# Hidden Words Statistics for Large Patterns

**Svante Janson**
Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
svante.janson@math.uu.se

**Wojciech Szpankowski**
Center for Science of Information, Department of Computer Science, Purdue University, West Lafayette, IN, USA
spa@cs.purdue.edu

## Abstract

We study here the so called *subsequence pattern matching* also known as *hidden pattern matching* in which one searches for a given pattern $w$ of length $m$ as a *subsequence* in a random text of length $n$. The quantity of interest is the number of occurrences of $w$ as a subsequence (i.e., occurring in *not* necessarily consecutive text locations). This problem finds many applications from intrusion detection, to trace reconstruction, to deletion channel, and to DNA-based storage systems. In all of these applications, the pattern $w$ is of variable length. To the best of our knowledge this problem was only tackled for a fixed length $m = O(1)$ [6]. In our main result Theorem 5 we prove that for $m = o(n^{1/3})$ the number of subsequence occurrences is normally distributed. In addition, in Theorem 6 we show that under some constraints on the structure of $w$ the asymptotic normality can be extended to $m = o(\sqrt{n})$. For a special pattern $w$ consisting of the same symbol, we indicate that for $m = o(n)$ the distribution of number of subsequences is either asymptotically normal or asymptotically log normal. We conjecture that this dichotomy is true for all patterns. We use Hoeffding's projection method for $U$-statistics to prove our findings.

**2012 ACM Subject Classification** Mathematics of computing → Probability and statistics

**Keywords and phrases** Hidden pattern matching, subsequences, probability, U-statistics, projection method

**Digital Object Identifier** 10.4230/LIPIcs.AofA.2020.1

**Funding** *Svante Janson*: Supported by the Knut and Alice Wallenberg Foundation.
*Wojciech Szpankowski*: This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grant CCF-1524312.


31st International Conference on Probabilistic, Combinatorial and Asympt otic Methods for the Analysis of Algorithms (AofA 2020).
Editors: Michael Drmota and Clemens Heuberger; Article No. 1; pp. 1:1–1:15
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

# 1 Introduction and Motivation

One of the most interesting and least studied problem in pattern matching is known as the *subsequence string matching* or the *hidden pattern matching* [11]. In this case, we search for a pattern $w = w_1 w_2 \cdots w_m$ of length $m$ in the text $\Xi^n = \xi_1 \ldots \xi_n$ of length $n$ as *subsequence*, that is, we are looking for indices $1 \le i_1 < i_2 < \cdots < i_m \le n$ such that $\xi_{i_1} = w_1, \xi_{i_2} = w_2, \ldots, \xi_{i_m} = w_m$. We say that $w$ is *hidden* in the text $\Xi^n$. We do not put any constraints on the gaps $i_{j+1} - i_j$, so in language of [6] this is known as the *unconstrained* hidden pattern matching. The most interesting quantity of such a problem is the number of subsequence occurrences in the text generated by a random source. In this paper, we study the limiting distribution of this quantity when $m$, the length of the pattern, grows with $n$.

Hereafter, we assume that a memoryless source generates the text $\Xi$, that is, all symbols are generated independently with probability $p_a$ for symbol $a \in \mathcal{A}$, where the alphabet $\mathcal{A}$ is assumed to be finite. We denote by $p_w = \prod_j p_{w_j}$ the probability of the pattern $w$. Our goal is to understand the probabilistic behavior, in particular, the limiting distribution of the number of subsequence occurrences that we denote by $Z := Z_\Xi(w)$. It is known that the behavior of $Z$ depends on the order of magnitude of the pattern length $m$. For example, for the *exact pattern matching* (i.e., the pattern $w$ must occur as a *string* in consecutive positions of the text), the limiting distribution is normal for $m = O(1)$ (more precisely, when $np_w \to \infty$, hence up to $m = O(\log n)$), but it becomes a Pólya–Aeppli distribution when $np_w \to \lambda > 0$ for some constant $\lambda$, and finally (conditioned on being non-zero) it turns into a geometric distribution when $np_w \to 0$ [11] (see also [1]). We might expect a similar behaviour for the subsequence pattern matching. In [6] it was proved by analytic combinatoric methods that the number of subsequence occurrences, $Z_\Xi(w)$, is asymptotically normal when $m = O(1)$, and not much is known beyond this regime. (See also [2]. Asymptotic normality for fixed $m$ follows also by general results for $U$-statistics [9].) However, in many applications – as discussed below – we need to consider patterns $w$ whose lengths grow with $n$. In this paper, we prove two main results. In Theorem 5 we establish that for $m = o(n^{1/3})$ the number of subsequence occurrences is normally distributed. Furthermore, in Theorem 6 we show that under some constraints on the structure of $w$, the asymptotic normality can be extended to $m = o(\sqrt{n})$. Moreover, for the special pattern $w = a^m$ consisting of the same symbol repeated, we show in Theorem 4 that for $m = o(\sqrt{n})$, the distribution of number of occurrences is asymptotically normal, while for larger $m$ (up to $cn$ for some $c > 0$) it is asymptotically log-normal. We conjecture that this dichotomy is true for a large class of patterns.

Regarding methodology, unlike [6] we use here probabilistic tools. We first observe that $Z$ can be represented as a $U$-statistic (see (2)). This suggests to apply the **(author?)** [9] projection method to prove asymptotic normality of $Z$ for some large patterns. Indeed, we first decompose $Z$ into a sum of orthogonal random variables with variances of decreasing order in $n$ (for $m$ not too large), and show that the variable of the largest variance converges to a normal distribution, proving our main results Theorems 5 and 6.

The hidden pattern matching problem, especially for large patterns, finds many applications from intrusion detection, to trace reconstruction, to deletion channel, to DNA-based storage systems [8; 5; 3; 11; 16]. Here we discuss below in some detail two of them, namely the deletion channel and the trace reconstruction problem.

A deletion channel [5; 3; 4; 13; 16; 17] with parameter $d$ takes a binary sequence $\Xi^n = \xi_1 \cdots \xi_n$ where $\xi_i \in \mathcal{A}$ as input and deletes each symbol in the sequence independently with probability $d$. The output of such a channel is then a *subsequence* $\zeta = \zeta(x) = \xi_{i_1} ... \xi_{i_M}$ of

$\Xi$, where $M$ follows the binomial distribution $\mathrm{Binom}(n,(1-d))$, and the indices $i_1,...,i_M$ correspond to the bits that are *not* deleted. Despite significant effort [3; 13; 14; 16; 17] the mutual information between the input and output of the deletion channel and its capacity are still unknown. We hope to provide a more detailed characterization of the mutual information for memoryless sources using results of this and forthcoming papers. Indeed, it turns out that the mutual information $I(\Xi^n;\zeta(\Xi^n))$ can be exactly formulated as the problem of the subsequence pattern matching. In [5] it was proved that

$$I(\Xi^n;\zeta(\Xi^n)) = \sum_w d^{n-|w|}(1-d)^{|w|}\left(\mathbb{E}[Z_{\Xi^n}(w)\log Z_{\Xi^n}(w)] - \mathbb{E}[Z_{\Xi^n}(w)]\log \mathbb{E}[Z_{\Xi^n}(w)]\right), \quad (1)$$

where the sum is over all binary sequences of length smaller than $n$ and $Z_{\Xi^n}(w)$ is the number of subsequence occurrences of $w$ in the text $\Xi^n$. As one can see, to find precise asymptotics of the mutual information we need to understand the probabilistic behavior of $Z$ for $m \le n$ and typical $w$, which is our long term goal. The trace reconstruction problem [**?** 10; 15; 18] is related to the deletion channel problem since we are asking how many copies of the output deletion channel we need to see until we can reconstruct the input sequence with high probability.

# 2 Main Results

In this section we formulate precisely our problem and present our main results. Proofs are delayed till the next section.

## 2.1 Problem formulation and notation

We consider a random string $\Xi^n = \xi_1 \dots \xi_n$ of length $n$. We assume that $\xi_1, \xi_2, \dots$ are i.i.d. random letters from a finite alphabet $\mathcal{A}$; each letter $\xi_i$ has the distribution $\mathbb{P}(\xi_i = a) = p_a$ where $a \in \mathcal{A}$, for some given vector $\mathbf{p} = (p_a)_{a\in\mathcal{A}}$; we assume $p_a > 0$, $a \in \mathcal{A}$.

Let $w = w_1 \cdots w_m$ be a fixed string of length $m$ over the same alphabet $\mathcal{A}$. We assume $n \ge m$. Let $p_w := \prod_{j=1}^m p_{w_j}$, which is the probability that $\xi_1 \cdots \xi_m$ equals $w$.

Let $Z = Z_{n,w}(\xi_1 \cdots \xi_n)$ be the number of occurrences of $w$ as a subsequence of $\xi_1 \cdots \xi_n$. For a set $\mathcal{S}$ (in our case $[n]$ or $[m]$) and $k \ge 0$, let $\binom{\mathcal{S}}{k}$ be the collection of sets $\alpha \subseteq \mathcal{S}$ with $|\alpha| = k$. Thus, $\left|\binom{\mathcal{S}}{k}\right| = \binom{|\mathcal{S}|}{k}$. For $k = 0$, $\binom{\mathcal{S}}{0}$ contains just the empty set $\emptyset$. For $k = 1$, we identify $\binom{\mathcal{S}}{1}$ and $\mathcal{S}$ in the obvious way. We write $\alpha \in \binom{[n]}{k}$ as $\{\alpha_1, \dots, \alpha_k\}$, where we assume that $\alpha_1 < \cdots < \alpha_k$. Then

$$Z = \sum_{\alpha\in\binom{[n]}{m}} I_\alpha, \quad \text{where} \quad I_\alpha = \prod_{j=1}^m \mathbf{1}\{\xi_{\alpha_j} = w_j\}, \quad \alpha_1 < \ldots < \alpha_m. \quad (2)$$

▶ Remark 1. In the limit theorems, we are studying the asymptotic distribution of $Z$. We then assume that $n \to \infty$ and (usually) $m \to \infty$; we thus implicitly consider a sequence of words $w^{(n)}$ of lengths $m_n = |w^{(n)}|$. But for simplicity we do not show this in the notation.

We have $\mathbb{E}\, I_\alpha = p_w$ for every $\alpha$. Hence,

$$\mathbb{E}\, Z = \sum_{\alpha\in\binom{[n]}{m}} \mathbb{E}\, I_\alpha = \binom{n}{m} p_w. \quad (3)$$

Further, let $Y_\alpha := p_w^{-1} I_\alpha$, so $\mathbb{E}\, Y_\alpha = 1$, and

$$Z^* := p_w^{-1} Z = \sum_{\alpha\in\binom{[n]}{m}} Y_\alpha, \quad (4)$$

so $\mathbb{E}\, Z^* = \binom{n}{m}$ and

$$Z^* - \mathbb{E}\, Z^* = p_w^{-1} Z - \binom{n}{m} = \sum_{\alpha \in \binom{[n]}{m}} \bigl(Y_\alpha - 1\bigr). \tag{5}$$

We also write $\|Y\|_p := \bigl(\mathbb{E}\,|Y|^p\bigr)^{1/p}$ for the $L^p$ norm of a random variable $Y$, while $\|\mathbf{x}\|$ is the usual Euclidean norm of a vector $\mathbf{x}$ in some $\mathbb{R}^m$. $C$ denotes constants that may be different at different occurrences; they may depend on the alphabet $\mathcal{A}$ and $(p_a)_{a\in\mathcal{A}}$, but not on $n$, $m$ or $w$. Finally, $\xrightarrow{\mathrm{d}}$ and $\xrightarrow{\mathrm{p}}$ mean convergence in distribution and probability, respectively.

We are now ready to present our main results regarding the limiting distribution of $Z$, the number of subsequence $w = a_1, \dots a_m$ occurrences when $m \to \infty$. We start with a simple example, namely, $w = a^m = a \cdots a$ for some $a \in \mathcal{A}$, and show that depending on whether $m = o(\sqrt{n})$ or not the number of subsequences will follow asymptotically either the normal distribution or the log-normal distribution.

Before we present our results we consider asymptotically normal and log-normal distributions in general, and discuss their relation.

## 2.2 Asymptotic normality and log-normality

If $X_n$ is a sequence of random variables and $a_n$ and $b_n$ are sequences of real numbers, with $b_n > 0$, then $X_n \sim \mathrm{AsN}(a_n, b_n)$ means that

$$\frac{X_n - a_n}{\sqrt{b_n}} \xrightarrow{\mathrm{d}} N(0,1). \tag{6}$$

We say that $X_n$ is *asymptotically normal* if $X_n \sim \mathrm{AsN}(a_n, b_n)$ for some $a_n$ and $b_n$, and *asymptotically log-normal* if $\ln X_n \sim \mathrm{AsN}(a_n, b_n)$ for some $a_n$ and $b_n$ (this assumes $X_n \ge 0$). Note that these notions are equivalent when the asymptotic variance $b_n$ is small, as made precise by the following lemma.

▶ **Lemma 2.** *If $b_n \to 0$, and $a_n$ are arbitrary, then*

$$\ln X_n \sim \mathrm{AsN}(a_n, b_n) \iff X_n \sim \mathrm{AsN}(e^{a_n}, b_n e^{2a_n}). \tag{7}$$

**Proof.** By replacing $X_n$ by $X_n/e^{a_n}$, we may assume that $a_n = 0$. If $\ln X_n \sim \mathrm{AsN}(0, b_n)$ with $b_n \to 0$, then $\ln X_n \xrightarrow{\mathrm{p}} 0$, and thus $X_n \xrightarrow{\mathrm{p}} 1$. It follows that $\ln X_n/(X_n - 1) \xrightarrow{\mathrm{p}} 1$ (with $0/0 := 1$), and thus

$$\frac{X_n - 1}{b_n^{1/2}} = \frac{X_n - 1}{\ln X_n} \frac{\ln X_n}{b_n^{1/2}} \xrightarrow{\mathrm{d}} N(0,1), \tag{8}$$

and thus $X_n \sim \mathrm{AsN}(1, b_n)$. The converse is proved by the same argument. ◀

▶ Remark 3. Lemma 2 is best possible. Suppose that $\ln X_n \sim \mathrm{AsN}(a_n, b_n)$. If $b_n \to b > 0$, then $\ln\bigl(X_n/e^{a_n}\bigr) = \ln X_n - a_n \xrightarrow{\mathrm{d}} N(0,b)$, and thus

$$X_n/e^{a_n} \xrightarrow{\mathrm{d}} e^{\zeta_b}, \qquad \zeta_b \sim N(0,b). \tag{9}$$

In this case (and only in this case), $X_n$ thus converges in distribution, after scaling, to a log-normal distribution. If $b_n \to \infty$, then no linear scaling of $X_n$ can converge in distribution to a non-degenerate limit, as is easily seen.

### 2.3 A simple example

We consider first a simple example where the asymptotic distribution can be found easily by explicit calculations. Fix $a \in \mathcal{A}$ and let $w = a^m = a \cdots a$, a string with $m$ identical letters. Then, if $N = N_a$ is the number of occurrences of $a$ in $\xi_1 \cdots \xi_n$, then

$$Z = \binom{N_a}{m}. \tag{10}$$

We will show that $Z$ is asymptotically normal if $m$ is small, and log-normal for larger $m$.

▶ **Theorem 4.** *Suppose that $m < np_a$, with $np_a - m \gg n^{1/2}$.*
(i) *Then*

$$\ln Z \sim \mathrm{AsN}\Bigl(\ln \binom{np_a}{m}, n\Bigl|\ln\Bigl(1-\frac{m}{np_a}\Bigr)\Bigr|^2 p_a(1-p_a)\Bigr). \tag{11}$$

(ii) *In particular, if $m = o(n)$, then*

$$\ln Z \sim \mathrm{AsN}\Bigl(\ln \binom{np_a}{m}, (p_a^{-1}-1)\frac{m^2}{n}\Bigr). \tag{12}$$

(iii) *If $m = o(n^{1/2})$, then this implies*

$$Z/\mathbb{E}\,Z \sim \mathrm{AsN}\Bigl(1, (p_a^{-1}-1)\frac{m^2}{n}\Bigr), \tag{13}$$

*and thus*

$$Z \sim \mathrm{AsN}\Bigl(\mathbb{E}\,Z, (p_a^{-1}-1)\frac{m^2}{n}(\mathbb{E}\,Z)^2\Bigr). \tag{14}$$

**Proof.** (i) We have $N_a \sim \mathrm{Bin}(n, p_a)$. Define $Y := N_a - np_a$. Then, by the Central Limit Theorem,

$$Y \sim \mathrm{AsN}\bigl(0, np_a(1-p_a)\bigr). \tag{15}$$

By (10), we have

$$\begin{aligned}\ln Z - \ln\binom{np_a}{m} &= \ln\binom{np_a+Y}{m} - \ln\binom{np_a}{m}\\ &= \ln\Gamma(np_a+Y+1) - \ln\Gamma(np_a+Y-m+1) - \ln m!\\ &\quad - \bigl(\ln\Gamma(np_a+1) - \ln\Gamma(np_a-m+1) - \ln m!\bigr)\\ &= \int_{y=0}^{Y}\int_{x=-m}^{0} (\ln\Gamma)''(np_a+x+y+1)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}\tag{16}$$

We fix a sequence $\omega_n \to \infty$ such that $np_a - m \gg \omega_n \gg n^{1/2}$; this is possible by the assumption. Note that (15) implies that $Y/\omega_n \xrightarrow{\mathrm{p}} 0$, and thus $\mathbb{P}(|Y| \le \omega_n) \to 1$. We may thus in the sequel assume $|Y| \le \omega_n$. We assume also that $n$ is so large that $np_a - m \ge 2\omega_n > 0$.

Stirling's formula implies, by taking the logarithm and differentiating twice (in the complex half-plane $\operatorname{Re} z > \frac{1}{2}$, say)

$$(\ln\Gamma)''(x) = \frac{1}{x} + O\Bigl(\frac{1}{x^2}\Bigr) = \frac{1}{x}\Bigl(1 + O\Bigl(\frac{1}{x}\Bigr)\Bigr), \qquad x \ge 1. \tag{17}$$

Consequently, (16) yields, noting the assumptions just made imply $|Y| \le \omega_n \le \frac{1}{2}(np_a - m)$,

$$\begin{aligned}
\ln Z - \ln \binom{np_a}{m} &= \int_{y=0}^{Y}\int_{x=-m}^{0} \frac{1}{np_a + x + y + 1}\Bigl(1 + O\Bigl(\frac{1}{np_a - m}\Bigr)\Bigr)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{y=0}^{Y}\int_{x=-m}^{0} \frac{1}{np_a + x}\Bigl(1 + O\Bigl(\frac{\omega_n}{np_a - m}\Bigr)\Bigr)\,\mathrm{d}x\,\mathrm{d}y \\
&= \Bigl(1 + O\Bigl(\frac{\omega_n}{np_a - m}\Bigr)\Bigr)Y\int_{x=-m}^{0} \frac{1}{np_a + x}\,\mathrm{d}x \\
&= \bigl(1 + o(1)\bigr)Y\ln\frac{np_a}{np_a - m}. \qquad (18)
\end{aligned}$$

Consequently, using also (15), we obtain

$$\frac{\ln Z - \ln\binom{np_a}{m}}{n^{1/2}\bigl|\ln\bigl(1 - \frac{m}{np_a}\bigr)\bigr|} = \bigl(1 + o_{\mathrm{p}}(1)\bigr)\frac{Y}{n^{1/2}} \xrightarrow{\mathrm{d}} N\bigl(0, p_a(1 - p_a)\bigr), \qquad (19)$$

which is equivalent to (11).

(ii) If $m = o(n)$, then $\bigl|\ln\bigl(1 - \frac{m}{np_a}\bigr)\bigr| \sim \frac{m}{np_a}$, and (12) follows.

(iii) If $m = o(n^{1/2})$, then (ii) applies, so (12) holds; hence Lemma 2 implies

$$Z/\binom{np_a}{m} \sim \mathrm{AsN}\Bigl(1, (p_a^{-1} - 1)\frac{m^2}{n}\Bigr). \qquad (20)$$

Furthermore,

$$\mathbb{E}\, Z = \binom{n}{m}p_a^m = \frac{n^m e^{O(m^2/n)}}{m!}p_a^m \sim \frac{n^m}{m!}p_a^m \qquad (21)$$

and, similarly, $\binom{np_a}{m} \sim \frac{n^m p_a^m}{m!}$. Hence, $\mathbb{E}\, Z \sim \binom{np_a}{m}$ and (13) follows from (20); (14) is an immediate consequence. ◀

## 2.4 General results

We now present our main results. However, first we discuss the road map of our approach. First, we observe that the representation (2) shows that $Z$ can be viewed as a $U$-statistic. For convenience, we consider $Z^*$ in (4), which differs from $Z$ by a constant factor only, and show in (41) that $Z^* - \mathbb{E}\, Z^*$ can be decomposed into a sum $\sum_{\ell=1}^{m} V_\ell$ of orthogonal random variables $V_\ell$ such that, when $m$ is not too large, $\mathrm{Var}\bigl(\sum_{\ell=2}^{m} V_\ell\bigr) = o(\mathrm{Var}\, V_1)$. Next, in Lemma 11 we prove that $V_1$ appropriately normalized converges to the standard normal distribution. This will allow us to conclude the asymptotic normality of $Z$.

In this paper, we only consider the region $m = o\bigl(n^{1/2}\bigr)$. First, for $m = o\bigl(n^{1/3}\bigr)$ we claim that the number of subsequence occurrences always is asymptotically normal.

▶ **Theorem 5.** *If $m = o\bigl(n^{1/3}\bigr)$, then*

$$Z \sim \mathrm{AsN}\Bigl(\binom{n}{m}p_w, \sigma_1^2 p_w^2\Bigr), \qquad (22)$$

*where*

$$\sigma_1^2 = \sum_{i=1}^{n}\sum_{a\in\mathcal{A}} p_a^{-1}\left(\sum_{j:\ w_j=a}\binom{i-1}{j-1}\binom{n-i}{m-j}\right)^2 - n\binom{n-1}{m-1}^2. \qquad (23)$$

*Furthermore, $\mathbb{E}\, Z = \binom{n}{m}p_w$ and $\mathrm{Var}\, Z \sim p_w^2\sigma_1^2$.*

In the second main result, we restrict the patterns $w$ to such that are not typical for the random text; however, we will allow $m = o(n^{1/2})$.

▶ **Theorem 6.** *Let $\mathbf{q} = (q_a)_{a\in\mathcal{A}}$ be the proportions of the letters in $w$, i.e., $q_a := \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}\{w_j = a\}$. Suppose that $\liminf_{n\to\infty} \|\mathbf{q}-\mathbf{p}\| > 0$. If further $m = o(n^{1/2})$, then the asymptotic normality (22) holds.*

# 3 Analysis and Proofs

In this section we will prove our main results. We start with some preliminaries.

## 3.1 Preliminaries and more notation

Let, for $a \in \mathcal{A}$,

$$\varphi_a(x) := p_a^{-1}\mathbf{1}\{x = a\} - 1. \tag{24}$$

Thus, letting $\xi$ be any random variable with the distribution of $\xi_i$,

$$\mathbb{E}\,\varphi_a(\xi) = 0, \qquad a \in \mathcal{A}. \tag{25}$$

Let $p_* := \min_a p_a$ and

$$B := p_*^{-1} - 1. \tag{26}$$

▶ **Lemma 7.** *Let $\varphi_a$ and $B$ be as above.*
(i) *For every $a \in \mathcal{A}$,*

$$\mathbb{E}\big[\varphi_a(\xi)^2\big] = p_a^{-1} - 1 \le B. \tag{27}$$

(ii) *For some $c_1 > 0$ and every $a \in \mathcal{A}$,*

$$\|\varphi_a(\xi)\|_2 = \big(p_a^{-1} - 1\big)^{1/2} \ge c_1. \tag{28}$$

(iii) *For any vector $\mathbf{r} = (r_a)_{a\in\mathcal{A}}$ with $\sum_a r_a = 1$,*

$$\Big\|\sum_{a\in\mathcal{A}} r_a\varphi_a(\xi)\Big\|_2 \ge \|\mathbf{r}-\mathbf{p}\| := \Big(\sum_{a\in\mathcal{A}} |r_\alpha - p_\alpha|^2\Big)^{1/2}. \tag{29}$$

**Proof.** The definition (24) yields

$$\mathbb{E}\big[\varphi_a(\xi)^2\big] = p_a^{-2}\operatorname{Var}\big[\mathbf{1}\{\xi = a\}\big] = p_a^{-2}p_a(1-p_a) = p_a^{-1} - 1. \tag{30}$$

Hence, (27) and (28) follow, with $B$ given by (26).

Finally, for every $x \in \mathcal{A}$, by (24) again,

$$\sum_{a\in\mathcal{A}} r_a\varphi_a(x) = r_x p_x^{-1} - \sum_{a\in\mathcal{A}} r_a = r_x/p_x - 1 \tag{31}$$

and thus

$$\mathbb{E}\Big(\sum_{a\in\mathcal{A}} r_a\varphi_a(\xi)\Big)^2 = \sum_{a\in\mathcal{A}} p_a\big(r_a/p_a - 1\big)^2 = \sum_{a\in\mathcal{A}} p_a^{-1}\big(r_a - p_a\big)^2 \tag{32}$$

and (29) follows. ◀

## 3.2 A decomposition

The representation (2) shows that $Z$ is a special case of a $U$-statistic. For fixed $m$, the general theory of **(author?)** [9] applies and yields asymptotic normality. (Cf. [12, Section 4] for a related problem.) For $m \to \infty$ (our main interest), we can still use the orthogonal decomposition of [9], which in our case takes the following form.

By the definitions in Section 2.1 and (24),

$$Y_\alpha = \prod_{j=1}^{m} \bigl(p_{w_j}^{-1} \mathbf{1}\{\xi_{\alpha_j} = w_j\}\bigr) = \prod_{j=1}^{m} \bigl(\varphi_{w_j}(\xi_{\alpha_j}) + 1\bigr). \tag{33}$$

By multiplying out this product, we obtain

$$Y_\alpha = \sum_{\gamma \subseteq [m]} \prod_{j \in \gamma} \varphi_{w_j}(\xi_{\alpha_j}). \tag{34}$$

Hence,

$$Z^* = \sum_{\alpha \in \binom{[n]}{m}} Y_\alpha = \sum_{\alpha \in \binom{[n]}{m}} \sum_{\gamma \subseteq [m]} \prod_{j \in \gamma} \varphi_{w_j}(\xi_{\alpha_j}) = \sum_{\alpha \in \binom{[n]}{m}} \sum_{\gamma \subseteq [m]} \prod_{k=1}^{|\gamma|} \varphi_{w_{\gamma_k}}(\xi_{\alpha_{\gamma_k}}). \tag{35}$$

We rearrange this sum. First, let $\ell := |\gamma| \in [m]$, and consider all terms with a given $\ell$. For each $\alpha$ and $\gamma$, with $|\gamma| = \ell$, let

$$\alpha_\gamma := \{\alpha_{\gamma_1}, \dots, \alpha_{\gamma_\ell}\} \in \binom{[n]}{\ell}. \tag{36}$$

For given $\gamma \in \binom{[m]}{\ell}$ and $\beta \in \binom{[n]}{\ell}$, the number of $\alpha \in \binom{[n]}{m}$ such that $\alpha_\gamma = \beta$ equals the number of ways to choose, for each $k \in [\ell+1]$, $\gamma_k - \gamma_{k-1} - 1$ elements of $\alpha$ in a gap of length $\beta_k - \beta_{k-1} - 1$, where we define $\beta_0 = \gamma_0 = 0$ and $\beta_{\ell+1} = n+1$, $\gamma_{\ell+1} = m+1$; this number is

$$c(\beta, \gamma) := \prod_{k=1}^{\ell+1} \binom{\beta_k - \beta_{k-1} - 1}{\gamma_k - \gamma_{k-1} - 1}. \tag{37}$$

Consequently, combining the terms in (35) with the same $\alpha_\gamma$,

$$Z^* = \sum_{\ell=0}^{m} \sum_{\gamma \in \binom{[m]}{\ell}} \sum_{\beta \in \binom{[n]}{\ell}} c(\beta, \gamma) \prod_{k=1}^{\ell} \varphi_{w_{\gamma_k}}(\xi_{\beta_k}). \tag{38}$$

We define, for $0 \le \ell \le m$ and $\beta \in \binom{[n]}{\ell}$,

$$V_{\ell,\beta} := \sum_{\gamma \in \binom{[m]}{\ell}} c(\beta, \gamma) \prod_{k=1}^{\ell} \varphi_{w_{\gamma_k}}(\xi_{\beta_k}) \tag{39}$$

and

$$V_\ell := \sum_{\beta \in \binom{[n]}{\ell}} V_{\ell,\beta}. \tag{40}$$

Thus (38) yields the decomposition

$$Z^* = \sum_{\ell=0}^{m} V_\ell. \tag{41}$$

For $\ell = 0$, $\binom{[n]}{0}$ contains only the empty set $\emptyset$, and

$$V_0 = V_{0,\emptyset} = \binom{n}{m} = \mathbb{E}\, Z^*. \tag{42}$$

Furthermore, note that two summands in (38) with different $\beta$ are orthogonal, as a consequence of (25) and independence of different $\xi_i$. Consequently, the variables $V_{\ell,\beta}$ ($\ell \in [m]$, $\beta \in \binom{[n]}{\ell}$) are orthogonal, and hence the variables $V_\ell$ ($\ell = 0, \dots, m$) are orthogonal.

Let

$$\sigma_\ell^2 := \operatorname{Var}(V_\ell) = \mathbb{E}\, V_\ell^2 = \sum_{\beta \in \binom{[m]}{\ell}} \mathbb{E}\, V_{\ell,\beta}^2, \qquad 1 \le \ell \le m. \tag{43}$$

Note also that by the combinatorial definition of $c(\beta,\gamma)$ given before (37), we see that

$$\sum_{\beta \in \binom{[n]}{\ell}} c(\beta,\gamma) = \binom{n}{m}, \tag{44}$$

since this is just the number of $\alpha \in \binom{[n]}{m}$, and

$$\sum_{\gamma \in \binom{[m]}{\ell}} c(\beta,\gamma) = \binom{n-\ell}{m-\ell}, \tag{45}$$

since this sum is the total number of ways to choose $m - \ell$ elements of the $n - \ell$ elements of $\alpha$ in the gaps.

## 3.3 The projection method

We use the projection method used by **(author?)** [9] to prove asymptotic normality for $U$-statistics. Translated to the present setting, the idea of the projection method is to approximate $Z^* - \mathbb{E}\, Z^* = Z^* - V_0$ by $V_1$, thus ignoring all terms with $\ell \ge 2$ in the sum in (41). In order to do this, we estimate variances.

First, by (27) and the independence of the $\xi_i$,

$$\Bigl\| \prod_{k=1}^{\ell} \varphi_{w_{\gamma_k}}(\xi_{\beta_k}) \Bigr\|_2 = \Bigl( \prod_{k=1}^{\ell} \mathbb{E} \bigl| \varphi_{w_{\gamma_k}}(\xi_{\beta_k}) \bigr|^2 \Bigr)^{1/2} \le B^{\ell/2}. \tag{46}$$

By Minkowski's inequality, (39), (46) and (45),

$$\bigl\| V_{\ell,\beta} \bigr\|_2 \le \sum_{\gamma \in \binom{[m]}{\ell}} c(\beta,\gamma) B^{\ell/2} = B^{\ell/2} \binom{n-\ell}{m-\ell} \tag{47}$$

or, equivalently,

$$\mathbb{E}\, V_{\ell,\beta}^2 \le B^\ell \binom{n-\ell}{m-\ell}^2. \tag{48}$$

This leads to the following estimates.

▶ **Lemma 8.** *For $1 \le \ell \le m$,*

$$\sigma_\ell^2 := \mathbb{E}\, V_\ell^2 \le \widehat{\sigma}_\ell^2 := B^\ell \binom{n}{\ell} \binom{n-\ell}{m-\ell}^2. \tag{49}$$

**Proof.** The definition of $V_\ell$ in (40) and (48) yield, since the summands $V_{\ell,\beta}$ are orthogonal,

$$\sigma_\ell^2 := \mathbb{E}\, V_\ell^2 = \sum_{\beta\in\binom{[n]}{\ell}} \mathbb{E}\, V_{\ell,\beta}^2 \le \binom{n}{\ell} B^\ell \binom{n-\ell}{m-\ell}^2, \tag{50}$$

as needed. ◀

Note that, for $1 \le \ell < m$,

$$\frac{\widehat{\sigma}_{\ell+1}^2}{\widehat{\sigma}_\ell^2} = B\frac{\binom{n}{\ell+1}\binom{n-\ell-1}{m-\ell-1}^2}{\binom{n}{\ell}\binom{n-\ell}{m-\ell}^2} = B\frac{n-\ell}{\ell+1}\Bigl(\frac{m-\ell}{n-\ell}\Bigr)^2 \le B\frac{m^2}{(\ell+1)n}. \tag{51}$$

▶ **Lemma 9.** *If $m \le B^{-1/2}n^{1/2}$, then*

$$\operatorname{Var}\bigl(Z^* - V_1\bigr) \le B^2 m^2 \binom{n-1}{m-1}^2. \tag{52}$$

**Proof.** By (51) and the assumption, for $1 \le \ell < m$,

$$\frac{\widehat{\sigma}_{\ell+1}^2}{\widehat{\sigma}_\ell^2} \le \frac{1}{\ell+1} \le \frac{1}{2}, \tag{53}$$

and thus, summing a geometric series,

$$\begin{aligned}\operatorname{Var}\bigl(Z^* - V_1\bigr) &= \sum_{\ell=2}^{m} \operatorname{Var}\bigl(V_\ell\bigr) \le \sum_{\ell=2}^{m} \widehat{\sigma}_\ell^2 \le \sum_{\ell=2}^{m} 2^{2-\ell}\widehat{\sigma}_2^2 \le 2\widehat{\sigma}_2^2 \\ &= B^2 n(n-1)\binom{n-2}{m-2}^2 \le B^2 m^2 \binom{n-1}{m-1}^2. \end{aligned} \tag{54}$$

◀

## 3.4 The first term $V_1$

For $\ell = 1$, we identify $\binom{[n]}{\ell}$ and $[n]$, and we write $V_{1,i} := V_{1,\{i\}}$. Note that, by (37),

$$c(i,j) := c\bigl(\{i\},\{j\}\bigr) = \binom{i-1}{j-1}\binom{n-i}{m-j}. \tag{55}$$

Thus (40) and (39) become

$$V_1 = \sum_{i=1}^{n} V_{1,i} \tag{56}$$

with, using (55),

$$V_{1,i} = \sum_{j=1}^{m} c(i,j)\varphi_{w_j}(\xi_i) = \sum_{j=1}^{m} \binom{i-1}{j-1}\binom{n-i}{m-j}\varphi_{w_j}(\xi_i). \tag{57}$$

Note that $V_{1,i}$ is a function of $\xi_i$, and thus the random variables $V_{1,i}$ are independent. Furthermore, (25) implies $\mathbb{E}\, V_{1,i} = 0$. Let $\tau_i^2 := \operatorname{Var} V_{1,i} = \mathbb{E}\, V_{1,i}^2$. Then, see (43),

$$\sigma_1^2 = \operatorname{Var} V_1 = \sum_{i=1}^{n} \operatorname{Var} V_{1,i} = \sum_{i=1}^{n} \tau_i^2. \tag{58}$$

Observe that it follows from (57) and (24) that

$$\tau_i^2 = \sum_{a\in\mathcal{A}} p_a^{-1}\Biggl(\sum_{j:\ w_j=a}\binom{i-1}{j-1}\binom{n-i}{m-j}\Biggr)^2 - \binom{n-1}{m-1}^2. \tag{59}$$

Taking $\ell = 1$ in (48) yields the upper bound

$$\tau_i^2 = \mathbb{E}\,V_{1,i}^2 \le B\binom{n-1}{m-1}^2, \qquad i\in[n]. \tag{60}$$

Summing over $i$, or using (49), we obtain

$$\sigma_1^2 := \mathbb{E}\,V_1^2 \le \widehat{\sigma}_1^2 := Bn\binom{n-1}{m-1}^2. \tag{61}$$

We notice that the upper bound is achievable. Indeed, for $w = a\cdots a$, by (59) and (58),

$$\tau_i^2 = (p_a^{-1}-1)\binom{n-1}{m-1}^2, \qquad\qquad \sigma_1^2 = n(p_a^{-1}-1)\binom{n-1}{m-1}^2. \tag{62}$$

We show also a general lower bound.

▶ **Lemma 10.** *There exists $c, c' > 0$ such that*

$$\sigma_1^2 \ge \frac{c}{m}\widehat{\sigma}_1^2 = c'\frac{n}{m}\binom{n-1}{m-1}^2. \tag{63}$$

**Proof.** We consider the first term in the sum in (57) separately, and write

$$V_{1,i} = c(i,1)\varphi_{w_1}(\xi_i) + V'_{1,i}, \tag{64}$$

where

$$V'_{1,i} := \sum_{j=2}^{m} c(i,j)\varphi_{w_j}(\xi_i). \tag{65}$$

We have, by (55), $c(i,1) = \binom{n-i}{m-1}$. Consequently, for any $i\in[n]$,

$$\begin{aligned}\frac{c(i,1)}{c(1,1)} &= \frac{\binom{n-i}{m-1}}{\binom{n-1}{m-1}} = \frac{\prod_{k=0}^{m-2}(n-i-k)}{\prod_{k=0}^{m-2}(n-1-k)} = \prod_{k=0}^{m-2}\Bigl(1-\frac{i-1}{n-1-k}\Bigr)\\ &\ge 1-\sum_{k=0}^{m-2}\frac{i-1}{n-1-k} \ge 1-\frac{m(i-1)}{n-m+1}.\end{aligned} \tag{66}$$

Let $\delta \le 1/4$ be a fixed small positive number, chosen later. Assume that $i \le 1+\delta n/m$. In particular, either $i=1$ or $m \le m(i-1) \le \delta n < n/2$, and thus (66) implies

$$\frac{c(i,1)}{c(1,1)} \ge 1-\frac{m(i-1)}{n-m} \ge 1-\frac{\delta n}{n/2} = 1-2\delta. \tag{67}$$

By (45), (67) implies

$$\sum_{j=2}^{m} c(i,j) = \binom{n-1}{m-1} - c(i,1) = c(1,1)-c(i,1) \le 2\delta c(1,1). \tag{68}$$

Hence, by (65), Minkowski's inequality and (27), cf. (47),

$$\|V'_{1,i}\|_2 \le \sum_{j=2}^{m} c(i,j)\|\varphi_{w_j}(\xi_i)\|_2 \le \sum_{j=2}^{m} c(i,j)B^{1/2} \le 2\delta B^{1/2}c(1,1). \tag{69}$$

Furthermore, (28) and (67) yield

$$\|c(i,1)\varphi_{w_1}(\xi_i)\|_2 \ge c(i,1)c_1 \ge c_1(1-2\delta)c(1,1) \ge \tfrac{1}{2}c_1c(1,1). \tag{70}$$

Finally, (64) and the triangle inequality yield, using (70) and (69),

$$\|V_{1,i}\|_2 \ge \|c(i,1)\varphi_{w_1}(\xi_i)\|_2 - \|V'_{1,i}\|_2 \ge \big(\tfrac{1}{2}c_1 - 2\delta B^{1/2}\big)c(1,1). \tag{71}$$

We now choose $\delta := c_1/(8B^{1/2})$, and find that for some $c_2 > 0$,

$$\tau_i^2 := \|V_{1,i}\|_2^2 \ge c_2c(1,1)^2, \qquad i \le 1 + \delta n/m. \tag{72}$$

Consequently, by (58),

$$\sigma_1^2 = \sum_{i=1}^{n} \tau_i^2 \ge \frac{\delta n}{m}c_2c(1,1)^2 = c_3\frac{n}{m}\binom{n-1}{m-1}^2. \tag{73}$$

This proves (63), with $c' := c_3$ and $c = c'/B$. ◀

The next lemma is proved in the Appendix in which we verify Lyapunov's condition to prove asymptotic normality of $V_1$.

▶ **Lemma 11.** *Suppose that $m = o(n)$. Then $V_1$ is asymptotically normal:*

$$V_1/\sigma_1 \xrightarrow{\mathrm{d}} N(0,1). \tag{74}$$

## 3.5 Proofs of Theorem 5 and 6

We next prove a general theorem showing asymptotic normality under some conditions.

▶ **Theorem 12.** *Suppose that $n \to \infty$ and that*

$$m^2\binom{n-1}{m-1}^2 = o\big(\sigma_1^2\big). \tag{75}$$

*Then*

$$\operatorname{Var} Z = p_w^2 \operatorname{Var} Z^* \sim p_w^2\sigma_1^2 \tag{76}$$

*and*

$$\frac{Z^* - \mathbb{E}\,Z^*}{\sigma_1} \xrightarrow{\mathrm{d}} N(0,1), \tag{77}$$

$$\frac{Z - \mathbb{E}\,Z}{(\operatorname{Var} Z)^{1/2}} = \frac{Z^* - \mathbb{E}\,Z^*}{(\operatorname{Var} Z^*)^{1/2}} \xrightarrow{\mathrm{d}} N(0,1). \tag{78}$$

**Proof.** By Lemma 9 and (75),

$$\operatorname{Var}\Big(\frac{Z^* - V_1}{\sigma_1}\Big) = \frac{\operatorname{Var}(Z^* - V_1)}{\sigma_1^2} \le B^2\frac{m^2\binom{n-1}{m-1}^2}{\sigma_1^2} = o(1). \tag{79}$$

Hence, recalling $\mathbb{E}\,V_1 = 0$,

$$\frac{Z^* - \mathbb{E}\,Z^* - V_1}{\sigma_1} \xrightarrow{\text{p}} 0. \tag{80}$$

Combining (74) and (80), we obtain (77).

Furthermore, by (79), and since the terms in (41) are orthogonal,

$$\operatorname{Var} Z^* = \operatorname{Var} V_1 + \operatorname{Var}\bigl(Z^* - V_1\bigr) = \sigma_1^2 + o(\sigma_1^2) \sim \sigma_1^2, \tag{81}$$

which yields (76), and also shows that we may replace $\sigma_1$ by $(\operatorname{Var} Z^*)^{1/2}$ in (77), which yields (78); the equality in (78) is a trivial consequence of (4). ◀

Now we are ready to prove our main results.

**Proof of Theorem 5.** By Lemma 10,

$$\frac{m^2\binom{n-1}{m-1}^2}{\sigma_1^2} \le C\frac{m^3}{n} = o(1). \tag{82}$$

Thus (75) holds, and the result follows by Theorem 12 together with (3) and (4). ◀

Recall that in Theorem 6, the range of $m$ is improved, assuming that $w$ is *not* typical for the random source with probabilities $\mathbf{p} = (p_a)_{a\in\mathcal{A}}$ that we consider.

**Proof of Theorem 6.** By Theorem 12, with (75) verified by Lemma 13 below. ◀

▶ **Lemma 13.** *Let $\mathbf{q} = (q_a)_{a\in\mathcal{A}}$ be the proportions of the letters in $w$. Then*

$$\sigma_1^2 \ge \frac{m^2}{n}\binom{n}{m}^2 \|\mathbf{q}-\mathbf{p}\|^2 = n\binom{n-1}{m-1}^2 \|\mathbf{q}-\mathbf{p}\|^2. \tag{83}$$

**Proof.** Let

$$\psi_i(x) := \sum_{j=1}^m c(i,j)\varphi_{w_j}(x). \tag{84}$$

Thus (57) is $V_{1,i} = \psi_i(\xi_i)$, and (58) is, since $\mathbb{E}\,\psi_i(\xi) = 0$,

$$\sigma_1^2 = \operatorname{Var} V_1 = \sum_{i=1}^n \mathbb{E}\bigl[\psi_i(\xi_i)^2\bigr] = \mathbb{E}\sum_{i=1}^n \psi_i(\xi)^2. \tag{85}$$

Hence, by the Cauchy–Schwarz inequality,

$$n\sigma_1^2 = n\,\mathbb{E}\sum_{i=1}^n \psi_i(\xi)^2 \ge \mathbb{E}\Bigl(\sum_{i=1}^n \psi_i(\xi)\Bigr)^2. \tag{86}$$

Furthermore, by (84) and (44)

$$\sum_{i=1}^n \psi_i(x) = \sum_{i=1}^n\sum_{j=1}^m c(i,j)\varphi_{w_j}(x) = \sum_{j=1}^m \binom{n}{m}\varphi_{w_j}(x) = \binom{n}{m}\sum_{a\in\mathcal{A}} mq_a\varphi_a(x). \tag{87}$$

Hence, (29) yields

$$\Bigl\|\sum_{i=1}^n \psi_i(\xi)\Bigr\|_2 = m\binom{n}{m}\Bigl\|\sum_{a\in\mathcal{A}} q_a\varphi_a(\xi)\Bigr\|_2 \ge m\binom{n}{m}\|\mathbf{q}-\mathbf{p}\|. \tag{88}$$

Combining (86) and (88) yields (83). ◀

## References


1. E. Bender and F. Kochman. The distribution of subword counts is usually normal. *European J. Combin.*, 14:265–275, 1993.
2. J. Bourdon and B. Vallée. Generalized pattern matching statistics. In *Mathematics and Computer Science II (Versailles, 2002)*, Trends. Math., pages 249–265. Birkhäuser, 2002.
3. S. Diggavi and M. Grossglauser. Information transmission over finite buffer channels. *IEEE Trans. Information Theory*, 52:1226–1237, 2006.
4. R. L. Dobrushin. Shannon's theorem for channels with synchronization errors. *Prob. Info. Trans.*, pages 18–36, 1967.
5. M. Drmota, K. Viswanathan, and W. Szpankowski. Mutual information for a deletion channel. In *IEEE International Symposium on Information Theory*, 2012.
6. P. Flajolet, W. Szpankowski, and B Vallée. Hidden word statistics. *J. ACM*, 53(1):147–183, 2006. doi:http://doi.acm.org/10.1145/1120582.1120586.
7. Allan Gut. *Probability: A Graduate Course.* Springer, New York, 2013.
8. R. Gwadera, M. Atallah, and W. Szpankowski. Reliable detection of episodes in event sequences. In *3rd IEEE Conf. on Data Mining*, pages 67–74. IEEE Computer Soc., 2003.
9. W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Mat. Statistics*, 19:293–325, 1984.
10. N. Holden and R. Lyones. Lower bounds for trace reconstruction. arXiv:1808.02336, 2018.
11. P. Jacquet and W. Szpankowski. *Analytic Pattern Matching: From DNA to Tiwitter*. Cambridge University Press, 2015.
12. S. Janson, B. Nakamura, and D. Zeilberger. On the asymptotic statistics of the number of occurrences of multiple permutation patterns. *J. Comb.*, 6:117–143, 2015.
13. A. Kalai, M. Mitzenmacher, and M. Sudan. Tight asymptotic bounds for the deletion channel with small deletion probabilities. In *IEEE International Symposium on Information Theory*, 2010.
14. Y. Kanoria and A. Montanari. On the deletion channel with small deletion probability. In *IEEE International Symposium on Information Theory*, 2010.
15. A. McGregor, E. Price, and S. Vorotnikova. Trace reconstruction revisisted. In *European Symposium on Algorithms*, pages 689–700, 2014.
16. M. Mitzenmacher. A survey of results for deletion channels and related synchronization channels. *Probab. Surveys*, pages 1–33, 2009.
17. R. Venkataramanan, S. Tatikonda, and K. Ramchandran. Achievable rates for channels with deletions and insertions. In *IEEE International Symposium on Information Theory*, 2011.
18. Y.Peres and A. Zhai. Average-case reconstruction for the deletion channel: subpolynomially many traces suffice. In *FOCS*. 2017.

# Appendix

## 3.6 Proof of Lemma 11

We show that the central limit theorem applies to the sum $V_1 = \sum_i V_{1,i}$ in (56). The terms $V_{1,i}$ are independent and have means $\mathbb{E}\, V_{1,i} = 0$. We verify Lyapunov's condition.

The random variable $\xi$ is defined on some probability space $(\Omega, \mathcal{F}, P)$ and takes values in the finite set $\mathcal{A}$. Thus the linear space $\mathcal{V}$ of functions $\Omega \to \mathbb{R}$ of the form $f(\xi)$ has finite dimension $|\mathcal{A}|$. Moreover, every function in $\mathcal{V}$ is bounded. The $L^2$ and $L^3$ norms $\|\cdot\|_2$ and $\|\cdot\|_3$ are thus finite on $\mathcal{V}$, and are thus both norms on the finite-dimensional vector space $\mathcal{V}$; hence there exists a constant $C$ such that for any function $f$,

$$\|f(\xi)\|_3 \le C\|f(\xi)\|_2. \tag{89}$$

In particular, since the definition (57) shows that $V_{1,i}$ is a function of $\xi_i \overset{\mathrm{d}}{=} \xi$,

$$\|V_{1,i}\|_3 \le C\|V_{1,i}\|_2 = C\tau_i, \qquad 1 \le i \le n. \tag{90}$$

Furthermore, by (60) and (63),

$$\frac{\max_i \tau_i^2}{\sigma_1^2} \le \frac{B\binom{n-1}{m-1}^2}{c' \frac{n}{m}\binom{n-1}{m-1}^2} = C\frac{m}{n} = o(1). \tag{91}$$

Consequently, using (90), (58) and (91),

$$\begin{aligned}\frac{\sum_{i=1}^n \mathbb{E}\,|V_{1,i}|^3}{\sigma_1^3} &= \frac{\sum_{i=1}^n \|V_{1,i}\|_3^3}{\sigma_1^3} \le \frac{C\sum_{i=1}^n \tau_i^3}{\sigma_1^3} \le C\frac{\max_i \tau_i \sum_{i=1}^n \tau_i^2}{\sigma_1^3} \\ &= C\frac{\max_i \tau_i}{\sigma_1} = o(1).\end{aligned} \tag{92}$$

This shows the Lyapunov condition, and thus a standard form of the central limit theorem, [7, Theorem 7.2.4 or 7.6.2], yields (74).