

Randomized Lower Bounds for Lopsided Set Disjointness

Mihai Pătraşcu

1 Lower Bound for Lopsided Set Disjointness

In the set disjointness problem, Alice and Bob receive sets S and T , and must determine whether $S \cap T = \emptyset$. We parameterize *lopsided set disjointness* (LSD) by the size of Alice’s set $|S| = N$, and B , the fraction between the universe and N . In other words, $S, T \subseteq [N \cdot B]$. Note that $|T|$ may be as large as $N \cdot B$.

Our goal here is to prove:

Theorem 1. *Let $\delta > 0$. In a bounded error protocol for LSD, either Alice sends at least $\delta N \lg B$ bits, or Bob sends at least $N \cdot B^{1-O(\delta)}$ bits.*

1.1 The Hard Instances

We image the universe to be partitioned into N blocks, each containing B elements. Alice’s set S will contain exactly one value from each block. Bob’s set T will contain $\frac{B}{2}$ values from each block; more precisely, it will contain one value from each pair $\{(j, 2k); (j, 2k + 1)\}$.

Let \mathcal{S} and \mathcal{T} be the possible choices for S and T according to these rules. Note that $|\mathcal{S}| = B^N$ and $|\mathcal{T}| = 2^{NB/2}$. We denote by S_i Alice’s set restricted to block i , and by T_i Bob’s set restricted to block i . Let \mathcal{S}_i and \mathcal{T}_i be the possible choices for S_i and T_i . We have $|\mathcal{S}_i| = B$ and $|\mathcal{T}_i| = 2^{B/2}$.

We now define \mathcal{D}_{YES} to be the uniform distribution on pairs $(S, T) \in \mathcal{S} \times \mathcal{T}$ with $S \cap T = \emptyset$. In each block i , there are two natural processes to generate $(S_i, T_i) \in \mathcal{S}_i \times \mathcal{T}_i$ subject to $S_i \cap T_i = \emptyset$:

1. Pick $T_i \in \mathcal{T}_i$ uniformly at random, i.e. independently pick one element from each pair $\{(i, 2k), (i, 2k + 1)\}$. Then, pick the singleton S_i uniformly at random from the complement of T_i . Note that $H(S_i | T_i) = \log_2(B/2)$.
2. Pick S_i to be a uniformly random element from block i . Then, pick T_i such that it doesn’t intersect S_i . Specifically, if $S_i \cap \{2k, 2k + 1\} = \emptyset$, T_i contains a random element among $2k$ and $2k + 1$. Otherwise, T_i gets the element not in S_i . Note that $H(T_i | S_i) = \frac{B}{2} - 1$.

To generate the distribution \mathcal{D}_{YES} , we will employ the following process. First, pick $q \in \{0, 1\}^N$ uniformly at random. For each $q_i = 0$, apply process 1. from above in block i ; for each $q_i = 1$, apply process 2. in block i . Now let Q be a random variable entailing: the vector q ; the value S_i for every i with $q_i = 0$; and the value T_i for every i with $q_i = 1$. Intuitively, Q describes the “first half” of each random process.

We now define distributions \mathcal{D}_k as follows. In block k (called *the designated block*), choose $(S_k, T_k) \in \mathcal{S}_k \times \mathcal{T}_k$ uniformly. Notice that $\Pr[S_k \cap T_k \neq \emptyset] = \frac{1}{2}$. In all other blocks $i \neq k$, choose $(S_i, T_i) \in \mathcal{S}_i \times \mathcal{T}_i$ as in the distribution \mathcal{D}_{YES} above. As above, we have a vector Q_{-k} , containing: q_i for $i \neq k$; all S_i such that $q_i = 0$; and all T_i such that $q_i = 1$.

We are going to prove that:

Theorem 2. Fix $\delta > 0$. If a protocol for LSD has error less than $\frac{1}{9999}$ on distribution $\frac{1}{N} \sum_{i=1}^N \mathcal{D}_i$, then either Alice sends at least $\delta N \lg B$ bits, or Bob sends at least $N \cdot B^{1-O(\delta)}$ bits.

The distribution \mathcal{D}_{YES} will be used to measure various entropies in the proof, which is convenient because the blocks are independent. However, the hard distribution on which we measure error is the mixture of \mathcal{D}_i 's. (Since \mathcal{D}_{YES} only has yes instances, measuring error on it would be meaningless.) While it may seem counterintuitive that we argue about entropies on one distribution and error on another, remember that \mathcal{D}_{YES} and \mathcal{D}_i are not too different: S and T are disjoint with probability $\frac{1}{2}$ when chosen by \mathcal{D}_i .

1.2 A Direct Sum Argument

We now wish to use a direct-sum argument to obtain a low-communication protocol for a single subproblem on $\mathcal{S}_i \times \mathcal{T}_i$. Intuitively, if the LSD problem is solved by a protocol in which Alice and Bob communicate a , respectively b bits, we might hope to obtain a protocol for some subproblem i in which Alice communicates $O(\frac{a}{N})$ bits and Bob communicates $O(\frac{b}{N})$ bits.

Let π be the transcript of the communication protocol. If Alice sends a bits and Bob b bits, we claim that $\mathbb{I}_{\mathcal{D}_{\text{YES}}}(S : \pi | Q) \leq a$ and $\mathbb{I}_{\mathcal{D}_{\text{YES}}}(T : \pi | Q) \leq b$. Indeed, once we condition on Q , S and T are independent random variables: in each block, either S is fixed and T is random, or vice versa. The independence implies that all information about S is given by Alice's messages, and all information about T by Bob's messages.

Because the \mathcal{S}_i 's are independent, we have $\mathbb{I}_{\mathcal{D}_{\text{YES}}}(S : \pi | Q) = \sum_{i=1}^N \mathbb{I}_{\mathcal{D}_{\text{YES}}}(S_i : \pi | Q)$. Similarly, $\mathbb{I}_{\mathcal{D}_{\text{YES}}}(T : \pi | Q) = \sum_{i=1}^N \mathbb{I}(T_i : \pi | Q)$. By averaging, it follows that for at least half of the values of i , we simultaneously have:

$$\mathbb{I}_{\mathcal{D}_{\text{YES}}}(S_i : \pi | Q) \leq \frac{4a}{N} \quad \text{and} \quad \mathbb{I}_{\mathcal{D}_{\text{YES}}}(T_i : \pi | Q) \leq \frac{4b}{N}. \quad (1)$$

Remember that the average error on $\frac{1}{N} \sum_i \mathcal{D}_i$ is $\frac{1}{9999}$. Then, there exists k among the half satisfying (1), such that the error on \mathcal{D}_k is at most $\frac{2}{9999}$. For the remainder of the proof, fix this k .

We can now reinterpret the original protocol for LSD as a new protocol for the disjointness problem in block k . This protocol has the following features:

Inputs: Alice and Bob receive $S_k \in \mathcal{S}_k$, respectively $T_k \in \mathcal{T}_k$.

Public coins: The protocol employs public coins to select Q_{-k} .

Private coins: Alice uses private coins to select S_i for all $i \neq k$ with $q_i = 0$. Bob uses private coins to select T_i for all $i \neq k$ with $q_i = 1$. As described above, S_i is chosen to be distinct from T_i (which is public knowledge, as part of Q_{-k}), and analogously for T_i .

Error: When S_k and T_k are chosen independently from $\mathcal{S}_k \times \mathcal{T}_k$, the protocol computes the disjointness of S_k and T_k with error at most $\frac{2}{9999}$. Indeed, the independent choice of S_k and T_k , and the public and private coins realize exactly the distribution \mathcal{D}_k .

Message sizes: Unfortunately, we cannot conclude that the protocol has small communication complexity in the regular sense, i.e. that the messages are small. We will only claim that the messages have small *information complexity*, namely that they satisfy (1).

1.3 Understanding Information Complexity

In normal communication lower bounds, one shows that if the protocol communicates too few bits, it must make a lot of errors. In our case, however, we must show that a protocol with small information complexity (but potentially large messages) must still make a lot of error.

Let us see what the information complexity of (1) implies. We have:

$$I_{\mathcal{D}_{\text{YES}}}(S_k : \pi | Q) = \frac{1}{2} \cdot I_{\mathcal{D}_{\text{YES}}}(S_k : \pi | q_k = 1, Q_{-k}, T_k) + \frac{1}{2} \cdot I_{\mathcal{D}_{\text{YES}}}(S_k : \pi | q_k = 0, Q_{-k}, S_k)$$

The second term is zero, since $H(S_k | S_k) = 0$. Thus, the old bound $I_{\mathcal{D}_{\text{YES}}}(S_k : \pi | Q) \leq \frac{4a}{N}$ can be rewritten as $I_{\mathcal{D}_{\text{YES}}}(S_k : \pi | q_k = 1, Q_{-k}, T_k) \leq \frac{8a}{N}$. We will now aim to simplify the left hand side of this expression.

First observe that we can eliminate $q_k = 1$ from the conditioning: $I_{\mathcal{D}_{\text{YES}}}(S_k : \pi | q_k = 1, Q_{-k}, T_k) = I_{\mathcal{D}_{\text{YES}}}(S_k : \pi | Q_{-k}, T_k)$. Indeed, π is a function of S and T alone. In other words, it is a function of the public coins Q_{-k} , the private coins, S_k , and T_k . But the distribution of the inputs is the same for $q_k = 1$ and $q_k = 0$. In particular, the two processes for generating S_k and T_k (one selected by $q_k = 0$, the other by $q_k = 1$) yield the the same distribution.

Now remember that \mathcal{D}_{YES} is simply \mathcal{D}_k conditioned on $S_k \cap T_k = \emptyset$. Thus, we can rewrite the information under the uniform distribution for S_k and T_k : $I(S_k : \pi | Q_{-k}, T_k, S_k \not\subset T_k) \leq \frac{8a}{N}$. (To alleviate notation, we drop subscripts for I and H whenever uniform distributions are used.) We are now measuring information under the same distribution used to measure the error.

Analogously, it follows that $I(T_k : \pi | Q_{-k}, S_k, S_k \not\subset T_k) \leq \frac{8b}{N}$. We can now apply three Markov bounds, and fix the public coins Q_{-k} such that all of the following hold:

1. the error of the protocol is at most $\frac{8}{9999}$;
2. $I(S_k : \pi | T_k, S_k \not\subset T_k) \leq \frac{32a}{N}$;
3. $I(T_k : \pi | S_k, S_k \not\subset T_k) \leq \frac{32b}{N}$.

To express the guarantee of 1., define a random variable \mathcal{E} which is one if the protocol makes an error, and zero otherwise. Note that \mathcal{E} is a function $\mathcal{E} : S_k \times \mathcal{C}_A \times T_k \times \mathcal{C}_B \rightarrow \{0, 1\}$, where we defined \mathcal{C}_A as the set of private coin outcomes for Alice and \mathcal{C}_B as the private coin outcomes for Bob. By 1., we have $\mathbb{E}[\mathcal{E}] \leq \frac{8}{9999}$.

We can rewrite 2. by expanding the definition of information:

$$\begin{aligned} I(S_k : \pi | T_k, S_k \not\subset T_k) &= H(S_k | T_k, S_k \not\subset T_k) - H(S_k | T_k, \pi, S_k \not\subset T_k) \\ &= \log_2 \frac{B}{2} - H(S_k | T_k, \pi, S_k \not\subset T_k) \end{aligned}$$

Applying a similar expansion to T_k , we conclude that:

$$\log_2 \frac{B}{2} - H(S_k | T_k, \pi, S_k \not\subset T_k) \leq \frac{32a}{N} \tag{2}$$

$$\left(\frac{B}{2} - 1\right) - H(S_k | T_k, \pi, S_k \not\subset T_k) \leq \frac{32b}{N} \tag{3}$$

Consider some transcript $\tilde{\pi}$ of the communication protocol. A standard observation in communication complexity is that the set of inputs for which $\pi = \tilde{\pi}$ is a combinatorial rectangle in the truth table of the protocol: one side is a subset of $S_k \times \mathcal{C}_A$, and the other a subset of $T_k \times \mathcal{C}_B$. In any rectangle, the output of the protocol is fixed.

Observe that the probability that the output of the protocol is “no” is at most $\frac{1}{2}$ (the probability that S_k and T_k intersect) plus $\frac{8}{9999}$ (the probability that the protocol makes an error). Discard all rectangles on which the output is “no.” Further discard all rectangles that fail to satisfy any of the following:

$$\begin{aligned}\mathbb{E}[\mathcal{E} \mid \pi = \tilde{\pi}] &\leq \frac{64}{9999} \\ \log_2 \frac{B}{2} - \mathbb{H}(S_k \mid T_k, S_k \not\subset T_k, \pi = \tilde{\pi}) &\leq \frac{256a}{N} \\ \left(\frac{B}{2} - 1\right) - \mathbb{H}(S_k \mid T_k, S_k \not\subset T_k, \pi = \tilde{\pi}) &\leq \frac{256b}{N}\end{aligned}$$

By the Markov bound, the mass of rectangles failing each one of these tests is at most $\frac{1}{8}$. In total, at most $\frac{1}{2} + \frac{8}{9999} + 3 \cdot \frac{1}{8} < 1$ of the mass got discarded. Thus, there exists a rectangle $\tilde{\pi}$ with answer “yes” that satisfies all three constraints.

Let σ be the distribution of S_k conditioned on $\pi = \tilde{\pi}$, and τ being the distribution of T_k conditioned on $\pi = \tilde{\pi}$. We this notation, we have:

1. $\mathbb{E}_{\sigma, \tau}[\mathcal{E}] \leq \frac{64}{9999}$, thus $\Pr_{\sigma, \tau}[S_k \cap T_k \neq \emptyset] \leq \frac{64}{9999}$.
2. $\mathbb{H}_{\sigma, \tau}(S_k \mid T_k, S_k \not\subset T_k) \geq \log_2 \frac{B}{2} - \frac{256a}{N}$.
3. $\mathbb{H}_{\sigma, \tau}(S_k \mid T_k, S_k \not\subset T_k) \geq \left(\frac{B}{2} - 1\right) - \frac{256b}{N}$.

In the next section, we shall prove that in every “large enough” rectangle (in the sense of entropy) the probability that S_k and T_k intersect is noticeable:

Lemma 3. *Let $\gamma > 0$. Consider probability distributions σ on support S_k , and τ on support T_k . The following cannot be simultaneously true:*

$$\Pr_{\sigma \times \tau}[S_k \cap T_k \neq \emptyset] \leq \frac{1}{42} \tag{4}$$

$$\mathbb{H}_{\sigma \times \tau}(S_k \mid T_k, S_k \not\subset T_k) \geq (1 - \gamma) \log_2 B \tag{5}$$

$$\mathbb{H}_{\sigma \times \tau}(T_k \mid S_k, S_k \not\subset T_k) \geq \frac{B}{2} - \frac{1}{840} \cdot B^{1-7\gamma} \tag{6}$$

Since $\frac{64}{9999} \leq \frac{1}{42}$, one of the following must hold:

$$\begin{aligned}\log_2 \frac{B}{2} - \frac{256a}{N} \leq (1 - \gamma) \log_2 B &\Rightarrow a \geq \frac{\gamma}{257} \cdot N \log_2 B \\ \left(\frac{B}{2} - 1\right) - \frac{256b}{N} \leq \frac{B}{2} - \frac{1}{840} \cdot B^{1-7\gamma} &\Rightarrow b \geq \frac{1}{216000} \cdot N \cdot B^{1-7\gamma}\end{aligned}$$

For N and B greater than a constant, it follows that either Alice sends at least $\delta N \lg B$ bits, or Bob must send at least $\frac{1}{216000} N \cdot B^{1-1799 \cdot \delta}$ bits.

1.4 Analyzing a Rectangle

The goal of this section is to show Lemma 3. Let μ_σ and μ_τ be the probability density functions of σ and τ . We define \mathcal{S}^* as the set of values of S_k that do not have unusually high probability according to σ : $\mathcal{S}^* = \{S_k \mid \mu_\sigma(S_k) \leq 1/B^{1-7\gamma}\}$. We first show that significant mass is left in \mathcal{S}^* :

Claim 4. $\mu_\sigma(\mathcal{S}^*) \geq \frac{1}{5}$.

Proof. Our proof will follow the following steps:

1. We find a column \widehat{T}_k in which the function is mostly one (i.e. typically $S_k \notin \widehat{T}_k$), and in which the entropy $H_\sigma(S_k | S_k \notin \widehat{T}_k)$ is large.
2. The mass of elements outside \mathcal{S}^* is bounded by the mass of elements outside \mathcal{S}^* and disjoint from \widehat{T}_k , plus the mass of elements intersecting \widehat{T}_k . The latter is small by point 1.
3. There are *few* elements outside \mathcal{S}^* and disjoint from \widehat{T}_k , because they each have high probability. Thus, if their total mass were large, their low entropy would drag down the entropy of $H_\sigma(S_k | S_k \notin \widehat{T}_k)$, contradiction.

To achieve step 1., we rewrite (4) and (5) as:

$$\begin{aligned} \Pr_{\sigma \times \tau}[S_k \subset T_k] &= \mathbb{E}_\tau \left[\Pr_\sigma[S_k \subset T_k] \right] && \leq \frac{1}{10} \\ \log_2 B - \mathbb{H}_{\sigma \times \tau}(S_k | T_k, S_k \notin T_k) &= \mathbb{E}_\tau \left[\log_2 B - \mathbb{H}_\sigma(S_k | S_k \notin T_k) \right] && \leq \gamma \log_2 B \end{aligned}$$

Applying two Markov bounds on T_k , we conclude that there exists some \widehat{T}_k such that:

$$\Pr_\sigma[S_k \subset \widehat{T}_k] \leq \frac{3}{10}; \quad \mathbb{H}_\sigma(S_k | S_k \notin \widehat{T}_k) \geq (1 - 3\gamma) \log_2(2B) \quad (7)$$

Define $\widehat{\sigma}$ to be the distribution σ conditioned on $S_k \notin \widehat{T}_k$.

We regard to step 2., we can write $\mu_\sigma(\mathcal{S}^*) \geq 1 - \Pr_\sigma[S_k \notin \mathcal{S}^* \wedge S_k \notin \widehat{T}_k] - \Pr_\sigma[S_k \subset \widehat{T}_k]$. The latter term is at most $\frac{3}{10}$. In step 3., we will upper bound the former term by $\frac{1}{2}$, implying $\mu_\sigma(\mathcal{S}^*) \geq \frac{1}{5}$.

For any variable X and event E , we can decompose:

$$\mathbb{H}(X) = \Pr[E] \cdot \mathbb{H}(X | E) + \Pr[\neg E] \cdot \mathbb{H}(X | \neg E) + H_b(\Pr[E]), \quad (8)$$

where $H_b(\cdot) \leq 1$ is the binary entropy function. We apply this relation to the variable S_k under the distribution $\widehat{\sigma}$, choosing \mathcal{S}^* as our event E . We obtain:

$$\mathbb{H}_{\widehat{\sigma}}(S_k) \leq \Pr_{\widehat{\sigma}}[\mathcal{S}^*] \cdot \mathbb{H}_{\widehat{\sigma}}(S_k | S_k \in \mathcal{S}^*) + \Pr_{\widehat{\sigma}}[\overline{\mathcal{S}^*}] \cdot \mathbb{H}_{\widehat{\sigma}}(S_k | S_k \notin \mathcal{S}^*) + 1$$

We have $\mathbb{H}_{\widehat{\sigma}}(S_k | S_k \in \mathcal{S}^*) \leq \log_2 \frac{B}{2}$ since there are at most $\frac{B}{2}$ choices for S_k disjoint from \widehat{T}_k . On the other hand, $\mathbb{H}_{\widehat{\sigma}}(S_k | S_k \notin \mathcal{S}^*) \leq (1 - 7\gamma) \log_2 B$. Indeed, there are at most $B^{1-7\gamma}$ distinct values outside \mathcal{S}^* , since each must have probability exceeding $1/B^{1-7\gamma}$. We thus obtain:

$$\mathbb{H}_{\widehat{\sigma}}(S_k) \leq \Pr_{\widehat{\sigma}}[\mathcal{S}^*] \cdot \log_2 \frac{B}{2} + \Pr_{\widehat{\sigma}}[\overline{\mathcal{S}^*}] \cdot (1 - 7\gamma) \log_2 B + 1$$

If we had $\Pr_{\widehat{\sigma}}[\overline{\mathcal{S}^*}] \geq \frac{1}{2}$, we would have $\mathbb{H}_{\widehat{\sigma}}(S_k) \leq (1 - 3.5\gamma) \log_2 B + 1 < (1 - 3\gamma) \log_2 B$ for large enough B . But this would contradict (7), which states that $\mathbb{H}_{\widehat{\sigma}}(S_k) \geq (1 - 3\gamma) \log_2 B$.

Since $\widehat{\sigma}$ was the distribution σ conditioned on $S_k \notin \widehat{T}_k$, Bayes' rule tells us that $\Pr_\sigma[S_k \notin \mathcal{S}^* \wedge S_k \notin \widehat{T}_k] \leq \Pr_\sigma[S_k \notin \mathcal{S}^*] \leq \frac{1}{2}$. \square

Let us now consider the function $f(T_k) = \mathbb{E}_\sigma[|S_k \cap T_k|]$. By linearity of expectation, $f(T_k) = \sum_{x \in T_k} \Pr_\sigma[x \in S_k] = \sum_{x \in T_k} \mu_\sigma(x)$, since S_k has a single element. Since $|S_k \cap T_k| \in \{0, 1\}$, we can write:

$$\Pr_{\sigma, \tau}[S_k \cap T_k \neq \emptyset] = \mathbb{E}_{\sigma, \tau}[|S_k \cap T_k|] = \mathbb{E}_\tau \left[\mathbb{E}_\sigma[|S_k \cap T_k|] \right] = \mathbb{E}_\tau[f(T_k)]$$

Thus, to reach a contradiction with (4), we must lower bound the expectation of $f(\cdot)$ over distribution τ . Since we do not have a good handle on τ , we will approach this goal indirectly: at first, we will completely ignore τ , and analyze the distribution of $f(T_k)$ when T_k is chosen uniformly at random from \mathcal{T}_k . After this, we will use the high entropy of τ , in the sense of (6), to argue that the behavior on τ cannot be too different from the behavior on the uniform distribution.

The expectation of $f(\cdot)$ over the uniform distribution is simple to calculate: $\mathbb{E}_{T_k \in \mathcal{T}_k}[f(T_k)] = \sum_x \Pr_{T_k \in \mathcal{T}_k}[x \in T_k] \cdot \mu_\sigma(x) = \sum_x \frac{1}{2} \mu_\sigma(x) = \frac{1}{2}$. In the sums, x ranges over elements in block k , each of which appears in T_k with probability $\frac{1}{2}$. Note that μ_σ is a probability density function, so $\sum_x \mu_\sigma(x) = 1$.

Our goal now is to show that when T_k is uniform in \mathcal{T}_k , the distribution of $f(\cdot)$ is tightly concentrated around its mean of $\frac{1}{2}$, and, in particular, away from zero. We will employ a Chernoff bound: we have $f(T_k) = \sum_{x \in T_k} \mu_\sigma(x)$, and each $x \in T_k$ is chosen independently among two distinct values. Thus, $f(T_k)$ is the sum of B random elements of μ_σ , each chosen independently.

The limitation in applying the Chernoff bound is the value of $\max_s \mu_\sigma(x)$, which bounds the variance of each sample. The set \mathcal{S}^* now comes handy, since we can restrict our attention to elements x with small μ_σ . Formally, consider $f^*(T_k) = \sum_{x \in T_k \cap \mathcal{S}^*} \mu_\sigma(x)$. Clearly $f^*(T_k)$ is a lower bound for $f(T_k)$.

The mean of $f^*(\cdot)$ is $\mathbb{E}_{T_k \in \mathcal{T}_k}[f^*(T_k)] = \sum_{x \in \mathcal{S}^*} \Pr_{T_k \in \mathcal{T}_k}[x \in T_k] \cdot \mu_\sigma(x) = \frac{1}{2} \mu_\sigma(\mathcal{S}^*) \geq \frac{1}{10}$. When T_k is uniform, $f^*(T_k)$ is the sum of B independent random variables, each of which is bounded by $1/B^{1-7\gamma}$. By the Chernoff bound,

$$\Pr_{T_k \in \mathcal{T}_k}[f^*(T_k) < \frac{1}{20}] < e^{-B^{1-7\gamma} \cdot \frac{1}{10} \cdot \frac{1}{8}} \leq e^{-B^{1-7\gamma}/80} \quad (9)$$

Now we are ready to switch back to distribution τ :

Claim 5. $\Pr_\tau[f^*(T_k) < \frac{1}{20}] \leq \frac{1}{2}$.

Proof. The main steps of our proof are:

1. As in the analysis of \mathcal{S}^* , we find a row \widehat{S}_k in which the function is mostly one (i.e. typically $\widehat{S}_k \not\subset T_k$), and in which the entropy $H_\tau(T_k | \widehat{S}_k \not\subset T_k)$ is large.
2. $\Pr_\tau[f^*(T_k) < \frac{1}{20}]$ is bounded by $\Pr_\tau[f^*(T_k) < \frac{1}{20} \wedge \widehat{S}_k \not\subset T_k]$, plus the probability that $\widehat{S}_k \subset T_k$. The latter is small by point 1.
3. There are few distinct values of T_k for which $f^*(T_k) < \frac{1}{20}$. If these values had a large mass conditioned on $\widehat{S}_k \not\subset T_k$, they would drag down the entropy of $H_\tau(T_k | \widehat{S}_k \not\subset T_k)$.

To achieve step 1., we rewrite (4) and (6) as:

$$\begin{aligned} \Pr_{\sigma \times \tau}[S_k \subset T_k] &= \mathbb{E}_\tau \left[\Pr_\sigma[S_k \subset T_k] \right] && \leq \frac{1}{10} \\ \frac{B}{2} - \mathbb{H}_{\sigma \times \tau}(T_k | S_k, S_k \not\subset T_k) &= \mathbb{E}_\sigma \left[B - \mathbb{H}_\tau(T_k | S_k \not\subset T_k) \right] && \leq \frac{1}{840} \cdot B^{1-7\gamma} \end{aligned}$$

Applying two Markov bounds on S_k , we conclude that there exists some \widehat{S}_k such that:

$$\Pr_{\tau}[\widehat{S}_k \subset T_k] \leq \frac{3}{10}; \quad \mathbb{H}_{\tau}(T_k \mid \widehat{S}_k \not\subset T_k) \geq \frac{B}{2} - \frac{1}{280} \cdot B^{1-7\gamma} \quad (10)$$

Define $\widehat{\tau}$ to be the distribution τ conditioned on $\widehat{S}_k \not\subset T_k$.

For step 2., we can write:

$$\begin{aligned} \Pr_{\tau} [f^*(T_k) < \frac{1}{20}] &= \Pr_{\tau} [f^*(T_k) < \frac{1}{20} \wedge \widehat{S}_k \not\subset T_k] + \Pr_{\tau} [f^*(T_k) < \frac{1}{20} \wedge \widehat{S}_k \subset T_k] \\ &\leq \Pr_{\tau} [f^*(T_k) < \frac{1}{20} \mid \widehat{S}_k \not\subset T_k] + \Pr_{\tau} [\widehat{S}_k \subset T_k] \leq \Pr_{\widehat{\tau}} [f^*(T_k) < \frac{1}{20}] + \frac{3}{10} \end{aligned}$$

We now wish to conclude by proving that $\Pr_{\widehat{\tau}}[f^*(T_k) < \frac{1}{20}] \leq \frac{1}{5}$. We apply the relation (8) to the variable T_k distributed according to $\widehat{\tau}$, with the event E chosen to be $f^*(T_k) < \frac{1}{20}$:

$$H_{\widehat{\tau}}(T_k) \leq \Pr_{\widehat{\tau}} [f^*(T_k) < \frac{1}{20}] \cdot H_{\widehat{\tau}}(T_k \mid f^*(T_k) < \frac{1}{20}) + \Pr_{\widehat{\tau}} [f^*(T_k) \geq \frac{1}{20}] \cdot B + 1$$

By (9), there are at most $2^{B/2}/e^{B^{1-7\gamma}/80}$ distinct choices of T_k such that $f^*(T_k) < \frac{1}{20}$. Thus, $H_{\widehat{\tau}}(T_k \mid f^*(T_k) < \frac{1}{20}) \leq \frac{B}{2} - B^{1-7\gamma} \cdot \frac{\log_2 e}{80}$.

If $\Pr_{\widehat{\tau}}[f^*(T_k) < \frac{1}{20}] \geq \frac{1}{5}$, then $H_{\widehat{\tau}}(T_k) \leq \frac{B}{2} - B^{1-7\gamma} \cdot \frac{\log_2 e}{400} + 1 < \frac{B}{2} - B^{1-7\gamma}/280$ for sufficiently large B . But this contradicts (10). \square

We have just shown that $\Pr_{\sigma, \tau}[S_k \cap T_k \neq \emptyset] = \mathbb{E}_{\tau}[f(T_k)] \geq \mathbb{E}_{\tau}[f^*(T_k)] \geq \frac{1}{20} \cdot \frac{1}{2} = \frac{1}{40}$. This contradicts (4). Thus, at least one of (4), (5), and (6) must be false.

This concludes the proof of Lemma 3 and of Theorem 2.