Annals of Mathematics **199** (2024), 1293–1321
https://doi.org/10.4007/annals.2024.199.3.6

On a conjecture of Talagrand on selector processes and a consequence on positive empirical processes

By Jinyoung Park and Huy Tuan Pham

Abstract

For appropriate Gaussian processes, as a corollary of the majorizing measure theorem, Michel Talagrand (1987) proved that the event that the supremum is significantly larger than its expectation can be covered by a set of half-spaces whose sum of measures is small. We prove a conjecture of Talagrand that is the analog of this result in the Bernoulli-p setting, and answer a question of Talagrand on the analogous result for general positive empirical processes.

1. Introduction

The study of suprema of stochastic processes is of central interest in probability theory, with influential applications in related areas. We refer the readers to [11], [17] for extensive discussions of various aspects of this subject. Through many fundamental developments, one now has a fairly good understanding of the suprema of centered Gaussian processes.[1] In particular, one can associate each Gaussian process $(Z_t)_{t\in T}$ indexed by a set T with a metric on T given by $d(t,s) := (\mathbb{E}[(Z_t - Z_s)^2])^{1/2}$, and Talagrand's celebrated generic chaining bound and majorizing measure theorem [12], [13], [17] determine the expectation of the supremum $\sup_{t\in T} Z_t$ (up to a constant factor) by a quantity depending only on the metric space (T,d). Via this fundamental result, one can obtain deep insights and characterizations of the suprema of Gaussian processes. One important example is Theorem 1.1 below, which gives a nice geometric characterization of large suprema of Gaussian processes: such an event must be contained in a union of half-spaces whose sum of measures is small.

Keywords: suprema of stochastic processes, selector processes, empirical processes, positive stochastic processes, Bernoulli processes

AMS Classification: Primary: 60G70, 60E15; Secondary: 60C05, 60G15, 60G17.

[1] Following [17], we always assume Gaussian processes are centered, i.e., $\mathbb{E}Z_t = 0$ for all $t \in T$.

THEOREM 1.1 (Talagrand, Theorem 2.12.2 in [17]). *There exists $L > 0$ such that the following holds. Let g be an M-dimensional standard Gaussian vector. For $\mathcal{T} \subseteq \mathbb{R}^M$, consider the process $Z_t = \langle t, g\rangle$ for $t \in \mathcal{T}$. Then one can find a sequence of half-spaces H_k of $\mathbb{R}^M$ with*

$$\left\{\sup_{t\in\mathcal{T}} Z_t \geq L\mathbb{E}\sup_{t\in\mathcal{T}} Z_t\right\} \subset \bigcup_{k\geq 1} H_k$$

and

$$\sum_{k\geq 1} \mathbb{P}(H_k) \leq \frac{1}{2}.$$

Our main contribution in this paper is the proof of a conjecture of Talagrand on selector processes (Theorem 1.2; originally [15, Prob. 4.1], [16, Conj. 5.7] and [17, Research Prob. 13.2.3]) and a result on positive empirical processes (Theorem 1.3; a question of Talagrand [18] and a problem posed in [15]), which are analogous to Theorem 1.1. We first quickly state our main results and then provide more context, definitions, and motivations for Talagrand's questions.

Given a finite set X, write 2^X for the power set of X. For $p \in [0,1]$, let μ_p be the product measure on 2^X given by $\mu_p(A) = p^{|A|}(1-p)^{|X\setminus A|}$. We use X_p for the random variable whose distribution is μ_p. For $S \subseteq X$, define the *upset generated by* S to be $\langle S\rangle := \{T : T \supseteq S\}$. Following [16], we say $\mathcal{F} \subseteq 2^X$ is *p-small* if there is $\mathcal{G} \subseteq 2^X$ such that

$$\mathcal{F} \subseteq \langle \mathcal{G}\rangle := \bigcup_{S\in\mathcal{G}} \langle S\rangle \tag{1}$$

and

$$\sum_{S\in\mathcal{G}} p^{|S|} \leq 1/2. \tag{2}$$

We say $\mathcal{G}$ is a *cover* of $\mathcal{F}$ if (1) holds.

Our first main result is the Bernoulli-p analog of Theorem 1.1.

THEOREM 1.2. *There exists $L > 0$ such that the following holds. Consider any $0 < p < 1$, any finite set X and any collection Λ of sequences $\boldsymbol{\lambda} = (\lambda_i)_{i\in X}$ with $\lambda_i \geq 0$. Then the family*

$$\left\{S \subseteq X : \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in S} \lambda_i \geq L\mathbb{E}\sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in X_p} \lambda_i\right\}$$

is p-small.

In [16], Talagrand explains the meaning of the above theorem this way: Conjecture 5.7 (now Theorem 1.2) shows that "if you are given a selector process, and would like to prove that, within a multiplicative factor, the quantity $\mathbb{E}\sup_{\boldsymbol{\lambda}\in\Lambda}\sum_{i\in X_p}\lambda_i \leq M$ for a constant M, there is in the end no other way

than to find the witnesses that the set $\{S \subseteq X : \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in S} \lambda_i \geq LM\}$ is small." In the same place, Talagrand suggests that this result "provides fundamental information."

Our second main result is the analog of Theorem 1.1 for positive empirical processes (see (4) and the discussion that follows).

THEOREM 1.3. *There exists $L > 0$ such that the following holds. For any $N > 0$ and i.i.d. random variables $Y_1, \dots, Y_N$ distributed according to a Borel probability measure ν on a Polish space $\mathbb{T}$, and any finite collection $\mathcal{F}$ of Borel functions $f : \mathbb{T} \to \mathbb{R}_{\geq 0}$ with $\mathcal{F} \subseteq L^\infty(\mathbb{T})$, consider the positive empirical process $Z_f = \frac{1}{N}\sum_{i=1}^N f(Y_i)$. Assume that $0 < \mathbb{E}[\sup_{f\in\mathcal{F}} Z_f] < \infty$. Then one can find a collection $\mathcal{C}$ of pairs (g,t), where g is a nonnegative function on $\mathbb{T}$ and $t > 0$, so that with $E_{g,t} := \{Z_g \geq t\}$, we have*

$$\left\{\sup_{f\in\mathcal{F}} Z_f \geq L\mathbb{E}\sup_{f\in\mathcal{F}} Z_f\right\} \subset \bigcup_{(g,t)\in\mathcal{C}} E_{g,t}$$

and

$$\sum_{(g,t)\in\mathcal{C}} \mathbb{P}(E_{g,t}) \leq \frac{1}{2}.$$

We remark that the conclusion of Theorem 1.3 readily extends to cases where $\mathcal{F}$ is not necessarily finite, for example when $\mathcal{F}$ is a totally bounded infinite subset of $L^\infty(\mathbb{T})$.

In [15], Talagrand proved versions of our main results for the special case where the class of functions consists of indicator of sets, and he posed the question of extending the results to general classes of functions as important open problems. This is fully addressed by our results in both the setting of selector processes and empirical processes. (Talagrand's result on empirical processes in [15] is stated slightly differently; see the remark at the end of Section 3 for how to obtain from our proof a stronger version of Talagrand's result in the general setting.)

More context and definitions. A Gaussian process can be described in the form $Z_t = \sum_{i=1}^\infty \xi_i t_i$, where $\xi_1, \xi_2, \dots$ are i.i.d. standard Gaussian random variables, and $t = \{t_i\}_{i\geq 1} \in T \subseteq \ell^2$ is a square-summable sequence. Alternatively, one can view Gaussian processes as random series (with i.i.d. Gaussian coefficients) of functions $f_i : T \to \mathbb{R}$ (where $f_i(t) = t_i$), an object of natural interest. Generalizing the coefficients beyond the Gaussian case immediately leads to substantially more challenging questions. In particular, in the case where the coefficients ξ_i are independent Rademacher random variables, a longstanding conjecture of Talagrand ("the Bernoulli Conjecture") suggests a precise way to control the supremum in expectation in the spirit of chaining. The conjecture was only resolved recently in a breakthrough by Bednorz and Latała [2].

The problem is even harder when ξ_i are centered Bernoulli-p random variables (to be contrasted with our focus later on the ordinary nonnegative Bernoulli-p random variables). In the "generalized Bernoulli Conjecture," which now is [17, Th. 11.12.1] and whose proof is inspired by work of Bednorz and Martynek [3], Talagrand showed that suprema of centered Bernoulli processes can be described in terms of quantities depending only on the metric structure of appropriate classes of functions, together with quantities depending on suprema of processes of the form $Z_t = \sum_{i=1}^{M} \xi_i t_i$ for $t : \mathbb{N} \to \mathbb{R}_{\geq 0}$, and ξ_i i.i.d. (ordinary) Bernoulli-p random variables. This way, the study of random sums of functions leads us naturally to *positive selector processes*, which we now define formally. Recall that X_p is the random variable whose distribution is μ_p. Given a collection Λ of sequences $\boldsymbol{\lambda} = (\lambda_i)_{i \in X}$, we define the *selector process* associated to Λ as the process indexed by Λ whose value at λ is given by

$$Z_\lambda := \sum_{i \in X_p} \lambda_i. \tag{3}$$

A selector process is *positive* if $\lambda_i \geq 0$ for all $\lambda \in \Lambda$ and $i \in X$.

Next, we motivate the study of *positive empirical processes*. Given i.i.d. random variables $Y_1, \ldots, Y_N$ distributed according to a Borel probability measure on a Polish space $\mathbb{T}$, and a class $\mathcal{F}$ of functions $f : \mathbb{T} \to \mathbb{R}$, an *empirical process* is a process indexed by $\mathcal{F}$ of the form

$$Z_f := \frac{1}{N} \sum_{i \leq N} f(Y_i). \tag{4}$$

We say that an empirical process is *positive* if the functions f in the class $\mathcal{F}$ are nonnegative. Empirical processes and their suprema form an important subject in probability theory and have a wide range of applications in computer science, statistics and machine learning [14], [17], [10], [6]. There, one is often interested in the suprema of the empirical process, $\sup_{f \in \mathcal{F}} Z_f$. While one is often interested in *centered* empirical processes, i.e., those indexed by zero mean functions, by a deep result of Talagrand [17, Th. 6.8.3] ("the fundamental theorem of empirical processes"), the supremum of general centered empirical processes can always be controlled, in a precise sense, by quantities depending only on the metric structure of an appropriate class of functions, and the supremum of an appropriate **positive** empirical process. Thus, the study of positive empirical processes is key to understanding centered empirical processes.

As we have discussed in the previous paragraphs, the study of suprema of centered stochastic processes naturally leads us to the study of their positive counterpart. While substantial advances in chaining allow us to understand precisely the reduction from centered processes to positive processes, suprema of positive processes are much less understood. As Talagrand explains in [17],

in many cases, we know that "chaining explains all the boundedness due to cancellation," but "what could we ask about boundedness of processes where no cancellation occurs?" Thus, while we have good understanding of the effect of cancellation on the suprema of stochastic processes, positive processes (where there is no cancellation to exploit) are much less understood and are essentially the last missing piece in the picture. In this context, our main theorems address the task of filling in this missing piece: Theorem 1.2 resolves a conjecture of Talagrand on large suprema of positive selector processes, one of the questions in [17, Ch. 13] on "Unfulfilled dreams;" and Theorem 1.3 answers a question of Talagrand on general positive empirical processes. We point out that our proof of Theorem 1.3 builds on a close connection between positive empirical processes and a version of selector process with multiplicities that has been informally observed in [16].

Roughly speaking, Theorem 1.1 shows that, for a Gaussian process, one can find simple geometric "witnesses" (half-spaces) that cover the event that the supremum of the process is large, and the sum of measures of these witnesses is small. In particular, even though it is a simple application of Markov's inequality to show that the probability of the event $\{\sup_{t\in T} Z_t \geq L\mathbb{E}\sup_{t\in T} Z_t\}$ is small, the simple geometric witnesses provide a much more refined structure on this event. Similarly, in Theorem 1.2, which is in the setting of positive selector processes, the role of the half-spaces is replaced by the upsets $\langle S\rangle$, and being p-small is an analog of admitting a cover by half-spaces with small total measure. The meaning of Theorem 1.3 can also be sketched in a similar way — it describes explicit "simple" witnesses (half-spaces of the empirical measure) that cover the tail event of $\sup_f Z_f$. The covering perspective, as observed by Talagrand [16], also provides striking connections between these sets of questions and the study of thresholds, specifically the Kahn-Kalai conjecture. Building on the insights in the present paper, particularly the notion of minimum fragment, in [9] we obtain the resolution of the Kahn-Kalai conjecture.

Finally, we mention that the "abstract setting" of [16, Conj. 5.7], which is [15, Prob. 4.2], [16, Conj. 7.1] and [17, Research Prob. 13.3.2], remains open.

Reformulations. We will prove Theorem 1.2 via the slightly more convenient equivalent reformulation below. As observed in [16], the following theorem is equivalent to Theorem 1.2:

THEOREM 1.4. *There exists $L' > 0$ such that the following holds. Consider any $0 < p < 1$, any finite set X, and any family $\mathcal{F} \subseteq 2^X$. Assume that for each $S \in \mathcal{F}$, we are given a sequence $\boldsymbol{\lambda}^S = (\lambda^S(i))_{i\in X}$ with $\lambda^S(i) \geq 0$ and*

$$\sum_{i\in S} \lambda^S(i) \geq 1. \tag{5}$$

Then if $\mathcal{F}$ is ***not*** *p-small, we have*

$$\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in X_p}\lambda^S(i)\geq 1/L'. \tag{6}$$

Note that we only have $\sup_{S\in\mathcal{F}}\mathbb{E}\sum_{i\in X_p}\lambda^S(i)\geq p$, so (6) suggests a nontrivial phenomenon. The theorem below implies Theorem 1.4. As usual, for $m\geq 0$, we denote by $\binom{X}{m}$ the collection of subsets of X of size m.

THEOREM 1.5. *There exists $K>0$ such that the following holds. Consider any $0<p<1$, any finite set X, and any family $\mathcal{F}\subseteq 2^X$. Assume that for each $S\in\mathcal{F}$, we are given a sequence $\boldsymbol{\lambda}^S=(\lambda^S(i))_{i\in S}$ with $\lambda^S(i)\geq 0$ and*

$$\sum_{i\in S}\lambda^S(i)\geq 1. \tag{7}$$

Suppose $\mathcal{F}$ is ***not*** *p-small. Then, for $\mathbf{W}$ chosen uniformly at random from $\binom{X}{\lfloor K|X|p\rfloor}$, we have*

$$\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in S\cap\mathbf{W}}\lambda^S(i)\geq 10^{-11}. \tag{8}$$

The derivation of Theorem 1.2 from Theorem 1.4 can be found in [16]. We include the simple proof of Theorem 1.4 from Theorem 1.5, and the derivation of Theorem 1.2 from Theorem 1.4 in Section 2.1 for completeness.

The general weighted setting of Theorems 1.4 and 3.1 (the main results leading toward Theorems 1.2 and 1.3) poses significant challenges, as one can anticipate from the statement: while the assumption on $\mathcal{F}$ is inherently combinatorial, the conclusion applies to general weight functions on the sets in $\mathcal{F}$. In the simplest unweighted case (where we restrict all $S\in\mathcal{F}$ to have the same size s, and $\lambda^S(i)=\frac{1}{s}\mathbb{I}(i\in S)$), our proof shares some inspiration with the argument in [1], [5], [7], and the recent improvement in [9], although in this special case there are also alternative approaches. In particular, as we mentioned earlier, in this case (and more generally in the special case where the class of functions consists of $\{0,1\}$-valued functions), Talagrand [15] gave a nice proof using a second moment argument. The treatment of the general weighted setting (where we place no restriction on $S\in\mathcal{F}$ and λ^S) requires a different set of ideas.

Firstly we work with elements at different dyadic scales, and one important idea is to choose to work with only certain "informative" scales. Secondly, a major difficulty of the general setting is that the weight of an element can vary with the set $S\in\mathcal{F}$. In order to address this problem, we need a much more delicate and involved notion of *fragments* (Definition 2.3), as well as the associated notion of *minimum fragment* and its analysis. The use of minimum fragment (associated to a simple notion of fragment), inspired by the present

work, is crucial in our recent resolution of the Kahn-Kalai conjecture [9]. We note that, whereas it is possible to establish the fractional version of the Kahn-Kalai conjecture without the idea of minimum fragment, it remains critical in our proof even if one only wants to establish the weaker fractional version of Theorem 1.2 ([16, Conj. 6.8]). In particular, the main difficulty of weight functions changing with $S \in \mathcal{F}$ remains in the fractional version of Theorem 1.2, and the full strength and flexibility of our generalized notion of fragment and minimum fragment is required to handle this problem, which is how we arrive at these notions.

As discussed later in Section 3, proving Theorem 1.3, especially in the general case where we do not impose that the underlying distribution ν is continuous, requires more ideas. In particular, we will need a version of Theorem 1.2 with multiplicities, Theorem 3.1, which itself admits an equivalent reformulation, Theorem 3.3. We emphasize that the core principles behind the proof of these results are the same as those behind the proof of Theorem 1.2.

Organization. The core part of the paper, the proof of Theorem 1.5, is contained in Section 2. The proof of Theorem 1.3 is given in Section 3, building on a version of Theorem 1.5 with multiplicities, Theorem 3.3. In the course, we have made no attempt to optimize absolute constants. Logarithms are in base e unless otherwise specified.

2. Proof of Theorem 1.2

2.1. *Reductions between Theorems 1.2, 1.4 and 1.5.* In this section, we give short reductions from Theorem 1.2 to Theorems 1.4 and 1.5.

The next derivation can be found in [16] for the "weakly p-small" version of Theorem 1.2.

Proof of Theorem 1.2 from Theorem 1.4. Consider a collection Λ of sequences $\boldsymbol{\lambda} = (\lambda_i)_{i\in X}$ with $\lambda_i \geq 0$, and with L to be determined, consider the collection

$$\mathcal{G} := \left\{ S \subseteq X : \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in S} \lambda_i \geq L\mathbb{E} \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in X_p} \lambda_i \right\}.$$

By definition of $\mathcal{G}$, for each $S \in \mathcal{G}$, there is $(\tau^S(i))_{i\in X} \in \Lambda$ for which

$$\sum_{i\in S} \tau^S(i) \geq L\mathbb{E} \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in X_p} \lambda_i. \tag{9}$$

Now for each $S \in \mathcal{G}$, define $\boldsymbol{\lambda}^S$:

$$\lambda^S(i) = \begin{cases} \tau^S(i) & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for any $Y \subseteq X$, we have

$$\sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in Y} \lambda_i \geq \sum_{i\in Y} \tau^S(i) \geq \sum_{i\in Y} \lambda^S(i)$$

and, hence, in particular

$$\mathbb{E} \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in X_p} \lambda_i \geq \mathbb{E} \sup_{S\in\mathcal{G}} \sum_{i\in X_p} \lambda^S(i). \tag{10}$$

Assume for contradiction that $\mathcal{G}$ is not p-small. Then by (9) and Theorem 1.4, we have

$$\mathbb{E} \sup_{S\in\mathcal{G}} \sum_{i\in X_p} \lambda^S(i) \geq (L/L')\mathbb{E} \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in X_p} \lambda_i.$$

Combining with (10), we have

$$\mathbb{E} \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in X_p} \lambda_i \geq (L/L')\mathbb{E} \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in X_p} \lambda_i.$$

This is a contradiction for $L > L'$. □

Proof of Theorem 1.4 from Theorem 1.5. Note that if $\mathcal{F}$ is not p-small, then $|X|p \geq 1/2$ (since $\{\{x\} : x \in X\}$ covers $\mathcal{F}$). With $N := \max\{\lfloor |X|p \rfloor, 1\}$, let $\mathbf{W}'$ be chosen uniformly at random from $\binom{X}{N}$. Then with $\zeta := N/\lfloor K|X|p \rfloor$, we can think of choosing $\mathbf{W}'$ as choosing $\mathbf{W}$ first and then picking a ζ-fraction of it. We consider two cases.

If $|X|p \geq 1$, then using the facts that $\zeta \geq 1/(2K)$ and that $\mathbb{P}(|X_p| \geq \lfloor |X|p \rfloor) \geq 1/2$ (see [8]),

$$\mathbb{E} \sup_{S\in\mathcal{F}} \sum_{i\in X_p} \lambda^S(i) \geq \frac{1}{2}\mathbb{E} \sup_{S\in\mathcal{F}} \sum_{i\in S\cap\mathbf{W}'} \lambda^S(i) \geq \frac{1}{4K}\mathbb{E} \sup_{S\in\mathcal{F}} \sum_{i\in S\cap\mathbf{W}} \lambda^S(i) \geq \frac{1}{4\cdot 10^{11}K}.$$

If $1/2 \leq |X|p < 1$, then using the facts that $\zeta \geq 1/K$ and that $\mathbb{P}(|X_p| \geq 1) \geq 1 - (1-p)^{|X|} \geq 1 - e^{-1/2} > 1/4$,

$$\mathbb{E} \sup_{S\in\mathcal{F}} \sum_{i\in X_p} \lambda^S(i) \geq \frac{1}{4}\mathbb{E} \sup_{S\in\mathcal{F}} \sum_{i\in S\cap\mathbf{W}'} \lambda^S(i) \geq \frac{1}{4K}\mathbb{E} \sup_{S\in\mathcal{F}} \sum_{i\in S\cap\mathbf{W}} \lambda^S(i) \geq \frac{1}{4\cdot 10^{11}K}.$$

Now, Theorem 1.4 follows by letting $L' = 4\cdot 10^{11}K$. □

2.2. *Proof of Theorem 1.5.* Given the previous reductions, the key step in the proof of Theorem 1.2 is the proof of Theorem 1.5, which is the focus of this section and the key result in the paper containing the main insights.

In this section, $p, X, \mathcal{F}$, and $\boldsymbol{\lambda}^S$ are as in Theorem 1.5, and K, a universal constant, is chosen sufficiently large to support our proof. We use n for $|X|$, and J and w are quantities that satisfy

$$\lfloor Knp \rfloor = Jnp = w;$$

S, S', S'' and $\hat{S}$ represent members of $\mathcal{F}$, and $W \in \binom{X}{w}$. Finally, for $m \in \mathbb{Z}^+$, $[m]$ denotes $\{0, 1, \ldots, m\}$.

We say W is *good* if

$$\max_{S \in \mathcal{F}} \sum_{i \in S \cap W} \lambda^S(i) \geq 10^{-10},$$

and *bad* otherwise. Note that (8) follows if a (1/10)-fraction of $\binom{X}{w}$ is good. Therefore, to prove Theorem 1.5, it suffices to show that

(11) if a (9/10)-fraction of $\binom{X}{w}$ is bad,

then

(12) $\mathcal{F}$ is p-small.

Before getting to the details of the proof, we first give an informal description of our overall strategy. Roughly speaking, a family $\mathcal{F}$ is p-small if $\mathcal{F}$ admits a "cheap" cover, where being cheap refers to the condition in (2). In order to derive (12), we will first construct a cover of $\mathcal{F}$, where the cover, $\mathcal{U} = \mathcal{U}(W)$, depends on the choice of W. We will show that the overall cost of the covers among *bad* W's is small, from which, combined with (11), the existence of a cheap cover is guaranteed.

We first need some pre-processing steps on the weights λ^S.

Observation 2.1. Let $\tau = \lfloor \log_{100} n \rfloor + 2$. It is sufficient to prove Theorem 1.5 assuming that, for all $S \in \mathcal{F}$ and $i \in X$,

(13) $$\lambda^S(i) = 100^{-j} \text{ for some } j \in \{0, 1, 2, \ldots, \tau\},$$

with (7) weakened to

(14) $$\sum_{i \in S} \lambda^S(i) \geq 100^{-2}.$$

Justification. We may first assume that

(15) $$\lambda^S(i) \leq 1 \text{ for all } S \text{ and } i,$$

by capping larger weights at 1. Under this assumption, for each S, let

$$S_j = \{i \in S : \lambda^S(i) \in [100^{-j}, 100^{-j+1})\} \quad (j = 0, 1, 2, \ldots).$$

By replacing the weights of elements in S_j with 100^{-j}, we can assume that $\lambda^S(i) = 100^{-j}$ for all $i \in S_j$ with (7) weakened to $\sum_{i \in S} \lambda^S(i) \geq 1/100$. Finally, note that

$$\sum_{j > \tau} \lambda^S(S_j) \leq |S| 10^{-(\tau+1)} \leq n 100^{-\log_{100} n - 2} = 100^{-2}.$$

Thus, by removing elements in S_j for $f > \tau$, we can assume that $S_j = \emptyset$ for all $j > \tau$ with (7) weakened to $\sum_{i \in S} \lambda^S(i) \geq 100^{-2}$. □

From now on, we assume (13). For $S \in \mathcal{F}$, we write s_j for $|S_j|$ and define the *profile* of S to be $\mathbf{s} = \mathbf{s}(S) = (s_0, \dots, s_\tau)$. In Observation 2.2 below, we use the trivial fact that

(16) a family $\mathcal{F} = \{S_1, S_2, \dots\}$ is p-small if $\mathcal{F}' = \{S_1', S_2', \dots\}$ is p-small and $S_i \supseteq S_i' \quad \forall i$.

Observation 2.2. It is sufficient to prove Theorem 1.5 assuming that for all $S \in \mathcal{F}$, in addition to (13), for each $j \in [\tau]$, either $s_j = 0$ or

$$100^j/(100^4(j+1)^2) \le s_j < 2 \cdot 100^j; \text{ and} \tag{17}$$

$$s_j \text{ is a power of } 100 \tag{18}$$

with (14) weakened to

$$\sum_{i \in S} \lambda^S(i) \ge 100^{-4}. \tag{19}$$

Justification. First, we can greedily remove elements from each S until $\lambda^S(S) < 2$. (Note that we have (15) under (13).) This gives the upper bound in (17) since $2 > \lambda^S(S) = \sum s_j 100^{-j}$.

Next, for each S and j, remove all elements in S_j for which

$$s_j < 100^j/(100^4(j+1)^2).$$

This process reduces the weight of S by at most

$$\sum_{j \ge 0} 100^j/(100^4(j+1)^2) \cdot 100^{-j} < 100^{-3}.$$

Finally, for each j with $|S_j| > 0$, remove elements from S_j to round $|S_j|$ to the largest power of 100 that is at most $|S_j|$. The weight of S_j remains at least a $1/100$ fraction of its weight before removal of elements, yielding (19). □

We say $\mathbf{s} = (s_0, s_1, \dots, s_\tau)$ is a *legal profile* if each nonzero $s_j \in \mathbf{s}$ satisfies

$$\max\{1, 100^j/(100^4(j+1)^2)\} \le s_j < 2 \cdot 100^j \tag{20}$$

and

$$s_j \text{ is a power of } 100. \tag{21}$$

We will also need the notion of a *partial legal profile* $\mathbf{s}_b := (s_j : j \le b)$, in which each nonzero s_j satisfies (20) and (21). Given $\mathbf{s}$ or $\mathbf{s}_b$, we use $\mathcal{F}_\mathbf{s}$ for the collection of S's whose profile is $\mathbf{s}$, and $\mathcal{F}_{\mathbf{s}_b}$ for the collection of S's whose partial profile is $\mathbf{s}_b$. Note that if $\mathbf{s}$ restricted on $\{j \le b\}$ is $\mathbf{s}_b$, then $\mathcal{F}_\mathbf{s} \subseteq \mathcal{F}_{\mathbf{s}_b}$.

The following definition is a key notion in our proof.

Definition 2.3. Given $S \in \mathcal{F}_{\mathbf{s}}$ and $W \in \binom{X}{w}$, we say $U \subseteq S \setminus W$ is an (S,W)-fragment (with index b) if the following holds: there are $b \in \{-1,0,\dots,\tau\}$ and $S' \in \mathcal{F}_{\mathbf{s}_b}$ such that

$$S'_j \subseteq W \cup U \quad \forall j \le b, \tag{22}$$

$$\sum_{j \ge b+1} \lambda^{S'}(S'_j \cap (W \cup U)) \ge .01 \left[\sum_{j \ge b+1} \lambda^{S'}(S'_j) - 100^{-b} \left(\sum_{j \le b} s_j - |U| \right) \right]. \tag{23}$$

For example, $S \setminus W$ is an (S,W)-fragment with index τ.

Remark 2.4. A similar notion of fragment is used in [1], [5], [7] and in the recent resolution of the Kahn–Kalai conjecture [9], in which the definition of fragment is much simpler. The more delicate definition of fragment, as in Definition 2.3, is crucial for the current setting of Theorem 1.5.

Given S and W, we denote by $T = T(S,W)$ the *minimum* (S,W)-fragment, where here minimum refers to the index b first (breaking ties arbitrarily), and then $|T|$ (again, breaking ties arbitrarily). We use $t = t(S,W)$ for $|T(S,W)|$. For any pair (S,W), its minimum fragment T and the corresponding index $b_T = b(S,W)$ are uniquely determined.

The following definition will be crucial in the construction of $\mathcal{U}(W)$, a cover of $\mathcal{F}$.

Definition 2.5. Given $Z \subseteq X$, $b \in [\tau]$, $t \ge 0$, and a partial profile $\mathbf{s}_b$, we say $S' \in \mathcal{F}_{\mathbf{s}_b}$ is $(Z, b, \mathbf{s}_b, t)$-feasible if

$$S'_j \subseteq Z \quad \forall j \le b, \tag{24}$$

$$\sum_{j \ge b+1} \lambda^{S'}(S'_j \cap Z) \ge .01 \left[\sum_{j \ge b+1} \lambda^{S'}(S'_j) - 100^{-b} \left(\sum_{j \le b} s_j - t \right) \right]. \tag{25}$$

The definitions of minimum fragment and feasibility are closely related, as shown in Propositions 2.6 and 2.7.

PROPOSITION 2.6. *Let $S \in \mathcal{F}_{\mathbf{s}}$. Let T be the minimum (S,W)-fragment with index b_T, and $t = |T|$. If S' is $(W \cup T, b_T, \mathbf{s}_{b_T}, t)$-feasible, then*

$$\Big(\bigcup_{j \le b_T} S'_j \Big) \setminus W = T. \tag{26}$$

Proof. Note that, by the definition of feasibility, S' satisfies (22) and (23) with $U = T$ and $b = b_T$; that $(\bigcup_{j \le b_T} S'_j) \setminus W \subseteq T$ follows from this definition. To show $T \subseteq \bigcup_{j \le b_T} S'_j$, first observe that by minimality of $|T|$, $T \subseteq S'$. Indeed, if $x \in T \setminus S'$, then replacing T by $T \setminus \{x\}$, one can check that S' still satisfies (22)

and (23). Suppose there is $x \in T \cap S'_{j_0}$ for some $j_0 \geq b_T + 1$. Note that

$$\lambda^{S'}(x) \leq 100^{-(b_T+1)}. \tag{27}$$

Then $T' := T \setminus \{x\}$ is an (S, W)-fragment, since S' trivially satisfies (22), and

$$\begin{aligned}
&\sum_{j \geq b_T+1} \lambda^{S'}(S'_j \cap (W \cup T')) \\
&\quad \geq \sum_{j \geq b_T+1} \lambda^{S'}(S'_j \cap (W \cup T)) - \lambda^{S'}(x) \\
&\quad \overset{(25),(27)}{\geq} .01 \left[\sum_{j \geq b_T+1} \lambda^{S'}(S'_j) - 100^{-b_T} \left(\sum_{j \leq b_T} s_j - |T| \right) - 100^{-b_T} \right] \\
&\quad = .01 \left[\sum_{j \geq b_T+1} \lambda^{S'}(S'_j) - 100^{-b_T} \left(\sum_{j \leq b_T} s_j - |T'| \right) \right].
\end{aligned}$$

This contradicts the minimality of T. □

PROPOSITION 2.7. *Let $S \in \mathcal{F}_{\mathbf{s}}$. Let T be the minimum (S, W)-fragment with index b_T, and $t = |T|$. If S' is $(W \cup T, b_T, \mathbf{s}_{b_T}, t)$-feasible, then*

$$t \geq .9 \sum_{j \leq b_T} s_j. \tag{28}$$

Proof. Let $T_j = S'_j \cap T$ and $t_j = |T_j|$ (so $t = \sum_{j \leq b} t_j$ by Proposition 2.6). We will prove the proposition by contradiction, showing that the failure of (28) violates the minimality of T.

First observe that, if we assume $t < .9 \sum_{j \leq b_T} s_j$, then there exists $b' \leq b_T$ such that

$$\sum_{b' \leq j \leq b_T} t_j 100^{-j} < .9 \sum_{b' \leq j \leq b_T} s_j 100^{-j}. \tag{29}$$

Indeed, if $(u_{b'} :=) \sum_{b' \leq j \leq b_T} 100^{-j}(t_j - .9 s_j) \geq 0$ for all $b' \leq b_T$, then

$$\begin{aligned}
\sum_{j \leq b_T} (t_j - .9 s_j) &= \sum_{i \leq b_T - 1} 100^i (u_i - u_{i+1}) + 100^{b_T} u_{b_T} \\
&= \sum_{i=1}^{b_T} (100^i - 100^{i-1}) u_i + u_0 \geq 0.
\end{aligned}$$

Note that (29) gives, by Proposition 2.6,

$$\sum_{b' \leq j \leq b_T} \lambda^{S'}(S'_j \cap W) \geq .1 \sum_{b' \leq j \leq b_T} \lambda^{S'}(S'_j). \tag{30}$$

Now we claim that

$$T' := \bigcup_{j \le b'-1} T_j \text{ is an } (S, W)\text{-fragment with index } b' - 1, \tag{31}$$

which contradicts the minimality of T.

Proof of (31). It is clear that S' satisfies property (22). For (23),

$$\begin{aligned}
&\sum_{j \ge b'} \lambda^{S'}(S'_j \cap (W \cup T')) \\
&\overset{(26)}{=} \sum_{j \ge b_T+1} \lambda^{S'}(S'_j \cap (W \cup T)) + \sum_{b' \le j \le b_T} \lambda^{S'}(S'_j \cap W) \\
&\overset{(23),(30)}{\ge} .01 \left[\sum_{j \ge b_T+1} \lambda^{S'}(S'_j) - 100^{-b_T} \left(\sum_{j \le b_T} s_j - t \right) \right] + .1 \sum_{b' \le j \le b_T} \lambda^{S'}(S'_j) \\
&= .01 \sum_{j \ge b'} \lambda^{S'}(S'_j) - 100^{-(b_T+1)} \left(\sum_{j \le b_T} s_j - t \right) + .09 \sum_{b' \le j \le b_T} s_j 100^{-j} \\
&\overset{(\dagger)}{\ge} .01 \left[\sum_{j \ge b'} \lambda^{S'}(S'_j) - 100^{-(b'-1)} \left(\sum_{j \le b'-1} s_j - t' \right) \right],
\end{aligned}$$

where (†) follows from the inequalities

$$.09 \sum_{b' \le j \le b_T} s_j 100^{-j} \ge 100^{-(b_T+1)} \sum_{b' \le j \le b_T} s_j$$

and

$$\begin{aligned}
(100^{-b'} - 100^{-(b_T+1)}) \sum_{j \le b'-1} s_j &\overset{(26)}{\ge} (100^{-b'} - 100^{-(b_T+1)}) t' \\
&\ge 100^{-b'} t' - 100^{-(b_T+1)} t. \qquad \square
\end{aligned}$$

This completes the proof of the proposition. $\square$

PROPOSITION 2.8. *If W is bad, then for any S, we have $b_T \ge 0$.*

Proof. If $b_T = -1$, then by Proposition 2.6 we have $T = \emptyset$. Then by (23), there is S' for which

$$\lambda^{S'}(S' \cap W) \ge .01 \lambda^{S'}(S') \ge 10^{-10},$$

which contradicts the fact that W is bad. $\square$

Now we construct $\mathcal{U}(W)$, the promised cover of $\mathcal{F}$. The following construction is valid for all $W \in \binom{X}{w}$, but in the end we will be interested only in bad W's.

For $W \in \binom{X}{w}$, define $\mathcal{U} = \mathcal{U}(W)$ to be

$$\mathcal{U}(W) := \{T(S,W) : S \in \mathcal{F}\}.$$

Note that $\mathcal{U}(W)$ covers $\mathcal{F}$ since $T(S,W) \subseteq S$ for each $S \in \mathcal{F}$.

The following lemma shows that if we only consider bad W's, then the overall cost (in (2)) for the covers is cheap. Recall that $n = |X|$ and $w = Jnp$.

LEMMA 2.9. *We have*

$$\sum_{W \text{ bad}} \sum_{U \in \mathcal{U}(W)} p^{|U|} \le J^{-c} \binom{n}{w}$$

for some constant $c > 0$.

Note that Lemma 2.9, combined with (11), implies that there is a bad W for which

$$\sum_{U \in \mathcal{U}(W)} p^{|U|} \le (10/9) J^{-c} < 1/2.$$

This gives (12) and thus concludes the proof of Theorem 1.5.

Proof of Lemma 2.9. Let a bad W be given. For each b, a legal partial profile $\mathbf{s}_b$ and $t \ge 0$, let

$$\mathcal{G}_W(\mathbf{s}_b, t) = \{S : S \in \mathcal{F}_{\mathbf{s}_b}, t(S,W) = t, b(S,W) = b\}$$

and

$$\mathcal{U}_W(\mathbf{s}_b, t) = \{T(S,W) : S \in \mathcal{G}_W(\mathbf{s}_b, t)\}.$$

Then with $n_b := \sum_{j \le b} s_j$ for given $\mathbf{s}_b$, we have

$$\sum_{W \text{ bad}} \sum_{U \in \mathcal{U}(W)} p^{|U|} \le \sum_{b \ge 0} \sum_{\mathbf{s}_b \text{ legal}} \sum_{.9 n_b \le t \le n_b} \sum_{W \text{ bad}} \sum_{U \in \mathcal{U}_W(\mathbf{s}_b, t)} p^{|U|}, \tag{32}$$

noting that the range $b \ge 0$ follows from Proposition 2.8, and the range of t is from Propositions 2.6 and 2.7.

Given $b \ge 0$, $\mathbf{s}_b$ and t, we bound $\sum_{W \text{ bad}} \sum_{U \in \mathcal{U}_W(\mathbf{s}_b, t)} p^{|U|}$ as follows:

Step 1. Pick $Z := W \cup T$. Since $|Z| = w + t$ (note W and T are always disjoint), the number of possibilities for Z is at most (recalling $w = Jpn$)

$$\binom{n}{w+t} = \binom{n}{w} \cdot \prod_{j=0}^{t-1} \frac{n-w-j}{w+j+1} \le \binom{n}{w} (Jp)^{-t}.$$

Step 2. Given Z, by the definition of fragment, there must exist a choice of $(Z, b, \mathbf{s}_b, t)$-feasible $\hat{S}$. Make a choice of $\hat{S}$ (arbitrarily) such that it only depends on Z. Then Proposition 2.6 enables us to specify T as a subset of $\bigcup_{j \le b} \hat{S}_j$, whose number of possibilities is at most 2^{n_b}.

Therefore, the right-hand side of (32) is at most

$$\sum_{b\ge 0}\sum_{\mathbf{s}_b \text{ legal}}\sum_{.9n_b\le t\le n_b}\sum_{W \text{ bad}}\sum_{U\in\mathcal{U}_W(\mathbf{s}_b,t)} p^t \le \sum_{b\ge 0}\sum_{\mathbf{s}_b \text{ legal}}\sum_{.9n_b\le t\le n_b}\binom{n}{w}J^{-t}2^{n_b}$$
$$\le \binom{n}{w}\sum_{b\ge 0}\sum_{\mathbf{s}_b \text{ legal}}(J/4)^{-.9n_b}.$$

Notice that, by (20) and (21), the number of possibilities for legal $\mathbf{s}_b$ is at most

$$\prod_{j\le b}\log_{100}(2\cdot 100^4(j+1)^2).$$

On the other hand, using the lower bound on s_j in (20),

$$n_b \ge \sum_{j\le b}\max\{1, 100^j/(100^4(j+1)^2)\}.$$

Therefore, with

$$a_j := \log_{100}[2\cdot 100^4(j+1)^2]\cdot(J/4)^{-.9\max\{1,100^j/(100^4(j+1)^2)\}}, \tag{33}$$

we have

$$\sum_{b\ge 0}\sum_{\mathbf{s}_b \text{ legal}}(J/4)^{-.9n_b} \le \prod_{j\in[\tau]}(1+a_j)-1 \le \exp\left(\sum_{j\in[\tau]}a_j\right)-1 \le J^{-c}$$

for some positive constant c, assuming that J is chosen sufficiently large. □

3. Proof of Theorem 1.3

3.1. *A version of Theorem 1.2 with multiplicities.* As a key tool in the proof of Theorem 1.3, we need a version of Theorem 1.2 for multisets. Given a finite set X, a *multiset* over X is a function $m : X \to \mathbb{N}_0$ (so $m(i)$ is the "multiplicity" of $i \in X$). Let $\mathcal{M}(X)$ denote the collection of multisets over X. The *size* of a multiset is given by $|m| := \sum_{i\in X} m(i)$. Let $\mathcal{M}_N(X)$ be the collection of multisets over X of size N. For multisets S and T, we say that T is a *subset* of S and denote $T \subseteq S$ if $T(x) \le S(x)$ for all x. Given a collection $\mathcal{F}$ of multisets over X, we say that $\mathcal{G}$ is a *cover* of $\mathcal{F}$ if for every $S \in \mathcal{F}$, there exists $T \in \mathcal{G}$ satisfying $T \subseteq S$. Given two multisets S and T, we define the *union* $S \vee T$ as the multiset with $(S \vee T)(i) = \max(S(i), T(i))$ and the *intersection* $S \wedge T$ as the multiset with $(S \wedge T)(i) = \min(S(i), T(i))$. We also define the *disjoint union* $S + T$ as the multiset with $(S + T)(i) = S(i) + T(i)$, and the *set difference* $S \setminus T$ as the multiset with $(S \setminus T)(i) = \max(S(i) - T(i), 0)$. For an element $x \in X$ and a multiset $S \in \mathcal{M}(X)$, we say that $x \in S$ if $S(x) > 0$ and $x \notin S$ if $S(x) = 0$.

We next give an overview of our proof of Theorem 1.3, which requires several additional ideas. First, we need the following multiset and weighted generalization of Theorem 1.2.

THEOREM 3.1. *There exist $L > 0$ with the following property. Consider any finite set X, and any collection Λ of sequences $\boldsymbol{\lambda} = (\lambda_i)_{i \in X}$ with $\lambda_i \geq 0$. Let $\mu : X \to \mathbb{R}_{\geq 0}$ be a probability measure on X. Let N be a positive integer. Let $\mathbf{W}$ be a random multiset of size N over X where for each multiset W of size N,*

$$\mathbb{P}[\mathbf{W} = W] = N! \prod_{x \in X} \frac{\mu(x)^{W(x)}}{W(x)!}.$$

Then the family

$$\left\{ S \in \mathcal{M}(X) : \sup_{\boldsymbol{\lambda} \in \Lambda} \sum_{i \in X} S(i)\lambda_i \geq L\mathbb{E} \sup_{\boldsymbol{\lambda} \in \Lambda} \sum_{i \in X} \mathbf{W}(i)\lambda_i \right\} \tag{34}$$

admits a cover $\mathcal{G}$ with

$$\sum_{G \in \mathcal{G}} \prod_{x \in X} \frac{(eN\mu(x))^{G(x)}}{G(x)!} \leq 1/2. \tag{35}$$

Note that for any positive integer N, the random multiset $\mathbf{W}$ induced by N independent samples from μ has distribution

$$\mathbb{P}[\mathbf{W} = W] = N! \prod_{x \in X} \frac{\mu(x)^{W(x)}}{W(x)!}.$$

In particular, for any positive integer N, the above distribution is a valid probability distribution.

In the next subsection, we discuss the reduction of Theorem 1.3 to Theorem 3.1. We first replace the distribution ν underlying the empirical process by an appropriate distribution μ over a finite set. For an appropriate large integer P, we partition the Polish space $\mathbb{T}$ into subsets $B_1, \ldots, B_P$ such that for each function $f \in \mathcal{F}$, f is approximately constant on each set. We then define the measure μ by setting $\mu(h) = \nu(B_h)$ for each $h \in [P]$. For each function in $\mathcal{F}$, we construct a sequence λ on $[P]$ by choosing an arbitrary element $x_h \in B_h$ and set $\lambda(h) = f(x_h)$. Let Λ be the collection of sequences λ corresponding to functions $f \in \mathcal{F}$.

To illustrate the connection to selector processes, observe that in the case ν is continuous, i.e., $\nu(\{x\}) = 0$ for any $x \in \mathbb{T}$, we can guarantee a partition as above in which $\mu(h) \in [1/(2P), 1/P]$ for all h. In this case, the empirical process given by N independent samples of ν is closely connected to the selector process on $[P]$ with parameter $p = N/P$ indexed by Λ. In a previous version of this paper, we give a proof of this special case following a simplified version of our strategy for Theorem 1.3 utilizing multisets. The follow-up work of Bednorz, Martynek and Meller [4] shows that this case can be treated without

resorting to multiset generalization. In particular, in this case, one observes that for P large a small cover can be constructed for the small probability event that an element appears more than once in N samples of μ.

In the general case where ν has atoms, this observation no longer holds: repeated appearance of elements in samples from ν is no longer a rare event. Multisets arise and as such, the multiset generalization (Theorem 3.1) of Theorem 1.2 is natural in this context. We prove Theorem 1.3 via the discretization of ν into μ and translate the desired conclusion into the multiset covering in Theorem 3.1. We emphasize that this entire plan presents no additional conceptual difficulty. Indeed, we shall present the entire proof of Theorem 3.1 in parallel with the proof of Theorem 1.5. One shall see that the only difference is the bookkeeping of combinatorial factors that arise due to the atomic nature of ν.

As will be evident in the proof of Theorem 1.3 from Theorem 3.1, it suffices in (34) to consider only multisets $S \in \mathcal{M}_N(X)$ of fixed size N. However, it turns out that it is not important to keep track of this restriction, and in fact, in proving Theorem 3.1, we will go through Theorem 3.3, which works instead with random multisets of size KN for an appropriate constant K. Thus, we find it more convenient and general to work with $S \in \mathcal{M}(X)$ in (34).

In Section 3.2, we prove Theorem 1.3 assuming Theorem 3.1, which requires careful construction of events $E_{g,t}$ in Theorem 1.3 from the cover $\mathcal{G}$ in Theorem 3.1. Finally, in Section 3.3, we give the full details of the proof of Theorem 3.1, which follows along the line of the proof of Theorem 1.5.

3.2. *Proof of Theorem 1.3 assuming Theorem 3.1.* Since $\mathbb{E}[\sup_{f\in\mathcal{F}} Z_f] \in (0,\infty)$, by renormalizing, we can assume that

$$\mathbb{E}[\sup_{f\in\mathcal{F}} Z_f] = 1. \tag{36}$$

Let $\epsilon > 0$ be sufficiently small. Let $\mathcal{F} = \{f_1, \dots, f_M\}$.

Let $U = \max_{i\le M} \sup f_i$, and note from (36) that $U \ge 1$. Partition $\mathbb{T}$ into sets of the form $I(k_1, \dots, k_M) := \bigcap_{i=1}^M f_i^{-1}([k_i\epsilon, (k_i{+}1)\epsilon))$, where $k_i \in [0, \lfloor U/\epsilon \rfloor]$ are integers. Define a new probability distribution μ on $\Omega := \{(k_1, \dots, k_M) : k_i \in [0, \lfloor U/\epsilon \rfloor]\}$ with $\mu(k_1, \dots, k_M) = \nu(I(k_1, \dots, k_M))$ and define $X_1, \dots, X_N$ as independent samples from the distribution μ. For each $i \in [M]$, define $\lambda^i : \Omega \to \mathbb{R}_{\ge 0}$ by $\lambda^i(k_1, \dots, k_M) = k_i\epsilon$, and let $\Lambda = \{\lambda^i : i \in [M]\}$. Note that for each $y \in I(k_1, \dots, k_M)$,

$$|f_i(y) - \lambda^i(k_1, \dots, k_M)| \le \epsilon. \tag{37}$$

Given independent samples $X_1, \dots, X_N$ from μ, the multiset given by $X_1, \dots, X_N$ is a multiset $\mathbf{W}$ of size N with distribution

$$\mathbb{P}[\mathbf{W} = W] = N! \prod_{x\in\Omega} \frac{\mu(x)^{W(x)}}{W(x)!}.$$

By Theorem 3.1, we can find a collection $\mathcal{G} \subseteq \mathcal{M}(\Omega)$ with

$$\sum_{G\in\mathcal{G}} \prod_{x\in\Omega} \frac{(eN\mu(x))^{G(x)}}{G(x)!} \leq 1/2, \tag{38}$$

and furthermore $\mathcal{G}$ covers

$$\left\{ S \in \mathcal{M}(\Omega) : \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{x\in\Omega} S(x)\lambda_x \geq L\mathbb{E} \sup_{\boldsymbol{\lambda}\in\Lambda} \sum_{x\in\Omega} \mathbf{W}(x)\lambda_x \right\}. \tag{39}$$

Notice that $\frac{(eN\mu(x))^{G(x)}}{G(x)!} \geq \left(\frac{eN\mu(x)}{G(x)}\right)^{G(x)}$. As such, by removing elements $x \in G$ in which $N\mu(x)/G(x) > 1$ from each $G \in \mathcal{G}$ (after which $\mathcal{G}$ remains a cover of (39) and the left-hand side of (38) decreases), we can assume without loss of generality that

$$N\mu(x)/G(x) \leq 1 \text{ for all } G \in \mathcal{G} \text{ and } x \in G. \tag{40}$$

For each $G \in \mathcal{G}$, let $\tilde{g}_G : \Omega \to \mathbb{R}$ be the function $\tilde{g}_G(x) = \log(N\mu(x)/G(x))$ for $x \in G$ and $\tilde{g}_G(x) = 0$ otherwise. Let $\tilde{t}_G = \sum_{x\in G} G(x)\log(N\mu(x)/G(x))$. Then, using (40), for any $(X_1, \ldots, X_N)$ for which the corresponding multiset W contains G, we have that

$$\prod_{i:X_i\in G} \frac{N\mu(X_i)}{G(X_i)} = \prod_{x\in G} \left(\frac{N\mu(x)}{G(x)}\right)^{W(x)} \leq \prod_{x\in G} \left(\frac{N\mu(x)}{G(x)}\right)^{G(x)},$$

so $(X_1, \ldots, X_N)$ is contained in the event

$$\mathcal{H}_{\tilde{g}_G,\tilde{t}_G} = \left\{ \sum_{i=1}^{N} \tilde{g}_G(X_i) \leq \tilde{t}_G \right\}. \tag{41}$$

We compute

$$\begin{aligned} \mathbb{E}\left[\exp\left(-\sum_{i=1}^{N} \tilde{g}_G(X_i)\right)\right] &= \left(\mathbb{E}\left[\exp(-\tilde{g}_G(X_1))\right]\right)^N \\ &= \left(\sum_{x\in G} \mu(x)\frac{G(x)}{N\mu(x)} + \sum_{x\notin G} \mu(x)\right)^N \\ &\leq (1 + |G|/N)^N. \end{aligned} \tag{42}$$

Note that $-\tilde{g}_G(x) \geq 0$ for all x. Hence, by Markov's inequality and (42),

$$\begin{aligned} \mathbb{P}[\mathcal{H}_{\tilde{g}_G,\tilde{t}_G}] &= \mathbb{P}\left[\exp\left(-\sum_{i=1}^{N} \tilde{g}_G(X_i)\right) \geq \exp(-\tilde{t}_G)\right] \\ &\leq \exp(\tilde{t}_G)(1 + |G|/N)^N \\ &= (1 + |G|/N)^N \prod_{x\in G} \left(\frac{N\mu(x)}{G(x)}\right)^{G(x)}. \end{aligned} \tag{43}$$

We also have $N\log(1+|G|/N) \le |G|$, and thus

$$\begin{aligned}\mathbb{P}[\mathcal{H}_{\tilde{g}_G,\tilde{t}_G}] &\le (1+|G|/N)^N \prod_{x\in G}\left(\frac{N\mu(x)}{G(x)}\right)^{G(x)} \\ &\le \prod_{x\in G} e^{G(x)}\left(\frac{N\mu(x)}{G(x)}\right)^{G(x)} \\ &\le \prod_{x\in G}\frac{(eN\mu(x))^{G(x)}}{G(x)!},\end{aligned} \tag{44}$$

where in the last inequality we use the estimate $G(x)! \le G(x)^{G(x)}$.

From (44) and (35), we have

$$\sum_{G\in\mathcal{G}} \mathbb{P}[\mathcal{H}_{\tilde{g}_G,\tilde{t}_G}] \le 1/2. \tag{45}$$

Finally, for $y \in \mathbb{T}$, define $\pi : \mathbb{T} \to \Omega$ by $\pi(y) = (k_1,\dots,k_M)$ for $y \in I(k_1,\dots,k_M)$. Observe that for independent samples $(Y_1,\dots,Y_N)$ from ν, defining $X_j = \pi(Y_j)$ for $j \in [N]$ and $\mathbf{W}$ the multiset given by $X_1,\dots,X_N$, then $(X_1,\dots,X_N)$ are independent samples from μ. Furthermore, by (37), for all $i \in [M]$,

$$\left|Z_{f_i}(Y_1,\dots,Y_N) - \frac{1}{N}\sum_{x\in\Omega}\mathbf{W}(x)\lambda^i(x)\right| \le \epsilon$$

and

$$\left|\mathbb{E}\left[\sup_{i\in[M]} Z_{f_i}(Y_1,\dots,Y_N)\right] - \frac{1}{N}\mathbb{E}\left[\sup_{i\in[M]}\sum_{x\in\Omega}\mathbf{W}(x)\lambda^i(x)\right]\right| \le \epsilon.$$

As such, assuming that $\sup_{f\in\mathcal{F}} Z_f(Y_1,\dots,Y_N) \ge 2L\mathbb{E}\sup_{f\in\mathcal{F}} Z_f$, recalling (36), for ϵ sufficiently small, we have that $\mathbf{W}$ is contained in (39) and hence covered by $\mathcal{G}$. Thus, by (41),

$$\left\{\sup_{f\in\mathcal{F}} Z_f \ge 2L\mathbb{E}\sup_{f\in\mathcal{F}} Z_f\right\} \subseteq \bigcup_{G\in\mathcal{G}} \pi^{-1}(\mathcal{H}_{\tilde{g}_G,\tilde{t}_G}).$$

Noting that $\pi^{-1}(\mathcal{H}_{\tilde{g}_G,\tilde{t}_G})$ is exactly the same as the event $E_{g,t}$ for $g = -N\tilde{g}_G\circ\pi$ and $t = -\tilde{t}_G$ for each $G \in \mathcal{G}$, this yields Theorem 1.3 (upon modifying the value of the constant L).

Remark 3.2. Theorem 3.1 also directly implies the following statement. For any $c > 0$, there exists $L > 0$ such that the following holds. We say that a subset of $\mathbb{T}^k$ is symmetric if it is invariant under coordinate permutations. For each $k \ge 1$, there exists a symmetric set $V_k \subseteq \mathbb{T}^k$ such that, if $(Y_1,\dots,Y_N)$ is contained in the event $\{\sup_{f\in\mathcal{F}} Z_f \ge L\mathbb{E}\sup_{f\in\mathcal{F}} Z_f\}$, then there exists $i_1 < \cdots < i_k$ such that $(Y_{i_1},\dots,Y_{i_k}) \in V_k$. Furthermore, for each $k \ge 1$, $\mathbb{P}((Y_1,\dots,Y_k) \in V_k) \le \frac{1}{2}(ck/N)^k$.

In particular, the above statement follows from Theorem 3.1 by taking $V_k = \bigcup_{G\in\mathcal{G},|G|=k} V_k(G)$, where $V_k(G)$ is the collection of tuples $(y_1,\dots,y_k)\in\mathbb{T}^k$ whose corresponding multiset is equal to G.

This implies (a stronger version of) the result in [15] on positive empirical processes in the special case where the class of functions only involves indicator functions of sets, and generalizes this result to the setting of general nonnegative functions.

3.3. *Proof of Theorem* 3.1. Theorem 3.1 follows from Theorem 3.3 below.

THEOREM 3.3. *There exist a positive integer K with the following property. Consider any finite set X and any family of multisets $\mathcal{F}\subseteq\mathcal{M}(X)$. Let $\mu: X\to\mathbb{R}_{\ge 0}$ be a probability measure on X. Let N be a positive integer. Let $\mathbf{W}$ be a random multiset of size KN over X where for each multiset W of size KN,*

$$\mathbb{P}[\mathbf{W}=W] = (KN)!\prod_{x\in X}\frac{\mu(x)^{W(x)}}{W(x)!}.$$

Assume that for each $S\in\mathcal{F}$ we are given a sequence $\boldsymbol{\lambda}^S=(\lambda^S(i))_{i\in X}$ with $\lambda^S(i)\ge 0$ and

$$\sum_{i\in X} S(i)\lambda^S(i)\ge 1. \tag{46}$$

Then if there is no cover $\mathcal{G}$ for $\mathcal{F}$ with

$$\sum_{G\in\mathcal{G}}\prod_{x\in X}\frac{(eN\mu(x))^{G(x)}}{G(x)!}\le 1/2,$$

we have

$$\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in X}\mathbf{W}(i)\lambda^S(i)\ge 10^{-11}. \tag{47}$$

Proof of Theorem 3.1 *from Theorem* 3.3. Under the assumptions in Theorem 3.3, a random multiset $\tilde{\mathbf{W}}$ over X of size KN with distribution

$$\mathbb{P}[\tilde{\mathbf{W}}=W] = (KN)!\prod_{x\in X}\frac{(\mu(x))^{W(x)}}{W(x)!}$$

satisfies

$$\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in X}\tilde{\mathbf{W}}(i)\lambda^S(i)\ge 10^{-11}.$$

Let $\mathbf{W}$ be a random multiset of size N with distribution

$$\mathbb{P}[\mathbf{W}=W] = N!\prod_{x\in X}\frac{(\mu(x))^{W(x)}}{W(x)!}.$$

Note that $\tilde{\mathbf{W}}$ has the distribution of a multiset induced by a tuple of KN independent samples from μ, and $\mathbf{W}$ has the distribution of a multiset induced

by a tuple of N independent samples from μ. Hence, we have

$$\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in X}\mathbf{W}(i)\lambda^S(i)\geq\frac{N}{KN}\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in X}\tilde{\mathbf{W}}(i)\lambda^S(i)\geq\frac{1}{K}\cdot 10^{-11}. \tag{48}$$

Theorem 3.1 then follows as in the proof of Theorem 1.2 from Theorem 1.4. In particular, letting

$$\mathcal{F}=\left\{S\in\mathcal{M}(X):\sup_{\boldsymbol{\lambda}\in\Lambda}\sum_{i\in X}S(i)\lambda_i\geq L\mathbb{E}\sup_{\boldsymbol{\lambda}\in\Lambda}\sum_{i\in X}\mathbf{W}(i)\lambda_i\right\},$$

for each $S\in\mathcal{F}$, there is $(\tau^S(i))_{i\in X}\in\Lambda$ with

$$\sum_{i\in X}S(i)\tau^S(i)\geq L\mathbb{E}\sup_{\boldsymbol{\lambda}\in\Lambda}\sum_{i\in X}\mathbf{W}(i)\lambda_i. \tag{49}$$

Define $\boldsymbol{\lambda}^S$ by $\lambda^S(i)=\tau^S(i)$ if $i\in S$ and $\lambda^S(i)=0$ otherwise. Then

$$\mathbb{E}\sup_{\boldsymbol{\lambda}\in\Lambda}\sum_{i\in X}\mathbf{W}(i)\lambda_i\geq\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in X}\mathbf{W}(i)\lambda^S(i). \tag{50}$$

However, assume for the sake of contradiction that $\mathcal{F}$ does not admit a cover satisfying (35). Then by (49) and (48),

$$\mathbb{E}\sup_{S\in\mathcal{F}}\sum_{i\in X}\mathbf{W}(i)\lambda^S(i)\geq\frac{1}{K}10^{-11}L\mathbb{E}\sup_{\boldsymbol{\lambda}\in\Lambda}\sum_{i\in X}\mathbf{W}(i)\lambda_i. \tag{51}$$

For L sufficiently large, (50) and (51) yield the desired contradiction. □

The proof of Theorem 3.3 follows along the main ideas behind the proof of Theorem 1.5. Below we give the full details to make transparent the parallel with the proof of Theorem 1.5. We emphasize that only direct adaptations are needed and there is no conceptual difficulty in following the argument upon having the proof of Theorem 1.5.

We follow the definitions in Section 3.1. In the following, denote $M=KN$, and note that $S,S',\hat{S}$ are always elements of $\mathcal{F}$. We say that $W\in\mathcal{M}_M(X)$ is *good* if $\max_{S\in\mathcal{F}}\sum_i W(i)\lambda^S(i)\geq 10^{-10}$, and bad otherwise.

By the same processing steps as in Section 2 (Observations 2.1 and 2.2), we can assume that $\lambda^S(i)=100^{-j}$ for some $j\in\mathbb{N}_0$, with the right-hand side of (46) weakened to 100^{-1}. Denote by S_j the sub-multiset of S consisting of all the elements i with $\lambda^S(i)=100^{-j}$. Let $s_j=|S_j|$. Then we can assume that for each $j\in\mathbb{N}_0$, either $s_j=0$ or $100^{j-4}/(j+1)^2\leq s_j<2\cdot 100^j$ and s_j is a power of 100, with the right-hand side of (46) weakened to 100^{-3}. Let $\tau=\max\{j:s_j\neq 0\}$. Note that $\tau<\infty$ since $\mathcal{F}$ is finite. The following definition is a straightforward generalization of Definition 2.3 to the multiset case.

Definition 3.4. Given $S \in \mathcal{F}_{\mathbf{s}}$ and $W \in \mathcal{M}_M(X)$, we say $U \subseteq S \setminus W$ is an (S,W)-fragment (with index b) if the following holds: there are $b \in \{-1, 0, \dots, \tau\}$ and $S' \in \mathcal{F}_{\mathbf{s}_b}$ such that

$$S'_j \subseteq W + U \quad \forall j \le b, \tag{52}$$

$$\sum_{j \ge b+1} \lambda^{S'}(S'_j \wedge (W+U)) \ge .01 \left[\sum_{j \ge b+1} \lambda^{S'}(S'_j) - 100^{-b} \left(\sum_{j \le b} s_j - |U| \right) \right]. \tag{53}$$

Given S and W, we denote by $T = T(S,W)$ the *minimum* (S,W)-fragment, where here minimum refers to the index b first (breaking ties arbitrarily), and then $|T|$ (again, breaking ties arbitrarily). We use $t = t(S,W)$ for $|T(S,W)|$. For any pair (S,W), its minimum fragment T and the corresponding index $b_T = b(S,W)$ are uniquely determined.

Definition 2.5 can be directly adapted to the current setting.

Definition 3.5. Given $Z \in \mathcal{M}(X)$, $b \in [\tau]$, $t \ge 0$, and a partial profile $\mathbf{s}_b$, we say $S' \in \mathcal{F}_{\mathbf{s}_b}$ is $(Z, b, \mathbf{s}_b, t)$-feasible if

$$S'_j \subseteq Z \quad \forall j \le b, \tag{54}$$

$$\sum_{j \ge b+1} \lambda^{S'}(S'_j \wedge Z) \ge .01 \left[\sum_{j \ge b+1} \lambda^{S'}(S'_j) - 100^{-b} \left(\sum_{j \le b} s_j - t \right) \right]. \tag{55}$$

The following propositions generalize Propositions 2.6, 2.7 and 2.8. Their proofs are direct adaptations of the proofs in Section 2 with little changes. We have included the details for completeness.

PROPOSITION 3.6. *For $S \in \mathcal{F}_{\mathbf{s}}$, let T be the minimum (S,W)-fragment with index b_T, and let $t = |T|$. If S' is $(W + T, b_T, \mathbf{s}_{b_T}, t)$-feasible, then*

$$\left(\bigvee_{j \le b_T} S'_j \right) \setminus W = T. \tag{56}$$

Proof. Note that, by the definition of feasibility, S' satisfies (52) and (53) with $U = T$ and $b = b_T$; that $(\bigvee_{j \le b_T} S'_j) \setminus W \subseteq T$ follows from this definition.

To show $T \subseteq (\bigvee_{j \le b_T} S'_j) \setminus W$, first observe that by minimality of $|T|$, $T \subseteq S' \setminus W$. Indeed, if $x \in T \setminus (S' \setminus W)$, then replacing T by $T \setminus \{x\}$, one can check that S' still satisfies (52) and (53). Suppose there is $x \in T \wedge S'_{j_0}$ for some $j_0 \ge b_T + 1$. Note that

$$\lambda^{S'}(x) \le 100^{-(b_T+1)}. \tag{57}$$

Then $T' := T \setminus \{x\}$ is an (S, W)-fragment, since S' trivially satisfies (52), and

$$\begin{aligned}
&\sum_{j \geq b_T+1} \lambda^{S'}(S'_j \wedge (W + T')) \\
&\quad \geq \sum_{j \geq b_T+1} \lambda^{S'}(S'_j \wedge (W + T)) - \lambda^{S'}(x) \\
&\quad \overset{(55),(57)}{\geq} .01 \left[\sum_{j \geq b_T+1} \lambda^{S'}(S'_j) - 100^{-b_T} \left(\sum_{j \leq b_T} s_j - |T| \right) - 100^{-b_T} \right] \\
&\quad = .01 \left[\sum_{j \geq b_T+1} \lambda^{S'}(S'_j) - 100^{-b_T} \left(\sum_{j \leq b_T} s_j - |T'| \right) \right].
\end{aligned}$$

This contradicts the minimality of T. □

PROPOSITION 3.7. *For $S \in \mathcal{F}_{\mathbf{s}}$, let T be the minimum (S, W)-fragment with index b_T, and let $t = |T|$. If S' is $(W + T, b_T, \mathbf{s}_{b_T}, t)$-feasible, then*

$$t \geq .9 \sum_{j \leq b_T} s_j. \tag{58}$$

Proof. Let $T_j = S'_j \wedge T$ and $t_j = |T_j|$ (so $t = \sum_{j \leq b_T} t_j$ by Proposition 3.6). We will prove the proposition by contradiction, showing that the failure of (58) violates the minimality of T.

First observe that, if we assume $t < .9 \sum_{j \leq b_T} s_j$, then there exists $b' \leq b_T$ such that

$$\sum_{b' \leq j \leq b_T} t_j 100^{-j} < .9 \sum_{b' \leq j \leq b_T} s_j 100^{-j}. \tag{59}$$

Indeed, if $(u_{b'} :=) \sum_{b' \leq j \leq b_T} 100^{-j}(t_j - .9 s_j) \geq 0$ for all $b' \leq b_T$, then

$$\begin{aligned}
\sum_{j \leq b_T} (t_j - .9 s_j) &= \sum_{i \leq b_T - 1} 100^i (u_i - u_{i+1}) + 100^{b_T} u_{b_T} \\
&= \sum_{i=1}^{b_T} (100^i - 100^{i-1}) u_i + u_0 \geq 0.
\end{aligned}$$

Note that by Proposition 3.6, (59) gives

$$\sum_{b' \leq j \leq b_T} \lambda^{S'}(S'_j \wedge W) \geq .1 \sum_{b' \leq j \leq b_T} \lambda^{S'}(S'_j). \tag{60}$$

Now we claim that

$$T' := \textstyle\bigvee_{j \leq b'-1} T_j \text{ is an } (S, W)\text{-fragment with index } b' - 1, \tag{61}$$

which contradicts the minimality of T.

Proof of (61). It is clear that S' satisfies property (52). For (53),

$$\begin{aligned}
&\sum_{j\geq b'} \lambda^{S'}(S'_j \wedge (W+T')) \\
&\overset{(56)}{=} \sum_{j\geq b_T+1} \lambda^{S'}(S'_j \wedge (W+T)) + \sum_{b'\leq j\leq b_T} \lambda^{S'}(S'_j \wedge W) \\
&\overset{(53),(60)}{\geq} .01\left[\sum_{j\geq b_T+1} \lambda^{S'}(S'_j) - 100^{-b_T}\left(\sum_{j\leq b_T} s_j - t\right)\right] + .1\sum_{b'\leq j\leq b_T} \lambda^{S'}(S'_j) \\
&= .01\sum_{j\geq b'} \lambda^{S'}(S'_j) - 100^{-(b_T+1)}\left(\sum_{j\leq b_T} s_j - t\right) + .09\sum_{b'\leq j\leq b_T} s_j 100^{-j} \\
&\overset{(\dagger)}{\geq} .01\left[\sum_{j\geq b'} \lambda^{S'}(S'_j) - 100^{-(b'-1)}\left(\sum_{j\leq b'-1} s_j - t'\right)\right],
\end{aligned}$$

where (†) follows from the inequalities

$$.09\sum_{b'\leq j\leq b_T} s_j 100^{-j} \geq 100^{-(b_T+1)}\sum_{b'\leq j\leq b_T} s_j$$

and

$$\begin{aligned}
(100^{-b'} - 100^{-(b_T+1)})\sum_{j\leq b'-1} s_j &\overset{(56)}{\geq} (100^{-b'} - 100^{-(b_T+1)})t' \\
&\geq 100^{-b'}t' - 100^{-(b_T+1)}t. \qquad \square
\end{aligned}$$

This completes the proof of the proposition. $\square$

PROPOSITION 3.8. *If W is bad, then for any S, we have $b_T \geq 0$.*

Proof. If $b_T = -1$, then by Proposition 3.6 we have $T = \emptyset$. Then by (53), there is S' for which

$$\lambda^{S'}(S' \wedge W) \geq .01\lambda^{S'}(S') \geq 10^{-10},$$

which contradicts the fact that W is bad. $\square$

For $W \in \mathcal{M}_M(X)$, define $\mathcal{U} = \mathcal{U}(W)$ to be

$$\mathcal{U}(W) := \{T(S,W) : S \in \mathcal{F}\}.$$

Note that $\mathcal{U}(W)$ covers $\mathcal{F}$ since $T(S,W) \subseteq S$ for each $S \in \mathcal{F}$.

We will also need the following simple observation.

PROPOSITION 3.9. *For $S \in \mathcal{F}_{\mathbf{s}}$, let T be the minimum (S,W)-fragment with index b_T, and let $t = |T|$. If S' is $(W+T, b_T, \mathbf{s}_{b_T}, t)$-feasible, then $W(x) < S'(x)$ for all x with $T(x) > 0$.*

Proof. Suppose that there is x with $T(x) > 0$ and $W(x) \geq S'(x)$. Then for T' with $T'(y) = T(y)$ for $y \neq x$ and $T'(x) = T(x) - 1$, we have that S' is $(W + T', b_T, \mathbf{s}_{b_T}, t-1)$-feasible, contradicting minimality of T. □

Choose an appropriate $c > 0$ and sufficiently large $J_0 > 0$ so that, with (cf. (33))

$$a_j := \log_{100}[2 \cdot 100^4 (j+1)^2] \cdot (e^{-4} J_0)^{-.9 \max\{1, 100^j/(100^4 (j+1)^2)\}}, \tag{62}$$

we have

$$\exp\left(\sum_{j \geq 0} a_j\right) - 1 \leq J_0^{-c}. \tag{63}$$

Now we choose K sufficiently large so that $K \geq J_0$.

LEMMA 3.10. *We have*

$$\sum_{W \text{ bad}} \mathbb{P}[\mathbf{W} = W] \sum_{U \in \mathcal{U}(W)} \prod_{x \in X} \frac{(eN\mu(x))^{U(x)}}{U(x)!} \leq J_0^{-c}$$

with c and J_0 as in (63).

Proof of Lemma 3.10. For a given bad $W \in \mathcal{M}_M(X)$ (recalling that $M = KN$), integer b, a partial profile $\mathbf{s}_b$ and $t \geq 0$, let

$$\mathcal{G}_W(\mathbf{s}_b, t) = \{S : S \in \mathcal{F}_{\mathbf{s}_b}, t(S, W) = t, b(S, W) = b\}$$

and

$$\mathcal{U}_W(\mathbf{s}_b, t) = \{T(S, W) : S \in \mathcal{G}_W(\mathbf{s}_b, t)\}.$$

Then we have, with $n_b := \sum_{j \leq b} s_j$ for given $\mathbf{s}_b$,

$$\begin{aligned} &\sum_{W \text{ bad}} \mathbb{P}[\mathbf{W} = W] \sum_{U \in \mathcal{U}(W)} \prod_{x \in X} \frac{(eN\mu(x))^{U(x)}}{U(x)!} \\ &\leq \sum_{b \geq 0} \sum_{\mathbf{s}_b \text{ legal}} \sum_{.9 n_b \leq t \leq n_b} \sum_{W \text{ bad}} \sum_{U \in \mathcal{U}_W(\mathbf{s}_b, t)} \mathbb{P}[\mathbf{W} = W] \prod_{x \in X} \frac{(eN\mu(x))^{U(x)}}{U(x)!}, \end{aligned} \tag{64}$$

noting that the range $b \geq 0$ follows from Proposition 3.8, and the range of t is from Propositions 3.6 and 3.7.

Fix $\mathbf{s}_b$ ($b \geq 0$) and t. Let $\mathbf{Z}$ be a random multiset in $\mathcal{M}_{M+t}(X)$ with

$$\mathbb{P}[\mathbf{Z} = Z] = (M+t)! \prod_{x \in X} \frac{\mu(x)^{Z(x)}}{Z(x)!}.$$

Let $\mathcal{M}_{(M,t)}(X)$ be the collection of $Z \in \mathcal{M}(X)$ that can be written as $W + T$ for $W \in \mathcal{M}_M(X)$ that is bad and $T \in \mathcal{U}_W(\mathbf{s}_b, t)$ has size t (so $|Z| = M + t$). By the definition of fragment, for $Z \in \mathcal{M}_{(M,t)}(X)$, there must exist a choice of $(Z, b, \mathbf{s}_b, t)$-feasible $\hat{S}$. Make a choice of $\hat{S} = \hat{S}(Z)$ (arbitrarily) such

that it only depends on Z. Then Proposition 3.6 shows that $T \subseteq \bigvee_{j\le b} \hat{S}_j$. In particular, we have

$$\begin{aligned}(65)\qquad &\sum_{W \text{ bad}} \sum_{U \in \mathcal{U}_W(\mathbf{s}_b,t)} \mathbb{P}[\mathbf{W}=W] \prod_{x\in X} \frac{(eN\mu(x))^{U(x)}}{U(x)!} \\ &\qquad \le \sum_{Z\in\mathcal{M}_{(M,t)}(X)} \sum_{T\subseteq \bigvee_{j\le b}\hat{S}_j} \mathbb{P}[\mathbf{W}=Z\setminus T] \prod_{x\in X} \frac{(eN\mu(x))^{T(x)}}{T(x)!}.\end{aligned}$$

It remains to bound the right-hand side:

$$\begin{aligned}&\sum_{Z\in\mathcal{M}_{(M,t)}(X)} \sum_{T\subseteq \bigvee_{j\le b}\hat{S}_j} \mathbb{P}[\mathbf{W}=Z\setminus T] \prod_{x\in X} \frac{(eN\mu(x))^{T(x)}}{T(x)!} \\ &= \sum_{Z\in\mathcal{M}_{(M,t)}(X)} \sum_{T\subseteq \bigvee_{j\le b}\hat{S}_j} \mathbb{P}[\mathbf{Z}=Z] \frac{\mathbb{P}[\mathbf{W}=Z\setminus T]}{\mathbb{P}[\mathbf{Z}=Z]} \cdot \prod_{x\in X} \frac{(eN\mu(x))^{T(x)}}{T(x)!} \\ &= \sum_{Z\in\mathcal{M}_{(M,t)}(X)} \mathbb{P}[\mathbf{Z}=Z] \sum_{T\subseteq \bigvee_{j\le b}\hat{S}_j} \frac{M!}{(M+t)!} \prod_{x\in X} \left(\mu(x)^{-T(x)} \frac{Z(x)!}{(Z(x)-T(x))!}\right) \\ &\quad \cdot N^t \prod_{x\in X} \frac{(e\mu(x))^{T(x)}}{T(x)!} \\ &\le \sum_{Z\in\mathcal{M}_{(M,t)}(X)} \mathbb{P}[\mathbf{Z}=Z] \frac{N^t}{M^t} e^t \sum_{T\subseteq \bigvee_{j\le b}\hat{S}_j} \prod_{x\in X} \binom{Z(x)}{T(x)}.\end{aligned}$$

Here, in the last inequality, we have used the trivial bound $M!/(M+t)! \le M^{-t}$. Furthermore, by Proposition 3.9, we have $\hat{S}(x) > W(x)$ for all x with $T(x) > 0$. In particular, for x with $T(x) > 0$, we have $Z(x) = W(x)+T(x) < \hat{S}(x)+T(x)$, and thus

$$\binom{Z(x)}{T(x)} \le \binom{\hat{S}(x)+T(x)}{T(x)} \le 2^{\hat{S}(x)+T(x)} \le 2^{2\hat{S}(x)}.$$

Hence, recalling that $M = KN$,

$$\begin{aligned}&\sum_{Z\in\mathcal{M}_{(M,t)}(X)} \mathbb{P}[\mathbf{Z}=Z] \frac{N^t}{M^t} e^t \sum_{T\subseteq \bigvee_{j\le b}\hat{S}_j} \prod_{x\in X} \binom{Z(x)}{T(x)} \\ &\le \sum_{Z\in\mathcal{M}_{(M,t)}(X)} \mathbb{P}[\mathbf{Z}=Z](e/K)^t \sum_{T\subseteq \bigvee_{j\le b}\hat{S}_j} \prod_{x\in \bigvee_{j\le b}\hat{S}_j} 2^{2\hat{S}(x)} \\ &\le \sum_{Z\in\mathcal{M}_{(M,t)}(X)} \mathbb{P}[\mathbf{Z}=Z](e/K)^t \cdot 2^{2n_b} \cdot 2^{n_b} \\ &= (e/K)^t \cdot 2^{3n_b}.\end{aligned}$$

Here, in the second inequality, we used that the number of $T \subseteq \bigvee_{j\le b} \hat{S}_j$ is at most 2^{n_b} and in the last step we used that $\sum_{Z\in\mathcal{M}_{(M,t)}(X)} \mathbb{P}[\mathbf{Z} = Z] \le 1$. Finally, we have

$$\begin{aligned}
&\sum_{b\ge 0} \sum_{\mathbf{s}_b \text{ legal}} \sum_{.9n_b\le t\le n_b} \sum_{W \text{ bad}} \sum_{U\in\mathcal{U}_W(\mathbf{s}_b,t)} \mathbb{P}[\mathbf{W} = W] \prod_{x\in X} \frac{(eN\mu(x))^{U(x)}}{U(x)!} \\
&\quad\le \sum_{b\ge 0} \sum_{\mathbf{s}_b \text{ legal}} \sum_{.9n_b\le t\le n_b} (e/K)^t \cdot 2^{3n_b} \\
&\quad\le \sum_{b\ge 0} \sum_{\mathbf{s}_b \text{ legal}} (e^{-4}J_0)^{-.9n_b}.
\end{aligned}$$

The rest of the proof proceeds identically to the proof of Theorem 1.5. As in (20) and (21), the number of possibilities for legal $\mathbf{s}_b$ is at most

$$\prod_{j\le b} \log_{100}(2\cdot 100^4(j+1)^2).$$

On the other hand, using the lower bound on s_j in (20),

$$n_b \ge \sum_{j\le b} \max\{1, 100^j/(100^4(j+1)^2)\}.$$

Therefore, with our choice of a_j from (62), we have

$$\sum_{b\ge 0} \sum_{\mathbf{s}_b \text{ legal}} (e^{-4}J_0)^{-.9n_b} \le \prod_{j\in[\tau]} (1+a_j) - 1 \le \exp\left(\sum_{j\in[\tau]} a_j\right) - 1 \le J_0^{-c}$$

for our choice of c and J_0 in (63). □

Theorem 3.3 follows immediately from Lemma 3.10.

Proof of Theorem 3.3. Since $\mathcal{F}$ does not have any cover $\mathcal{G}$ with

$$\sum_{U\in\mathcal{G}} \prod_{x\in X} \frac{(eN\mu(x))^{G(x)}}{G(x)!} \le 1/2,$$

we have for all bad W that

$$\sum_{U\in\mathcal{U}(W)} \prod_{x\in X} \frac{(eN\mu(x))^{U(x)}}{U(x)!} > 1/2.$$

Thus, Lemma 3.10 implies that

$$\mathbb{P}[\mathbf{W} \text{ is bad}] \le 2J_0^{-c}.$$

Hence,

$$\mathbb{E}\sup_{S\in\mathcal{F}} \sum_{i\in X} \mathbf{W}(i)\lambda^S(i) \ge (1-2J_0^{-c})10^{-10} \ge 10^{-11}.$$ □

Acknowledgements. We would like to thank Michel Talagrand for raising the question on general positive empirical processes and for many helpful discussions. We are grateful to David Conlon, Amir Dembo, Jacob Fox, and Jeff Kahn for their support and useful comments. We would also like to thank the anonymous referees for insightful comments that improve the paper and prompt us to realize an oversight in a previous version of the paper. Huy Tuan Pham is supported by a Two Sigma Fellowship, a Clay Research Fellowship and a Stanford Science Fellowship. Jinyoung Park is supported by NSF grants DMS-2153844 and DMS-2324978.

References

[1] R. Alweiss, S. Lovett, K. Wu, and J. Zhang, Improved bounds for the sunflower lemma, *Ann. of Math.* (2) **194** no. 3 (2021), 795–815. MR 4334977. Zbl 1479.05343. https://doi.org/10.4007/annals.2021.194.3.5.

[2] W. Bednorz and R. Latała, On the boundedness of Bernoulli processes, *Ann. of Math.* (2) **180** no. 3 (2014), 1167–1203. MR 3245015. Zbl 1309.60053. https://doi.org/10.4007/annals.2014.180.3.8.

[3] W. Bednorz and R. Martynek, The suprema of infinitely divisible processes, *Ann. Probab.* **50** no. 1 (2022), 397–417. MR 4385829. Zbl 1498.60146. https://doi.org/10.1214/21-aop1546.

[4] W. Bednorz, R. Martynek, and R. Meller, The suprema of selector processes with the application to positive infinitely divisible processes, 2022. arXiv 2212.14636.

[5] K. Frankston, J. Kahn, B. Narayanan, and J. Park, Thresholds versus fractional expectation-thresholds, *Ann. of Math.* (2) **194** no. 2 (2021), 475–495. MR 4298747. Zbl 1472.05132. https://doi.org/10.4007/annals.2021.194.2.2.

[6] S. A. van de Geer, *Empirical Processes in M-estimation*, *Cambridge Ser. Statistical Probabilistic Math.* **6**, Cambridge Univ. Press, Cambridge, 2000. Zbl 0953.62049.

[7] J. Kahn, B. Narayanan, and J. Park, The threshold for the square of a Hamilton cycle, *Proc. Amer. Math. Soc.* **149** no. 8 (2021), 3201–3208. MR 4273128. Zbl 1465.05094. https://doi.org/10.1090/proc/15419.

[8] N. Lord, Binomial averages when the mean is an integer, *Math. Gaz.* **94** no. 530 (2010), 331–332. https://doi.org/10.1017/S0025557200006690.

[9] J. Park and H. T. Pham, A proof of the Kahn–Kalai conjecture, *J. Amer. Math. Soc.* **37** no. 1 (2024), 235–243. MR 4654612. Zbl 1522.05442. https://doi.org/10.1090/jams/1028.

[10] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning*: *From Theory to Algorithms*, Cambridge Univ. Press, Cambridge, 2014. Zbl 1305.68005. https://doi.org/10.1017/CBO9781107298019.

[11] M. Talagrand, Donsker classes and random geometry, *Ann. Probab.* **15** no. 4 (1987), 1327–1338. MR 0905334. Zbl 0637.60040. https://doi.org/10.1214/aop/1176991979.

[12] M. Talagrand, Regularity of Gaussian processes, *Acta Math.* **159** no. 1-2 (1987), 99–149. MR 0906527. Zbl 0712.60044. https://doi.org/10.1007/BF02392556.

[13] M. Talagrand, A simple proof of the majorizing measure theorem, *Geom. Funct. Anal.* **2** no. 1 (1992), 118–125. MR 1143666. Zbl 0760.60043. https://doi.org/10.1007/BF01895708.

[14] M. Talagrand, *The Generic Chaining. Upper and Lower Bounds of Stochastic Processes*, *Springer Monogr. Math.*, Springer-Verlag, Berlin, 2005. MR 2133757. Zbl 1075.60001. https://doi.org/10.1007/3-540-27499-5.

[15] M. Talagrand, Selector processes on classes of sets, *Probab. Theory Related Fields* **135** no. 4 (2006), 471–486. MR 2240697. Zbl 1097.60019. https://doi.org/10.1007/s00440-004-0350-2.

[16] M. Talagrand, Are many small sets explicitly small?, in *STOC'*10—*Proceedings of the Forty-Second ACMSymposium on Theory of Computing*, ACM, New York, 2010, pp. 13–35. MR 2743011. Zbl 1293.60014. https://doi.org/10.1145/1806689.1806693.

[17] M. Talagrand, *Upper and Lower Bounds for Stochastic Processes. Modern Methods and Classical Problems*, *Ergeb. Math. Grenzgeb.* **60**, Springer, Heidelberg, 2014. MR 3184689. Zbl 1293.60001. https://doi.org/10.1007/978-3-642-54075-2.

[18] M. Talagrand, 2022, personal communication.

(Received: May 3, 2022)
(Revised: January 2, 2024)

Courant Institute of Mathematical Sciences, New York University, New York, NY
E-mail: jinyoungpark@nyu.edu

Stanford University, Stanford, CA
E-mail: huypham@stanford.edu