquad (4.1)$$

where

$$\widehat{\Lambda}^{m,r}(x) := \log \int_{\mathbb{R}^r} \widehat{\nu}^{m,r}(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy$$

is the empirical version of $\Lambda^{m,r}$ obtained by replacing $\nu^{m,r}$ by $\widehat{\nu}^{m,r}$. The two curves are computed for the two examples from Figure 4.1 and drawn as a function of λ_1.

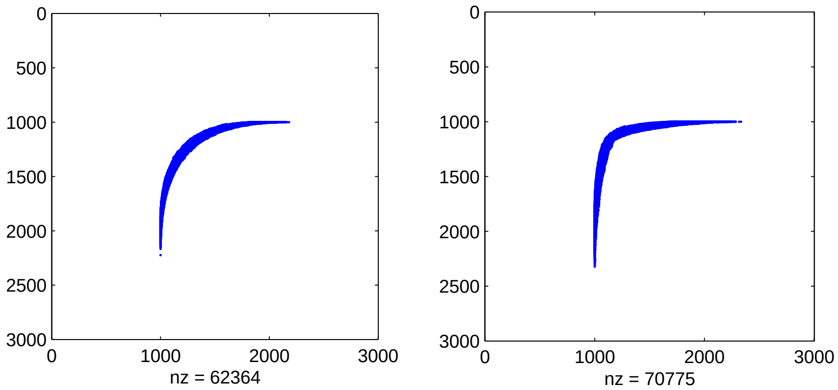

FIG. 4.1. *Sparsity patterns for* $\widehat{\nu}^{1000,2}$ *after* $10,000$ *simulation runs, using* μ_1 *and* μ_2 *respectively for the distribution of the random sequences.*

While it is clearly the case that

$$\gamma(\lambda) \stackrel{\lambda_1 \to 0}{\longrightarrow} 0,$$
$$q_m(\lambda) \stackrel{\lambda_1 \to 0}{\longrightarrow} 0,$$

the numerically computed values of $\widehat{q}_{1000}(\lambda)$ do not tend to zero. This is to be expected, as the optimization problem (4.1) becomes very ill-conditioned and the reduction of q has to be stopped prematurely if the criterion

$$\inf\left\{\widehat{\Lambda}^{m,r}(x) - (rm/q)\langle\lambda, x\rangle : x \in \mathbb{R}^r_-\right\} < 0$$

is to hold to more than machine precision. This limits the computation of tight bounds for values of λ near the boundary $\partial\Delta_r$ of the simplex.

The curve $\widehat{q}_m(\lambda)$ as defined above is an estimate of the upper bound curve $q_m(\lambda)$

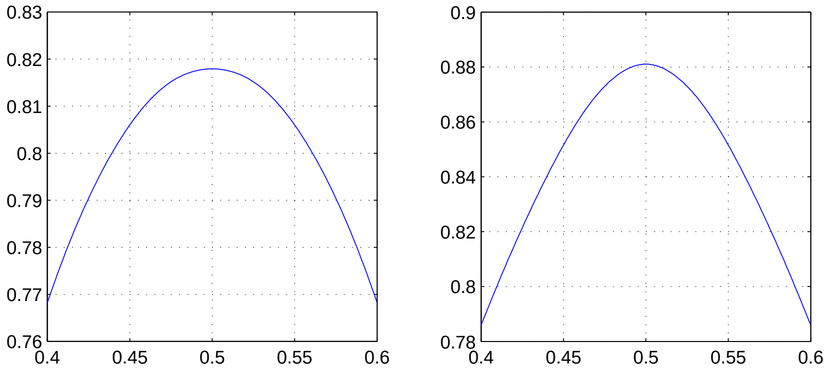

FIG. 4.2. *Numerically computed upper bounds $\widehat{q}_m(\lambda)$ for the examples of Figure 4.1. The plot on the left corresponds to μ_1, the one on the right to μ_2. The accuracy of approximation gets gradually worse for λ near the boundary.*

derived in Section 3. Note however that we did not yet determine a confidence level for the event that $\widehat{q}_m(\lambda)$ is an upper bound on $\gamma(\lambda)$. To determine such a confidence level we need to simulate $\widehat{\nu}^{m,r}$ several times independently. In our experiments we computed 20 independent copies $\widehat{\nu}_1, \ldots, \widehat{\nu}_{20}$ of $\widehat{\nu}^{1000,2}$ with $\ell_0 = 500$ simulation runs each. Each of the measures $\widehat{\nu}_k$ defines an empirical analogue

$$\widehat{\Lambda}_k(x) := \log \int_{\mathbb{R}^r} \widehat{\nu}_k(y)\, \mathrm{e}^{\langle y,x\rangle}\, dy$$

of $\Lambda^{m,r}$. For fixed values of q and λ this defines i.i.d. random variables

$$I_k(q,\lambda) := \inf \left\{ \widehat{\Lambda}_k(x) - (rm/q)\langle \lambda, x\rangle \,:\, x \in \mathbb{R}^r_- \right\}$$

that are close to normally distributed but with a slightly tighter concentration of mass around the mean. To illustrate this, Figure 4.3 shows the cumulative fraction plot of

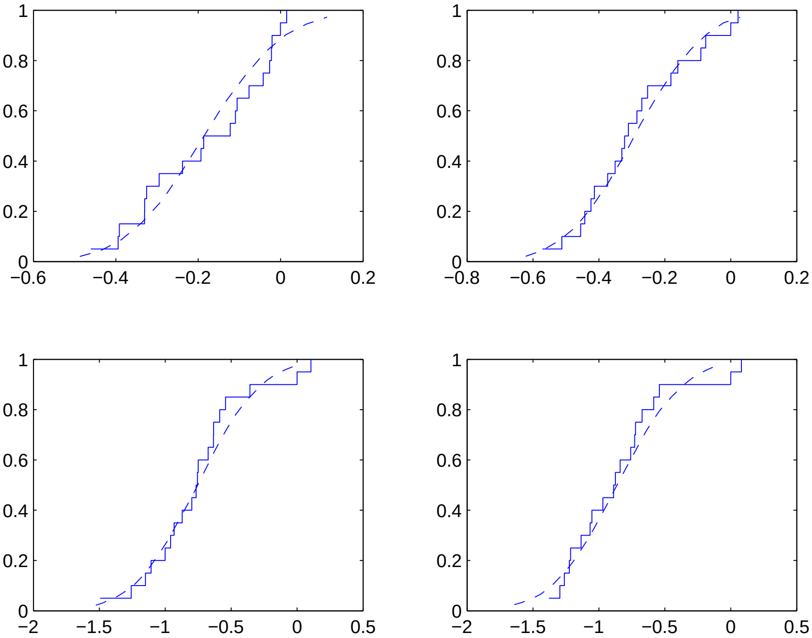

FIG. 4.3. *Cumulative fraction plots of $I_k(q,\lambda)$ versus the cumulative distribution of normal distributions with the estimated mean and variance. The plots of column 1 are based on μ_1 and those of column 2 on μ_2. In row 1 the value $\lambda_1 = 0.47$ was used, whereas $\lambda_1 = 0.5$ was used in the second row. In all four plots the computed value of $\widehat{q}_m^{95\%}(\lambda)$ was used for q.*

$I_k(q,\lambda)$ $(k = 1,\ldots,20)$ for several values of q and λ, both for μ_1 and μ_2 (as introduced above), along with the cumulative distribution plot of $N\big(\mathrm{mean}(I_1,\ldots,I_{20}),\mathrm{var}(I_1,\ldots,I_{20})\big)$, where mean and var are the empirical mean and variance respectively.

Let $J(q,\lambda)$ denote a random variable with law

$$\mathscr{L}(J(q,\lambda)) = N\big(\mathrm{mean}(I_1,\ldots,I_{20}),\mathrm{var}(I_1,\ldots,I_{20})\big).$$

An approximate 95%-confidence-level upper bound on $\gamma(\lambda)$ can be computed as follows,

$$\widehat{q}_m^{95\%}(\lambda) = \inf\{q > 0 : \ |\{k : I_k < 0\}| \geq 19\}.$$

Figure 4.4 shows the 95%-confidence-level upper bounds $\widehat{q}_{1000}^{95\%}$ (solid curve) versus the

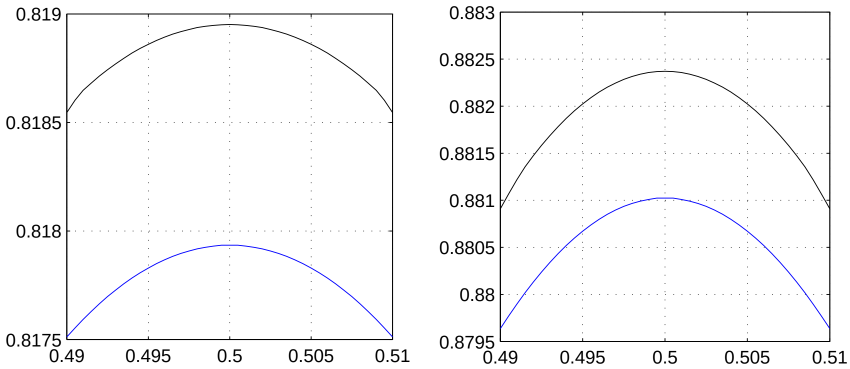

FIG. 4.4. *95%-confidence-level upper bounds (solid curve) versus estimated upper bounds (dashed curve) on $\gamma(\lambda)$ as a function of λ_1 for random sequences governed by μ_1 (figure on the left) and μ_2 (figure on the right).*

upper bound estimates $\widehat{q}_{1000}$ (dashed curve) for the two examples discussed above (distributions μ_1 and μ_2 over $\mathbb{A} = \{0, 1\}$, 20 independent copies of $\widehat{\nu}_k$ based on $\ell_0 = 500$ simulation runs each).

Steele [15] conjectured that in the case of two random sequences with i.i.d. uniformly distributed characters over a finite alphabet $\mathbb{A}$ it be the case that $\gamma(0.5, 0.5) = 2/(1 + \sqrt{|\mathbb{A}|})$. However, in the case of a binary alphabet this would mean that $\gamma(0.5, 0.5) = 0.8284$, whereas Figure 4.4 shows that we are 95% confident that $\gamma(0.5, 0.5) \leq 0.81895$. Therefore, we believe that the Steele conjecture is wrong.

REFERENCES

[1] Kenneth S. Alexander. The rate of convergence of the mean length of the longest common

subsequence. *Ann. Appl. Probab.*, 4(4):1074–1082, 1994.

[2] Richard Arratia and Michael S. Waterman. A phase transition for the score in matching random sequences allowing deletions. *Ann. Appl. Probab.*, 4(1):200–225, 1994.

[3] K. Azuma. Weighted sums of certain dependent random variables. *Tohuku Math. J.*, 19:357–367, 1967.

[4] R.A. Baeza-Yates, R. Gavaldà, G. Navarro, and R. Scheihing. Bounding the expected length of longest common subsequences and forests. *Theory Comput. Syst.*, 32(4):435–452, 1999.

[5] J.M. Borwein and A.S. Lewis. *Convex Analysis and Nonlinear Optimization*. Springer, New York, 2000.

[6] Václáv Chvatal and David Sankoff. Longest common subsequences of two random sequences. *J. Appl. Probability*, 12:306–315, 1975.

[7] Vlado Dančík and Mike Paterson. Upper bounds for the expected length of a longest common subsequence of two binary sequences. *Random Structures Algorithms*, 6(4):449–458, 1995.

[8] Quentin Decouvelaere. *Upper Bounds for the LCS Problem*. MSc Thesis, Oxford University Computing Laboratory, 2003.

[9] Joseph G. Deken. Some limit results for longest common subsequences. *Discrete Math.*, 26(1):17–31, 1979.

[10] Frank den Hollander. *Large Deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

[11] Raphael Hauser, Servet Martinez, and Heinrich Matzinger. Large deviation based upper bounds for the lcs problem. Technical Report NA03/13 Oxford University Computing Laboratory, 2003.

[12] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statis. Assoc.*, 58:13–30, 1963.

[13] Mike Paterson and Vlado Dančík. Longest common subsequences. In *Mathematical foundations of computer science 1994 (Košice, 1994)*, volume 841 of *Lecture Notes in Comput. Sci.*, pages 127–142. Springer, Berlin, 1994.

[14] H. Robbins. A remark on stirling's formula. *Amer. Math. Monthly*, 62:26–29, 1955.

[15] Michael J. Steele. An efron-stein inequality for nonsymmetric statistics. *Ann. Statist.*, 14:753–758, 1986.

[16] Robert A. Wagner and Michael J. Fischer. The sring-to-string correction problem. *J. Ass. Comp. Mach.*, 21:168–173, 1974.

[17] M. Waterman. *Introduction to Computational Biology*. Chapman & Hall, 1995.