Communications in Information and Systems
Volume 22, Number 1, 1–38, 2022

# Sequential universal modeling for non-binary sequences with constrained distributions*

Michael Drmota, Gil I. Shamir, and Wojciech Szpankowski

Sequential probability assignment and universal compression go hand in hand. We propose sequential probability assignment for non-binary (and large alphabet) sequences with distributions whose parameters are known to be bounded within a limited interval. Sequential probability assignment algorithms are essential in many applications that require fast and accurate estimation of the maximizing sequence probability. These applications include learning, regression, channel estimation and decoding, prediction, and universal compression. On the other hand, constrained distributions introduce interesting theoretical twists that must be overcome in order to present efficient sequential algorithms. Here, we focus on universal compression for memoryless sources, and present a precise analysis for the maximal minimax and the (asymptotic) average minimax redundancy for constrained distributions. We show that our sequential algorithm based on modified Krichevsky-Trofimov (KT) estimator is asymptotically optimal up to $O(1)$ for both maximal and average redundancies. In addition, we provide precise asymptotics of the minimax redundancy for monotone distributions which is a special case of the constrained distribution. This paper follows and addresses some challenges presented in [17] that suggested 'results for the binary case lay the foundation to studying larger alphabets".

Keywords and phrases: Information theory, minimax redundancy, constrained distribution, analytic combinatorics.

## 1. Introduction

Universal coding and universal modeling (probability assignments) are two driving forces of information theory, model selection, and statistical infer-

*This work was supported in part by the Austrian Science Foundation grant FWF Grant SFB F50-02, and in part by NSF Center on Science of Information Grant CCF-0939370 and NSF Grants CCF-2006440, CCF-2007238, and Google Research Award.

ence. In universal coding one is to construct a code for data sequences generated by an unknown source from a known family such that, as the length of the sequence increases, the average code length approaches the entropy of whatever processes in the family has generated the data. In seminal works of Davisson [4], Rissanen [12], Krichevsky and Trofimov [9], and Shtarkov [13] it was shown how to construct such codes for finite alphabet sources. Universal codes are often characterized by the average *minimax* redundancy which is the excess over the entropy of the *best* code from a class of decodable codes for the worst process in the family.

As pointed out by Rissanen [12], over years universal coding evolved into *universal modeling* where the purpose is no longer restricted to just coding but rather to learn optimal models [12]. The central question of interest in universal modeling seems to be in universal codes achievable for *individual* sequences. The burning question is how to measure its performance. The *worst case* minimax redundancy became handy since it measures the worst case excess of the best code maximized over the processes in the family. Unfortunately, low-complexity universal codes that are optimal for the worst case minimax are not easily implementable (it would require to implement the maximum likelihood distribution). Therefore, we design a sequential algorithm based on the KT-estimator that is asymptotically optimal on average (i.e., for the average minimax redundancy), and show that both redundancies differ by a small constant.

In this paper we focus on universal compression and probability assignment/learning for a class of memoryless sources with *constrained distributions*. Let us start with some definitions and notation. We define a code $C_n : \mathcal{A}^n \to \{0,1\}^*$ as a mapping from the set $\mathcal{A}^n$ of all sequences $x^n = (x_1, \dots, x_n)$ of length $n$ over the finite alphabet $\mathcal{A} = \{1, \dots, m\}$ of size $m$ to the set $\{0,1\}^*$ of all binary sequences. Given a probabilistic source model, we let $P(x^n)$ be the probability of the message $x^n$; given a code $C_n$, we let $L(C_n, x^n)$ be the code length for $x^n$. However, in practice the probability distribution (i.e., source) $P$ is unknown, and one looks for *universal codes* for which the redundancy is $o(n)$ for all $P \in \mathcal{S}$ where $\mathcal{S}$ is a class of source models (distributions). It is convenient to ignore the integer nature of the code length and replace it by its best distributional guess, say $Q(x^n)$. In other words, we just write $L(C_n, x^n) = -\log Q(x^n)$ and use it throughout the paper. The question is how well $Q$ approximates $P$ within the class $\mathcal{S}$. Minimax redundancy enters. Usually, we consider two types of minimax redundancy, namely *average* and *maximal* or *worst case* defined, respectively, as

$$\overline{R}_n(\mathcal{S}) = \min_Q \sup_{P\in\mathcal{S}} \mathbf{E}[\log P(X^n)/Q(X^n)], \tag{1}$$

$$R_n^*(\mathcal{S}) = \min_Q \sup_{P\in\mathcal{S}} \mathbf{max}_{x^n}[\log P(x^n)/Q(x^n)]. \tag{2}$$

In this paper we analyze precisely both redundancies for *memoryless sources* over $m$-ary alphabet $\mathcal{A} = \{1,\dots,m\}$ with *restricted symbol probability* $\theta_i$, that is, we assume that $\boldsymbol{\theta} \in \mathcal{S}$, where $\mathcal{S}$ is a proper subset of

$$\Theta = \{\boldsymbol{\theta} : \theta_i \geq 0 \ (1 \leq i \leq m), \ \theta_1 + \cdots + \theta_m = 1\}.$$

We will assume that $\mathcal{S}$ is a convex polytope. As a special case we have the *interval* restriction $0 \leq a_i \leq \theta_i \leq b_i \leq 1$ for $i = 1,\dots m-1$, where $\sum_{i=1}^{m-1} b_i \leq 1$ (this ensures that $\theta_m = 1 - \sum_{i=1}^{m-1}\theta_i$ is always well defined). Also, a class of monotone distributions [16] defined as

$$\mathcal{M} = \{\boldsymbol{\theta} : 0 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_m, \ \theta_1 + \cdots + \theta_m = 1\}$$

is a special case of the constrained distributions.

Here, we present a sequential algorithm that estimates asymptotically the optimal probability $P(x^n)$ for all $x^n$. It turns out that restricting the set of parameters is important from a practical point of view and at the same time introduces new interesting theoretical twists that we explore in this paper. We first prove in Theorem 2.1 that (for fixed $m$ that can still be large)

$$\overline{R}_n(\mathcal{S}) = R_n^*(\mathcal{S}) + O(1) = \frac{m-1}{2}\log(n) + O(1)$$

where the constant implied by the $O$-term depends on $m$ and on the set $\mathcal{S}$. Second we provide in Theorem 2.2 precise asymptotics for $\overline{R}_n(\Theta)$ and $R_n^*(\Theta)$ if $m = o(n)$. While the leading terms of these redundancies are known [10, 14, 15, 20], we derive here *precise* asymptotics up to $O(m^{3/2}/\sqrt{n})$ term in a uniform manner that can be used to extend our analysis to the constrained case in this regime. This allows us in Theorem 2.3 to provide the best asymptotic expansions for the minimax redundancy for monotone distributions. Finally, we present in Corollary 1 a sequential add-1/2 KT-like estimator to compute $P(x_{n+1}|x^n)$ for the constrained distributions that is asymptotically optimal up to a constant for both the maximal and average redundancy. This final result has been wanting since [17] which suggested that "results for the binary case lay the foundation to studying larger alphabets".

This paper is organized as follows. In the next section we present our main results, including the sequential algorithm that directly generalizes the add-1/2-KT estimator. The proofs are discussed in the last section, with some technical proofs collected in Section 4.

## 2. Main results

In this section we present our main results including asymptotically optimal probability estimation for the class $\mathcal{S} \subset \Theta$ of memoryless sources with constrained distributions.

We start with the *worst case redundancy* defined in (2). We recall that the distribution $P(x^n)$ is of the form

$$P(x^n) = \prod_{i=1}^{m} \theta_i^{k_i}, \qquad \theta_i \geq 0,\ \sum_{i=1}^{m} \theta_i = 1,$$

where $k_i$ is the number of symbol $i \in \mathcal{A}$ in the sequence $x^n$. The probabilities $\theta_i$ are unknown to us except that we restrict them to the subset $\mathcal{S} \subseteq \Theta$. Following Shtarkov [13] and [5] we can re-write the worst case redundancy for $\mathcal{S}$, by noting that max and sup commute, hence

$$\begin{aligned} R_n^*(\mathcal{S}) &= \min_Q \sup_{P \in \mathcal{S}} \max_{x^n} (-\log Q(x^n) + \log P(x^n)) \\ &= \min_Q \max_{x^n} [-\log Q(x^n) + \sup_{P \in \mathcal{S}} \log P(x^n)] \\ &= \min_Q \max_{x_n} [\log Q^{-1}(x^n) + \log P^*(x^n) + \log \sum_{z^n} \sup_{P \in \mathcal{S}} P(z^n)] \\ &= \log \sum_{x^n} \sup_{P \in \mathcal{S}} P(x^n) \end{aligned}$$

where $P^*(x^n)$ is

$$P^*(x^n) := \frac{\sup_{P \in \mathcal{S}} P(x^n)}{\sum_{z_1^n} \sup_{P \in \mathcal{S}} P(z_1^n)} \tag{3}$$

is the *maximum-likelihood distribution* and we set $Q(x^n) = P^*(x^n)$ for attaining the minimum. In this context the distribution $P^*$ is also called the

Shtarkov distribution and the sum

$$D_n = \sum_{x^n} \sup_{P\in\mathcal{S}} P(x^n)$$

is called Shtarkov sum. Note that $R_n^*(\mathcal{S}) = \log D_n$.

If we define the worst case redundancy with the help of code lengths $L(C_n, x^n)$ instead of $-\log Q(x^n)$ – that we denote by $\widetilde{R}_n^*$ – then we would get a similar expression of the form $\widetilde{R}_n^*(\mathcal{S}) = \log D_n + \widetilde{R}_n^*(P^*)$. Using Shannon's code Shtarkov immediately noticed that $0 < \widetilde{R}_n^*(P^*) < 1$. More precisely, in [5] it was proved that asymptotically for the unconstrained case

$$\widetilde{R}_n^*(P^*) = \frac{\log\left(\frac{1}{m-1}\log m\right)}{\log m} + o(1).$$

From now on we shall ignore this *correction term* and analyze $R_n^*(\mathcal{S}) = \log D_n$.

To estimate $D_n = \sum_{x^n} \sup_P P(x^n)$ we need first to find $\sup \prod_{i=1}^m \theta_i^{k_i}$ when $\boldsymbol{\theta} \in \mathcal{S}$. For the unrestricted case ($\mathcal{S} = \Theta$) we know that the optimal $\theta_i = k_i/n$. The situation is more complicated in the constrained case. For example, if we assume an interval restriction $a_i \le \theta_i \le b_i$, $i = 1, \ldots, m-1$ with $\theta_m = 1 - \theta_1 - \cdots - \theta_{m-1}$, then for $k_i < na_i$ or $k_i > nb_i$ the optimal $\theta_i$ may be $a_i$ or $b_i$, respectively. Fortunately, in the next sections we prove that the main contribution to $D_n$ comes from those $\mathbf{k} = (k_1, \ldots, k_m)$ for which $\mathbf{k}/n \in \mathcal{S}$. So we are led to analyze the following sum

$$D_n^{(\mathcal{S})} = \sum_{\mathbf{k}\in n\mathcal{S}} \binom{n}{k_1, \cdots k_m} \prod_{i=1}^m \left(\frac{k_i}{n}\right)^{k_i}$$

which is of order $n^{\frac{m-1}{2}}$ as in (26) and proved in Section 4.2. The contribution of the remaining terms is typically of order $O(n^{\frac{m-2}{2}})$.

It is our goal to present a sequential low-complexity algorithm for the probability assignment, that is, an iterative procedure to compute $P(x_{n+1}|x^n)$. Unfortunately, the maximum-likelihood distribution (3) is not well suited for it. To find one, we switch to the *average* minimax redundancy (1) and we re-cast it in the Bayesian framework.

Before we discuss the average minimax redundancy, we need to introduce one more notation element. Let us define the Dirichlet density as

$$\mathrm{Dir}(\theta_1, \ldots, \theta_m; \alpha_1, \ldots, \alpha_m) = \frac{1}{B(\alpha_1, \ldots, \alpha_m)} \prod_{i=1}^m \theta_i^{\alpha_i - 1},$$

where $\sum_{i=1}^{m}\theta_i = 1$ and

$$B(\alpha_1,\ldots,\alpha_m) = B_m(\alpha_1,\ldots,\alpha_m) = \frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_m)}{\Gamma(\alpha_1+\cdots+\alpha_m)}$$

is the beta function. We shall write $\boldsymbol{\alpha} = (\alpha_1,\ldots\alpha_m)$ and $\boldsymbol{\theta} = (\theta_1,\ldots\theta_m)$ with $\sum_{i=1}^{m}\theta_i = 1$. Finally, we set for $\mathcal{S}\subset\Theta$

$$\mathrm{Dir}(\mathcal{S};\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})}\int_{\mathcal{S}}\boldsymbol{\theta}^{\boldsymbol{\alpha}-1}\,d\boldsymbol{\theta}.$$

Let $\mathcal{S}\subseteq\Theta$. Then the *average minimax* problem is

$$\overline{R}_n(\mathcal{S}) = \inf_{Q}\sup_{\theta\in\mathcal{S}} D_n(P^\theta\|Q)$$

where $D(P^\theta\|Q)$ is the Kullback-Leibler divergence. In the Bayesian framework, one assumes that the parameter $\theta$ is generated by the density $w(\theta)$ and the mixture $M_n^w(x^n)$ is

$$M_n^w(x^n) = \int_{\mathcal{S}} P^\theta(x^n)w(d\theta).$$

Observe now

$$\begin{aligned}\inf_Q \mathbf{E}_w[D_n(P^\theta\|Q)] &= \inf_Q\int_{\mathcal{S}} D_n(P^\theta\|Q)dw(\theta)\\ &= \int_{\mathcal{S}} D_n(P^\theta\|M_n^w)dw(\theta),\end{aligned}$$

where we use the fact that

$$\min_Q\sum_i P_i\log 1/Q_i = \sum_i P_i\log 1/P_i.$$

As pointed out by Gallager [7], and Davisson [4] the minimax theorem of game theory entitles us to conclude that

$$\overline{R}_n(\mathcal{S}) = \inf_Q\sup_{\theta\in\mathcal{S}} D_n(P^\theta\|Q) = \sup_w\inf_Q \mathbf{E}_w[D_n(P^\theta\|Q)]$$

leading to

$$\overline{R}_n(\mathcal{S}) = \int_{\mathcal{S}} D(P^\theta\|M_n^{w^*})dw^*(\theta) \tag{4}$$

where $w^*(\theta)$ is the maximizing prior distribution. Bernardo [1] proved that asymptotically the maximizing density is proportional to the square root of the determinant of the Fisher information $I(\boldsymbol{\theta})$, the so-called Jeffrey prior. This leads to the density

$$\widetilde{w}^*(\boldsymbol{\theta}) = \frac{1}{C(\mathcal{S}) \cdot B(\mathbf{1/2})} \frac{1}{\sqrt{\theta_1 \cdots \theta_m}}, \tag{5}$$

where $C(\mathcal{S})$ defined as

$$C(\mathcal{S}) = \mathrm{Dir}(\mathcal{S}; \mathbf{1/2}) = \frac{1}{B(\mathbf{1/2})} \int_{\mathcal{S}} \frac{d\boldsymbol{\theta}}{\sqrt{\theta_1 \cdots \theta_m}} \tag{6}$$

is the probability that the Dirichlet distribution with $\alpha_i = 1/2$ falls into the subset $\mathcal{S}$. For example, Clarke and Barron [2] showed (under proper regularity conditions such as the finiteness of the determinant of Fisher information and $\mathcal{S}$ being a compact subset of the interior of $\Theta$) that

$$\begin{aligned}
&\lim_{n\to\infty} \left( \int_{\mathcal{S}} D(P^{\theta} \| M_n^{w^*}) dw^*(\theta) - \frac{m-1}{2} \log \frac{n}{2\pi e} \right) \\
&= \lim_{n\to\infty} \left( \int_{\mathcal{S}} D(P^{\theta} \| M_n^{\widetilde{w}^*}) d\widetilde{w}^*(\theta) - \frac{m-1}{2} \log \frac{n}{2\pi e} \right) = \log \int_{\mathcal{S}} \sqrt{\det I(\boldsymbol{\theta})}\, d\boldsymbol{\theta}.
\end{aligned}$$

Barron and Xie [22] extended this result to the unconstrained case $\mathcal{S} = \Theta$. We note that $C(\Theta) = 1$ for the unconstrained case.

This discussion leads us to the following notation of the *asymptotic average minimax redundancy*

$$\overline{R}_n^{\mathrm{asymp}}(\mathcal{S}) = \int_{\mathcal{S}} D(P^{\boldsymbol{\theta}} \| M_n^{\widetilde{w}^*}) d\widetilde{w}^*(\boldsymbol{\theta}).$$

The mixture distribution $M_n^{\widetilde{w}^*}(x^n)$ can be calculated as follows

$$\begin{aligned}
M_n^{\widetilde{w}^*}(x^n) &= \frac{1}{C(\mathcal{S}) \cdot B(\mathbf{1/2})} \int_{\mathcal{S}} \prod_{i=1}^{m} \theta_i^{k_i - 1/2} \\
&= \frac{1}{C(\mathcal{S}) \cdot B(\mathbf{1/2})} B(k_1 + 1/2, \cdots, k_m + 1/2) \\
&\quad \cdot \frac{1}{B(k_1 + 1/2, \cdots, k_m + 1/2)} \int_{\mathcal{S}} \prod_{i=1}^{m} \theta_i^{k_i - 1/2} \\
&= \frac{1}{C(\mathcal{S}) \cdot B(\mathbf{1/2})} B(k_1 + 1/2, \cdots, k_m + 1/2) \cdot \mathrm{Dir}(\mathcal{S} : \mathbf{k} + \mathbf{1/2}). \qquad (7)
\end{aligned}$$

Observe again that for the unconstrained case $\mathrm{Dir}(\Theta; \mathbf{k}+1/2) = 1$. In summary

$$D_n(P^\theta \| M_n^{\widetilde{w}^*}) = \log\left(C(\mathcal{S})B(\mathbf{1/2})\right) + \tag{8}$$

$$+\sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \prod_{i=1}^m \theta_i^{k_i} \log \frac{\prod_{i=1}^m \theta_i^{k_i}}{B(\mathbf{k}+\mathbf{1/2})\mathrm{Dir}(\mathcal{S}; \mathbf{k}+\mathbf{1/2})}. \tag{9}$$

We are now ready to formulate our first main result that reads as follows. We prove it in the next section and delay some technical derivations to Section 4.

**Theorem 2.1.** *Consider a memoryless constrained source $\mathcal{S} \subset \Theta$ with fixed but arbitrarily large $m \geq 2$ where $\mathcal{S}$ is a convex polytope. Then the worst case minimax redundancy for $\mathcal{S}$ is*

$$R_n^*(\mathcal{S}) = \frac{m-1}{2} \log \frac{n}{2\pi} + \log \frac{\Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2})} + \log C(\mathcal{S}) + O(1/\sqrt{n}), \tag{10}$$

*and the corresponding asymptotic average minimax redundancy is*

$$\overline{R}_n^{\mathrm{asymp}}(\mathcal{S}) = \frac{m-1}{2} \log \frac{n}{2\pi e} + \log \frac{\Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2})} + \log C(\mathcal{S}) + O(1/\sqrt{n}) \tag{11}$$

*where, we recall,*

$$C(\mathcal{S}) = \mathrm{Dir}(\mathcal{S}; \mathbf{1/2}) = \frac{1}{B(\mathbf{1/2})} \int_{\mathcal{S}} \frac{d\boldsymbol{\theta}}{\sqrt{\theta_1 \cdots \theta_m}}$$

*as defined above in (6) with $C(\Theta) = 1$.*

We observe that $R_n^*(\mathcal{S})$ and $\overline{R}_n^{\mathrm{asymp}}(\mathcal{S})$ differ approximately by $\frac{m-1}{2}$. This fact should be compared with a general results of [5, Theorem 6] where it was proved that for a large class of sources

$$|\overline{R}_n(\mathcal{S}) - \overline{R}_n^*(\mathcal{S})| \leq c_n(\mathcal{S})$$

where

$$c_n(\mathcal{S}) = \sup_{P \in \mathcal{S}} \sum_{x^n} P(x_1^n) \lg \frac{\sup_{P \in \mathcal{S}} P(x_1^n)}{P(x_1^n)}.$$

Actually, for binary memoryless sources $c_n(\mathcal{S}) \leq 1$ and $c_n(\mathcal{S}) \leq m-1$ for $m$-ary memoryless sources ([5, Lemma 8] extends directly to the $m$-ary case).

In Theorem 2.1 we assumed that $m$ is fixed to avoid complications with constrains $\mathcal{S}_m$ that may depend on $m$. We now present our results for large $m = o(n)$ and $\mathcal{S} = \Theta$. While the leading terms, especially for the maximal minimax redundancy, were known before (see [10, 11, 14, 15, 20, 21, 22]), our results are derived in a novel way that allows us to apply our methodology to obtain in Theorem 2.3 best redundancy results for monotone distributions over large alphabets (see [16]).

**Theorem 2.2.** *Consider a memoryless unconstrained source* $\Theta$ *with* $m = o(n)$*. Then the unconstrained maximal redundancy is*

$$R_n^*(\Theta) = \frac{m-1}{2}\log\frac{n}{2\pi} + \log\frac{\Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2})} + O\left(m^{3/2}/\sqrt{n}\right)$$

$$= \frac{m-1}{2}\log\left(\frac{e\,n}{m}\right) + \frac{1}{2}(1-\log 2) + O(1/m) + O(m^{3/2}/\sqrt{n}) \tag{12}$$

*and the unconstrained asymptotic average redundancy becomes*

$$\overline{R}_n^{\text{asymp}}(\Theta) = \frac{m-1}{2}\log\frac{n}{2\pi e} + \log\frac{\Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2})} + O\left(m^{3/2}/\sqrt{n}\right)$$

$$= \frac{m-1}{2}\log\left(\frac{n}{m}\right) + \frac{1}{2}(1-\log 2) + O(1/m) + O(m^{3/2}/\sqrt{n}). \tag{13}$$

**Remark 1.** We see that Theorems 2.1 and 2.2 match for constant $m$, since $C(\Theta) = 1$. Indeed, we can use Stirling's formula for $\Gamma(m/2)$ to replace $\log\Gamma(m/2)$ in Theorem 2.1 by

$$\log\Gamma(m/2) = \frac{m}{2}\log\frac{m}{2} - \frac{m}{2}\log e + \log\sqrt{2\pi} - \frac{1}{2}\log\frac{m}{2} + O(1/m)$$

if $m \to \infty$.

**Monotone Distributions**. As an application of Theorems 2.1–2.2 we provide precise asymptotics for monotone distributions which can be viewed as a special case of the constrained distribution. Indeed, let again

$$\mathcal{M} = \{\boldsymbol{\theta} \in \Theta : \theta_1 \le \theta_2 \le \cdots \le \theta_m\}.$$

Then the unconstrained set $\Theta$ can be divided into $m!$ subsets

$$\mathcal{M}_\pi = \{\boldsymbol{\theta} \in \Theta : \theta_{\pi(1)} \le \theta_{\pi(2)} \le ... \le \theta_{\pi(m)}\}$$

for any permutation $\pi$ of $\{1,2,...m\}$. Furthermore, for all symmetric functionals $f$ it is easy to see that

$$\int_{\mathcal{M}} f(\theta_1,...,\theta_m)d\theta_1...d\theta_m = \frac{1}{m!}\int_{\Theta} f(\theta_1,...,\theta_m)d\theta_1...d\theta_m.$$

More precisely, let $D_n(\Theta)$ and $D_n(\mathcal{M})$ denote the Shtarkov sums for the unconstrained distribution and monotone distribution, respectively. Then in Section 3.3 we show (see (31))

$$D_n(\mathcal{M}) = \frac{1}{m!}D_n(\Theta)+O\left((2\pi)^{-\frac{m-1}{2}}m^{3/2}n^{\frac{m-2}{2}}B_m(\mathbf{1/2})\exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)\right).$$

This leads to our next main result regarding the redundancy for monotone distribution and large alphabet. To the best of our knowledge this is the most precise asymptotic expansion (up to $O(1)$ term) for $m = o(\log n/\log\log n)$ (cf. [16]).

**Theorem 2.3.** *Consider a class of monotone distributions $\mathcal{M}$. Then $R_n^*(\mathcal{M}) = R_n^*(\Theta) - \log m! + O(m!\, m^{3/2}/\sqrt{n})$. In particular, for $m = o(n)$ we have*

$$\begin{aligned} R_n^*(\mathcal{M}) &= \frac{m}{2}\log\left(\frac{n}{m^3}\right) + \frac{3}{2}m\log e + \frac{1}{2}\log\frac{2\pi m}{e} + \frac{1}{2}(1-\log 2) \\ &\quad + O(1/m) + O(m!\, m^{3/2}/\sqrt{n}) \end{aligned} \tag{14}$$

*for large $n$.*

**Sequential Probability Assignment**. Now, we are ready to present our probability assignment algorithm. We start with formula (7) on the mixture $M_n(x^n)$. Then we observe that $M_n(x_{n+1}|x^n) = M_n(x^{n+1})/M_n(x^n)$. For example, if assume that $x_{n+1}$ symbol is $i \in \mathcal{A}$. Thus

$$\begin{aligned} M_n(x^{n+1}) &= \frac{B(k_1+1/2,\cdots,k_i+3/2,\cdots,k_m+1/2)}{C(\mathcal{S})\cdot B(\mathbf{1/2})} \\ &\quad \cdot \mathrm{Dir}(\mathcal{S};k_1+1/2,\cdots,k_i+3/2,\cdots,k_m+1/2). \end{aligned}$$

Using the functional equation of the gamma function, namely $\Gamma(x+1) = x\Gamma(x)$ allows us to write a simple sequential update algorithm that we present next.

**Corollary 1.** *Suppose that $m$ is fixed and that $\mathcal{S} \subseteq \Theta$ is a convex polytope. Let $N_i(x^n)$ be the number of symbol $i$ in $x^n$. Then*

$$M_n(x_{n+1}|x^n) = \frac{N_{x_{n+1}}(x^n)+1/2}{n+m/2} \cdot \cdot \frac{\mathrm{Dir}(\mathcal{S}; N_i(x^n)+1/2+1(x_{n+1}=i),\ i=1\cdots m)}{\mathrm{Dir}(\mathcal{S}; N_i(x^n)+1/2,\ i=1\cdots m)} \tag{15}$$

*which is the generalized add-1/2-KT estimator.*

Observe that for the unconstrained case $\mathrm{Dir}(\Theta; N_i(x^n)+1/2+1(x_{n+1} = i),\ i = 1\cdots m) = \mathrm{Dir}(\Theta; N_i(x^n)+1/2,\ i = 1\cdots m) = 1$, and then our estimation algorithm reduces to the KT-estimator, that is,

$$M_n(x_{n+1}|X^n) = \frac{N_{x_{n+1}}(x^n)+1/2}{n+m/2}. \tag{16}$$

We also observe that for the binary alphabet we recover the update from [17], namely

$$\begin{aligned} M_n(x_{n+1}|x^n) &= \frac{N_{x_{n+1}}(x^n)+1/2}{n+1} + \\ &+(2x_{n+1}-1)\frac{a_1^{N_1(x^n)+1/2}(1-a_1N_0(x^n)+1/2}{C(\mathcal{S})(n+1)} \\ &-(2x_{n+1}-1)\frac{b_1^{N_1(x^n)+1/2}(1-b_1N_0(x^n)+1/2}{C(\mathcal{S})(n+1)} \end{aligned}$$

where $C(\mathcal{S})$ is defined in (6). We should point out that the binary sequential probability assignment as above was derived in [17] using a different technique that seems to be working only for binary sequences.

## 3. Analysis and proofs

In this section we prove our main results Theorems 2.1–2.3. Some technical details are delayed till the last section. We start with Theorem 2.2 since we will use some ideas and calculations of the proof of Theorem 2.2 in the proof of Theorem 2.1.

### 3.1. Proof of Theorem 2.2

We start with the worst case minimax redundancy $R_n^*(\Theta) = \log D_n$, where

$$D_n = \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_i}$$

and the sum is taken over all non-negative integer vectors $\mathbf{k} = (k_1, \ldots, k_m)$ with $\sum_i k_i = n$. If we set

$$S_m(n) = \frac{(2\pi)^{(m-1)/2}}{\sqrt{n}} \sum_{\mathbf{k}}{}' \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_i} \tag{17}$$

in which sum $\sum'$ is taken over all $m$-dimensional integer vectors $\mathbf{k} = (k_1, \ldots, k_m)$ with $k_j \geq 1$ $(1 \leq j \leq m)$ and $k_1 + \cdots + k_m = n$, then we can represent $D_n$ as

$$D_n = (2\pi)^{-\frac{m-1}{2}} \sqrt{n} \sum_{r=0}^{m-1} \binom{m}{r} (2\pi)^{r/2} S_{m-r}(n),$$

that is, $S_{m-r}(n)$ takes care of those $\mathbf{k}$, where precisely $r$ components are zero.

The asymptotic properties of $S_m(n)$ are summarized in the following lemma that will be proved in Section 4.1. Recall that $B_m(\mathbf{1/2}) = B(\mathbf{1/2}) = \Gamma\left(\frac{1}{2}\right)^m / \Gamma\left(\frac{m}{2}\right)$.

**Lemma 1.** *The terms $S_m(n)$ satisfy*

$$S_m(n) = n^{\frac{m}{2}-1} B_m(\mathbf{1/2}) \left(1 + O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right) \tag{18}$$

*uniformly as $n, m \to \infty$ and $m \leq n$.*

With the help of Lemma 1 and the relation

$$B_{m-r}(\mathbf{1/2}) = O\left((m/\pi)^{r/2} B_m(\mathbf{1/2})\right)$$

we obtain

$$\begin{aligned} D_n &= (2\pi)^{-\frac{m-1}{2}} n^{\frac{m-1}{2}} B_m(\mathbf{1/2}) \left(1 + O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right) \\ &\quad + (2\pi)^{-\frac{m-1}{2}} \sqrt{n} \sum_{r=1}^{m-1} \binom{m}{r} O\left((2\pi)^{r/2} B_{m-r}(\mathbf{1/2}) n^{-r/2} \left(1 + O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)\right) \\ &\quad = (2\pi)^{-\frac{m-1}{2}} n^{\frac{m-1}{2}} B_m(\mathbf{1/2}) \left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right)^m \left(1 + O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right) \\ &\quad + O\left((2\pi)^{-\frac{m-1}{2}} n^{\frac{m-1}{2}} B_m(\mathbf{1/2}) \frac{m^{3/2}}{\sqrt{n}}\right) \end{aligned}$$

(19)

$$= (2\pi)^{-\frac{m-1}{2}} n^{\frac{m-1}{2}} B_m(\mathbf{1/2}) \exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right).$$

Since

$$\log B_m(\mathbf{1/2}) = m \log \Gamma(1/2) - \log \Gamma(m/2)m)$$

we directly derive the proposed representation (12) for $R_n^*(\Theta) = \log D_n$.

For the asymptotic average minimax $\overline{R}_n^{\text{asymp}}(\Theta)$ our starting point is

$$\overline{R}_n^{\text{asymp}}(\Theta) = \frac{1}{B(\mathbf{1/2})} \int_\Theta \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \boldsymbol{\theta}^{\mathbf{k}-\mathbf{1/2}} \log\left(\frac{\boldsymbol{\theta}^{\mathbf{k}} B(\mathbf{1/2})}{B(\mathbf{k}+\mathbf{1/2})}\right) \tag{20}$$

where we write $\boldsymbol{\theta}^{\mathbf{k}-\mathbf{1/2}} := \prod_i \theta_i^{k_i - 1/2}$. We need to estimate different parts of the above sum. We first observe that

$$\begin{aligned}\sum_{\mathbf{k}} \binom{n}{\mathbf{k}} B(\mathbf{k}+\mathbf{1/2}) &= \int_\Theta \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \boldsymbol{\theta}^{\mathbf{k}-\mathbf{1/2}} d\boldsymbol{\theta} \\ &= \int_\Theta \boldsymbol{\theta}^{-\mathbf{1/2}} d\boldsymbol{\theta} = B(\mathbf{1/2}).\end{aligned}$$

More importantly we notice that

$$\begin{aligned}\int_\Theta \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \boldsymbol{\theta}^{\mathbf{k}-\mathbf{1/2}} \log \boldsymbol{\theta}^{\mathbf{k}} d\boldsymbol{\theta} &= \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \sum_{i=1}^m k_i \int_\Theta \boldsymbol{\theta}^{\mathbf{k}-\mathbf{1/2}} \log \theta_i d\boldsymbol{\theta} \\ &= \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \sum_{i=1}^m k_i \frac{\partial}{\partial k_i} B(\mathbf{k}+\mathbf{1/2}),\end{aligned}$$

$$\int_\Theta \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \boldsymbol{\theta}^{\mathbf{k}-\mathbf{1/2}} \log B(\mathbf{k}+\mathbf{1/2}) = \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} B(\mathbf{k}+\mathbf{1/2}) \log B(\mathbf{k}+1/2).$$

Thus

(21)

$$\overline{R}_n^{\text{asymp}}(\Theta) = \log B(\mathbf{1/2})$$

$$+\frac{1}{B(\mathbf{1/2})} \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \cdot \left( \sum_{i=1}^m k_i \frac{\partial}{\partial k_i} B(\mathbf{k}+\mathbf{1/2}) - B(\mathbf{k}+\mathbf{1/2}) \log B(\mathbf{k}+\mathbf{1/2}) \right).$$

To deal with sums like (21) we use the relation between the beta function, the gamma function, and the psi function [19]. For example

$$\frac{\partial}{\partial k_i} B(\mathbf{k} + \mathbf{1/2}) = \left(\Psi(k_i + 1/2) - \Psi(n + m/2)\right) B(\mathbf{k} + \mathbf{1/2}),$$

where $\Psi(x) = \Gamma'(x)/\Gamma(x)$. Asymptotically we have

$$\begin{aligned} \Psi(x + 1/2) &= \log x + 1/(12x) + O(1/x^3), \\ \Psi(x + m/2) &= \log x + (m-1)/(2x) + O(1/x^2). \end{aligned}$$

Using this and Stirling's formula we finally find

$$\begin{aligned} \log \Gamma(x + 1/2) &= x \log x - x + \log \sqrt{2\pi} - \frac{1}{24x} + O(1/x^2), \\ \log \Gamma(x + m/2) &= x \log x - x + \frac{m-1}{2} \log(x + m/2) \\ &\qquad + \log \sqrt{2\pi} + \left(\frac{1}{12} - \frac{m^2}{8}\right) \frac{1}{x} + O(m^3/x^2) \end{aligned}$$

leading to

$$\begin{aligned} &\sum_{i=1}^{m} k_i \frac{\partial}{\partial k_i} B(\mathbf{k} + \mathbf{1/2}) - B(\mathbf{k} + \mathbf{1/2}) \log B(\mathbf{k} + \mathbf{1/2}) = \\ &B(\mathbf{k} + \mathbf{1/2}) \left( \frac{m-1}{2} \left(\log(n + m/2) - 1 - \log(2\pi)\right) + O(m^2/n) \right. \\ &\left. + O\left( \sum_i (k_i + 1)^{-1} \right) \right). \end{aligned}$$

We also obtain, similarly to (37) derived in the proof of Lemma 1 presented in Section 4.1,

$$\sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \frac{B(\mathbf{k} + \mathbf{1/2})}{k_i + 1} = O\left( \frac{\sqrt{m} B(\mathbf{1/2})}{\sqrt{n}} \right).$$

Summing up we arrive at

$$\overline{R}_n^{\text{asmp}}(\Theta) = \frac{m-1}{2} \log(n/2\pi e) + \log \frac{\Gamma^m(1/2)}{\Gamma(m/2)} + O(m^{3/2}/\sqrt{n}). \tag{22}$$

We now use Stirling's formula for $\Gamma(m/2)$ and complete the proof.

### 3.2. Proof of Theorem 2.1

We start with the worst case minimax redundancy $R_n^*(\mathcal{S}) = \log D_n$, where

$$D_n = \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \sup_{\theta \in \mathcal{S}} \prod_{i=1}^{m} \theta_i^{k_i}.$$

The problem is now that we have to distinguish between the case, where $\mathbf{k}/n \in \mathcal{S}$ and the case, where $\mathbf{k}/n \notin \mathcal{S}$. If $\mathbf{k}/n \in \mathcal{S}$ then we have

$$\sup_{\theta \in \mathcal{S}} \prod_{i=1}^{m} \theta_i^{k_i} = \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_i} \tag{23}$$

as in the unconstrained case. If $\mathbf{k}/n \notin \mathcal{S}$ then we have

$$\sup_{\theta \in \mathcal{S}} \prod_{i=1}^{m} \theta_i^{k_i} = \prod_{i=1}^{m} \theta_{i,\mathrm{opt}}^{k_i}.$$

By concavity of $\prod_{i=1}^{m} \theta_i^{k_i}$ it follows that the optimal choice $\theta_{i,\mathrm{opt}}$ has to be on the boundary of $\mathcal{S}$. We will use this fact in the sequel.

Let us first assume that $\mathbf{k}/n \in \mathcal{S}$ so that (23) holds. Then by a standard analysis (note that the dimension $m$ is fixed) we directly obtain

$$\begin{aligned}
D_n^{(\mathcal{S})} &:= \sum_{\mathbf{k}/n \in \mathcal{S}} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_i} \\
&= \frac{\sqrt{n}}{(2\pi)^{\frac{m-1}{2}}} \sum_{\mathbf{k}/n \in \mathcal{S}} \frac{1}{\sqrt{k_1 k_2 \cdots k_m}} \left(1 + O\left(\sum_{i=1}^{m} \frac{1}{k_i}\right)\right) \\
&= \left(\frac{n}{2\pi}\right)^{\frac{m-1}{2}} \int_{\mathcal{S}} \frac{d\boldsymbol{\theta}}{\sqrt{\theta_1 \theta_2 \cdots \theta_m}} \left(1 + O(1/\sqrt{n})\right) \qquad (24) \\
&= \left(\frac{n}{2\pi}\right)^{\frac{m-1}{2}} C(\mathcal{S}) B(\mathbf{1/2}) \left(1 + O(1/\sqrt{n})\right). \qquad (25)
\end{aligned}$$

The sum over $\mathbf{k}/n \notin \mathcal{S}$ is more difficult to handle. However, if $\mathcal{S}$ is a convex polytope this can be handled and in Section 4.2 we prove the following lemma.

**Lemma 2.** *Suppose that $\mathcal{S}$ is a convex polytope contained in $\Theta$. Then we have*

$$D_n - D_n^{(\mathcal{S})} = O\left(n^{\frac{m}{2}-1}\right) \tag{26}$$

*for large $n$ and $m = O(1)$.*

Clearly by using $B(\mathbf{1/2}) = \Gamma(1/2)^m/\Gamma(m/2)$ and $\Gamma(1/2) = \sqrt{\pi}$ and (25) and (26) we directly obtain (10).

To prove the second statement of Theorem 2.1, the starting point for the asymptotic average redundancy is (8), however, we rewrite it in terms of $\mathcal{S} \subseteq \Theta$ as follows

$$\overline{R}_n^{\text{asympt}}(\mathcal{S}) = \frac{1}{B_{\mathcal{S}}(\mathbf{1/2})} \int_{\mathcal{S}} \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \boldsymbol{\theta}^{\mathbf{k}-\mathbf{1/2}} \log\left(\frac{\boldsymbol{\theta}^{\mathbf{k}} B_{\mathcal{S}}(\mathbf{1/2})}{B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2})}\right), \tag{27}$$

where we use the short hand notation

$$B_{\mathcal{S}}(\boldsymbol{\alpha}) = \int_{\mathcal{S}} \boldsymbol{\theta}^{\boldsymbol{\alpha}-1}\, d\boldsymbol{\theta} = \mathrm{Dir}(\mathcal{S};\boldsymbol{\alpha})B(\boldsymbol{\alpha}).$$

As in the proof of Theorem 2.2 we obtain

$$\overline{R}_n^{\text{asympt}}(\mathcal{S}) = \log B_{\mathcal{S}}(\mathbf{1/2}) + \\ \frac{1}{B_{\mathcal{S}}(\mathbf{1/2})} \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \left( \sum_{i=1}^m k_i \frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) - B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) \log B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) \right).$$

Again we split the summation over $\mathbf{k}$ into several parts. If $\mathbf{k}/n \in \mathcal{S}^-$, where $\mathcal{S}^-$ denotes all points in the interior of $\mathcal{S}$ with distance $\geq n^{-1/2+\varepsilon}$ to the boundary (for some $\varepsilon > 0$), then the saddle point $\theta_i = k_i/n$ of the integrand $\boldsymbol{\theta}^{\mathbf{k}}$ of the integral of $B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2})$ or $\frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2})$, respectively, is contained in $\mathcal{S}^-$. The precise statement is as follows with the proof given in Section 4.3.

**Lemma 3.** *Suppose that $\mathbf{k}/n \in \mathcal{S}^-$ then for every fixed integer $L \geq 0$ we have*

$$B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) = B(\mathbf{k}+\mathbf{1/2}) \left(1 + O\left(\sum_{i=1}^m k_i^{-L-1}\right)\right)$$

*and*

$$\sum_{i=1}^{m} k_i \frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) - B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) \log B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) =$$
$$= \sum_{i=1}^{m} k_i \frac{\partial}{\partial k_i} B(\mathbf{k}+\mathbf{1/2}) - B(\mathbf{k}+\mathbf{1/2}) \log B(\mathbf{k}+\mathbf{1/2}) +$$
$$+ O\left( B(\mathbf{k}+\mathbf{1/2}) \sum_{i=1}^{m} k_i^{-L-1} \right)$$

*for large $n$.*

Note that $\mathbf{k}/n \in \mathcal{S}^-$ implies that $k_i \ge c\,n^{\frac{1}{2}+\varepsilon}$ for some constant $c > 0$ and all $i = 1, \ldots, m$. Thus, the error term can be also stated in terms of negative powers of $n$.

With the help of Lemma 3 we obtain for any large $L' > 0$

$$\sum_{\mathbf{k}/n \in \mathcal{S}^-} \binom{n}{\mathbf{k}} \left( \sum_{i=1}^{m} k_i \frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) - B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) \log B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) \right) =$$
$$\sum_{\mathbf{k}/n \in \mathcal{S}^-} \binom{n}{\mathbf{k}} \left( \sum_{i=1}^{m} k_i \frac{\partial}{\partial k_i} B(\mathbf{k}+\mathbf{1/2}) - B(\mathbf{k}+\mathbf{1/2}) \log B(\mathbf{k}+\mathbf{1/2}) \right) + O(n^{-L'})$$
$$= \sum_{\mathbf{k}/n \in \mathcal{S}^-} \binom{n}{\mathbf{k}} B(\mathbf{k}+\mathbf{1/2}) \cdot \left( \frac{m-1}{2} \log \frac{n}{2\pi e} + O\left( \sum_{i=1}^{m} 1/(k_i+1) \right) \right) + O(n^{-L'})$$
$$= \left( \frac{m-1}{2} \log \frac{n}{2\pi e} + O\left(1/\sqrt{n}\right) \right) B_{\mathcal{S}}(\mathbf{1/2}).$$

Finally, to simplify our presentation, we now explain the other parts of the summation over $\mathbf{k}$ only for $m = 2$ and $\mathcal{S} = \{(\theta, 1-\theta) : \theta \in [a,b]\}$. Suppose that $|k_1 - nb| \le n^{1/2+\varepsilon}$, that is $(k_1/n, 1-k_1/n)$ is at distance $\le n^{-1/2+\varepsilon}$ from the boundary of $\mathcal{S}$. Here we have

$$B_{\mathcal{S}}(k_1+1/2, n-k_1+1/2) = \sqrt{\frac{2\pi}{n}} \left(\frac{k_1}{n}\right)^{k_1} \left(\frac{n-k_1}{n}\right)^{n-k_1}$$
$$\cdot \left( \Phi\left( \frac{nb-k_1}{\sqrt{nb(1-b)}} \right) + O(1/\sqrt{n}) \right),$$

where $\Phi(u)$ denotes the normal distribution function. A similar representation holds for the derivatives $\frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2})$. A standard analysis yields

$$\sum_{|k_1-nb|\le n^{1/2+\varepsilon}} \binom{n}{\mathbf{k}} \left( \sum_{i=1}^{2} k_i \frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) - B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) \cdot \log B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) \right) \\ = O(1/\sqrt{n}).$$

The summation for $nb + n^{1/2+\varepsilon} < k_1 \le n$ is much easier to handle, so we skip it.

For dimension $m > 2$ one has to handle multivariate Gaussian approximations. This is just technical and more involved but there is no substantial problem (see Remark 2 at the end of Section 4.2). Summing up, in all cases the remainder is of order $O(1/\sqrt{n})$. This completes the proof of Theorem 2.1.

### 3.3. Proof of Theorem 2.3

In order to prove Theorem 2.3 we estimate the sum $D_n = D_n(\mathcal{M})$:

$$\begin{aligned} D_n &= \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \max_{\theta_1 \le \theta_2 \le \cdots \le \theta_m} \theta_1^{k_1} \theta_2^{k_2} \cdots \theta_m^{k_m} \\ &= \sum_{\mathbf{k} \in n\mathcal{M}} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_1} + \sum_{\mathbf{k} \notin n\mathcal{M}} \binom{n}{\mathbf{k}} \max_{\theta_1 \le \theta_2 \le \cdots \le \theta_m} \theta_1^{k_1} \theta_2^{k_2} \cdots \theta_m^{k_m} \\ &= D_n^{(\mathcal{M})} + D_n^{(\Theta \setminus \mathcal{M})}. \end{aligned}$$

First let us consider the first part $D_n^{(\mathcal{M})}$ that we (again) partition into two parts. We set

$$\mathcal{M}^{<} = \{\mathbf{k} \in n\mathcal{M} : k_1 < k_2 < \cdots k_m\}$$

and then

$$D_n^{(\mathcal{M})} = D_n^{(\mathcal{M}^{<})} + D_n^{(\mathcal{M} \setminus \mathcal{M}^{<})}.$$

By using the fact that the functional $\prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_1}$ is symmetric it follows that the unconstrained sum (that we denote by $D_n(\Theta)$) is given by

$$D_n(\Theta) = \sum_{\mathbf{k}} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_1} = m!\, D_n^{(\mathcal{M}^{<})} + \sum_{k_i = k_j \text{ for some } i \ne j} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_1}.$$

Now we have

$$\sum_{k_i=k_j \text{ for some } i\neq j} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_i} \leq \binom{m}{2} \sum_{k_1=k_2} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_i}.$$

With the help of the approximation for $k_i \geq 1$ and $k_1 = k_2$:

$$\binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_i} \approx (2\pi)^{-\frac{m-1}{2}} \sqrt{\frac{n}{k_1^2 k_3 \cdots k_m}},$$

we arrive at

$$\sum_{\mathbf{k}}{}' \frac{1}{\sqrt{k_1^2 k_3 \cdots k_m}} = \sum_{k=1}^{n/2} \frac{1}{k} S^{(1)}_{m-2}(n-2k), \tag{28}$$

where $S^{(1)}_n(m)$ is defined as follows

$$S^{(1)}_m(n) = \sum_{\mathbf{k}}{}' \frac{1}{\sqrt{k_1 \cdots k_m}}, \tag{29}$$

with the sum taken over all $m$-dimensional integer vectors $\mathbf{k} = (k_1, \ldots, k_m)$ with $k_j \geq 1$ $(1 \leq j \leq m)$ and $k_1 + \cdots + k_m = n$. Using the upper bound (18) it follows that

$$\sum_{k_1=k_2} \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \left(\frac{k_i}{n}\right)^{k_1} = \tag{30}$$
$$= O\left((2\pi)^{-\frac{m-1}{2}} \sqrt{m}\, n^{\frac{m-2}{2}} \log n \, B_m(\mathbf{1/2}) \exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)\right).$$

Finally, by using (19) we find

$$D_n^{(\mathcal{M}^<)} = \frac{1}{m!} (2\pi)^{-\frac{m-1}{2}} n^{\frac{m-1}{2}} B_m(\mathbf{1/2}) \exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)$$
$$\times \left(1 + O\left(\frac{m^{5/2} \log n}{\sqrt{n}} \exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)\right)\right).$$

For the sum $D_n^{(\mathcal{M}\setminus\mathcal{M}^<)}$ we use a simple estimate

$$D_n^{(\mathcal{M}\setminus\mathcal{M}^<)} \leq (m-1) \sum_{k_1=k_2} \binom{n}{\mathbf{k}} \left(\frac{k_i}{n}\right)^{k_i}$$

that can upper bounded by (30).

It remains to consider the sum $D_n^{(\Theta\setminus\mathcal{M})}$, where we sum over integer vectors $(k_1,\ldots,k_m)$ for which there exists $i$ with $k_i > k_{k+1}$. Due to the concavity property of the term $\prod_i \theta_i^{k_i}$ (that has optimum outside of $\mathcal{M}$) it follows that the optimum $(\theta_1^{\mathrm{opt}},\ldots,\theta_m^{\mathrm{opt}}) \in \mathcal{M}$ of

$$\prod_{i=1}^m (\theta_1^{\mathrm{opt}})^{k_i} = \max_{(\theta_1,\ldots,\theta_m)\in\mathcal{M}} \prod_{i=1}^m \theta_i^{k_i}$$

has to be on the boundary of $\mathcal{M}$, so we either have $\theta_1 = 0$ or $\theta_j = \theta_{j+1}$ for some $j = 1,\ldots,m-1$. Hence we trivially have

$$\max_{(\theta_1,\ldots,\theta_m)\in\mathcal{M}} \prod_{i=1}^m \theta_i^{k_i} \le 0^{k_1} \max_{\theta_2+\cdots+\theta_{m-1}=1} \prod_{i=2}^m \theta_i^{k_i}$$
$$+ \sum_{j=1}^{m-1} \max_{\theta_1+\cdots+\theta_{j-1}+2\theta_j+\theta_{j+2}+\cdots\theta_m=1} \prod_{i<j} \theta_i^{k_i} \cdot \theta_j^{k_j+k_{j+1}} \cdot \prod_{i>j+1} \theta_i^{k_i}.$$

The first part is only non-zero if $k_1 = 0$. So it simplifies to the $(m-1)$-dimensional case. For the second part we note that

$$\sum_{k_j+k_{j+1}=K} \binom{n}{k_1 \cdots k_m} = 2^K \binom{n}{k_1 \cdots k_{j-1}\, K\, k_{j+2} \cdots k_m}.$$

Thus, we are led to the optimize

$$\max_{\theta_1+\cdots+\theta_{j-1}+2\theta_j+\theta_{j+2}+\cdots\theta_m=1} \prod_{i<j} (\theta_i)^{k_i} \cdot (2\theta_j)^K \cdot \prod_{i>j+1} \theta_i^{k_i}$$
$$= \prod_{i<j} \left(\frac{k_i}{n}\right)^{k_i} \cdot \left(\frac{K}{n}\right)^K \cdot \prod_{i>j+1} \left(\frac{k_i}{n}\right)^{k_i}.$$

So this case simplifies to the $m-1$-dimensional case, too. Summing up we have

$$D_n^{(\Theta\setminus\mathcal{M})} = O\left((2\pi)^{-\frac{m-1}{2}} m^{3/2} n^{\frac{m-2}{2}} B_m(\mathbf{1/2}) \exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)\right). \tag{31}$$

Putting all together, we find

$$D_n = \frac{1}{m!}(2\pi)^{-\frac{m-1}{2}} n^{\frac{m-1}{2}} B_m(\mathbf{1/2}) \exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)$$
$$+ O\left((2\pi)^{-\frac{m-1}{2}} m^{3/2} n^{\frac{m-2}{2}} B_m(\mathbf{1/2}) \exp\left(O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right)\right).$$

This completes the proof of Theorem 2.3.

## 4. Proofs of technical lemmas

In this technical section we provide the proof of Lemmas 1–3.

### 4.1. Proof of Lemma 1

**4.1.1. Upper bounds for $S_m^{(1)}(n)$ and $|S_m(n) - S_m^{(1)}(n)|$** We recall that $S_m(n)$ and $S_m^{(1)}(n)$ are defined in (17) and (29), respectively. In a first step we prove uniform upper bounds for $S_m^{(1)}(n)$ and for the difference $|S_m(n) - S_m^{(1)}(n)|$ for $m \geq 2$ and $n \geq 1$:

$$S_m^{(1)}(n) = O\left(n^{\frac{m-2}{2}} B_m(\mathbf{1/2})\right) \tag{32}$$

$$\left|S_m(n) - S_m^{(1)}(n)\right| = O\left(B(\mathbf{1/2}) m^{\frac{3}{2}} n^{\frac{m}{2}-\frac{3}{2}}\right). \tag{33}$$

By definition (29) we have the recurrence

$$S_m^{(1)}(n) = \sum_{k=1}^{n-1} \frac{1}{\sqrt{k}} S_{m-1}^{(1)}(n-k). \tag{34}$$

Note also that

$$B(\mathbf{1/2}) = B_m(\mathbf{1/2}) = \frac{\Gamma(1/2)^m}{\Gamma(m/2)} = \frac{\pi^{m/2}}{\Gamma(m/2)}$$

gives the relation

$$B_{m-1}(\mathbf{1/2}) B(1/2,(m-1)/2) = B_m(\mathbf{1/2}).$$

Furthermore, for $m \geq 3$ the function $x \mapsto x^{-\frac{1}{2}}(n-x)^{\frac{m-3}{2}}$ is decreasing which leads to the upper bound

$$\sum_{k=1}^{n-1} \frac{1}{\sqrt{k}} (n-k)^{\frac{m-3}{2}} \leq \int_0^n \frac{1}{\sqrt{x}} (n-x)^{\frac{m-3}{2}} dx = n^{\frac{m-2}{2}} B(1/2,(m-1)/2). \tag{35}$$

It is now easy to obtain (32) by induction. Clearly we have

$$\sum_{k=1}^{n-1} \frac{1}{\sqrt{k(n-k)}} \leq C_1$$

with a proper constant $C_1 > 0$. Thus, (32) holds for $m = 2$. Furthermore, by using (34) and (35) we get inductively (for $m \ge 3$)

$$\begin{aligned} S_m^{(1)}(n) &\le \sum_{k=1}^{n-1} \frac{1}{\sqrt{k}} C_1 (n-k)^{\frac{m-3}{2}} B_{m-1}(\mathbf{1/2}) \\ &\le C_1 n^{\frac{m-2}{2}} B(1/2,(m-1)/2) B_{m-1}(\mathbf{1/2}) \\ &= C_1 n^{\frac{m-2}{2}} B_m(\mathbf{1/2}). \end{aligned}$$

This completes the proof of (32).

By Stirling's formula we have $k^k e^{-k}\sqrt{2\pi k} e^{1/(12k+1)} < k! < k^k e^{-k}\sqrt{2\pi k}$ $e^{1/(12k)}$ for all $k \ge 1$. Hence $S_m(n)$ satisfies

$$\sum_{\mathbf{k}}{}' \frac{1}{\sqrt{k_1 \cdots k_m}} e^{-\sum_{i=1}^m \frac{1}{12k_i+1}} < S_m(n) < \sum_{\mathbf{k}}{}' \frac{1}{\sqrt{k_1 \cdots k_m}} e^{-\sum_{i=1}^m \frac{1}{12k_i} + \frac{1}{12n+1}}.$$

By applying the inequality $1 - e^{-x} \le x$ we, thus, have

$$\begin{aligned} \left| S_m(n) - S_m^{(1)}(n) \right| &\le \sum_{\mathbf{k}}{}' \frac{1}{\sqrt{k_1 \cdots k_m}} \left( \sum_{i=1}^m \frac{1}{12k_i} + \frac{1}{12n} \right) \\ &= O\left( m \sum_{\mathbf{k}}{}' \frac{1}{\sqrt{k_1 \cdots k_m}} \frac{1}{k_1} \right). \end{aligned} \tag{36}$$

We apply now the upper bound (32) that leads to

$$\begin{aligned} \sum_{\mathbf{k}} \frac{1}{\sqrt{k_1 \cdots k_m}} \frac{1}{k_1} &= \sum_{k=1}^{n-1} \frac{1}{k\sqrt{k}} S_{m-1}^{(1)}(n-k) \\ &= O\left( B_{m-1}(\mathbf{1/2}) \sum_{k=1}^{n-1} \frac{1}{k\sqrt{k}} (n-k)^{\frac{m-3}{2}} \right). \end{aligned}$$

Since $B_{m-1}(\mathbf{1/2}) = O\left(\sqrt{m}\, B_m(\mathbf{1/2})\right)$ and

$$\sum_{k=1}^{n-1} \frac{1}{k\sqrt{k}} (n-k)^{\frac{m-3}{2}} = O\left( n^{\frac{m-3}{2}} \right)$$

we obtain

$$\sum_{\mathbf{k}} \frac{1}{\sqrt{k_1 \cdots k_m}} \frac{1}{k_1} = O\left( B(\mathbf{1/2}) m^{\frac{1}{2}} n^{\frac{m}{2} - \frac{3}{2}} \right). \tag{37}$$

Together with (36) we verify (33)

$$\left|S_m(n) - S_m^{(1)}(n)\right| = O\left(B(\mathbf{1/2}) m^{\frac{3}{2}} n^{\frac{m}{2}-\frac{3}{2}}\right).$$

**4.1.2. Asymptotic relation for $S_m^{(1)}(n)$** In a next step we prove an asymptotic relation for $S_m^{(1)}(n)$ that holds uniformly for $2 \le m \le n$:

$$S_m^{(1)}(n) = n^{\frac{m}{2}-1} B_m(\mathbf{1/2}) \left(1 + O\left(\frac{m^{3/2}}{\sqrt{n}}\right)\right). \tag{38}$$

First we note that the upper bound (32) shows that (38) is true if $m \ge cn^{1/3}$ (for any positive constant $c > 0$. Thus, it will remain to consider the case $m < cn^{1/3}$.

It remains to show that (38) holds for $m < cn^{1/3}$. For this purpose we use (again) induction and the simplest form of the Euler-MacLaurin formula [19], namely

$$\sum_{n=a}^{b} f(n) = \int_a^b f(x)\,dx + \frac{f(a)+f(b)}{2} + \int_a^b \left(x - \lfloor x \rfloor - \frac{1}{2}\right) f'(x)\,dx,$$

where $f(x)$ is continuously differentiable and $a < b$ are integers. For example, we obtain the asymptotic relation

$$\begin{aligned} S_2^{(1)}(n) &= \sum_{k=1}^{n-1} \frac{1}{\sqrt{k(n-k)}} \\ &= \int_1^{n-1} \frac{dx}{\sqrt{x(n-x)}} + O\left(n^{-1/2}\right) + O\left(\int_1^{n-1} \frac{dx}{\sqrt{x^3(n-x)}}\right) \\ &= B(1/2, 1/2) + O\left(n^{-1/2}\right) \end{aligned} \tag{39}$$

that is in accordance with (38).

Furthermore we find for $m \ge 3$

$$\begin{aligned} \sum_{k=1}^{n-1} \frac{1}{\sqrt{k}} (n-k)^{\frac{m-3}{2}} &= \int_1^{n-1} \frac{1}{\sqrt{x}} (n-x)^{\frac{m-3}{2}}\,dx + O\left(n^{\frac{m-3}{2}}\right) \\ &+ O\left(\int_1^{n-1} x^{-3/2} (n-x)^{\frac{m-3}{2}}\,dx\right) + O\left(m \int_1^{n-1} x^{-1/2} (n-x)^{\frac{m-5}{2}}\,dx\right) \end{aligned}$$

$$= n^{\frac{m-2}{2}} B(1/2,(m-1)/2) + O\left(n^{\frac{m-3}{2}}\right) + O\left(\sqrt{m}\, n^{\frac{m-4}{2}}\right)$$
$$= n^{\frac{m-2}{2}} B(1/2,(m-1)/2)\left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right).$$

With the help of these expansions we will prove (38) by induction on $m$. For notational convenience we write $O_C(X)$ for a term that is absolutely bounded by $\le C\,|X|$, that is, we specify the implicit constant.

We already mentioned that (38) holds for $m = 2$, see (39). Actually we can prove (38) for every fixed $m \ge 2$ (we just have to apply the inductive method a finite number of times). Hence, we can assume that (38) holds for $m \le m_0$, where $m_0 \ge 2$ is a fixed but arbitrary integer. Now suppose that $m > m_0$ and that (38) holds for $m-1$:

$$S^{(1)}_{m-1}(n) = n^{\frac{m-3}{2}} B_{m-1}(\mathbf{1/2})\left(1 + O_C\left(\frac{(m-1)^{3/2}}{\sqrt{n}}\right)\right).$$

By (34) we, thus, obtain

$$S^{(1)}_{m}(n) = \sum_{k=1}^{n-1} \frac{1}{\sqrt{k}}(n-k)^{\frac{m-3}{2}} B_{m-1}(\mathbf{1/2})\left(1 + O_C\left(\frac{(m-1)^{3/2}}{\sqrt{n-k}}\right)\right)$$
$$= n^{\frac{m-2}{2}} B(1/2,(m-1)/2) B_{m-1}(\mathbf{1/2})\left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right)$$
$$+ O_C\left((m-1)^{3/2} n^{\frac{m-3}{2}} B(1/2,(m-2)/2) B_{m-1}(\mathbf{1/2})\left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right)\right).$$

Now note that $(m-1)^{3/2} \le m^{3/2}(1 - c_1/m)$, where the constant $c_1 > 0$ is certainly $\le \frac{3}{2}$. However, if $m$ is sufficiently large (say $m \ge m_0$) then we can assume that $c_1 > 1$. Furthermore we have

$$B(1/2,(m-2)/2) B_{m-1}(\mathbf{1/2}) \le B_m(\mathbf{1/2})\,(1 + c_2/m),$$

where $c_2 > \frac{1}{2}$. However, if $m$ is sufficiently large then we can assume that $c_2 < \frac{2}{3}$. Consequently we have

$$S^{(1)}_{m}(n) = n^{\frac{m-2}{2}} B_m(\mathbf{1/2})\left(1 + O\left(\sqrt{\frac{m}{n}}\right) + \right.$$
$$\left. + O_C\left(\frac{m^{3/2}}{\sqrt{n}}\left(1 - \frac{c_1}{m}\right)\left(1 + \frac{c_2}{m}\right)\left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right)\right)\right).$$

Recall that we only have to consider the case, where $m < cn^{3/2}$, where $c > 0$ can be arbitrarily chosen. Since $c_1 > 1$ and $c_2 < \frac{2}{3}$ we can choose $c > 0$ properly such that

$$\left(1 - \frac{c_1}{m}\right)\left(1 + \frac{c_2}{m}\right)\left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right) \le 1 - \frac{1}{4m}$$

for $m_0 \le m < cn^{3/2}$. Finally, by adjusting $C$ we also assure that

$$O_C\left(\frac{m^{3/2}}{\sqrt{n}}\left(1 - \frac{1}{4m}\right)\right) + O\left(\sqrt{\frac{m}{n}}\right) = O_C\left(\frac{m^{3/2}}{\sqrt{n}}\right)$$

for $m < cn^{3/2}$. Hence, (38) follows for all $m < cn^{3/2}$ and we are done.

*4.1.2.1. Proof of (18)* By combining (33) and (38) we immediately obtain (18) which completes the proof of Lemma 1.

## 4.2. Proof of Lemma 2

**4.2.1. Reduction to a half space** We recall that a convex polytope is the intersection of finitely many half spaces. Thus, in order to cover $\Theta \setminus \mathcal{S}$ and estimate $D_n - D_n^{(\mathcal{S})}$ as in (26) of Lemma 2, we just have to consider the union of finitely many half spaces. Furthermore, by concavity of $\prod_i \theta_i^{k_i}$, it follows that the optimal choice of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ has to be on the boundary of $\mathcal{S}$. In other words, $\mathbf{k}/n \notin \mathcal{S}$ implies that there exists a half space $H$ such that $\mathbf{k} \in n(H \cap \Theta)$. Therefore it is sufficient to consider just one half space $H$ and take the optimum $\boldsymbol{\theta}$ on the boundary $\partial H$.

Without loss of generality we can assume that the boundary $\partial H$ is given by the equation

$$a_1\theta_1 + a_2\theta_2 + \cdots + a_{m-1}\theta_{m-1} = 1,$$

where $a_{m-1} \neq 0$. (If all coefficients are non-zero we can use the equation $\theta_1 + \cdots + \theta_m = 1$ to eliminate one variable.) Of course we only consider half spaces with the property that $H \cap \Theta$ is a $(m-2)$-dimensional polytope of positive $(m-2)$-dimensional volume.

The difference $D_n - D_n^{(\mathcal{S})}$ that we need to estimate is then consisting of the following sums

$$S = \sum_{\mathbf{k} \in n(H \cap \Theta)} \binom{n}{\mathbf{k}} \max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^{m} \theta_i^{k_i}. \tag{40}$$

In what follows we assume that $k_i > 0$ for all $i = 1, \ldots, m$. If $k_i = 0$ then the problem is reduced to a lower dimensional one, where we can assume by induction that the corresponding contribution is of smaller order.

It is also convenient to consider $\mathbf{k} = (k_1, \ldots, k_m)$ as continuous non-negative variables with $k_1 + \cdots + k_m = n$. Clearly, $\binom{n}{\mathbf{k}}$ as well as $\prod_{i=1}^m \theta_i^{k_i}$ are then well defined, too.

If $\mathbf{k}_0 = (k_{1,0}, \ldots, k_{m,0}) \in n(\partial H \cap \Theta)$, that is, if

$$\sum_{i=1}^{m} k_{i,0} = n \quad \text{and} \quad \sum_{i=1}^{m-1} a_i k_{i,0} = n \tag{41}$$

then

$$\max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^{m} \theta_i^{k_{i,0}} = \prod_{i=1}^{m} \left( \frac{k_{i,0}}{n} \right)^{k_{i,0}}.$$

**4.2.2. A proper parametrization** The next goal is to represent $\mathbf{k} \in n\Theta$ as $\mathbf{k} = \mathbf{k}_0 + \ell$, where $\mathbf{k}_0 \in n(\partial H \cap \Theta)$ such that we can control $\max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^m \theta_i^{k_i}$.

More precisely, let's fix some $\mathbf{k}_0 = (k_{1,0}, \ldots, k_{m,0}) \in n(\partial H \cap \Theta)$ with $k_{i,0} > 0$ and set $\theta_{i,0} = k_{i,0}/n$. We are searching for all non-negative $\mathbf{k} \in n\Theta$ (that is, $k_1 + \cdots + k_m = n$ and $k_i \geq 0$) for which

$$\max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^{m} \theta_i^{k_i} = \prod_{i=1}^{m} \theta_{i,0}^{k_i}. \tag{42}$$

We will show that all solutions can be written as

$$k_i = k_{i,0} + \ell_i$$

where $\ell_i$, $1 \leq i \leq m-1$, are parametrized by $\ell_m$:

$$\ell_i = \frac{k_{i,0}}{k_{m,0}} (1 - a_i) \ell_m, \quad 1 \leq i \leq m-1. \tag{43}$$

Conversely for every $\mathbf{k} \in n\Theta$ (with positive coordinates) there is a unique vector $\mathbf{k}_0 \in \partial H \cap \Theta$ with these properties.

We parametrize $\boldsymbol{\theta} \in \partial H \cap \Theta$ by $\theta_1, \ldots, \theta_{m-2}$ so that $\theta_{m-1}$ and $\theta_m$ are given by

$$\theta_{m-1} = \frac{1}{a_{m-1}} - \sum_{i=1}^{m-2} \frac{a_i}{a_{m-1}} \theta_i,$$

$$\theta_m = \frac{a_{m-1}-1}{a_{m-1}} + \sum_{i=1}^{m-2} \frac{a_i - a_{m-1}}{a_{m-1}}\theta_i.$$

Since our optimum is assumed to be in the interior of $\partial H \cap \Theta$ and by the concavity property of the mapping $\boldsymbol{\theta} \mapsto \sum_i k_i \log\theta_i$ it follows that $\mathbf{k}$ (with (42)) has to satisfy (besides (41)) the system of equations

$$\frac{\partial}{\partial\theta_i}\sum_{i=1}^{m} k_i \log\theta_{i,0} = \frac{k_i}{\theta_{i,0}} - \frac{k_{m-1}}{\theta_{m-1,0}}\frac{a_i}{a_{m-1}} + \frac{k_m}{\theta_{m,0}}\frac{a_i - a_{m-1}}{a_{m-1}} = 0, \quad 1 \le i \le m-2.$$

Clearly $k_i = k_{i,0}$ satisfy this system of equations. If we set $\ell_i = k_i - k_{i,0}$, $1 \le i \le m$, then we certainly have $\sum_i \ell_i = 0$ and

$$\frac{\ell_i}{\theta_{i,0}} - \frac{\ell_{m-1}}{\theta_{m-1,0}}\frac{a_i}{a_{m-1}} + \frac{\ell_m}{\theta_{m,0}}\frac{a_i - a_{m-1}}{a_{m-1}} = 0, \quad 1 \le i \le m-2,$$

that is, we have $m-1$ homogeneous equations for $m$ variables. It is easy to check that the one dimensional solution is just (43) and this completes the proof of (43).

Conversely, it is not difficult to observe that for every $\mathbf{k} \in n\Theta^\circ$ there is a uniquely defined $\mathbf{k}_0 = \mathbf{k}_0(k_1,\ldots,k_{m-1}) \in n(\partial H \cap \Theta)$ such that

$$k_i = k_{i,0}\left(1 + \frac{k_m - k_{m,0}}{k_{m,0}}(1-a_i)\right), \quad 1 \le i \le m.$$

By setting $\lambda = \ell_m / k_{m,0}$ one finds $k_i = k_{i,0}(1+\lambda(1-a_i))$ and consequently we have one equation

$$\sum_{i=1}^{m-1} \frac{a_i k_i}{1+\lambda(1-a_i)} = n \tag{44}$$

for computing $\lambda = \lambda(k_1,\ldots,k_{m-1})$ from which we obtain

$$k_{i,0} = k_{i,0}(k_1,\ldots,k_{m-1}) = k_i/(1+\lambda(1-a_i)).$$

We note again that $k_i$ and $k_{i,0}$ are considered as continuous variables. We also note that for $\mathbf{k} \in n(\partial H \cap \Theta)$ we have $\lambda = 0$ or equivalently $k_{i,0} = k_i$.

**4.2.3. Asymptotic concentration** Next we show that the terms $\binom{n}{k}\prod_{i=1}^m \theta_{i,0}^{k_i}$ are highly concentrated for those $\mathbf{k}$ that are close to the corresponding $\mathbf{k}_0 \in (\partial H \cap \Theta)$.

Let $\mathcal{K}_0$ denote the set of $\mathbf{k} = (k_1, \ldots, k_m) \in n\Theta$ with $k_i > 0$, for which $|\ell_i|/k_{i,0} \le |1 - a_i| c_0$ (for some sufficiently small constant $c_0 > 0$ that will be fixed in the sequel) and $\mathcal{K}_1$ the remaining $\mathbf{k} \in n\Theta$ with $k_i > 0$. We use the representation

$$\binom{n}{\mathbf{k}} \max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^{m} \theta_i^{k_i} = \binom{n}{\mathbf{k}} \prod_{i=1}^{m} \theta_{i,0}^{k_i}$$
$$= \binom{n}{\mathbf{k}_0} \prod_{i=1}^{m} \theta_{i,0}^{k_{i,0}} \prod_{i=1}^{m} \frac{\Gamma(k_{i,0}+1) k_{i,0}^{k_i - k_{i,0}}}{\Gamma(k_i+1)},$$

where $\theta_{i,0} = k_{i,0}/n$. If $\mathbf{k} \in \mathcal{K}_0$ we get uniformly (for some constant $c_1 > 0$)

$$\binom{n}{\mathbf{k}} \max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^{m} \theta_i^{k_i} = O\left( \sqrt{\frac{n}{k_{1,0} \cdots k_{m,0}}} \exp\left( -c_1 \sum_{i=1}^{m} \frac{\ell_i^2}{k_{i,0}} \right) \right)$$
$$= O\left( \sqrt{\frac{n}{k_{1,0} \cdots k_{m,0}}} \exp\left( -c_1 \frac{\ell_m^2}{k_{m,0}^2} \sum_{i=1}^{m} k_{i,0} (1 - a_i)^2 \right) \right),$$

whereas for $\mathbf{k} \in \mathcal{K}_1$ we have uniformly

$$\binom{n}{\mathbf{k}} \max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^{m} \theta_i^{k_i} = O\left( \sqrt{\frac{n}{k_{1,0} \cdots k_{m,0}}} \exp\left( -c_1 c_0^2 \sum_{i=1}^{m} k_{i,0} (1 - a_i)^2 \right) \right).$$

**4.2.4. Upper bounds for $S_0$** First we study those situations, where $k_i$ and $k_{i,0}$ do not differ too much, that is, if $\mathbf{k} \in \mathcal{K}_0$. We set

$$S_0 = \sum_{\mathbf{k} \in \mathcal{K}_0} \binom{n}{\mathbf{k}} \max_{\boldsymbol{\theta} \in \partial H \cap \Theta} \prod_{i=1}^{m} \theta_i^{k_i}$$
$$= O\left( \sum_{\mathbf{k} \in \mathcal{K}_0} \sqrt{\frac{n}{k_{1,0} \cdots k_{m,0}}} \exp\left( -c_1 \frac{\ell_m^2}{k_{m,0}^2} \sum_{i=1}^{m} k_{i,0} (1 - a_i)^2 \right) \right).$$

We will next show that the derivatives $\frac{\partial}{\partial k_j} k_{i,0}$ are bounded if $c_0$ is sufficiently small. By implicit differentiation of the equation (44) we have

$$\frac{\partial \lambda}{\partial k_j} = a_j (1 + \lambda(1 - a_j)) \left( \sum_{i=1}^{m-1} \frac{a_i (1 - a_i) k_i}{(1 + \lambda(1 - a_i))^2} \right)^{-1}.$$

Note that (with the convention $a_m = 0$)

$$\sum_{i=1}^{m-1} a_i(1-a_i)k_{i,0} = -\sum_{i=1}^{m}(1-a_i)^2 k_{i,0} \neq 0.$$

Consequently (and by using the property $|\lambda| \le c_0$) the relation

$$\begin{aligned}
\sum_{i=1}^{m-1} \frac{a_i(1-a_i)k_i}{(1+\lambda(1-a_i))^2} &= \sum_{i=1}^{m-1} \frac{a_i(1-a_i)k_{i,0}}{1+\lambda(1-a_i)} \\
&= \sum_{i=1}^{m-1} a_i(1-a_i)k_{i,0}(1+O(c_0(1-a_i)) \\
&= -\sum_{i=1}^{m}(1-a_i)^2 k_{i,0} + O\left(c_0 \sum_{i=1}^{m-1} \max_j |a_j|(1-a_i)^2 k_{i,0}\right) \\
&= -\sum_{i=1}^{m}(1-a_i)^2 k_{i,0}\,(1+O(c_0)))
\end{aligned}$$

leads to

$$\frac{\partial \lambda}{\partial k_j} = O\left(\left(\sum_{i=1}^{m}(1-a_i)^2 k_{i,0}\right)^{-1}\right).$$

Thus, if $a_i \neq 1$ we have

$$\begin{aligned}
\frac{\partial k_{i,0}}{\partial k_j} &= \frac{\delta_{ij}}{1+\lambda(1-a_i)} - \frac{k_i(1-a_i)}{(1+\lambda(1-a_i))^2}\frac{\partial \lambda}{\partial k_j} \\
&= O(1) + O\left(\frac{k_{i_0}(1-a_i)}{\sum_{i=1}^{m-1}(1-a_i)^2 k_{i,0}}\right) \\
&= O(1) + O((1-a_i)^{-1}) = O(1).
\end{aligned}$$

If $a_i = 1$ then we trivially have $\frac{\partial}{\partial k_j} k_{i,0} = O(1)$.

At this step we can replace the sum $S_0$ by an integral since we know that the derivatives $\frac{\partial}{\partial k_j} k_{i,0}$ are all uniformly bounded and the upper bound is (up to constant) stable if we replace $k_{i,0}$ by $k_{i,0} + O(1)$. Consequently we have

$$S_0 = O\left(\int_{\mathcal{K}_0} \sqrt{\frac{n}{k_{1,0}\cdots k_{m,0}}} \exp\left(-c_1 \frac{\ell_m^2}{k_{m,0}^2} \sum_{i=1}^{m} k_{i,0}(1-a_i)^2\right) dk_1 \cdots dk_{m-1}\right),$$

where $k_{i,0}$ and $\ell_m$ are considered as functions in $k_1, \ldots, k_{m-1}$.

In a next step we substitute $k_1, \ldots, k_{m-1}$ by $k_{1,0}, \ldots, k_{m-2,0}$ and $\ell_m$. Recall that $k_{m-1,0}$ and $k_{m,0}$ are well defined if $k_{1,0}, \ldots, k_{m-2,0}$ are given:

$$k_{m-1,0} = \frac{n}{a_{m-1}} - \sum_{j=1}^{m-2} \frac{a_j}{a_{m-1}} k_{j,0},$$
$$k_{m,0} = n\frac{a_{m-1}-1}{a_{m-1}} + \sum_{j=1}^{m-2} \frac{a_j - a_{m-1}}{a_{m-1}} k_{j,0}.$$

Thus, this substitution mapping is given by

$$k_i = k_{i,0}\left(1 + \ell_m \frac{1-a_i}{k_{m,0}}\right) = k_{i,0}\left(1 + \ell_m \frac{1-a_i}{n\frac{a_{m-1}-1}{a_{m-1}} + \sum_{j=1}^{m-2} \frac{a_j - a_{m-1}}{a_{m-1}} k_{j,0}}\right)$$

for $1 \le i \le m-2$ and by

$$\begin{aligned} k_{m-1} &= k_{m-1,0}\left(1 + \ell_m \frac{1-a_{m-1}}{k_{m,0}}\right) \\ &= \left(\frac{n}{a_{m-1}} - \sum_{j=1}^{m-2} \frac{a_j}{a_{m-1}} k_{j,0}\right)\left(1 + \ell_m \frac{1-a_{m-1}}{n\frac{a_{m-1}-1}{a_{m-1}} + \sum_{j=1}^{m-2} \frac{a_j - a_{m-1}}{a_{m-1}} k_{j,0}}\right). \end{aligned}$$

It is an easy (and nice) exercise to show that the functional determinant of this mapping is given by

$$\frac{\partial(k_1, \ldots, k_{m-1})}{\partial(k_{1,0}, \ldots, k_{m-2,0}, \ell_m)} = -\frac{1}{a_{m-1}k_{m,0}} \sum_{i=1}^{m} k_{i,0}(1-a_i)^2 \cdot \left(1 + O\left(\ell_m/k_{m,0}\right)\right).$$

Hence, by assuming that $\lambda = \ell_m/k_{m,0}$ is sufficiently small, the sum $S_0$ is also upper bounded by

$$S_0 = O\left(\int_{\mathcal{K}_{0,0}} \sqrt{\frac{n}{k_{1,0}\cdots k_{m,0}}} \exp\left(-c_1 \frac{\ell_m^2}{k_{m,0}^2} \sum_{i=1}^{m} k_{i,0}(1-a_i)^2\right) \cdot \frac{\sum_{i=1}^{m} k_{i,0}(1-a_i)^2}{k_{m,0}}\, dk_{1,0}\cdots dk_{m-2,0}\, d\ell_m\right)$$

where $\mathcal{K}_{0,0}$ is the corresponding image set. We are now ready to integrate.

Since

$$\int_{|\ell_m|\le c_0 k_{m,0}} \exp\left(-c_1\frac{\ell_m^2}{k_{m,0}^2}\sum_{i=1}^m k_{i,0}(1-a_i)^2\right)\,d\ell_m$$
$$= O\left(\frac{k_{m,0}}{\sqrt{\sum_{i=1}^m k_{i,0}(1-a_i)^2}}\right),$$
$$\sum_{i=1}^m k_{i,0}(1-a_i)^2 = O(n),$$

and

$$\int_{\partial H\cap\Theta}\sqrt{\frac{1}{k_{1,0}\cdots k_{m,0}}}\,dk_{1,0}\cdots dk_{m-2,0} = O\left(n^{\frac{m}{2}-2}\right)$$

we arrive at

$$S_0 = O\left(n^{\frac{m}{2}-1}\right).$$

**4.2.5. Upper bounds for $S_1$** The remaining sum

$$S_1 = \sum_{\mathbf{k}\in\mathcal{K}_1}\binom{n}{\mathbf{k}}\max_{\boldsymbol{\theta}\in\partial H\cap\Theta}\prod_{i=1}^m \theta_i^{k_i}$$

is more easy to handle. Let us suppose first that $a_i\ne 1$ for all $1\le i\le m-1$. Then we have

$$\sum_{i=1}^m k_{i,0}(1-a_i)^2 \ge \min_{1\le i\le m}(1-a_i)^2\sum_{i=1}^m k_{i,0} = \min_{1\le i\le m}(1-a_i)^2\,n$$

and, thus, for $\mathbf{k}\in\mathcal{K}_1$

$$\binom{n}{\mathbf{k}}\max_{\boldsymbol{\theta}\in\partial H\cap\Theta}\prod_{i=1}^m\theta_i^{k_i} = O\left(\sqrt{\frac{n}{k_{1,0}\cdots k_{m,0}}}\exp\left(-c_2 n\right)\right).$$

This directly implies that

$$S_1 = O\left(n^{m-1}e^{-c_2 n}\right) = O\left(n^{\frac{m}{2}-1}\right).$$

It remains to consider the case, where there exists $i$ with $a_i = 1$. Let $I_1$ denote the set of indices, where this occurs and $I_2 = \{1,2,\ldots,m\}\setminus I_1$. Note that $I_2 \neq \emptyset$ (since $a_m = 0$). Here we only have the lower bound

$$\sum_{i=1}^{m} k_{i,0}(1-a_i)^2 \ge \min_{i\in I_2}(1-a_i)^2 \sum_{i\in I_2} k_{i,0}.$$

Furthermore if $a_i = 1$ then $k_i = k_{i,0}$ and consequently

$$\sum_{i\in I_1} k_i = \sum_{i\in I_1} k_{i,0} \quad \text{and} \quad \sum_{j\in I_2} k_j = \sum_{j\in I_2} k_{j,0}.$$

We now use the upper bound

$$\sum_{\sum_{i\in I_1} k_1} \frac{1}{\sqrt{\prod_{i\in I_1} k_i}} = O\left(K_1^{\frac{|I_1|}{2}-1}\right)$$

to conclude that

$$\begin{aligned} S_1 &= O\left(\sqrt{n}\sum_{K_1=1}^{n} K_1^{\frac{|I_1|}{2}-1}(n-K_1)^{m-|I_1|}e^{-c_3(n-K_1)}\right) \\ &= O\left(n^{\frac{|I_1|-1}{2}}\right) \\ &= O\left(n^{\frac{m}{2}-1}\right). \end{aligned}$$

**4.2.6. Completion of the proof of Lemma 2** Summing up, we have shown that $S = S_0 + S_1 = O\left(n^{\frac{m}{2}-1}\right)$. Since the complement $\Theta \setminus \mathcal{S}$ is the union of finitely many half spaces this completes the proof of Lemma 2.

### 4.3. Proof of Lemma 3

Before we start with the proof of Lemma 3 we collect some comments on

$$B(\mathbf{k}+\mathbf{1/2}) = \frac{\Gamma\left(k_1+\frac{1}{2}\right)\cdots\Gamma\left(k_m+\frac{1}{2}\right)}{\Gamma\left(k_1+\cdots+k_m+\frac{m}{2}\right)}.$$

First this formula follows from the following substitution on $m$-dimensional integrals:

$$\begin{aligned} &\Gamma\left(k_1+\cdots+k_m+\frac{m}{2}\right)B(\mathbf{k}+\mathbf{1/2}) = \\ &= \int_0^\infty e^{-z} z^{k_1+\cdots+k_m+-\frac{m}{2}} z^{m-1}\,dz \int_\Theta \prod_{i=1}^{m} \theta_i^{k_i-\frac{1}{2}}\,d\boldsymbol{\theta} \end{aligned}$$

$$= \int_0^\infty \int_\Theta \prod_{i=1}^m \left( e^{-z\theta_i} (z\theta_i)^{k_i - \frac{1}{2}} \right) z^{m-1}\, dz\, d\boldsymbol{\theta}$$
$$= \prod_{i=1}^m \int_0^\infty e^{-z_i} z_i^{k_i - \frac{1}{2}}\, dz_i,$$

where $\theta_m$ abbreviates $\theta_m = 1 - (\theta_1 + \cdots + \theta_{m-1})$ and the substitution

$$z_i = z\theta_i \quad (1 \le i \le m-1), \quad z_m = z(1 - (\theta_1 + \cdots + \theta_{m-1})$$

has determinant $z^{m-1}$.

Second, we can use Stirling's formula to obtain the asymptotic formula

$$B(\mathbf{k} + \mathbf{1/2}) = \frac{(2\pi)^{\frac{m-1}{2}}}{n^{\frac{m-1}{2}}} \prod_{i=1}^m \left( \frac{k_i}{n} \right)^{k_i} \left( 1 + O\left( \sum_{i=1}^m \frac{1}{k_i + 1} \right) \right).$$

Note that we only have to work with $\Gamma$-integral, that can be asymptotically evaluated by

$$\Gamma(k + 1/2) = \int_0^\infty e^{-z} z^{k-\frac{1}{2}}\, dz$$
$$= \int_{|z-k| \le c\, k^{\frac{1}{2}+\varepsilon}} e^{-z} z^{k-\frac{1}{2}}\, dz + O\left( e^{-k} k^k e^{-c^2 k^{2\varepsilon}/2} \right)$$
$$= \sqrt{2\pi} e^{-k} k^k \left( 1 + \sum_{\ell=1}^L c_\ell k^{-\ell} + O(k^{-L-1}) \right)$$

for every given integer $L \ge 0$, for every $\varepsilon > 0$, and for every constant $c > 0$, and where $c_\ell$ are certain real constants. We just have to use the saddle point $z_0 = k$ of the function $e^{-z} z^k$ and the local expansion

$$e^{-z} z^{k-\frac{1}{2}} = e^{-k} k^{k-\frac{1}{2}} e^{-(z-k)^2/(2k)} \left( 1 - \frac{z-k}{2k} + \frac{(z-k)^3}{3k^2} + \cdots \right).$$

In particular, it is sufficient to consider the integral on the interval $[k - c\, k^\varepsilon, k + c\, k^\varepsilon]$. The remaining part of the integral is negligible.

**4.3.1. Asymptotics for $B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2})$** We adapt the above approach to the asymptotic evaluation of $B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2})$. By using the same substitution as above we have

$$\begin{aligned}\Gamma\left(k_1+\cdots+k_m+\frac{m}{2}\right) B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) &= \\ &= \int_0^\infty e^{-z} z^{k_1+\cdots+k_m+-\frac{m}{2}}\, z^{m-1}\, dz \int_{\mathcal{S}} \prod_{i=1}^m \theta_i^{k_i-\frac{1}{2}}\, d\boldsymbol{\theta} \\ &= \int_0^\infty \int_{\mathcal{S}} \prod_{i=1}^m \left(e^{-z\theta_i}(z\theta_i)^{k_i-\frac{1}{2}}\right) z^{m-1}\, dz\, d\boldsymbol{\theta} \\ &= \int_{\mathrm{cone}(\mathcal{S})} \prod_{i=1}^m e^{-z_i} z_i^{k_i-\frac{1}{2}}\, dz_1 \cdots dz_m, \qquad (45)\end{aligned}$$

where

$$\mathrm{cone}(\mathcal{S}) = \{z\boldsymbol{\theta} : z \ge 0,\ \boldsymbol{\theta} \in \mathcal{S}\}.$$

By assumption we have $\mathbf{k}/n \in \mathcal{S}^-$, which implies (for a properly chosen constant $c > 0$) that

$$\prod_{i=1}^m \left[k_i - c\, n^{\frac{1}{2}+\varepsilon}, k_i + c\, n^{\frac{1}{2}+\varepsilon}\right] \subseteq \mathrm{cone}(\mathcal{S}).$$

Since $k_i \le n$ it also follows that

$$\prod_{i=1}^m \left[k_i - c\, k_i^{\frac{1}{2}+\varepsilon}, k_i + c\, k_i^{\frac{1}{2}+\varepsilon}\right] \subseteq \mathrm{cone}(\mathcal{S}).$$

This implies that

$$\begin{aligned}\Gamma\left(k_1+\cdots+k_m+\frac{m}{2}\right) B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) = \Gamma\left(k_1+\cdots+k_m+\frac{m}{2}\right) B(\mathbf{k}+\mathbf{1/2}) \\ + O\left(e^{-k_1-\ldots-k_m} \prod_{i=1}^m k_i^{k_i} \cdot \sum_{i=1}^m k_i^{-L-1}\right)\end{aligned}$$

and consequently

$$\begin{aligned}B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) &= B(\mathbf{k}+\mathbf{1/2}) + O\left(n^{-\frac{m-1}{2}} \prod_{i=1}^m \left(\frac{k_i}{n}\right)^{k_i} \sum_{i=1}^m k_i^{-L-1}\right) \\ &= B(\mathbf{k}+\mathbf{1/2})\left(1 + O\left(\sum_{i=1}^m k_i^{-L-1}\right)\right).\end{aligned}$$

This proves the first part of Lemma 3.

**4.3.2. Asymptotic comparison** It remains to consider the second part of Lemma 3. We start with the integral representation (45) for

$$\Gamma\left(k_1 + \cdots + k_m + \frac{m}{2}\right) B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2})$$

and consider the derivative with respect to $k_i$. By the product rule this leads to in integral representation for

$$\begin{aligned}
\frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2}) &= \frac{1}{\Gamma\left(n + \frac{m}{2}\right)} \frac{\partial}{\partial k_i} \left(\Gamma\left(k_1 + \cdots + k_m + \frac{m}{2}\right) B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2})\right) \\
&\quad - \frac{\Gamma'\left(n + \frac{m}{2}\right)}{\Gamma\left(n + \frac{m}{2}\right)^2} \Gamma\left(n + \frac{m}{2}\right) B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2}) \\
&= \frac{1}{\Gamma\left(n + \frac{m}{2}\right)} \int_{\mathrm{cone}(\mathcal{S})} \log z_i \prod_{i=1}^{m} e^{-z_i} z_i^{k_i - \frac{1}{2}} \, dz_1 \cdots dz_m \\
&\quad - \frac{\Gamma'\left(n + \frac{m}{2}\right)}{\Gamma\left(n + \frac{m}{2}\right)^2} \int_{\mathrm{cone}(\mathcal{S})} \prod_{i=1}^{m} e^{-z_i} z_i^{k_i - \frac{1}{2}} \, dz_1 \cdots dz_m.
\end{aligned}$$

Next we consider

$$\sum_{i=1}^{m} k_i \frac{\partial}{\partial k_i} B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2}) - B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2}) \log B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2}),$$

where we replace $\log B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2})$ by $\log B(\mathbf{k} + \mathbf{1/2}) + O\left(\sum_{i=1}^{m} k_i^{-L-1}\right)$. The appearing integrals over $\mathrm{cone}(\mathcal{S})$ can be safely replaced by integrals over $[0, \infty)^m$ since the dominating part of the integral is (again) contained in $\mathcal{S}$. This means that we can switch between $\mathcal{S}$ and $\Theta$ without changing the leading asymptotic behavior. Actually the error term from the integration can be upper bounded by

$$O\left(B(\mathbf{k} + \mathbf{1/2}) \exp\left(-c^2 \sum_{k=1}^{m} k_i^{\varepsilon}\right)\right).$$

Together with the error from the approximation of $\log B_{\mathcal{S}}(\mathbf{k} + \mathbf{1/2})$ by $\log B(\mathbf{k} + \mathbf{1/2})$, that is of the form

$$O\left(B(\mathbf{k} + \mathbf{1/2}) \sum_{i=1}^{m} k_i^{-L-1}\right)$$

we end up with the proposed relation. This completes the proof of Lemma 3.

**Remark 2.** The above proof method can be extended to derive asymptotic relations for $B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2})$ if $\mathbf{k}/n \notin \mathcal{S}^-$. Here we can distinguish between two cases. The first one is the case, where $\mathbf{k}/n$ is not contained in $\mathcal{S}$ and has distance $\ge n^{-\frac{1}{2}+\varepsilon}$ to $\mathcal{S}$. In this case the box

$$Q := \prod_{i=1}^{m} \left[k_i - c\,n^{\frac{1}{2}+\varepsilon}, k_i + c\,n^{\frac{1}{2}+\varepsilon}\right]$$

is not contained in cone$(\mathcal{S})$ which implies that

$$B_{\mathcal{S}}(\mathbf{k}+\mathbf{1/2}) = O\left(B(\mathbf{k}+\mathbf{1/2}) \exp\left(-c^2 \sum_{k=1}^{m} k_i^{\varepsilon}\right)\right).$$

Finally if $\mathbf{k}/n$ has distance $\le n^{-\frac{1}{2}+\varepsilon}$ to the boundary of $\mathcal{S}$ then the *Gaussian integral*

$$\int\limits_{Q\cap \mathrm{cone}(\mathcal{S})} \prod_{i=1}^{m} e^{-z_i} z_i^{k_i-\frac{1}{2}}\, dz_1 \cdots dz_m \sim$$
$$\sim e^{-n} \prod_{i=1}^{m} k_i^{k_i-\frac{1}{2}} \int\limits_{Q\cap \mathrm{cone}(\mathcal{S})} \exp\left(-\sum_{i=1}^{m} \frac{(z_i-k_i)^2}{2k_i}\right) dz_1 \cdots dz_m$$

can be expressed in terms of the distribution function $\Phi$ of the standard normal distribution provided that boundary of $\mathcal{S}$ is (locally) a hyperplane. Since we only need upper bounds in this case the general case can be reduced to this one.

## References


[1] J. Bernardo, Reference Posterior Distributions for Bayesian Inference, *J. Roy. Stat. Soc. B.*, 41, 113–147, 1979. MR0547240

[2] B. S. Clarke and A. R. Barron, Asymptotic of Bayes Information's Theoretic Asymptotics of Bayes Methods, *IEEE Trans. Inform. Theory*, 36, 453–471, 1990. MR1053841

[3] T. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991. MR1122806

[4] L. Davisson, Universal Noiseless Coding, *IEEE Trans. Inform. Theory*, 19, 783–795, 1973. MR0465519

[5] M. Drmota and W. Szpankowski, Precise Minimax Redundancy and Regrets, *IEEE Trans. Information Theory*, 50, 2686–2707, 2004. MR2096987

[6] P. Flajolet, Singularity analysis and asymptotics of Bernoulli sums, *Theoretical Computer Science*, 215, 371–381, 1999. MR1678788

[7] R. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.

[8] P. Jacquet and W. Szpankowski, Entropy Computations via Analytic Depoissonization, *IEEE Information Theory*, 45, 1072–1081, 1999. MR1686243

[9] R. Krichevsky and V. Trofimov, The Performance of Universal Coding, *IEEE Trans. Information Theory*, 27, 199–207, 1983. MR0633417

[10] A. Orlitsky and N. P. Santhanam, Speaking of Infinity, *IEEE Transactions on Information Theory*, 50, 2215–2230, 2004 MR2097043

[11] A. Orlitsky, P. Santhanam, and J. Zhang, Universal Compression of Memoryless Sources over Unknown Alphabets, *IEEE Trans. Information Theory*, 50, 1469–1481, 2004. MR2095850

[12] J. Rissanen, Fisher Information and Stochastic Complexity, *IEEE Trans. Information Theory*, 42, 40–47, 1996. MR1375327

[13] Y. Shtarkov, Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, 23, 175–186, 1987. MR0914346

[14] G. Shamir, Universal Lossless Compression with Unknown Alphabets – The Average Case, *IEEE Transactions on Information*, 52, 4915–4944, 2006. MR2300364

[15] G. Shamir, On the MDL Principle for i.i.d. Sources With Large Alphabets, *IEEE Transactions on Information*, 52, 1939–1955, 2006. MR2234457

[16] G. Shamir, Universal Source Coding for Monotonic and Fast Decaying Monotonic Distributions, *IEEE Transactions on Information Theory*, 59, 7194–7211, 2013. MR3124637

[17] G. Shamir, T. Tjalkens and F. Willems, Low-Complexity Sequential Probability Estimation and Universal Compression for Binary Sequences with Constrained Distributions, *ISIT*, 2008.

[18] W. Szpankowski, On Asymptotics of Certain Recurrences Arising in Universal Coding, *Problems of Information Transmission*, 34, 55–61, 1998.

[19] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001. MR1816272

[20] W. Szpankowski and M. Weinberger, Minimax Pointwise Redundancy for Memoryless Models over Large Alphabets, *IEEE Trans. Information Theory*, 58, 4094–4104, 2012. MR2943077

[21] Q. Xie, A. Barron, Minimax Redundancy for the Class of Memoryless Sources, *IEEE Trans. Information Theory*, 43, 647–657, 1997.

[22] Q. Xie, A. Barron, Asymptotic Minimax Regret for Data Compression, Gambling, and Prediction, *IEEE Trans. Information Theory*, 46, 431–445, 2000. MR1748979

Michael Drmota
TU Wien
Austria
*E-mail address:* drmota@tuwien.ac.at

Gil I. Shamir
Google Inc
U.S.A.
*E-mail address:* gshamir@ieee.org

Wojciech Szpankowski
Purdue University
U.S.A.
*E-mail address:* szpan@purdue.edu

Received June 7, 2020