A Tight Parallel Repetition Theorem for Partially Simulatable Interactive Arguments via Smooth KL-Divergence

Itay Berman* Iftach Haitner†‡ Eliad Tsfadia†

June 2, 2020

Abstract

Hardness amplification is a central problem in the study of interactive protocols. While "natural" parallel repetition transformation is known to reduce the soundness error of some special cases of interactive arguments: three-message protocols (Bellare, Impagliazzo, and Naor [FOCS '97]) and public-coin protocols (Håstad, Pass, Wikström, and Pietrzak [TCC '10], Chung and Liu [TCC '10] and Chung and Pass [TCC '15]), it fails to do so in the general case (the above Bellare, Impagliazzo, and Naor; also Pietrzak and Wikström [TCC '07]).

The only known round-preserving approach that applies to all interactive arguments is Haitner's *random-terminating* transformation [SICOMP '13], who showed that the parallel repetition of the transformed protocol reduces the soundness error at a *weak* exponential rate: if the original m-round protocol has soundness error $1-\varepsilon$, then the n-parallel repetition of its random-terminating variant has soundness error $(1-\varepsilon)^{\varepsilon n/m^4}$ (omitting constant factors). Håstad *et al.* have generalized this result to *partially simulatable interactive arguments*, showing that the n-fold repetition of an m-round δ-simulatable argument of soundness error $1-\varepsilon$ has soundness error $(1-\varepsilon)^{\varepsilon\delta^2 n/m^2}$. When applied to random-terminating arguments, the Håstad *et al.* bound matches that of Haitner.

In this work we prove that parallel repetition of random-terminating arguments reduces the soundness error at a much stronger exponential rate: the soundness error of the n parallel repetition is $(1-\varepsilon)^{n/m}$, only an m factor from the optimal rate of $(1-\varepsilon)^n$ achievable in public-coin and three-message arguments. The result generalizes to δ-simulatable arguments, for which we prove a bound of $(1-\varepsilon)^{\delta n/m}$. This is achieved by presenting a tight bound on a relaxed variant of the KL-divergence between the distribution induced by our reduction and its ideal variant, a result whose scope extends beyond parallel repetition proofs. We prove the tightness of the above bound for random-terminating arguments, by presenting a matching protocol.

Keywords: parallel repetition; interactive argument; partially simulatable; smooth KL-divergence

*MIT. E-mail:itayberm@mit.edu. Research supported in part by NSF Grants CNS-1413920 and CNS-1350619, and by the Defense Advanced Research Projects Agency (DARPA) and the U.S. Army Research Office under contracts W911NF-15-C-0226 and W911NF-15-C-0236.

†School of Computer Science, Tel Aviv University. E-mail:{iftachh@cs.tau.ac.il, eliadtsf@tau.ac.il}. Research supported by ERC starting grant 638121 and Israel Science Foundation grant 666/19.

‡Director of the Check Point Institute for Information Security.

Contents

1 Introduction

Hardness amplification is a central question in the study of computation: can a somewhat secure primitive be made fully secure, and, if so, can this be accomplished without loss (i.e., while preserving certain desirable properties the original primitive may have). In this paper we focus on better understanding the above question with respect to interactive arguments (also known as, computationally sound proofs). In an interactive argument, a prover tries to convince a verifier in the validity of a statement. The basic properties of such proofs are *completeness* and *soundness*. Completeness means that the prover, typically using some extra information, convinces the verifier to accept valid statements with high probability. Soundness means that a cheating *polynomial-time* prover cannot convince the verifier to accept invalid statements, except with small probability. Interactive arguments should be compared with the related notion of *interactive proofs*, whose soundness should hold against *unbounded* provers. Interactive argument are important for being "sufficiently secure" proof system that sometimes achieve properties (e.g., compactness) that are beyond the reach of interactive proofs. Furthermore, the security of many cryptographic protocols (e.g., binding of a computationally binding commitment) can be cast as the soundness of a related interactive argument, but (being computational) cannot be cast as the soundness of a related interactive proof.

The question of hardness amplification with respect to interactive arguments is whether an argument with *non-negligible* soundness error, i.e., a cheating prover can convince the verifier to accept false statements with some non-negligible probability, can be transformed into a new argument, with similar properties, of negligible soundness error (i.e., the verifier almost never accepts false statements). The most common paradigm to obtain such an amplification is via *repetition*: repeat the protocol multiple times with independent randomness, and the verifier accepts only if the verifiers of the original protocol accept in *all* executions. Such repetitions can be done in two different ways, sequentially (known as *sequential repetition*), where the $(i+1)$ execution of the protocol starts only after the i^{th} execution has finished, or in parallel (known as *parallel repetition*), where the executions are all simultaneous. Sequential repetition is known to reduce the soundness error in most computational models (cf., Damgärd and Pfitzmann [DP98]), but has the undesired effect of increasing the round complexity of the protocol. Parallel repetition, on the other hand, does preserve the round complexity, and reduces the soundness error for (single-prover) interactive proofs (Goldreich [Gol99]) and two-prover interactive proofs (Raz [Raz98], Holenstein [Hol09], and Rao [Rao11]). Parallel repetition was also shown to reduce the soundness error in three-message arguments ([BIN97]) and public-coin arguments (Håstad, Pass, Wikström, and Pietrzak [HPWP10], Chung and Lu [CL02], and Chung and Pass [CP15]). Unfortunately, as shown by Bellare, Impagliazzo, and Naor [BIN97], and by Pietrzak and Wikström [PW12], parallel repetition *might not* reduce the soundness error of any interactive argument: assuming common cryptographic assumptions, [PW12] presented an 8-message interactive proof with constant soundness error, whose parallel repetition, for *any* polynomial number of repetitions, still has a constant soundness error.

Faced with the above barrier, Haitner [Hai13] presented a simple method for transforming any interactive argument π into a slightly modified protocol $\widetilde{\pi}$, such that the parallel repetition of $\widetilde{\pi}$ does reduce the soundness error. Given any m-round interactive protocol $\pi = (\mathrm{P}, \mathrm{V})$, let $\widetilde{\mathrm{V}}$ be the following *random-terminating variant* of V: in each round, $\widetilde{\mathrm{V}}$ flips a coin that takes one with probability $1/m$ and zero otherwise. If the coin outcome is one, $\widetilde{\mathrm{V}}$ accepts and aborts the execution. Otherwise, $\widetilde{\mathrm{V}}$ acts as V would, and continues to the next round. At the end of the

prescribed execution, if reached, $\widetilde{\mathrm{V}}$ accepts if and only if V would. Observe that if the original protocol π has soundness error $1-\varepsilon$, then the new protocol $\widetilde{\pi} = (\mathrm{P}, \widetilde{\mathrm{V}})$ has soundness error $1-\varepsilon/4$ (i.e., only slightly closer to one). Haitner [Hai13] proved that the parallel repetition of $\widetilde{\pi}$ does reduce the soundness error (for any protocol π). Håstad, Pass, Wikström, and Pietrzak [HPWP10] have generalized the above to *partially-simulatable interactive arguments*, a family of interactive arguments that contains the random-terminating variant protocols as a special case. An interactive argument $\pi = (\mathrm{P}, \mathrm{V})$ is δ-simulatable if given any partial view v of an efficient prover P^* interacting with V, the verifier's future messages in $(\mathrm{P}^*, \mathrm{V})$ can be simulated with probability δ. This means that one can efficiently sample a random continuation of the execution conditioned on an event of density δ over V's coins consistent with v. It is easy to see that the random-terminating variant of any protocol is $1/m$ simulatable. Unfortunately, the soundness bound proved by Haitner [Hai13] and Håstad *et al.* [HPWP10] lags way behind what one might have hoped for, making parallel repetition impractical in many typical settings. Assuming a δ-simulatable argument π has soundness error is $1-\varepsilon$, then π^n, the n-parallel repetition of π, was shown to have soundness error $(1-\varepsilon)^{\varepsilon\delta^2 n/m^2}$ (equals $(1-\varepsilon)^{\varepsilon n/m^4}$ if π is a random-terminating variant), to be compared with the $(1-\varepsilon)^n$ bound achieved by parallel repetition of interactive proofs, and by three-message and public-coin interactive arguments.[1] Apart from the intellectual challenge, improving the above bound is important since repeating the random-termination variant in parallel is the *only* known unconditional round-preserving amplification method for arbitrary interactive arguments.

1.1 Proving Parallel Repetition

Let $\pi = (\mathrm{P}, \mathrm{V})$ be an interactive argument with assumed soundness error $1-\varepsilon$, i.e., a polynomial time prover cannot make the verifier accept a false statement with probability larger than $1-\varepsilon$. Proving amplification theorems for such proof systems is done via reduction: assuming the existence of a cheating prover P^{n*} making all the n verifiers in n-fold protocol $\pi^n = (\mathrm{P}^n, \mathrm{V}^n)$ accept a false statement "too well" (e.g., more than $(1-\varepsilon)^n$), this prover is used to construct a cheating prover P^* making V accept this false statement with probability larger than $1-\varepsilon$, yielding a contradiction. Typically, the cheating prover P^* emulates an execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$ while *embedding* the (real) verifier V as one of the n verifiers (i.e., by embedding its messages). Analyzing the success probability of this P^* is directly reduced to bounding the "distance" (typically statistical distance or KL-divergence) between the following Winning and Attacking distributions: the Winning distribution is the n verifiers' messages distribution in a winning (all verifiers accept) execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$. The Attacking distribution is the n verifiers' messages distribution in the emulated execution done by P^* (when interacting with V).

If the verifier is public-coin, or if the prover is unrestricted (as in single-prover interactive proofs), an optimal strategy for P^* is sampling the emulated verifiers messages uniformly at random conditioned on all verifiers accept, and the messages so far. Håstad *et al.* [HPWP10] have bounded the statistical distance between the induced Winning and Attacking distributions in such a case, while Chung and Pass [CP15] gave a tight bound for the KL-divergence between these distributions, yielding an optimal result for public-coin arguments.

For non public-coin protocols, however, a computationally bounded prover cannot always perform the above sampling task (indeed, this inability underneath the counter examples for parallel

[1] As in all known amplifications of computational hardness, and proven to be an inherent limitation (at least to some extent) in Dodis *et al.* [DJMW12], the improvement in the soundness error does not go below negligible. We ignore this subtly in the introduction. We also ignore constant factors in the exponent.

repetition of such arguments). However, if the argument is random terminating, the cheating prover can sample the following "skewed" variant of the desired distribution: it samples as described above, but conditioned that the real verifier *aborts at the end of the current round*, making the simulation of its future messages trivial. More generally, for partially-simulatable arguments, the cheating prover samples the future messages of the real verifier using the built-in mechanism for sampling a skewed sample of its coins. Analyzing the prover success probability for such an attack, and thus upper-bounding the soundness error of the parallel repetition of such arguments, reduces to understanding the (many-round) skewed distributions induced by the above attack. This will be discussed in the next section.

1.2 Skewed Distributions

The Attacking distribution induced by the security proof of parallel repetition of partially-simulatable arguments discussed in Section 1.1, gives rise to the following notion of (many-round) skewed distributions. Let $P = P_X$ be a distribution over an $m \times n$ size matrices, letting P_{X_i} and P_{X^j} denoting the induced distribution over the i^{th} row and j^{th} column of X, respectively. For an event W, let $\widetilde{P} = P|W$. The following distribution $Q_{X,J}$ is a skewed variant of $\widetilde{P}$ induced by an event family $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ over P: let $Q_J = U_{[n]}$, and let

$$Q_{X|J} = \prod_{i=1}^{m} P_{X_{i,J}|X_{<i,J}} \widetilde{P}_{X_{i,-J}|X_{<i},X_{i,J},E_{i,J}} \tag{1}$$

for $X_{<i} = (X_1, \ldots, X_{i-1})$, $X_{<i,j} = (X_{<i})^j = (X_{1,j}, \ldots, X_{i-1,j})$ and $X_{i,-j} = X_{i,[n]\setminus\{j\}}$. That is, Q induced by first sampling $J \in [n]$ uniformly at random, and then sampling the following skewed variant of $\widetilde{P}$: At round i

1. Sample $X_{i,J}$ according to $P_{X_{i,J}|X_{<i,J}}$ (rather than $P_{X_{i,J}|X_{<i},W}$ as in $\widetilde{P}$),
2. Sample $X_{i,-J}$ according $\widetilde{P}_{X_{i,-J}|X_{<i},X_{i,J},E_{i,J}}$ (rather than $\widetilde{P}_{X_{i,J}|X_{<i},X_{i,J}}$).

At a first glance, the distribution Q looks somewhat arbitrary. Nevertheless, as we explain below, it naturally arises in the analysis of parallel repetition theorem of partially-simulatable interactive arguments, and thus of random-terminating variants. Somewhat similar skewed distributions also come up when proving parallel repetition of two-prover proofs, though there we only care for single round distributions, i.e., $m = 1$.

The distributions $\widetilde{P}$ and Q relate to the Winning and Attacking distributions described in Section 1.1 in the following way: let $\pi = (\mathrm{P}, \mathrm{V})$ be an m-round δ-simulatable argument, and let P^{n*} be an efficient (for simplicity) deterministic cheating prover for π^n. Let P to be the distribution of the n verifiers messages in a random execution of π^n, and let W be the event that P^{n*} wins in $(\mathrm{P}^{n*}, \mathrm{V}^n)$. By definition, $\widetilde{P} = P|W$ is just the Winning distribution. Assume for sake of simplicity that V is a random-termination variant (halts at the end of each round with probability $1/m$), let $E_{i,j}$ be the set of coins in which the j^{th} verifier halts at the end of the i^{th} round of $(\mathrm{P}^n, \mathrm{V}^n)$, and let $Q = Q(P, W, \{E_{i,j}\})$ be according to Equation (1). Then, ignoring some efficiency concerns, Q is just the Attacking distribution. Consequently, a bound on the soundness error of π^n can be proved via the following result:

Lemma 1.1 (informal)**.** *Let π be a partially simulatable argument of soundness error $(1-\varepsilon)$. Assume that for every efficient cheating prover for π^n and every event T, it holds that*

$$\Pr_{Q_X}[T] \leq \Pr_{\widetilde{P}_X}[T] + \gamma$$

where W, $\widetilde{P}$ and Q are as defined above with respect to this adversary, and that Q is efficiently samplable. Then π^n has soundness error $(1-\varepsilon)^{\log(1/P[W])/\gamma}$.

It follows that proving a parallel repetition theorem for partially simulatable arguments, reduces to proving that low probability events in $\widetilde{P}_X$ have low probability in Q_X (for the sake of the introduction, we ignore the less fundamental samplability condition assumed for Q). One can try to prove the latter, as implicitly done in [Hai13; HPWP10], by bounding the *statistical distance* between $\widetilde{P}$ and Q (recall that $\mathrm{SD}(P,Q) = \max_E(\Pr_P[E] - \Pr_Q[E])$). This approach, however, seems doomed to give non-tight bounds for several reasons: first, statistical distance is not geared to bound non-product distributions (i.e., iterative processes) as the one defined by Q, and one is forced to use a wasteful hybrid argument in order to bound the statistical distance of such distributions. A second reason is that statistical distance bounds the difference in probability between the two distributions for *any* event, where we only care that this difference is small for low (alternatively, high) probability events. In many settings, achieving this (unneeded) stronger guarantee inherently yields a weaker bound.

What seems to be a more promising approach is bounding the *KL-divergence* between $\widetilde{P}$ and Q (recall that $D(P||Q) = \mathrm{E}_{x\sim P} \log \frac{P(x)}{Q(x)}$). Having a chain rule, KL-divergence is typically an excellent choice for non-product distributions. In particular, bounding it only requires understanding the non-product nature (i.e., the dependency between the different entries) of the left-hand-side distribution. This makes KL-divergence a very useful measure in settings where the iterative nature of the right-hand-side distribution is much more complicated. Furthermore, a small KL-divergence guarantees that low probability events in $\widetilde{P}$ happen with almost the same probability in Q, but it only guarantees a weaker guarantee for other events (so it has the potential to yield a tighter result). Chung and Pass [CP15] took advantage of this observation for proving their tight bound on parallel repetition of public-coin argument by bounding the KL-divergence between their variants of $\widetilde{P}$ and Q. Unfortunately, for partially simulatable (and for random terminating) arguments, the KL-divergence between these distributions might be infinite.

Faced with the above difficulty, we propose a relaxed variant of KL-divergence that we name *smooth KL-divergence*. On the one hand, this measure has the properties of KL-divergence that make it suitable for our settings. However, on the other hand, it is less fragile (i.e., oblivious to events of small probability), allowing us to tightly bound its value for the distributions under consideration.

1.3 Smooth KL-divergence

The KL-divergence between distributions P and Q is a very sensitive distance measure: an event x with $P(x) \gg Q(x)$ might make $D(P||Q)$ huge even if $P(x)$ is tiny (e.g., $P(x) > 0 = Q(x)$ implies $D(P||Q) = \infty$). While events of tiny probability are important in some settings, they have no impact in ours. So we seek a less sensitive measure that enjoys the major properties of KL-divergence, most notably having chain-rule and mapping low probability events to low probability events. A natural attempt would be to define it as $\inf_{P',Q'}\{D(P'||Q')\}$, where the infimum is over all pairs of distributions such that both $\mathrm{SD}(P,P')$ and $\mathrm{SD}(Q,Q')$ are small. This relaxation,

however, requires an upper bound on the probability of events with respect to Q, which in our case is the complicated skewed distribution Q. Unfortunately, bounding the probability of events with respect to the distribution Q is exactly the issue in hand.

Instead, we take advantage of the asymmetric nature of the KL-divergence to propose a relaxation that only requires upper-bounding events with respect to P, which in our case is the much simpler $\widetilde{P}$ distribution. Assume P and Q are over a domain $\mathcal{U}$. Then the *α-smooth KL-divergence of P and Q* is defined by

$$D^{\alpha}(P||Q) = \inf_{(F_P,F_Q)\in\mathcal{F}}\{D(F_P(P)||F_Q(Q))\}$$

for $\mathcal{F}$ being the set of randomized function pairs, such that for every $(F_P, F_Q) \in \mathcal{F}$:

1. $\Pr_{x\sim P}[F_P(x) \neq x] \leq \alpha$.
2. $\forall x \in \mathcal{U}$ and $C \in \{P, Q\}$: $F_C(x) \in \{x\} \cup \overline{\mathcal{U}}$.

Note that for any pair $(F_P, F_Q) \in \mathcal{F}$ and any event B over $\mathcal{U}$, it holds that $\Pr_Q[B] \geq \Pr_{F_Q(Q)}[B]$, and $\Pr_{F_P(P)}[B] \geq \Pr_P[B] - \alpha$. Thus, if $\Pr_P[B]$ is low, a bound on $D(F_P(P)||F_Q(Q))$ implies that $\Pr_Q[B]$ is also low. Namely, low probability events in P happen with low probability also in Q.

Bounding smooth KL-divergence. Like the (standard) notion of KL-divergence, the power of smooth KL-divergence is best manifested when applied to non-product distributions. Let P and Q be two distributions for which we would like to prove that small events in $P_{X=(X_1,\ldots,X_m)}$ are small in $Q_{X=(X_1,\ldots,X_m)}$ (as a running example, let P and Q be the distributions $\widetilde{P}_X$ and $Q_{X,J}$ from the previous section, respectively). By chain rule of KL-divergence, it suffices to show that for some events $B_1, \ldots, B_m$ over Q (e.g., B_i is the event that $J|X_{<i}$ has high min entropy) it holds that

$$\sum_{i=1}^{m} D(P_{X_i}||Q_{X_i|B_{\leq i}} \mid P_{X_{<i}}) \qquad \left(\text{i.e.,} \sum_{i=1}^{m} \mathrm{E}_{x\leftarrow P_{X_{<i}}}\left[D\Big(P_{X_i|X_{<i}=x}||Q_{X_i|X_{<i}=x,B_{\leq i}}\Big)\right]\right) \tag{2}$$

is small, and $Q[B_{\leq m}]$ is large. Bounding Equation (2) only requires understanding P and simplified variants of Q (in which all but the i^{th} entry is sampled according to P). Unfortunately, bounding $Q[B_{\leq m}]$ might be hard since it requires a good understanding of the distribution Q itself. We would have liked to relate the desired bound to $P[B_{\leq m}]$, but the events $\{B_i\}$ might not even be defined over P (in the above example, P has no J part). However, smooth KL-divergence gives us the means to do almost that.

Lemma 1.2 (Bounding smooth KL-divergence, informal)**.** *Let P, Q and $\{B_i\}$ be as above. Associate the events $\{\widetilde{B}_i\}$ with P, each $\widetilde{B}_i$ (independently) occur with probability $Q[B_i \mid B_{<i}, X_{<i}]$. Then*

$$D^{1-P[\widetilde{B}_{\leq m}]}(P_X||Q_X) \leq \sum_{i=1}^{m} D\Big(P_{X_i}||Q_{X_i|B_{\leq i}} \mid P_{X_{<i}|\widetilde{B}_{\leq i}}\Big).$$

Namely, $\{\widetilde{B}_i\}$ mimics the events $\{B_i\}$, defined over Q, in (an extension of) P. It follows that bounding the smooth KL-divergence of P_X and Q_X (and thus guarantee that small events in P_X are small in Q_X), is reduced to understanding P and *simplified* variants of Q.

1.4 Main Results

We prove the following results (in addition to Lemmas 1.1 and 1.2). The first result, which is the main technical contribution of this paper, is the following bound on the smooth KL-divergence between a distribution and its many-round skewed variant.

Theorem 1.3 (Smooth KL-divergence for skewed distributions, informal)**.** *Let $P = P_X$ be a distribution over an $m \times n$ matrices with independent columns, and let W and $\mathcal{E} = \{E_{i,j}\}$ be events over P. Let $\widetilde{P} = P|W$ and let $Q = Q(P, W, \mathcal{E})$ be the skewed variant of $\widetilde{P}$ defined in Equation (1). Assume $\forall (i,j) \in [m] \times [n]$: (1) $E_{i,j}$ is determined by X^j and (2) There exists $\delta_{i,j} \in (0,1]$ such that $P[E_{i,j}|X_{\leq i,j}] = \delta_{i,j}$ for any fixing of $X_{\leq i,j}$. Then (ignoring constant factors, and under some restrictions on n and $P[W]$)*

$$D^{\varepsilon m + 1/\delta n}(\widetilde{P}_X || Q_X) \leq \varepsilon m + m/\delta n$$

for $\delta = \min_{i,j}\{\delta_{i,j}\}$ and $\varepsilon = \log(\frac{1}{P[W]})/\delta n$. In a special case where $E_{i,j}$ is determined by $X_{\leq i+1,j}$, it holds that

$$D^{\varepsilon + 1/\delta n}(\widetilde{P}_X || Q_X) \leq \varepsilon + m/\delta n.$$

Combining Lemma 1.1 and Theorem 1.3 yields the following bound on parallel repetition of partially simulatable arguments. We give separate bounds for partially simulatable argument and for *partially prefix-simulatable arguments*: a δ-simulatable argument is δ-prefix-simulatable if for any i-round view, the event E guaranteed by the simulatable property for this view is determined by the coins used in the first $i+1$ rounds. It is clear that the random-termination variant of an m-round argument is $1/m$-prefix-simulatable.

Theorem 1.4 (Parallel repetition for partially simulatable arguments, informal)**.** *Let π be an m-round δ-simulatable interactive argument with soundness error $1-\varepsilon$, and let $n \in \mathbb{N}$. Then π^n has soundness error $(1-\varepsilon)^{\delta n/m}$. Furthermore, if π is δ-prefix-simulatable, then π^n has soundness error $(1-\varepsilon)^{\delta n}$.*[2]

A subtlety that arises when proving Theorem 1.4 is that a direct composition of Lemma 1.1 and Theorem 1.3 only yields the desired result when the number of repetitions n is "sufficiently" large compared to the number of rounds m (roughly, this is because we need the additive term $m/\delta n$ in Theorem 1.3 to be smaller than ε). We bridge this gap by presenting a sort of upward-self reduction from a few repetitions to many repetitions. The idea underlying this reduction is rather general and applies to other proofs of this type, and in particular to those of [HPWP10; Hai13; CL10].[3]

We complete the picture by showing that an δ factor in the exponent in Theorem 1.4 is unavoidable.

[2] Throughout, we assume that the protocol transcript contains the verifier's Accept/Reject decision (which is without loss of generality for random-terminating variants). We deffer the more general case for the next version.

[3] Upward-self reductions trivially exist for interactive proof: assume the existence of a cheating prover P^{n*} breaking the α soundness error of π^n, then $(\mathrm{P}^{n*})^\ell$, i.e., the prover using P^{n*}in parallel for ℓ times, violates the assumed α^ℓ soundness error of $\pi^{n\ell}$. However, when considering interactive arguments, for which we cannot guarantee a soundness error below negligible (see Footnote 1), this approach breaks down when α^ℓ is negligible.

Theorem 1.5 (lower bound, informal)**.** *Under suitable cryptographic assumptions, for any $n, m \in \mathbb{N}$ and $\varepsilon \in [0,1]$, there exists an m-round δ-prefix-simulatable interactive argument π with soundness error $1-\varepsilon$, such that π^n has soundness error at least $(1-\varepsilon)^{\delta n}$. Furthermore, protocol π is a random-terminating variant of an interactive argument.*

It follows that our bound for partially prefix-simulatable arguments and random-termination variants, given in Theorem 1.4, is tight.

1.4.1 Proving Theorem 1.3

We highlight some details about the proof of Theorem 1.3. Using Lemma 1.2, we prove the theorem by showing that the following holds for a carefully chosen events $\{B_i\}$ over $Q_{X,J}$:

- $\sum_{i=1}^{m} D\left(\widetilde{P}_{X_i} || Q_{X_i|B_{\leq i}} \mid \widetilde{P}_{X_{<i}|\widetilde{B}_{\leq i}}\right)$ is small, and
- $\widetilde{P}[\widetilde{B}_{\leq m}]$ is large,

where $\{\widetilde{B}_i\}$ are events over (extension of) $\widetilde{P}$, with $\widetilde{B}_i$ taking the value 1 with probability $Q[B_i \mid B_{<i}, X_{<i}]$. We chose the events $\{B_i\}$ so that we have the following guarantees on $Q_{X_i,J|B_{\leq i},X_{<i}}$:

1. $J|X_{<i}$ has high entropy (like it has without any conditioning), and
2. $P[W \mid X_{<i}, X_{i,J}, E_{i,J}] \geq P[W|X_{<i}]/2$.

Very roughly, these guarantees make the task of bounding the required KL-divergence much simpler since they guarantee that the skewing induced by Q does not divert it too much (compared to $\widetilde{P}$). The remaining challenge is therefore lower-bounding $\widetilde{P}[\widetilde{B}_{\leq m}]$. We bound the latter distribution by associating a martingale sequence with the distribution Winning. In order to bound this sequence, we prove a new concentration bound for "slowly evolving" martingale sequences, Lemma 2.18, that we believe to be of independent interest.

1.5 Related Work

1.5.1 Interactive Arguments

Positive results. Bellare, Impagliazzo, and Naor [BIN97] proved that the parallel repetition of three-message interactive arguments reduces the soundness error at an exponential, but not optimal, rate. Canetti, Halevi, and Steiner [CHS05] later showed that parallel repetition does achieve an optimal exponential decay in the soundness error for such arguments. Pass and Venkitasubramaniam [PV12] have proved the same for constant-round public-coin arguments. For public-coin arguments of any (polynomial) round complexity, Håstad *et al.* [HPWP10] were the first to show that parallel repetition reduces the soundness error exponentially, but not at an optimal rate. The first optimal analysis of parallel repetition in public-coin arguments was that of Chung and Liu [CL10], who showed that the soundness error of the k repetitions improves to $(1-\varepsilon)^k$. Chung and Pass [CP15] proved the same bound using KL-divergence. For non-public coin argument (of any round complexity), Haitner [Hai13] introduced the random-terminating variant of a protocol, and proved that the parallel repetition of these variants improves the soundness error at a weak exponential rate. Håstad *et al.* [HPWP10] proved the same, with essentially the same parameters, for partially-simulatable arguments, that contain random-terminating protocols as a special

case. All the above results extend to "threshold verifiers" where the parallel repetition is considered accepting if the number of accepting verifiers is above a certain threshold. Our result rather easily extends to such verifiers, but we defer the tedious details to the next version. Chung and Pass [CP11] proved that full independence of the parallel executions is not necessary to improve the soundness of public-coin arguments, and that the verifier can save randomness by carefully correlating the different executions. It is unknown whether similar savings in randomness can be achieved for random-terminating arguments. Finally, the only known round-preserving alternative to the random-terminating transformation is the elegant approach of Chung and Liu [CL10], who showed that a fully-homomorphic encryption (FHE) can be used to compile any interactive argument to a one (with the same soundness error) for which parallel repetition improves the soundness error at ideal rate, i.e., $(1-\varepsilon)^n$. However, in addition to being conditional (and currently it is only known how to construct FHE assuming hardness of learning with errors [BV14]), the compiled protocol might lack some of the guarantees of the original protocol (e.g., fairness). Furthermore, the reduction is *non* black box (the parties homomorphically evaluate *each* of the protocol's gates), making the resulting protocol highly impractical, and preventing the use of this approach when only black-box access is available (e.g., the weak protocol is given as a DLL or implemented in hardware).

Negative results. Bellare, Impagliazzo, and Naor [BIN97] presented for any $n \in \mathbb{N}$, a four-message interactive argument of soundness error $1/2$, whose n-parallel repetition soundness remains $1/2$. Pietrzak and Wikström [PW12] ruled out the possibility that enough repetitions will eventually improve the soundness of an interactive argument. They presented a *single* 8-message argument for which the above phenomenon holds for all polynomial n simultaneously. Both results hold under common cryptographic assumptions.

1.5.2 Two-Prover Interactive Proofs

The techniques used in analyzing parallel-repetition of interactive arguments are closely related to those for analyzing parallel repetition of two-prover one-round games. Briefly, in such a game, two unbounded *isolated* provers try to convince a verifier in the validity of a statement. Given a game of soundness error $(1-\varepsilon)$, one might expect the soundness error of its n parallel repetition to be $(1-\varepsilon)^n$, but as in the case of interactive arguments, this turned out to be false [Fei91; FV02; FRS90]. Nonetheless, Raz [Raz98] showed that parallel repetition does achieve an exponential decay for any two-prover one-round game, and in particular reduces the soundness error to $(1-\varepsilon)^{\varepsilon^{O(1)}n/s}$, where s is the provers' answer length. These parameters were later improved by Holenstein [Hol09], and improved further for certain types of games by Rao [Rao11], Dinur and Steurer [DS14], and Moshkovitz [Mos14]. The core challenge in the analysis of parallel repetition of interactive arguments and of multi-prover one-round games is very similar: how to simulate a random accepting execution of the proof/game given the verifier messages. In interactive arguments, this is difficult since the prover lacks computational power. In multi-prover one-round games, the issue is that the different provers cannot communicate.

Open Questions

While our bound for the parallel repetition of partially prefix-simulatable arguments is tight, this question for (non prefix) partially simulatable arguments is still open (there is a $1/m$ gap in the

exponent). A more important challenge is to develop a better (unconditional) round-preserving amplification technique for arbitrary interactive arguments (which cannot be via random termination), or alternatively to prove that such an amplification does not exist.

Paper Organization

Basic notations, definitions and tools used throughout the paper are stated and proved in Section 2. The definition of smooth KL-divergence and some properties of this measure are given in Section 3. The definition of many-round skewed distributions and our main bound for such distributions are given in Section 4. The aforementioned bound is proven in Section 6, and is used in Section 5 for proving our bound on the parallel repetition of partially simulatable arguments. The matching lower bound on such parallel repetition, along with an intuitive explanation of why random-termination helps to beat [BIN97]'s counterexample, is given in Section 7. Missing proofs can be found in Section 8.

2 Preliminaries

2.1 Notation

We use calligraphic letters to denote sets, uppercase for random variables, and lowercase for values and functions. All logarithms considered here are natural logarithms (i.e., in base e). For $n \in \mathbb{N}$, let $[n] := \{1, \ldots, n\}$ and $(n) := \{0, \ldots, n\}$. Given a vector $v \in \Sigma^m$, we let v_i denote its i^{th} entry, and for ordered $\mathcal{S} = (s_1, \ldots, s_k) \subseteq [n]$ let $c_{\mathcal{S}} = (v_{s_1}, \ldots, v_{s_k})$. In particular, $v_{<i} = v_{1,\ldots,i-1}$ and $v_{\leq i} = v_{1,\ldots,i}$. For $v \in \{0,1\}^n$, let $1_v = \{i \in [n] \colon v_i = 1\}$. For $m \times n$ matrix x, let x_i and x^j denote their i^{th} row and j^{th} column respectively, and defined $x_{<i}$, $x_{\leq i}$, $x^{<j}$ and $x^{\leq j}$ respectively. Given a Boolean statement S (e.g., $X \geq 5$), let $\mathbb{1}_S$ be the indicator function that outputs 1 if S is a true statement and 0 otherwise. For $a \in \mathrm{R}$ and $b \geq 0$, let $a \pm b$ stand for the interval $[a - b, a + b]$.

Let poly denote the set of all polynomials, PPT denote for probabilistic polynomial time, and PPTM denote a PPT algorithm (Turing machine). A function $\nu \colon \mathbb{N} \to [0,1]$ is *negligible*, denoted $\nu(n) = \mathrm{neg}(n)$, if $\nu(n) < 1/p(n)$ for every $p \in \mathrm{poly}$ and large enough n. Function ν is *noticeable*, denoted $\nu(n) \geq 1/\mathrm{poly}(n)$, if exists $p \in \mathrm{poly}$ such that $\nu(n) \geq 1/p(n)$ for all n.

2.2 Distributions and Random Variables

A discrete random variable X over $\mathcal{X}$ is sometimes defined by its probability mass function (pmf) P_X (P is an arbitrary symbol). A conditional probability distribution is a function $P_{Y|X}(\cdot|\cdot)$ such that for any $x \in \mathcal{X}$, $P_{Y|X}(\cdot|x)$ is a pmf over $\mathcal{Y}$. The joint pmf P_{XY} can be written the product $P_X P_{Y|X}$, where $(P_X P_{Y|X})(x,y) = P_X(x) P_{Y|X}(y|x) = P_{XY}(xy)$. The marginal pmf P_Y can be written as the composition $P_{Y|X} \circ P_X$, where $(P_{Y|X} \circ P_X)(y) = \sum_{x \in \mathcal{X}} P_{Y|X}(y|x) P_X(x) = P_Y(y)$. We sometimes write $P_{\cdot,Y}$ to denote a pmf $P_{X,Y}$ for which we do not care about the random variable X. We denote by $P_X[W]$ the probability that an event W over P_X occurs, and given a set $\mathcal{S} \subseteq \mathcal{X}$ we define $P_X(\mathcal{S}) = P_X[X \in \mathcal{S}]$. Distribution P'_{XY} is an extension of P_X if $P'_X \equiv P_X$. Random variables and events defined over P_X are defined over the extension P'_{XY} by ignoring the value of Y. We sometimes abuse notation and say that P_{XY} is an extension of P_X.

The support of a distribution P over a finite set $\mathcal{X}$, denoted $\mathrm{Supp}(P)$, is defined as $\{x \in \mathcal{X} : P(x) > 0\}$. The *statistical distance* of two distributions P and Q over a finite set $\mathcal{X}$, denoted

as $\mathrm{SD}(P,Q)$, is defined as $\max_{\mathcal{S}\subseteq\mathcal{X}}|P(\mathcal{S})-Q(\mathcal{S})| = \frac{1}{2}\sum_{x\in\mathcal{S}}|P(x)-Q(x)|$. Given a set $\mathcal{S}$, let $U_{\mathcal{S}}$ denote the uniform distribution over the elements of $\mathcal{S}$. We sometimes write $x \sim \mathcal{S}$ or $x \leftarrow \mathcal{S}$, meaning that x is uniformly drawn from $\mathcal{S}$. For $p \in [0,1]$, let $\mathrm{Bern}(p)$ be the Bernoulli distribution over $\{0,1\}$, taking the value 1 with probability p.

2.3 KL-Divergence

Definition 2.1. *The* KL-divergence *(also known as, Kullback-Leibler divergence and relative entropy) between two distributions P, Q on a discrete alphabet $\mathcal{X}$ is*

$$D(P||Q) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)} = \mathrm{E}_{x\sim P}\log\frac{P(x)}{Q(x)},$$

where $0\cdot\log\frac{0}{0}=0$ and if $\exists x\in\mathcal{X}$ such that $P(x)>0=Q(x)$ then $D(P||Q)=\infty$.

Definition 2.2. *Let P_{XY} and Q_{XY} be two probability distributions over $\mathcal{X}\times\mathcal{Y}$. The* conditional divergence *between $P_{Y|X}$ and $Q_{Y|X}$ is*

$$D(P_{Y|X}||Q_{Y|X}|P_X) = \mathrm{E}_{x\sim P_X}[D(P_{Y|X=x}||Q_{Y|X=x})] = \sum_{x\in\mathcal{X}} P_X(x)D(P_{Y|X=x}||Q_{Y|X=x}).$$

Fact 2.3 (Properties of divergence)**.** *P_{XY} and Q_{XY} be two probability distributions over $\mathcal{X}\times\mathcal{Y}$. It holds that:*

1. *(Information inequality) $D(P_X||Q_X)\geq 0$, with equality holds iff $P_X = Q_X$.*
2. *(Monotonicity) $D(P_{XY}||Q_{XY}) \geq D(P_Y||Q_Y)$.*
3. *(Chain rule) $D(P_{X_1\cdots X_n}||Q_{X_1\cdots X_n}) = \sum_{i=1}^n D(P_{X_i|X_{<i}}||Q_{X_i|X_{<i}}|P_{X_{<i}})$.*
   If $Q_{X_1\cdots X_n} = \prod_{i=1}^n Q_{X_i}$ then

$$D(P_{X_1\cdots X_n}||Q_{X_1\cdots X_n}) = D(P_{X_1\cdots X_n}||P_{X_1}P_{X_2}\cdots P_{X_n}) + \sum_{i=1}^n D(P_{X_i}||Q_{X_i}).$$

4. *(Conditioning increases divergence) If $Q_Y = Q_{Y|X}\circ P_X$ (and $P_Y = P_{Y|X}\circ P_X$), then $D(P_Y||Q_Y)\leq D(P_{Y|X}||Q_{Y|X}|P_X)$.*
5. *(Data-processing) If $Q_Y = P_{Y|X}\circ Q_X$ (and $P_Y = P_{Y|X}\circ P_X$), it holds that $D(P_Y||Q_Y)\leq D(P_X||Q_X)$.*

Fact 2.4. *Let X be random variable drawn from P and let W be an event defined over P. Then*

$$D\big(P_{X|W}||P_X\big) \leq \log\frac{1}{P[W]}.$$

Fact 2.5. *Let X, Y be random variables drawn from either P or Q and let W be an event defined over P. It holds that*

$$\mathrm{E}_{x\sim P_{X|W}} D(P_{Y|X=x}||Q_{Y|X=x}) \leq \frac{1}{P[W]}\cdot D(P_{Y|X}||Q_{Y|X}||P_X).$$

Proof.

$$\begin{aligned}
\mathrm{E}_{x\sim P_{X|W}} D(P_{Y|X=x}||Q_{Y|X=x}) &= \sum_x P_{X|W}(x) D(P_{Y|X=x}||Q_{Y|X=x}) \\
&= \sum_x \frac{P[X=x,W]}{P[W]} D(P_{Y|X=x}||Q_{Y|X=x}) \\
&\leq \sum_x \frac{P_X(x)}{P[W]} D(P_{Y|X=x}||Q_{Y|X=x}) \\
&= \frac{1}{P[W]} \cdot D(P_{Y|X}||Q_{Y|X}||P_X),
\end{aligned}$$

where the inequality follows since $P[X=x,W] \leq P_X(x)$ and $D(\cdot||\cdot) \geq 0$. □

Fact 2.6. *Let X be a random variable over $\mathcal{X}$ drawn form either P_X or Q_X and let $\mathcal{S} \subseteq \mathcal{X}$. It holds that*

$$D(P_{X|X\in\mathcal{S}}||Q_X) \leq \frac{1}{P_X(\mathcal{S})} \cdot \left(D(P_X||Q_X) + \frac{1}{e} + 1 \right).$$

Proof. If $D(P_X||Q_X) = \infty$, then the statement holds trivially. Assume that $D(P_X||Q_X) < \infty$ and compute

$$\begin{aligned}
D(P_{X|X\in\mathcal{S}}||Q_X) &= \sum_{x\in\mathcal{S}} P_{X|X\in\mathcal{S}}(x) \log \frac{P_{X|X\in\mathcal{S}}(x)}{Q_X(x)} \\
&= \sum_{x\in\mathcal{S}} \frac{P_X(x)}{P_X(\mathcal{S})} \log \frac{P_X(x)/P_X(\mathcal{S})}{Q_X(x)} \\
&= \sum_{x\in\mathcal{S}} \frac{P_X(x)}{P_X(\mathcal{S})} \log \frac{1}{P_X(\mathcal{S})} + \sum_{x\in\mathcal{S}} \frac{P_X(x)}{P_X(\mathcal{S})} \log \frac{P_X(x)}{Q_X(x)}.
\end{aligned}$$

To bound the left sum, compute

$$\begin{aligned}
\sum_{x\in\mathcal{S}} \frac{P_X(x)}{P_X(\mathcal{S})} \log \frac{1}{P_X(\mathcal{S})} &\leq \sum_{x\in\mathcal{S}} \frac{P_X(x)}{P_X(\mathcal{S})} \cdot \frac{1}{P_X(\mathcal{S})} \\
&\leq \frac{1}{P_X(\mathcal{S})},
\end{aligned}$$

where the first inequality follows since $\log(x) \leq x$ for all x.

To bound the right sum, compute

$$\begin{aligned}
\sum_{x\in\mathcal{S}} \frac{P_X(x)}{P_X(\mathcal{S})} \log \frac{P_X(x)}{Q_X(x)} &= \frac{1}{P_X(\mathcal{S})} \left(\sum_{x\in\mathcal{S}} P_X(x) \log \frac{P_X(x)}{Q_X(x)} + \sum_{x\notin\mathcal{S}} P_X(x) \log \frac{P_X(x)}{Q_X(x)} - \sum_{x\notin\mathcal{S}} P_X(x) \log \frac{P_X(x)}{Q_X(x)} \right) \\
&= \frac{1}{P_X(\mathcal{S})} \left(D(P_X||Q_X) - \sum_{x\notin\mathcal{S}} P_X(x) \log \frac{P_X(x)}{Q_X(x)} \right).
\end{aligned}$$

The following calculation completes the proof:

$$\begin{aligned}\sum_{x\notin\mathcal{S}} P_X(x)\log\frac{P_X(x)}{Q_X(x)} &= \sum_{x\notin\mathcal{S}} Q_X(x)\frac{P_X(x)}{Q_X(x)}\log\frac{P_X(x)}{Q_X(x)}\\ &\geq \sum_{x\notin\mathcal{S}} Q_X(x)(-e^{-1})\\ &\geq -e^{-1},\end{aligned}$$

where the first inequlity holds since $x\log(x)\geq -e^{-1}$ for all $x>0$. □

Definition 2.7. *For $p,q\in[0,1]$ let $D(p||q) := D(\mathrm{Bern}(p)||\,\mathrm{Bern}(q))$.*

Fact 2.8 ([Mul, Implicit in Corollary 3.2 to 3.4])**.** *For any $p\in[0,1]$ it holds that*

1. *$D((1-\delta)p||p)\geq\delta^2 p/2$ for any $\delta\in[0,1]$.*
2. *$D((1+\delta)p||p)\geq\min\{\delta,\delta^2\}p/4$ for any $\delta\in[0,\frac{1}{p}-1]$.*

The proof of the following proposition, which relies on Donsker and Varadhan [DV83]'s inequality, is given in Section 8.3.

Proposition 2.9. *Let X be a random variable drawn form either P or Q. Assume that $\Pr_P[|X|\leq 1]=1$ (i.e., if X is drawn from P then $|X|\leq 1$ almost surely) and that there exist $\varepsilon,\sigma^2,K_1,K_2>0$ such that $\Pr_Q[|X|\leq 1]\geq 1-\varepsilon$ and*

$$\Pr_Q[|X|\geq t]\leq K_2\cdot\exp\left(-\frac{t^2}{K_1\sigma^2}\right)\quad \text{for all } 0\leq t\leq 1.$$

Then, there exists $K_3=K_3(K_1,K_2,\varepsilon)>0$ such that

$$\mathrm{E}_P[X^2]\leq K_3\cdot\sigma^2\cdot(D(P||Q)+1).$$

2.4 Interactive Arguments

Definition 2.10 (Interactive arguments)**.** *A* PPT *protocol* (P,V) *is an* interactive argument *for a language* $\mathrm{L}\in\mathrm{NP}$ *with* completeness α *and* soundness error β*, if the following holds:*

- $\Pr[(\mathrm{P}(w),\mathrm{V})(x)=1]\geq\alpha(|x|)$ *for any* $(x,w)\in R_{\mathrm{L}}$.
- $\Pr[(\mathrm{P}^*,\mathrm{V})(x)=1]\leq\max\{\beta(|x|),\mathrm{neg}(|x|)\}$ *for any* PPT P^* *and large enough* $x\notin\mathrm{L}$.

We refer to party P *as the* prover*, and to* V *as the* verifier*.*

Soundness against *non-uniform* provers is analogously defined, and all the results in this paper readily extend to this model.

Since in our analysis we only care about soundness amplification, in the following we fix L to be the empty language, and assume the input to the protocol is just a string of ones, which we refer to as the *security parameter*, a parameter we omit when cleared from the context.

2.4.1 Random-Terminating Variant

Definition 2.11 (Random-terminating variant, [Hai13])**.** *Let* V *be a* m*-round randomized interactive algorithm. The* random-terminating variant of V*, denoted* $\widetilde{\mathrm{V}}$*, is defined as follows: algorithm* V *acts exactly as* V *does, but adds the following step at the end of each communication round: it tosses an* $(1-1/m, 1/m)$ *biased coin (i.e.,* 1 *is tossed with probability* $1/m$*), if the outcome is one then it outputs* 1 *(i.e., accept) and halts. Otherwise, it continues as* V *would.*

For a protocol $\pi = (\mathrm{P}, \mathrm{V})$*, the protocol* $\widetilde{\pi} = (\mathrm{P}, \widetilde{\mathrm{V}})$ *is referred to as the random-terminating variant of* π*.*

2.4.2 Partially Simulatable Interactive Arguments

Definition 2.12 (Partially simulatable protocols, [HPWP10])**.** *A randomized interactive algorithm* V *is* δ-simulatable*, if there exists an oracle-aided* S *(simulator) such that the following holds: for every strategy* P^* *and a partial view* v *of* P^* *in an interaction of* $(\mathrm{P}^*, \mathrm{V})(1^\kappa)$*, the output of* $\mathsf{S}^{\mathrm{P}^*}(1^\kappa, v)$ *is* P^**'s view in a random continuation of* $(\mathrm{P}^*, \mathrm{V})(1^\kappa)$ *conditioned on* v *and* Δ*, for* Δ *being a* δ*-dense subset of the coins of* V *that are consistent with* v*. The running time of* $\mathsf{S}^{\mathrm{P}^*}(1^\kappa, v)$ *is polynomial in* κ *and the running time of* $\mathrm{P}^*(1^\kappa)$*.*

Algorithm V *is* δ-prefix-simulatable *if membership in the guaranteed event* Δ *is determined by the coins* V *uses in the first* $\mathrm{round}(v)+1$ *rounds.*[4]

An interactive argument (P, V) *is* δ-simulatable/ δ-prefix-simulatable, if V is.

It is clear that random termination variant of an m-round interactive argument is $1/m$-prefix-simulatable.

Remark 2.13. *One can relax the above definition and allow a different (non-black) simulator per* P^**, and then only require it to exists for poly-time* P^**. While our proof readily extends to this relaxation, we prefer to use the above definition for presentation clarity.*

2.4.3 Parallel Repetition

Definition 2.14 (Parallel repetition)**.** *Let* (P, V) *be an interactive protocol, and let* $n \in \mathbb{N}$*. We define the* n-parallel-repetition *of* (P, V) *to be the protocol* $(\mathrm{P}^n, \mathrm{V}^n)$ *in which* P^n *and* V^n *execute* n *copies of* (P, V) *in parallel, and at the end of the execution,* V^n *accepts if all copies accept.*

Black-box soundness reduction. As in most such proofs, our proof for the parallel repetition of partially-simulatable arguments has the following black-box form.

Definition 2.15 (Black-box reduction for parallel repetition)**.** *Let* $\pi = (\mathrm{P}, \mathrm{V})$ *be an interactive argument. An oracle-aided algorithm* R *is a* black-box reduction for the g-soundness of the parallel repetition of π*, if the following holds for any poly-bounded* n*: let* $\kappa \in \mathbb{N}$ *and* P^{n*} *be deterministic cheating prover breaking the soundness of* $\pi^{n=n(\kappa)}(1^\kappa)$ *with probability* $\varepsilon' \geq g(n, \varepsilon = \varepsilon(\kappa))$*. Then*

Sucesss probability. $\mathsf{R} = \mathsf{R}^{\mathrm{P}^{n*}}(1^\kappa, 1^n)$ *breaks the soundness of* π *with probability at least* $1 - \varepsilon/3$*.*

Running time. *Except with probability* $\varepsilon/3$*, the running time of* R *is polynomial in* κ*, the running time of* $\mathrm{P}^{n*}(1^\kappa)$ *and* $1/\varepsilon'$*.*

[4] $\Delta = \Delta_1 \times \Delta_2$, for Δ_1 being a (δ-dense) subset of the possible values for first $\mathrm{round}(v)+1$ round coins, and Δ_2 is the set of all possible values for the coins used in rounds $\mathrm{round}(v)+2, \ldots, m$, for m being the round complexity of V.

We use the following fact.

Proposition 2.16. *Assume there exists a black-box reduction for the g-soundness of the parallel repetition of any δ-simulatable [resp., δ-prefix-simulatable] interactive argument, then for any poly-bounded n, the soundness error of the n-fold repetition of any such argument is bounded by $g(n, \varepsilon)$.*

Proof. The only non-trivial part is how to handle randomized cheating provers (the above definition of black-box reduction only considers deterministic provers). Let $\pi = (\mathrm{P}, \mathrm{V})$ be a δ-simulatable interactive argument (the proof for δ-prefix-simulatable arguments follows the same lines). Let P^{n*} be an efficient randomized cheating prover violating the $g(n, \varepsilon)$ soundness error of π, and let $r(\kappa)$ be a bound in the number of coins it uses. Let $\widehat{\mathrm{V}}$ be the variant of V that appends $r(\kappa)$ uniform coins to its first message. It is clear that if $\widehat{\mathrm{V}}$ is also δ-simulatable. Consider the deterministic cheating prover $\mathrm{P}^{n*\prime}$ that attack $\widehat{\mathrm{V}}^n$ by acting as P^{n*} whose random coins set to the randomness appended to the first message of the first verifier. It is clear that $\mathrm{P}^{n*\prime}$ success probability (when attacking $\widehat{\mathrm{V}}^n$) equals that of P^{n*} (when attacking V^n). Hence, the existence of a black-box reduction for the g-soundness of $(\mathrm{P}, \widehat{\mathrm{V}})^n$, yields an efficient attacker $\mathrm{P}^{*\prime}$ breaking the $(1-\varepsilon)$ soundness of $(\mathrm{P}, \widehat{\mathrm{V}})$. This attacker can be easily modified to create an efficient attacker breaking the $(1-\varepsilon)$ soundness of π. □

2.5 Martingales

Definition 2.17. *A sequence of random variables $Y_0, Y_1, \ldots, Y_n$ is called a **martingale sequence with respect to** a sequence $X_0, X_1, \ldots, X_n$, if for all $i \in [n]$: (1) Y_i is a deterministic function of $X_0, \ldots, X_i$, and (2) $\mathrm{E}[Y_i \mid X_0, \ldots, X_{i-1}] = Y_{i-1}$.*

The following lemma (proven in Section 8.4) is a new concentration bound on "slowly evolving" martingales.

Lemma 2.18 (A bound on slowly evolving martingales)**.** *Let $Y_0 = 1, Y_1, \ldots, Y_n$ be a martingale w.r.t $X_0, X_1, \ldots, X_n$ and assume that $Y_i \geq 0$ for all $i \in [n]$. Then for every $\lambda \in (0, \frac{1}{4}]$ it holds that*

$$\Pr[\exists i \in [n] \text{ s.t. } |Y_i - 1| \geq \lambda] \leq \frac{23 \cdot \mathrm{E}\left[\sum_{i=1}^{n} \min\{|R_i|, R_i^2\}\right]}{\lambda^2}$$

for $R_i = \frac{Y_i}{Y_{i-1}} - 1$, letting $R_i = 0$ in case $Y_{i-1} = Y_i = 0$.

That is, if Y_i is unlikely to be far from Y_{i-1} in a multiplicative manner, then the sequence is unlikely to get far from 1. We use the following corollary of Lemma 2.18 (proven in Section 8.5).

Proposition 2.19. *Let $Y_0 = 1, Y_1, \ldots, Y_n$ be a martingale w.r.t $X_0, X_1, \ldots, X_n$ where $Y_i \geq 0$ for all $i \in [n]$. Let $Z_1, \ldots, Z_n$ and $T_1, \ldots, T_n$ be sequences of random variables satisfying for all $i \in [n]$: (1) $Y_i = Y_{i-1} \cdot (1 + Z_i)/(1 + T_i)$, and (2) T_i is a deterministic function of $X_0, X_1, \ldots, X_{i-1}$. Then*

$$\Pr[\exists i \in [n] \text{ s.t. } |Y_i - 1| \geq \lambda] \leq \frac{150 \cdot \mathrm{E}\left[\sum_{i=1}^{n} \left(\min\{|Z_i|, Z_i^2\} + \min\{|T_i|, T_i^2\}\right)\right]}{\lambda^2}$$

2.6 Additional Fact and Concentration Bounds

We use the following fact.

Fact 2.20 ([Hai13], Proposition 2.5)**.** *Let $P_{X_1,\ldots,X_m}$ be a distribution and let W be an event over P. Then for every $i \in [m]$ it holds that $\mathrm{E}_{x_{<i} \sim P_{X_{<i}|W}}[1/P[W \mid X_{<i} = x_{<i}]] = 1/P[W]$.*

2.6.1 Sum of Independent Random Variables

Fact 2.21 (Hoeffding's inequality)**.** *Let $X = X_1 + \cdots + X_n$ be the sum of independent random variables such that $X_i \in [a_i, b_i]$. Then for all $t \geq 0$:*

1. $\Pr[X - \mathrm{E}[X] \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$.
2. $\Pr[|X - \mathrm{E}[X]| \geq t] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$.

Fact 2.22 ([CL02, Lemma 2.1])**.** *Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \sim \mathrm{Bern}(p_i)$. Let $X = \sum_{i=1}^n b_i X_i$ with $b_i > 0$, and let $v = \sum_{i=1}^n b_i^2 p_i$. Then for all $t \geq 0$:*

$$\Pr[|X - \mathrm{E}[X]| \geq t] \leq 2\exp\left(-\frac{t^2}{2(v + bt/3)}\right)$$

for $b = \max\{b_1, b_2, \ldots, b_n\}$.

We use the following fact.

Fact 2.23. *Let $L_1, \ldots, L_n$ be independent random variables over $\mathbb{R}$ with $|L_i| \leq \ell$ for all $i \in [n]$ and let $Z_i = (L_i/p_i) \cdot \mathrm{Bern}(p_i)$ with $p_i > 0$ for all $i \in [n]$. Let $L = \sum_{i=1}^n L_i$, let $Z = \sum_{i=1}^n Z_i$, let $\mu = \mathrm{E}[L]$ and let $p = \min_{i \in [n]}\{p_i\}$. Finally, let $\Gamma = Z/\mu - 1$. Then for any $\gamma \in [0, 1]$ it holds that*

$$\Pr[|\Gamma| \geq \gamma] \leq 4\exp\left(-\frac{p\mu^2\gamma^2}{5\ell^2 n}\right)$$

Proof. Note that

$$\begin{aligned}\Pr[|\Gamma| \geq \gamma] &= \Pr[|Z - \mu| \geq \mu\gamma] \\ &\leq \Pr[|Z - L| \geq \mu\gamma/2] + \Pr[|L - \mu| \geq \mu\gamma/2] \end{aligned} \tag{3}$$

We bound each term in Equation (3) separately. For the right-hand side term, we use Hoeffding's inequality (Fact 2.21) to get

$$\Pr[|L - \mu| \geq \mu\gamma/2] \leq 2\exp\left(-\frac{2(\mu\gamma/2)^2}{\ell^2 \cdot n}\right) \leq 2\exp\left(-\frac{\mu^2\gamma^2}{\ell^2 n}\right), \tag{4}$$

We now focus on bounding the left-hand side term. The following holds for any fixing of $L_1, \ldots, L_n$. Since $p_i > 0$ for all $i \in [n]$, it holds that $\mathrm{E}[Z_i] = L_i \implies \mathrm{E}[Z] = L$. Moreover, the Z_i's are independent random variables such that $Z_i = b_i \cdot \mathrm{Bern}(p_i)$ for $b_i = L_i/p_i$, where $b = \max\{b_1, \ldots, b_n\} \leq \ell/p$ and $v = \sum_{i=1}^n b_i^2 p_i \leq \ell^2 n/p$. Fact 2.22 yields that

$$\begin{aligned}\Pr[|Z - L| \geq \mu\gamma/2] &\leq 2\exp\left(-\frac{(\mu\gamma/2)^2}{2(v + b\mu\gamma/6)}\right) \leq 2\exp\left(-\frac{\mu^2\gamma^2}{4(\ell^2 n/p + \ell\mu\gamma/6p)}\right) \\ &\leq 2\exp\left(-\frac{p\mu^2\gamma^2}{5\ell^2 n}\right), \end{aligned} \tag{5}$$

where the last inequality holds since $\mu \leq \ell n$ and $\gamma \leq 1$. The proof follows by Equations (3) to (5). □

3 Smooth KL-Divergence

In this section we formally define the notion of smooth KL-divergence, state some basic properties of this measure in Section 3.1, and develop a tool to help bounding it in Section 3.2.

Definition 3.1 (α-*smooth* divergence)**.** *Let P and Q be two distributions over a universe $\mathcal{U}$ and let $\alpha \in [0,1]$. The* α-smooth divergence of P and Q*, denoted $D^{\alpha}(P||Q)$, is defined as $\inf_{(F_P,F_Q)\in\mathcal{F}}\{D(F_P(P)||F_Q(Q))\}$, for $\mathcal{F}$ being the set of randomized functions pairs such that for every $(F_P,F_Q)\in\mathcal{F}$:*

1. $\Pr_{x\sim P}[F_P(x)\neq x]\leq\alpha$*, where the probability is also over the coins of F_P.*
2. $\forall x\in\mathcal{U}$*:* $\mathrm{Supp}(F_P(x))\cap\mathcal{U}\subseteq\{x\}$ *and* $\mathrm{Supp}(F_Q(x))\cap\mathcal{U}\subseteq\{x\}$*.*

Remark 3.2 (comparison to H-Technique)**.** *At least syntactically, the above notion of smooth KL-divergence is similar to the distance measure used by the (coefficients) H-Technique tool, introduced by Patarin [Pat90], for upper-bounding* statistical distance*. Consider the following alternative definition of statistical distance:* $\mathrm{SD}(A,B)=\mathrm{E}_{x\sim A}\max\{0,1-\frac{B(x)}{A(x)}\}$*. The H-Technique approach considers a smooth variant of the above formulation: small events with respect to A are ignored. However, while smooth KL-divergence is useful in settings when the actual KL-divergence might be* unbounded*, as in our settings, the above smooth variant of statistical distance is always very close to the actual statistical distance, and as such, it is more of a tool for bounding statistical distance than a measure of interest for its own sake.*

3.1 Basic Properties

The following proposition (proven in Section 8.1) states that small smooth KL-divergence guarantees that small events with respect to the left-hand-side distribution are also small with respect to the right-hand-side distribution.

Proposition 3.3. *Let P and Q be two distributions over $\mathcal{U}$ with $D^{\alpha}(P||Q)<\beta$. Then for every event E over $\mathcal{U}$, it holds that $Q[E]<2\cdot\max\{\alpha+P[E],4\beta\}$.*

Like any useful distribution measure, smooth KL-divergence posses a data-processing property. The following proposition is proven in Section 8.2.

Proposition 3.4 (Data processing of smooth KL-divergence)**.** *Let P and Q be two distributions over a universe $\mathcal{U}$, let $\alpha\in[0,1]$ and let H be a randomized function over $\mathcal{U}$. Then $D^{\alpha}(H(P)||H(Q))\leq D^{\alpha}(P||Q)$.*

3.2 Bounding Smooth KL-Divergence

The following lemma allow us to bound the smooth KL-divergence between P and Q, while only analyzing simpler variants of Q.

Lemma 3.5 (Bounding smooth KL-Divergence, restatement of Lemma 1.2)**.** *Let P and Q be distributions with P_X and Q_X being over universe $\mathcal{U}^m$, and let $A_1,\ldots,A_m$ and $B_1,\ldots,B_m$ be two sets of events over P and Q respectively. Let $P_{\cdot,XY}$ be an extension of $P=P_{\cdot,X}$ defined*

by $P_{Y|\cdot,X} = \prod_i P_{Y_i|X}$ *for* $P_{Y_i|X} = \mathrm{Bern}(P[A_i \mid X, A_{<i}] \cdot Q[B_i \mid X_{<i}, B_{<i}])$*, letting* $P_{Y_i|X} = 0$ *if* $P[A_{<i} \mid X] = 0$ *or* $Q[B_{<i} \mid X_{<i}] = 0$*, and let* $C_i = \{Y_i = 1\}$*. Then*[5]

$$D^{1-P[C_{\leq m}]}(P_X||Q_X) \leq \sum_{i=1}^{m} D(P_{X_i|A_{\leq i}}||Q_{X_i|B_{\leq i}} \mid P_{X_{<i}|C_{\leq i}}).$$

Proof. Let $Q_{\cdot,XY}$ be an extension of $Q = Q_{\cdot,X}$ defined by $Q_{Y|\cdot,X} = \prod_i Q_{Y_i|X}$ for $Q_{Y_i|X} = \mathrm{Bern}(P[A_i \mid X_{<i}, A_{<i}] \cdot Q[B_i \mid X, B_{<i}])$, letting $Q_{Y_i|X} = 0$ if $P[A_{<i} \mid X_{<i}] = 0$ or $Q[B_{<i} \mid X] = 0$. Our goal is to show that

$$D^{1-P[C_{\leq m}]}(P_{Y_1,X_1,\ldots,Y_m,X_m}||Q_{Y_1,X_1,\ldots,Y_m,X_m}) \leq \sum_{i=1}^{m} D(P_{X_i|A_{\leq i}}||Q_{X_i|B_{\leq i}} \mid P_{X_{<i}|C_{\leq i}}) \tag{6}$$

The proof then follows by data processing of smooth KL-divergence (Proposition 3.4). By definition, for any $i \in [m]$:

$$P_{X_{<i}|Y_{\leq i}=1^i} \equiv P_{X_{<i}|C_{\leq i}} \tag{7}$$

and for any fixing of $x_{<i} \in \mathrm{Supp}(P_{X_{<i}|Y_{\leq i}=1^i})$:

$$P_{X_i|Y_{\leq i}=1^i, X_{<i}=x_{<i}} \equiv P_{X_i|X_{<i},A_{\leq i}} \tag{8}$$

$$Q_{X_i|Y_{\leq i}=1^i, X_{<i}=x_{<i}} \equiv Q_{X_i|X_{<i},B_{\leq i}} \tag{9}$$

and for any fixing of $x_{<i} \in \mathrm{Supp}(P_{X_{<i}|Y_{<i}=1^{i-1}})$:

$$\begin{aligned} &P_{Y_i|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}(1) \\ &\equiv \mathrm{E}_{x \leftarrow P_{X|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}}[P[A_i \mid X = x, A_{<i}] \cdot Q[B_i \mid X_{<i} = x_{<i}, B_{<i}]] \\ &\equiv P[A_i \mid X_{<i} = x_{<i}, A_{<i}] \cdot Q[B_i \mid X_{<i} = x_{<i}, B_{<i}] \\ &\equiv \mathrm{E}_{x \leftarrow Q_{X|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}}[P[A_i \mid X_{<i} = x_{<i}, A_{<i}] \cdot Q[B_i \mid X = x, B_{<i}]] \\ &\equiv Q_{Y_i|Y_{<i}=1^{i-1}, X_{<i}=x_{<i}}(1). \end{aligned} \tag{10}$$

By Equations (7) to (9):

$$\mathrm{E}_{P_{X_{<i}|Y_{\leq i}=1^i}}\left[D\left(P_{X_i|X_{<i},Y_{\leq i}=1^i}||Q_{X_i|X_{<i},Y_{\leq i}=1^i}\right)\right] = \mathrm{E}_{P_{X_{<i}|C_{\leq i}}}\left[D\left(P_{X_i|X_{<i},A_{\leq i}}||Q_{X_i|X_{<i},B_{\leq i}}\right)\right] \tag{11}$$

and by Equation (10), for any fixing of $x \in \mathrm{Supp}(P_{X_{<i}|Y_{<i}=1^{i-1}})$:

$$D\left(P_{Y_i|X_{<i}=x,Y_{<i}=1^{i-1}}||Q_{Y_i|X_{<i}=x,Y_{<i}=1^{i-1}}\right) = 0 \tag{12}$$

We use Equations (11) and (12) for proving Equation (6), by applying on both distributions a function that "cuts" all values after the first appearance of $Y_i = 0$. Let $f_{\mathsf{cut}}(y_1, x_1, \ldots y_m, x_m) = (y_1, x_1, \ldots y_m, x_m)$ if $y = (y_1, \ldots, y_m) = 1^m$, and $f_{\mathsf{cut}}(y_1, x_1, \ldots y_m, x_m) = (y_1, x_1, \ldots y_{i-1}, x_{i-1}, y_i, \perp^{2n-2i+1})$ otherwise, where i is the minimal index with $y_i = 0$, and $\perp$ is an arbitrary symbol $\notin \mathcal{U}$. By definition,

$$\Pr_{s \sim P_{Y_1,X_1,\ldots,Y_m,X_m}}[f_{\mathsf{cut}}(s) \neq s] = P[Y \neq 1^m] = 1 - P[C_{\leq m}],$$

[5]Note that Lemma 1.2 is a special case of Lemma 3.5 that holds when choosing $A_1, \ldots, A_m$ with $P[A_{\leq m}] = 1$.

and by Equations (11) and (12) along with data-processing of standard KL-divergence (Fact 2.3(3)),

$$D(f_{\mathsf{cut}}(P_{Y_1,X_1,\ldots,Y_m,X_m})||f_{\mathsf{cut}}(Q_{Y_1,X_1,\ldots,Y_m,X_m})) \leq \sum_{i=1}^{m} D(P_{X_i|A_{\leq i}}||Q_{X_i|B_{\leq i}} \mid P_{X_{<i}|C_{\leq i}}).$$

That is, f_{cut} is the function realizing the stated bound on the smooth KL-divergence of P_X and Q_X.

□

4 Skewed Distributions

In this section we formally define the notion of many-round skewed distributions and state our main result for such distributions.

Definition 4.1 (The skewed distribution Q)**.** *Let P be a distribution with P_X being a distribution over $m \times n$ matrices, and let W and $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ be events over P. We define the skewed distribution $Q_{X,J} = Q(P, W, \mathcal{E})$ of $\widetilde{P}_X = P|W$, by $Q_J = U_{[n]}$ and*

$$Q_{X|J} = \prod_{i=1}^{m} P_{X_{i,J}|X_{<i,J}} \widetilde{P}_{X_{i,-J}|X_{<i},X_{i,J},E_{i,J}}$$

Definition 4.2 (dense and prefix events)**.** *Let P_X be a distribution over $m \times n$ matrices, and let $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ be an event family over P_X such that $E_{i,j}$, for each i, j, is determined by X^j. The family $\mathcal{E}$ has* density *δ if $\forall(i,j) \in [m] \times [n]$ and for any fixing of $X_{\leq i,j}$, it holds that $P[E_{i,j}|X_{\leq i,j}] = \delta_{i,j} \geq \delta$. The family $\mathcal{E}$ is a* prefix family *if $\forall(i,j) \in [m] \times [n]$ the event $E_{i,j}$ is determined by $X_{\leq i+1,j}$.*

Bounding smooth KL-divergence of smooth distributions. The following theorem states our main result for skewed distributions. In Section 6.1 we give a proof sketch of Theorem 4.3, and in Section 6.2 we give the full details.

Theorem 4.3. *Let P be a distribution with P_X being a distribution over $m \times n$ matrices with independent columns, let W be an event over P and let $\mathcal{E} = \{E_{i,j}\}$ be a δ-dense event family over P_X. Let $\widetilde{P} = P|W$ and let $Q_{X,J} = Q(P, W, \mathcal{E})$ be the skewed variant of $\widetilde{P}$ defined in Definition 4.1. Let $Y_i = (Y_{i,1}, \ldots, Y_{i,n})$ for $Y_{i,j}$ being the indicator for $E_{i,j}$, and let $d = \sum_{i=1}^{m} D(\widetilde{P}_{X_iY_i}||P_{X_iY_i}|\widetilde{P}_{X_{<i}})$. Assuming $n \geq c \cdot m/\delta$ and $d \leq \delta n/c$, for a universal constant $c > 0$, then*

$$D^{\frac{c}{\delta n}(d+1)}(\widetilde{P}||Q) \leq \frac{c}{\delta n}(d+m).$$

We now prove that Theorem 1.3 is an immediate corollary of Theorem 4.3.

Corollary 4.4 (Restatement of Theorem 1.3)**.** *Let $P, \widetilde{P}, Q, W, \mathcal{E}, \delta$ and c be as in Theorem 4.3, and let $\varepsilon = \log(\frac{1}{P[W]})/\delta n$. Then the following hold assuming $n \geq c \cdot m/\delta$:*

- *if $P[W] \geq \exp(-\delta n/cm)$, then $D^{c\cdot(\varepsilon m+1/\delta n)}(\widetilde{P}||Q) \leq c \cdot (\varepsilon m + m/\delta n)$, and*
- *if $P[W] \geq \exp(-\delta n/2c)$ and $\mathcal{E}$ is a prefix family, then $D^{2c\cdot(\varepsilon+1/\delta n)}(\widetilde{P}||Q) \leq 2c \cdot (\varepsilon + m/\delta n)$.*

Proof. Let $\{Y_{i,j}\}$ be as in Theorem 4.3. Note that for each $i \in [m]$:

$$D(\widetilde{P}_{X_iY_i}||P_{X_iY_i} \mid \widetilde{P}_{X_{<i}}) \leq D(\widetilde{P}_{X_{\geq i}}||P_{X_{\geq i}} \mid \widetilde{P}_{X_{<j}}) \leq D(\widetilde{P}_X||P_X) \leq \log \frac{1}{P[W]}.$$

The first inequality holds by data-processing of KL-divergence (Fact 2.3(5)). The second inequality holds by chain-rule of KL-divergence (Fact 2.3(3)). The last inequality holds by Fact 2.4. Assuming $P[W] \geq \exp(-\delta n/cm)$, it holds that

$$d \leq m \cdot \log \frac{1}{P[W]} \leq \delta n/c,$$

concluding the proof of the first part.

Assuming $P[W] \geq \exp(-\delta n/c)$ and $\mathcal{E}$ is a prefix family (i.e., $E_{i,j}$ is a function of $X_{\leq i+1}$), then

$$\begin{aligned}
d &\leq \sum_{i=1}^{m-1} D(\widetilde{P}_{X_iX_{i+1}}||P_{X_iX_{i+1}} \mid \widetilde{P}_{X_{<i}}) + D(\widetilde{P}_{X_m}||P_{X_m} \mid \widetilde{P}_{X_{<m}}) \\
&= \sum_{i \in [m-1] \cap \mathbb{N}_{even}} D(\widetilde{P}_{X_iX_{i+1}}||P_{X_iX_{i+1}} \mid \widetilde{P}_{X_{<i}}) + \sum_{i \in [m-1] \cap \mathbb{N}_{odd}} D(\widetilde{P}_{X_iX_{i+1}}||P_{X_iX_{i+1}} \mid \widetilde{P}_{X_{<i}}) \\
&+ D(\widetilde{P}_{X_m}||P_{X_m} \mid \widetilde{P}_{X_{<m}}) \leq 2 \cdot D(\widetilde{P}_X||P_X) \\
&\leq 2 \cdot \log \frac{1}{P[W]} \leq \delta n/c,
\end{aligned}$$

concluding the proof of the second part. The first inequality holds by data-processing of KL-divergence, and the second one holds by chain-rule and data-processing of KL-divergence. □

In order to show that the attacking distribution Q can be carried out efficiently, it suffice to show that with high probability over $(x, j) \sim Q_{X,J}$, we have for all $i \in [m]$ that $P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}]$ is not much smaller than $P[W]$. The following lemma (proven in Section 6.2) states that the above holds under $\widetilde{P}_X$. Namely, when sampling $x \sim \widetilde{P}_X$ (instead of $x \sim Q_X$) and then $j \sim Q_{J|X=x}$, then $P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}]$ is indeed not too low.

Lemma 4.5. *Let $P, \widetilde{P}, Q, W, \mathcal{E}, \delta, d$ be as in Theorem 4.3, let $t > 0$ and let*

$$p_t := \Pr_{x \sim \widetilde{P}_X \,;\, j \sim Q_{J|X=x}}[\exists i \in [m] : P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] < P[W]/t]$$

Assuming $n \geq c \cdot m/\delta$ and $d \leq \delta n/c$, for a universal constant $c > 0$, then

$$p_t \leq 2m/t + c(d+1)/(\delta n).$$

As an immediate corollary, we get the following result.

Corollary 4.6. *Let $P, \widetilde{P}, Q, W, \mathcal{E}, \delta$ be as in Theorem 4.3, let $\varepsilon = \log(\frac{1}{P[W]})/\delta n$, let $t > 0$ and let c and p_t as in Lemma 4.5. Assuming $n \geq c \cdot m/\delta$, it holds that*

- *if $P[W] \geq \exp(-\delta n/cm)$, then $p_t \leq 2m/t + c \cdot (\varepsilon m + 1/\delta n)$.*
- *if $P[W] \geq \exp(-\delta n/2c)$ and $\mathcal{E}$ is a prefix family, then $p_t \leq 2m/t + 2c \cdot (\varepsilon + 1/\delta n)$.*

5 The Parallel Repetition Theorem

In this section, we use Theorem 4.3 for prove Theorem 1.4, restated below.

Theorem 5.1 (Parallel repetition for partially simulatable arguments, restatement of Theorem 1.4)**.** *Let π be an m-round δ-simulatable [resp., prefix δ-simulatable] interactive argument of soundness error $1-\varepsilon$. Then π^n has soundness error $(1-\varepsilon)^{cn\delta/m}$ [resp., $(1-\varepsilon)^{cn\delta}$], for a universal constant $c>0$.*

Since the random terminating variant of an m-round interactive argument is $1/m$-prefix-simulatable, the (tight) result for such protocols immediately follows. The proof of Theorem 5.1 follows from our bound on the smooth KL-divergence of skewed distributions, Theorem 4.3, and Lemma 5.3, stated and proven below.

Definition 5.2 (bounding function for many-round skewing)**.** *A function f is a* bounding function for many-round skewing *if there exists a polynomial $p(\cdot,\cdot)$ such that the following holds for every $\delta\in(0,1]$ and every $m,n\in\mathbb{N}$ with $n>p(m,1/\delta)$: let P be a distribution with P_X being a column independent distribution over $m\times n$ matrices. Let W be an event and let $\mathcal{E}$ be a δ-dense [resp., prefix δ-dense] event family over P (see Definition 4.2). Let $\widetilde{P}=P|W$ and let $Q=Q(P,W,\mathcal{E})$ be according to Definition 4.1. Then the following holds for $\gamma=\log(1/P[W])/f(n,m,\delta)$:*

1. *$Q_X[T]\le 2\cdot\widetilde{P}_X[T]+\gamma$ for every event T,*[6] *and*
2. *$\Pr_{x\sim\widetilde{P}_X\,;\,j\sim Q_{J|X=x}}[(x,j)\in\mathsf{Bad}_t] \quad\le\quad p(m,1/\delta)/t+\gamma$ for every $t>0$, letting*

$$\mathsf{Bad}_t:=\{(x,j)\colon\exists i\in[m]:P[W\mid(X_{<i},X_{i,j})=(x_{<i},x_{i,j}),E_{i,j}]<P[W]/t\}.$$

Lemma 5.3 (Restatement of Lemma 1.1)**.** *Let π be an m-round δ-simulatable [resp., prefix δ-simulatable] interactive argument of soundness error $1-\varepsilon$, let f be a bounding function for many-round skewing (according to Definition 5.2). Then π^n has soundness error $(1-\varepsilon)^{f(n,m,\delta)/160}$.*

That is, Lemma 5.3 tells us that the task of maximizing the decreasing rate of π^n is directly reduces to the task of maximizing a bounding function for many-round skewing. A larger bounding function yields a smaller γ in Definition 5.2. This γ both defines an additive bound on the difference between a small event in $\widetilde{P}$ to a small event in Q, and bounds a specific event in $\widetilde{P}$ that captures the cases in which an attack can be performed efficiently.

We first prove Theorem 5.1 using Lemma 5.3.

Proof of Theorem 5.1.

Proof. We prove for δ-simulatable arguments, the proof for δ-prefix-simulatable arguments follows accordingly. Let m,n, P, δ, $\mathcal{E}$, W, $\widetilde{P}$ and Q be as in Lemma 5.3, where $\mathcal{E}$ is δ-dense, and let $c=\max\{c',c''\}$ where c' is the constant from Corollary 4.4 and c'' is the constant from Corollary 4.6. By Corollary 4.4, if $n\ge c\cdot m/\delta$ and $P[W]\ge\exp(-\delta n/cm)$, then

$$D^{3cm\mu}(\widetilde{P}||Q)\le 3cm\mu \tag{13}$$

[6] The constant 2 can be replaced with any other constant without changing (up to a constant factor) the decreasing rate which is promised by Lemma 5.3.

for $\mu = \log(1/P[W])/\delta n$, where we assumed without loss of generality that $P[W] \leq 1/2$. Hence, assuming that $n \geq c \cdot m/\delta$ and $P[W] \geq \exp(-\delta n/cm)$, Proposition 3.3 and Equation (13) yields that for every event T:

$$Q[T] \leq 2 \cdot \widetilde{P}[T] + \gamma, \tag{14}$$

where $\gamma = \log(1/P[W])/f(n, m, \delta)$ for $f(n, m, \delta) = \delta n/(24cm)$. For event W of smaller probability, it holds that $\gamma \geq 24$, and therefore Equation (14) trivially holds for such events. In addition, by Corollary 4.6, if $n \geq c \cdot m/\delta$ and $P[W] \geq \exp(-\delta n/cm)$, then

$$\Pr_{x \sim \widetilde{P}_X \,;\, j \sim Q_{J|X=x}}[\exists i \in [m] : P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] < P[W]/t] \leq 2m/t + \gamma, \tag{15}$$

where for event W of smaller probability, Equation (15) trivially holds. By Equations (14) and (15), f is a bounding function for many-round skewing with the polynomial $p(m, 1/\delta) = c \cdot m/\delta$. Therefore, Lemma 5.3 yields that the soundness error of π^n is bounded by $(1-\varepsilon)^{f(n,m,\delta)/80} = (1-\varepsilon)^{\delta n/(c'm)}$, for $c' = 1920c$. □

5.1 Proving Lemma 5.3

Let f be a bounding function for many-round skewing with the polynomial $p(\cdot, \cdot) \in \text{poly}$. We first prove the case when the number of repetition n is at least $p(m, 1/\delta)$, and then show how to extend the proof for the general case.

Many repetitions case.

Proof of Lemma 5.3, many repeitions. Fix an m-round δ-simulatable interactive argument $\pi = (\mathrm{P}, \mathrm{V})$ of soundness error $1-\varepsilon$ (the proof of the δ-prefix-simulatable case follows the same lines), and let $n = n(\kappa) > p(m(\kappa), 1/\delta(\kappa))$. Note that without loss of generality $\varepsilon(\kappa) \geq 1/\text{poly}(\kappa)$.

Our proof is a black-box reduction according to Definition 2.15: we present an oracle-aided algorithm that given access to a deterministic cheating prover for π^n violating the claimed soundness of π^n, uses it to break the assumed soundness of π while not running for too long. The lemma then follows by Proposition 2.16.

Let S be the oracle-aided simulator guaranteed by the δ-simulatablily of V. For a cheating prover P^{n*} for π^n, let P^* be the cheating prover that for interacting with V, emulates a random execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$, letting V plays one of the n verifiers, at a random location. (Clearly, P^* only requires oracle access to P^{n*}.) Assume without loss of generality that in each round V flips $t = t(\kappa)$ coins. The oracle-aided algorithm P^* is defined as follows.

Algorithm 5.4 (P^*)**.**

Input: 1^κ, $m = m(\kappa)$ *and* $n = n(\kappa)$.

Oracles: cheating prover P^{n*} *for* π^n.

Operation:

1. *Let* $j \leftarrow [n]$.
2. *For* $i = 1$ *to* m *do:*
   (a) *Let* a_i *be the* i^{th} *message sent by* V.

(b) Do the following ("rejection continuation"):

i. Let $x_{i,-j} \leftarrow (\{0,1\}^t)^{n-1}$

ii. Let $v = \mathsf{S}^{\mathrm{P}^{n*}}(1^\kappa, (j, x_{\leq i,-j}, a_{\leq i}))$.

iii. If all n *verifiers accept in* v*, break the inner loop.*

(c) Send to V *the* i^{th} *message* P^{n*} *sends in* v.

Fix a cheating prover P^{n*}. We also fix $\kappa \in \mathbb{N}$, and omit it from the notation. Let $P = P_X$ denotes the coins V^n use in a uniform execution of $(\mathrm{P}^{n*}, \mathrm{V}^n)$. (Hence P_X is uniformly distributed over $m \times n$ matrices.) Let W be the event over P that P^{n*} wins in $(\mathrm{P}^{n*}, \mathrm{V}^n)$ (i.e., all verifiers accept), and let $\widetilde{P}_X = P_X|W$. For an i rounds view $v = (j, \cdot)$ of P^{n*} in $(\mathrm{P}^{n*}, \mathrm{V})$, let Δ_v be the δ-dense subset of V's coins describing the output distribution of $\mathsf{S}^{\mathrm{P}^{n*}}(v)$. Let $\mathcal{T}_{i,j}$ be all possible i round views of P^{n*} in $(\mathrm{P}^{n*}, \mathrm{V})$ that are starting with j. Finally, let $\mathcal{E} = \{E_{i,j}\}_{i\in[m],j\in[n]}$ be the event family over P defined by $E_{i,j} = \bigcup_{v\in\mathcal{T}_{i,j}} \Delta_v$, and let $Q_{X,J}$ be the e (skewed) distribution described in Definition 4.1 with respect to $P, W, \mathcal{E}$. By inspection, Q describes the distribution of $(j, x_{\leq m})$ in a random execution of $(\mathrm{P}^*, \mathrm{V}^n)$, where $x_{\leq m,j}$ denotes the coins of V, and $x_{\leq m,-j}$ denote the final value of this term in the execution. Assume

$$\Pr[(\mathrm{P}^{n*}, \mathrm{V}^n) = 1] = P[W] > (1-\varepsilon)^{f(n,m,\delta)/80}, \tag{16}$$

and let $\gamma = \log(1/P[W])/f(n, m, \delta)$. By Equation (16) it holds that

$$\gamma < -\log(1-\varepsilon)/80 \leq \varepsilon/80 \tag{17}$$

Since $\widetilde{P}[W] = 1$, we deduce by Property 5.2(1) of f on the event $\neg W$ that

$$\Pr[(\mathrm{P}^*, \mathrm{V}) = 1] \geq Q_X[W] > 1 - \gamma > 1 - \varepsilon/80 \tag{18}$$

So it is left to argue about the running time of P^*. By Property 5.2(2) of f on $t = 80 \cdot p(m, 1/\delta)/\varepsilon$ it holds that

$$\Pr_{x\sim\widetilde{P}_X\ ;\ j\sim Q_{J|X=x}}[(x,j) \in \mathsf{Bad}_t] \quad \leq \quad p(m, 1/\delta)/t + \gamma < \varepsilon/40$$

Therefore, we now can apply Property 5.2(1) of f on the following event "Given x, choose $j \sim Q_{J|X=x}$ and check whether $(x, j) \in \mathsf{Bad}_t$" (note that this event defined over an extension of $\widetilde{P}$ that additionally samples j according to $Q_{J|X}$). This yields that

$$\Pr_{x\sim Q_X\ ;\ j\sim Q_{J|X=x}}[(x,j) \in \mathsf{Bad}_t] \quad \leq \quad 2\varepsilon/40 + \gamma < \varepsilon/10 \tag{19}$$

By Equations (18) and (19) we obtain that

$$\Pr_{(x,j)\sim Q_{X,J}}[W \wedge ((x,j) \notin \mathsf{Bad}_t)] > 1 - \varepsilon/5 \tag{20}$$

Namely, with probability larger than $1 - \varepsilon/5$, the attacker P^* wins and its expected running time in each round is bounded by $O(t/P[W]) \leq \mathrm{poly}(\kappa)$. This contradicts the soundness guaranty of π. □

Any number of repetitions. The assertions of the function f in Equations (18) and (19) only guarantee to hold if $n > p(m, 1/\delta)$ (for some $p(\cdot,\cdot) \in$ poly). We now prove the lemma for smaller values of repetitions. As mentioned in the introduction, for interactive arguments (and unlike interactive proofs), there is no generic reduction from large to small number of repetitions. Assume

$$\alpha := \Pr[(\mathrm{P}^{n*}, \mathrm{V}^n) = 1] > (1-\varepsilon)^{f(m,n,\delta)/80} \tag{21}$$

and let $\ell \in$ poly be such that $\ell n \geq p(m, 1/\delta)$. It is immediate that

$$\alpha_\ell := \Pr\Big[((\mathrm{P}^{n*})^\ell, (\mathrm{V}^n)^\ell) = 1\Big] = \alpha^\ell > (1-\varepsilon)^{\ell \cdot f(m,n,\delta)/80} \tag{22}$$

for $(\mathrm{P}^{n*})^\ell$ and $(\mathrm{V}^n)^\ell$ being the ℓ repetition of P^{n*} and V^n respectively. Therefore, the same lines as the proof above yields that the cheating prover $\mathrm{P}^{*(\mathrm{P}^{n*})^\ell}$ breaks the soundness of π with probability $1-\varepsilon/80$. The problem is that the running time of $\mathrm{P}^{*(\mathrm{P}^{n*})^\ell}$ is proportional to $1/\alpha_\ell$ and not to $1/\alpha$, and in particular is not polynomial even if $\alpha > 1/\mathrm{poly}$. We overcome this difficulty by giving a different (efficient) implementation of $\mathrm{P}^{*(\mathrm{P}^{n*})^\ell}$ that takes advantage of the parallel nature of $(\mathrm{P}^{n*})^\ell$.

Proof of Lemma 5.3, small number of repetitions. Let π, P^{n*}, P^* and S be as in the proof for the many repetitions case. Let $\ell \in$ poly be such that $\ell n \geq p(m, 1/\delta)$, and for $q \in [\ell]$ let $\mathcal{Z}^q = \{(q-1)n+1, \ldots, qn\}$. The oracle-aided algorithm $\widehat{\mathrm{P}}$ is defined as follows.

Algorithm 5.5 ($\widehat{\mathrm{P}}$)**.**

Input: 1^κ, $m = m(\kappa)$, $n = n(\kappa)$ *and* $\ell = \ell(\kappa)$.

Oracles: cheating prover P^{n*} *for* π^n.

Operation:

1. *Let* $j \leftarrow [n\ell]$.
2. *For* $i = 1$ *to* m *do:*
   (a) *Let* a_i *be the* i^{th} *message sent by* V.
   (b) *For* $q = 1$ *to* ℓ *do the following ("rejection continuation"):*
   If $j \in \mathcal{Z}^q$*:*
      i. *Let* $x_{i,\mathcal{Z}^q \setminus \{j\}} \leftarrow (\{0,1\}^t)^{n-1}$.
      ii. *Let* $v = \mathsf{S}^{\mathrm{P}^{n*}}(1^\kappa, (j \bmod n, x_{i,\mathcal{Z}^q\setminus\{j\}}, a_{\leq i}))$.
      iii. *If all* n *verifiers accept in* v*, break the inner loop.*
   Else,
      i. *Let* $x_{>i,\mathcal{Z}^q} \leftarrow (\{0,1\}^t)^n$
      ii. *If all* n *verifiers accept in* $x_{\mathcal{Z}^q}$*, break the inner loop.*
   (c) *Send* V *the* i^{th} *message* P^{n*} *sends in* v.

Namely, $\widehat{\mathrm{P}}^{\mathrm{P}^{n*}}$ emulates $\mathrm{P}^{*(\mathrm{P}^{n*})^\ell}$, for $(\mathrm{P}^{n*})^\ell$ being the ℓ parallel repetition of P^{n*}, while exploiting the product nature of $(\mathrm{P}^{n*})^\ell$ for separately sampling the coins of each the ℓ groups of verifiers.

Fix a cheating prover P^{n*} and $\kappa \in \mathbb{N}$, and define $P = P_X$, W, $Q_{X,J}$ with respect to a random execution of $((\mathrm{P}^{n*})^\ell, (\mathrm{V}^n)^\ell)$ as done in the proof for large number of repetition. Assume

$$\Pr[(\mathrm{P}^{n*}, \mathrm{V}^n) = 1] > (1-\varepsilon)^{f(m,n,\delta)/80} \tag{23}$$

then

$$P[W] = \Pr\Big[((\mathrm{P}^{n*})^\ell, (\mathrm{V}^n)^\ell) = 1\Big] > (1-\varepsilon)^{\ell \cdot f(m,n,\delta)/80} \tag{24}$$

Equation (18) yield that

$$\Pr\Big[(\widehat{\mathrm{P}}, \mathrm{V}) = 1\Big] > 1 - \varepsilon/80 \tag{25}$$

So it is left to argue about the running time $\widehat{\mathrm{P}}$. For $q \in [\ell]$, let W_q be the event that all verifiers in $\mathcal{Z}^q$ accept in P_X. Note that $P[W_q] = \Pr[(\mathrm{P}^{n*}, \mathrm{V}^n) = 1] = \alpha$ and that $P[W] = \alpha^\ell$. Moreover, For $j \in [n\ell]$, let q_j be the (unique) value $q \in [\ell]$ such that $j \in \mathcal{Z}^q$. By Equation (19) it holds that

$$\Pr_{(j,x)\sim Q_{J,X}}[(x,j) \in \mathsf{Bad}_t] \quad < \varepsilon/10 \tag{26}$$

for $t = 80 \cdot p(m, 1/\delta)/\varepsilon$, where recall that

$$\mathsf{Bad}_t = \{(x,j) \colon \exists i \in [m] : P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] < P[W]/t\}.$$

Note that by construction, it holds that

$$\begin{aligned} &P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] \\ &= P[W_{q_j} \mid X_{<i,\mathcal{Z}^{q_j}} = x_{<i,\mathcal{Z}^{q_j}}, X_{i,j} = x_{i,j}, E_{i,j}] \cdot \prod_{q \in [\ell]\setminus\{q_j\}} P[W_q \mid X_{<i,\mathcal{Z}^q} = x_{<i,\mathcal{Z}^q}]. \end{aligned} \tag{27}$$

Moreover, by Markov inequality we have

$$\Pr_{(j,x)\sim Q_{J,X}}\left[\prod_{q \in [\ell]\setminus\{q_j\}} P[W_q \mid X_{<i,\mathcal{Z}^q} = x_{<i,\mathcal{Z}^q}] > 10 \cdot \alpha^{\ell-1}/\varepsilon\right] < \varepsilon/10. \tag{28}$$

Recall that $P[W] = \alpha^\ell$. Therefore, by Equations (26) to (28) we deduce that

$$\Pr_{(j,x)\sim Q_{J,X}}\left[\exists i \in [m] : P[W_{q_j} \mid X_{<i,\mathcal{Z}^{q_j}} = x_{<i,\mathcal{Z}^{q_j}}, X_{i,j} = x_{i,j}, E_{i,j}] < \varepsilon\alpha/(10t)\right] \quad < \varepsilon/5 \tag{29}$$

Moreover, by Fact 2.20 along with Markov inequality and a union bound, we have

$$\Pr_{(j,x)\sim Q_{J,X}}[\exists (i,q) \in [m] \times ([\ell] \setminus \{q_j\}) : P[W_q \mid X_{<i,\mathcal{Z}^q} = x_{<i,\mathcal{Z}^q}] < \varepsilon\alpha/(5m)] \quad < \varepsilon/5 \tag{30}$$

Hence, Equations (29) and (30) yields that with probability $> 1-\varepsilon/2$ it holds that at the beginning of each inner round of $\widehat{\mathrm{P}}$, the expected running time of it is bounded by $\max\{10t/(\varepsilon\alpha), 5m/(\varepsilon\alpha)\} \leq \mathrm{poly}(\kappa)$. This (along with Equation (25)) contradicts the soundness guarantee of π.

□

6 Bounding Smooth KL-Divergence of Skewed Distributions

In this section we prove Theorem 4.3. As a warmup, we give in Section 6.1 a proof sketch and explain the difficulties that arise. In Section 6.2 we define conditional variants of $\widetilde{P}$ and Q, and use Lemma 3.5 to prove the theorem assuming that (1) the standard KL-divergence of these variants is small and (2) these variants are not too far, in the sense that allows us to use Lemma 3.5, from their origin. We prove (1) in Section 6.3, and prove (2), which is the most challenging part, in Section 6.4.

In the following we fix distribution P with P_X being a distribution over $\mathcal{U}^{m\times n}$ matrices with independent columns, event W over P and δ-dense event family $\mathcal{E} = \{E_{i,j}\}$ over P_X. We let $\widetilde{P} = P|W$ and let $Q_{X,J} = Q(P, W, \mathcal{E})$ be the skewed variant of $\widetilde{P}$ defined in Definition 4.1. Let $Y_i = (Y_{i,1}, \ldots, Y_{i,n})$ for $Y_{i,j}$ be the indicator for $E_{i,j}$, and let $d = \sum_{i=1}^{m} D(\widetilde{P}_{X_iY_i}||P_{X_iY_i}|\widetilde{P}_{X_{<i}})$.

6.1 Warmup

In this section we give a rather detailed proof sketch (more accurately, an attempt proof sketch) for Theorem 4.3. Specifically, we try to bound the divergence between $\widetilde{P}$ and Q; That is, to show that

$$D(\widetilde{P}||Q) \leq O\left(\frac{1}{\delta n}\right) \cdot (d+m) \tag{31}$$

We try to do so by showing that for every $i \in [m]$ it holds that

$$D(\widetilde{P}_{X_i}||Q_{X_i}|\widetilde{P}_{X_{<i}}) \leq O\left(\frac{1}{\delta n}\right) \cdot (d_i+1) \tag{32}$$

for $d_i = D(\widetilde{P}_{X_iY_i}||P_{X_iY_i}|\widetilde{P}_{X_{<i}})$, and applying chain-rule of KL-divergence for deducing Equation (31). By data-processing of KL-divergence (Fact 2.3(5)), it holds that

$$D(\widetilde{P}_{X_i}||Q_{X_i}|\widetilde{P}_{X_{<i}}) \leq D(\widetilde{P}_{X_iY_i}||Q'_{X_iY_i}|\widetilde{P}_{X_{<i}}), \tag{33}$$

where

$$Q'_{X_iY_i|X_{<i}} = \widetilde{P}_{X_iY_i|X_{<i},X_{i,J},Y_{i,J}=1} \circ Q_{J,X_{i,J}|X_{<i}} \equiv P_{X_{i,J}|X_{<i}} \widetilde{P}_{X_iY_i|X_{<i},X_{i,J},Y_{i,J}=1} \circ Q_{J|X_{<i}}$$

(note that $Q'_{X_i} \equiv Q_{X_i}$ and that $P_{X_{i,J}|X_{<i}} \equiv P_{X_{i,J}|X_{<i,J}}$ because the columns under P are independent). By definition of Q', for any fixing of $x_{\leq i}y_i \in \mathrm{Supp}(\widetilde{P}_{X_{\leq i}Y_i})$ it holds that

$$\begin{aligned}
Q'_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i) &= \mathrm{E}_{j\sim Q_{J|X_{<i}=x_{<i}}}\left[P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j}) \cdot \widetilde{P}_{X_iY_i|X_{<i}=x_{<i},X_{i,j}=x_{i,j},Y_{i,j}=1}(x_iy_i)\right] \qquad (34)\\
&= \sum_{j=1}^{n} Q_{J|X_{<i}=x_{<i}}(j) \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j}) \cdot \frac{\widetilde{P}_{X_iY_iX_{i,j}Y_{i,j}|X_{<i}=x_{<i}}(x_iy_ix_{i,j}1)}{\widetilde{P}_{X_{i,j},Y_{i,j}|X_{<i}=x_{<i}}(x_{i,j},1)}\\
&= \sum_{j\in 1_{y_i}} Q_{J|X_{<i}=x_{<i}}(j) \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j}) \cdot \frac{\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)}{\widetilde{P}_{X_{i,j},Y_{i,j}|X_{<i}=x_{<i}}(x_{i,j},1)}\\
&= \sum_{j\in 1_{y_i}} Q_{J|X_{<i}=x_{<i}}(j) \cdot \frac{\beta_{i,j}(x_{i,j}) \cdot \widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)}{\tilde{\delta}_{i,j}},
\end{aligned}$$

for $\beta_{i,j}(x_{i,j}) = \beta_{i,j}(x_{i,j}; x_{<i}) = \frac{P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})}{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}(x_{i,j})}$ and $\tilde{\delta}_{i,j} = \tilde{\delta}_{i,j}(x_{<i}) = \widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}}(1)$ $(= \widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}])$, where recall that we denote $1_{y_i} = \{j \in [n] \colon y_{i,j} = 1\}$. In addition, note that

$$\begin{aligned} Q_{J|X_{<i}=x_{<i}}(j) &= \frac{Q[X_{<i} = x_{<i} \mid J = j] \cdot Q[J = j]}{Q[X_{<i} = x_{<i}]} = \frac{Q[X_{<i} = x_{<i} \mid J = j] \cdot Q[J = j]}{\sum_{t=1}^{n} Q[J = t] Q[X_{<i} = x_{<i} \mid J = t]} \\ &= 1/\left(\sum_{t=1}^{n} \frac{Q[X_{<i} = x_{<i} \mid J = t]}{Q[X_{<i} = x_{<i} \mid J = j]}\right). \end{aligned} \tag{35}$$

Since for all $t \in [n]$ it holds that

$$\begin{aligned} &Q[X_{<i} = x_{<i} \mid J = t] = \prod_{s=1}^{i-1} P[X_{s,t} = x_{s,t} \mid X_{<s} = x_{<s}] \cdot \widetilde{P}[X_s = x_s \mid X_{<s} = x_{<s}, X_{s,t} = x_{s,t}, E_{s,t}] \\ &= \prod_{s=1}^{i-1} P[X_{s,t} = x_{s,t} \mid X_{<s} = x_{<s}] \cdot \frac{\widetilde{P}[X_{s,t} = x_{s,t}, E_{s,t} \mid X_{\leq s} = x_{\leq s}] \cdot \widetilde{P}[X_s = x_s \mid X_{<s} = x_{<s}]}{\widetilde{P}[E_{s,t} \mid X_{<s} = x_{<s}] \cdot \widetilde{P}[X_{s,t} = x_{s,t} \mid X_{<s} = x_{<s}, E_{s,t}]} \\ &= \prod_{s=1}^{i-1} \frac{P[X_{s,t} = x_{s,t} \mid X_{<s} = x_{<s}]}{\widetilde{P}[X_{s,t} = x_{s,t} \mid X_{<s} = x_{<s}, E_{s,t}]} \cdot \frac{\widetilde{P}[E_{s,t} \mid X_{\leq s} = x_{\leq s}]}{\widetilde{P}[E_{s,t} \mid X_{<s} = x_{<s}]} \cdot \widetilde{P}[X_s = x_s \mid X_{<s} = x_{<s}] \end{aligned} \tag{36}$$

we deduce from Equations (35) and (36) that

$$Q_{J|X_{<i}=x_{<i}}(j) = \frac{\omega_{i,j}}{\sum_{t=1}^{n} \omega_{i,t}}, \tag{37}$$

where

$$\begin{aligned} \omega_{i,j} &= \omega_{i,j}(x_{<i}) \\ &= \frac{n}{\sum_{t=1}^{n} \omega'_{i,t}} \cdot \prod_{s=1}^{i-1} \frac{P[X_{s,j} = x_{s,j} \mid X_{<s} = x_{<s}]}{\widetilde{P}[X_{s,j} = x_{s,j} \mid X_{<s} = x_{<s}, E_{s,j}]} \cdot \frac{\widetilde{P}[E_{s,j} \mid X_{\leq s} = x_{\leq s}]}{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}]} \\ &= \frac{n}{\sum_{t=1}^{n} \omega'_{i,t}} \cdot \prod_{s=1}^{i-1} \frac{P[X_{s,j} = x_{s,j} \mid X_{<s} = x_{<s}]}{\widetilde{P}[X_{s,j} = x_{s,j} \mid X_{<s} = x_{<s}]} \cdot \frac{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}]}{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}, X_{s,j} = x_{s,j}]} \cdot \frac{\widetilde{P}[E_{s,j} \mid X_{\leq s} = x_{\leq s}]}{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}]} \\ &= \frac{n \cdot \omega'_{i,j}}{\sum_{t=1}^{n} \omega'_{i,t}} \cdot \prod_{s=1}^{i-1} \frac{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}]}{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}, X_{s,j} = x_{s,j}]} \cdot \frac{\widetilde{P}[E_{s,j} \mid X_{\leq s} = x_{\leq s}]}{\widetilde{P}[E_{s,j} \mid X_{<s} = x_{<s}]} \end{aligned}$$

for $\omega'_{i,j} = \omega'_{i,j}(x_{<i}) = \prod_{s=1}^{i-1} \frac{P[X_{s,j}=x_{s,j}|X_{<s}=x_{<s}]}{\widetilde{P}[X_{s,j}=x_{s,j}|X_{<s}=x_{<s}]}$. Note that $\omega_{i,j}$ is basically a relative "weight" for the column j, where a large $\omega_{i,j}$ with respect to the other $\omega_{i,t}$'s means that $Q_{J|X_{<i}=x_{<i}}(j)$ is higher. In an extreme case it is possible that $\omega_{i,j} = \infty$, meaning that $Q_{J|X_{<i}=x_{<i}}(j) = 1$. However, we assume for now that all $\omega_{i,j} < \infty$. Later in this proof attempt we even assume that all the terms are close to 1, meaning that $Q_{J|X_{<i}=x_{<i}}$ has high min entropy (assumptions that are eliminated in Section 6.2). As a side note, observe that $\omega_{1,j} = 1$ for all $j \in [n]$ (meaning that Q_J is the uniform distribution over $[n]$). At this point, we just mention that we added (the same) multiplicative factor

of $\frac{n}{\sum_{t=1}^{n}\omega'_{i,t}}$ to all $\{\omega_{i,j}\}_{j=1}^{n}$. On the one hand this does not change the relative weight, but on the other hand it will help us to claim in the coming sections that these $\omega_{i,j}$'s are indeed close to 1. By Equations (33), (34) and (37), it holds that

$$\begin{aligned} D(\widetilde{P}_{X_i}||Q_{X_i}|\widetilde{P}_{X_{<i}}) &\leq D(\widetilde{P}_{X_iY_i}||Q'_{X_iY_i}|\widetilde{P}_{X_{<i}}) \\ &= \mathrm{E}_{x_{<i}\sim X_{<i}}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\log\frac{\widetilde{P}_{X_iY_i|X_{<i}=X_{<i}}(x_iy_i)}{Q_{X_iY_i|X_{<i}=X_{<i}}(x_iy_i)}\right] \\ &= \mathrm{E}_{x_{<i}\sim X_{<i}}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\log\frac{\sum_{j=1}^{n}\omega_{i,j}}{\sum_{j\in 1_{y_i}}\frac{\omega_{i,j}\cdot\beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}}\right] \\ &= \mathrm{E}_{x_{<i}\sim X_{<i}}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}[-\log(1+\gamma_i(x_iy_i))], \end{aligned} \tag{38}$$

for

$$\gamma_i(x_iy_i) = \gamma_i(x_iy_i;x_{<i}) = \left(\sum_{j\in 1_{y_i}}\frac{\omega_{i,j}\cdot\beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right)/\left(\sum_{j=1}^{n}\omega_{i,j}\right) - 1 \tag{39}$$

Naturally, we would like to approximate the logarithm in the above equation with a low-degree polynomial. However, we can only do if γ_i is far away from -1. In particular, if $\widetilde{P}[\gamma_i(X_iY_i;X_{<i}) = -1] > 0$ (which happens if the event W allows for none of the events $\{E_{i,j}\}_{i=1}^{n}$ to occur), the above expectation is unbounded. At that point, we only show how to bound Equation (38) under simplifying assumptions, while in Section 6.2 we present how to eliminate the assumptions via smooth KL-divergence. We now assume that for any $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{X_{<i}})$ and any $j \in [n]$, the following holds:

Assumption 6.1.

1. $|\gamma_i(x_iy_i)| \leq 1/2$ *for any* $x_iy_i \in \mathrm{Supp}(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}})$.
2. $\tilde{\delta}_{i,j} \geq 0.9\delta_{i,j}$ *(recall that* $\delta_{i,j} = P[E_{i,j}] = P[E_{i,j} \mid X_{\leq i}]$ *for any fixing of* $X_{\leq i}$*).*
3. $\omega_{i,j} \in 1 \pm 0.1$.
4. $\mathrm{Supp}(P_{X_{i,j}|X_{<i}=x_{<i}}) \subseteq \mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1})$.
5. $\beta_{i,j}(x_{i,j}) \leq 1.1$ *for any* $x_{i,j} \in \mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i}})$.

Note that Assumption 3 implies that $Q_{J|X_{<i}}$ has high min-entropy, and Assumptions 2 along with 5 imply that for all j:

$$\begin{aligned} &P[W \mid (X_{<i},X_{i,j}) = (x_{<i},x_{i,j}), E_{i,j}] \\ &= \frac{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j})}{P_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j})}\cdot\frac{\widetilde{P}[E_{i,j} \mid X_{<i}=x_{<i}]}{P[E_{i,j} \mid X_{<i}=x_{<i}]}\cdot P[W \mid X_{<i}=x_{<i}] \\ &= \beta_{i,j}(x_{i,j})\cdot\left(\tilde{\delta}_{i,j}/\delta_{i,j}\right)\cdot P[W \mid X_{<i}=x_{<i}] \geq P[W \mid X_{<i}=x_{<i}]/2, \end{aligned}$$

which fits the explanation in Section 1.4.1 (note that in the second equality we used the fact that $P_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j}) = P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})$ by assumption).
By Equation (38), note that in order to prove Equation (32), it is enough to show that for any $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{x_{<i}})$ it holds that

$$\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}[-\log(1+\gamma_i(x_i y_i))] \leq O\left(\frac{1}{\delta n}\right) \cdot \left(D(\widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}} || P_{X_i Y_i | X_{<i}=x_{<i}}) + 1\right) \tag{40}$$

In the following, fix $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{x_{<i}})$. We now focus on proving Equation (40). Using the inequality $-\log(1+x) \leq -x + x^2$ for $|x| \leq \frac{1}{2}$, we deduce from Assumption 1 that

$$\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}[-\log(1+\gamma_i(x_i y_i))] \leq \mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}\left[-\gamma_i(x_i y_i) + \gamma_i(x_i y_i)^2\right] \tag{41}$$

Note that

$$\begin{aligned}
&\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}\left[\sum_{j \in 1_{y_i}} \frac{\omega_{i,j} \cdot \beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right] = \sum_{j=1}^{n} \mathrm{E}_{x_{i,j} y_{i,j} \sim \widetilde{P}_{X_{i,j} Y_{i,j} | X_{<i}=x_{<i}}}\left[y_{i,j} \cdot \frac{\omega_{i,j} \cdot \beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right] \qquad (42)\\
&= \sum_{j=1}^{n} \omega_{i,j} \cdot \mathrm{E}_{x_{i,j} \sim \widetilde{P}_{X_{i,j} | X_{<i}=x_{<i}, Y_{i,j}=1}}[\beta_{i,j}(x_{i,j})] = \sum_{j=1}^{n} \omega_{i,j} \cdot \mathrm{E}_{x_{i,j} \sim \widetilde{P}_{X_{i,j} | X_{<i}=x_{<i}, Y_{i,j}=1}}\left[\frac{P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})}{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}(x_{i,j})}\right]\\
&= \sum_{j=1}^{n} \omega_{i,j} \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1})) = \sum_{j=1}^{n} \omega_{i,j}.
\end{aligned}$$

The second equality holds since $y_{i,j} \in \{0,1\}$ and since Assumption 2 implies that $\widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}}(1) = \tilde{\delta}_{i,j} > 0$ for all $j \in [n]$, and the last equality holds by Assumption 4. Therefore, we deduce from Equation (42) that

$$\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}[\gamma_i(x_i y_i)] = \left(\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}\left[\sum_{j \in 1_{y_i}} \frac{\omega_{i,j} \cdot \beta_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right]\right) / \left(\sum_{j=1}^{n} \omega_{i,j}\right) - 1 = 0. \tag{43}$$

Hence, in order to prove Equation (40), we deduce from Equations (41) and (43) that it is left to prove that

$$\mathrm{E}_{x_i y_i \sim \widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}}\left[\gamma_i(x_i y_i)^2\right] \leq O\left(\frac{1}{\delta n}\right) \cdot \left(D(\widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}} || P_{X_i Y_i | X_{<i}=x_{<i}}) + 1\right) \tag{44}$$

In the following, rather than directly bounding the expected value of $\gamma_i(x_i y_i)^2$ under $\widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}$, we show that under the product of the marginals of $\widetilde{P}_{X_i Y_i | X_{<i}=x_{<i}}$ (namely, under the distribution $\prod_{j=1}^{n} \widetilde{P}_{X_{i,j} Y_{i,j} | X_{<i}=x_{<i}}$), the value of $\gamma_i(x_i y_i)$ is well concentrated around its mean (i.e., zero), and the proof will follow by Proposition 2.9. More formally, let Γ be the value of $\gamma_i(x_i y_i)$ when $x_i y_i$ is

drawn from either $\widetilde{P} = \widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}$ or $\widetilde{P}^{\Pi} = \prod_{j=1}^{n} \widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$. We prove that there exist two constants $K_1, K_2 > 0$ such that for any $\gamma \in [0,1]$:

$$\widetilde{P}^{\Pi}[|\Gamma| \geq \gamma] \leq K_2 \cdot \exp\left(-\frac{\gamma^2}{K_1 \cdot \sigma^2}\right) \tag{45}$$

for $\sigma^2 = 1/\delta n$. Using Equation (45) and the fact that $|\Gamma| \leq 1$ (Assumption 1), Proposition 2.9 yields that

$$\begin{aligned} \mathrm{E}_{x_iy_i \sim \widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\gamma_i(x_iy_i)^2\right] = \mathrm{E}_{\widetilde{P}}\left[\Gamma^2\right] &\leq \frac{K_3}{\delta n} \cdot \left(D(\widetilde{P}||\widetilde{P}^{\Pi}) + 1\right) \\ &= \frac{K_3}{\delta n} \cdot \left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||\prod_{j=1}^{n} \widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}) + 1\right) \\ &\leq \frac{K_3}{\delta n} \cdot \left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||P_{X_iY_i|X_{<i}=x_{<i}}) + 1\right). \end{aligned} \tag{46}$$

The last inequality holds by chain rule of KL-divergence when the right-hand side distribution is product (Fact 2.3(3), where recall that $P_{X_iY_i|X_{<i}=x_{<i}} = \prod_{j=1}^{n} P_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$). This concludes the proof of Equation (44). It is left to prove Equation (45). In the following, given x_iy_i which are drawn from either $\widetilde{P}^{\Pi} = \prod_{j=1}^{n} \widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$ or $\widetilde{P}^{\Pi'} = \prod_{j=1}^{n} \widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}} \cdot \widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}$, we define the random variables L_j,Z_j,L and Z (in addition to Γ), where L_j is the value of $\omega_j \cdot \beta_j(x_{i,j})$, $L = \sum_{j=1}^{n} L_j$, $Z_j = \begin{cases} L_j/\tilde{\delta}_j & y_{i,j} = 1 \\ 0 & y_{i,j} = 0 \end{cases}$ and $Z = \sum_{j=1}^{n} Z_j$, letting $\omega_j = \omega_{i,j}$, $\beta_j(\cdot) = \beta_{i,j}(\cdot)$ and $\tilde{\delta}_j = \tilde{\delta}_{i,j}$. Note that by definition, $Z = (1+\Gamma)\mu$ for $\mu = \sum_{j=1}^{n} \omega_j$. Namely, Γ measures how far Z is from its expected value μ (follows by Equation (42) that calculates $\mathrm{E}_{\widetilde{P}}[Z]$, which also equals to $\mathrm{E}_{\widetilde{P}^{\Pi}}[Z]$ and $\mathrm{E}_{\widetilde{P}^{\Pi'}}[Z]$). Note that the distribution of Z and Γ when x_iy_i is drawn from $\widetilde{P}^{\Pi}$ is identical to the distribution of Z and Γ (respectively) when x_iy_i is drawn from $\widetilde{P}^{\Pi'}$. Therefore, in particular it holds that

$$\widetilde{P}^{\Pi}[|\Gamma| \geq \gamma] = \widetilde{P}^{\Pi'}[|\Gamma| \geq \gamma] \tag{47}$$

Under $\widetilde{P}^{\Pi'}$, the L_j's are independent random variables with $\mathrm{E}_{\widetilde{P}^{\Pi'}}[L_j] = \omega_j$ and $\mathrm{E}_{\widetilde{P}^{\Pi'}}[L] = \mu$ where $\mu = \sum_{j=1}^{n} \omega_j \geq n/2$ and $|L_j| \leq 2$ (by Assumptions 3 and 5). Moreover, for all $j \in [n]$, $Z_j = (L_j/\tilde{\delta}_j) \cdot \mathrm{Bern}(\tilde{\delta}_j)$ where $\tilde{\delta}_j \geq 0.9\delta_{i,j} \geq 0.9\delta$ (by Assumption 2). Hence, Fact 2.23 yields that

$$\widetilde{P}^{\Pi'}[|\Gamma| \geq \gamma] \leq 4\exp\left(-\frac{\delta n \gamma^2}{100}\right) \tag{48}$$

The proof of Equation (45) now follows by Equations (47) and (48), which ends the proof of Theorem 4.3 under the assumptions in 6.1.

6.1.1 Eliminating the Assumptions

The assumptions we made in 6.1 may seem unjustified at first glance. For instance, even for $j = 1$, there could be "bad" columns $j \in [n]$ with $\tilde{\delta}_{1,j} < 0.9\delta_{1,j}$. We claim, however, that the probability

that a uniform J (chosen by Q) will hit such a "bad" column j is low. For showing that, let $\mathcal{B}_1 = \{j \in [n] \colon \tilde{\delta}_{1,j} < 0.9\delta_{1,j}\}$ be the set of "bad" columns $j \in [n]$ for $i = 1$. A simple calculation yields that

$$\begin{aligned} d_1 &= D(\widetilde{P}_{X_1Y_1}||P_{X_1Y_1}) \geq D(\widetilde{P}_{Y_1}||P_{Y_1}) \geq \sum_{j=1}^{n} D(\widetilde{P}_{Y_{1,j}}||P_{Y_{1,j}}) \\ &= \sum_{j=1}^{n} D(\tilde{\delta}_{1,j}||\delta_{1,j}) \geq \sum_{j\in\mathcal{B}_1} D(\tilde{\delta}_{1,j}||\delta_{1,j}) \geq \sum_{j\in\mathcal{B}_1} \delta_{1,j}/200 \geq |\mathcal{B}_1| \cdot \delta/200. \end{aligned}$$

The second inequality holds by chain-rule of KL-divergence when the right-hand side distribution is product (Fact 2.3(3))) and the penultimate inequality holds by Fact 2.8(1). This implies that $|\mathcal{B}_1| \leq 200d_1/\delta$, and hence, $Q_J[J \in \mathcal{B}_1] < 200d_1/(\delta n)$. Extending the above argument for a row $i > 1$ is a much harder task. As we saw in Equation (37), the conditional distribution $Q_{J|X_{<i}}$ is much more complicated, and it also seems not clear how to bound $|\mathcal{B}_i|$ (now a function of $X_{<i}$) as we did for $i = 1$, when $X_{<i}$ is drawn from Q. Yet, we show in the next sections that when $X_{<i}$ is drawn from $\widetilde{P}$ (and not from Q), then we are able to understand $Q_{J|X_{<i}}$ and $\mathcal{B}_i(X_{<i})$ better and bound by $O(d/(\delta n))$ the probability of hitting a "bad" column for all $i \in [m]$. This is done by relating martingale sequences for each sequence $\{\omega_{i,j}\}_{i=1}^{m}$ under $\widetilde{P}$, and by showing (using Lemma 2.18) that with high probability, the sequences of most $j \in [n]$ remain around 1.

6.2 The Conditional Distributions

Following the above discussion, the high level plan of our proof is to define the "good" events $A_1, \ldots, A_n$ for $\widetilde{P}$ and $B_1, \ldots, B_n$ for Q such that for all $i \in [m]$, the conditional distributions $\widetilde{P}_{X_i|A_{\leq i}}$ and $Q_{X_i|B_{\leq i}}$ satisfies the assumptions in 6.1. Then, by only bounding the probability of "bad" events under $\widetilde{P}$, the proof of Theorem 4.3 will follow by Lemma 3.5. We start with notations.

Notation 6.2.

- $\omega'_{i,j} = \omega'_{i,j}(x_{<i}) = \prod_{s=1}^{i-1} \frac{P_{X_{s,j}|X_{<s}=x_{<s}}(x_{s,j})}{\widetilde{P}_{X_{s,j}|X_{<s}=x_{<s}}(x_{s,j})}$.
- $\omega_{i,j} = \omega_{i,j}(x_{<i}) = \frac{n\cdot\omega'_{i,j}}{\sum_{t=1}^{n}\omega'_{i,t}} \cdot \prod_{s=1}^{i-1} \frac{\widetilde{P}[E_{s,j}|X_{<s}=x_{<s}]}{\widetilde{P}[E_{s,j}|X_{<s}=x_{<s}, X_{s,j}=x_{s,j}]} \cdot \frac{\widetilde{P}[E_{s,j}|X_{\leq s}=x_{\leq s}]}{\widetilde{P}[E_{s,j}|X_{<s}=x_{<s}]}$
- $\beta_{i,j}(x_{i,j}) = \beta_{i,j}(x_{i,j}; x_{<i}) = P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})/\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j})$
- $\tilde{\delta}_{i,j} = \tilde{\delta}_{i,j}(x_{<i}) = \widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}]$
- $\mathcal{X}_{i,j} = \mathcal{X}_{i,j}(x_{<i}) = \{x_{i,j} \in \mathrm{Supp}(P_{X_{i,j}|X_{<i}=x_{<i}}) \colon \beta_{i,j}(x_{i,j}) \leq 1.1\}$.
- $\mathcal{J}_i = \mathcal{J}_i(x_{<i}) = \{j \in [n] \colon \left(\tilde{\delta}_{i,j} \geq 0.9\delta_{i,j}\right) \wedge (\omega_{i,j} \in 1 \pm 0.1) \wedge \left(P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j}) \geq 0.9\right)\}$.
- $\mathcal{G}_i(x_i) = \mathcal{G}_i(x_i; x_{<i}) = \{j \in [n] \colon \bigwedge_{s=1}^{i}(j \in \mathcal{J}_s \wedge x_{s,j} \in \mathcal{X}_{s,j})\}$, *letting* $\mathcal{G}_0 = [n]$.
- $\mathcal{S}_i = \mathcal{S}_i(x_{<i}) = \mathcal{G}_{i-1} \bigcap \mathcal{J}_i$.
- $\beta'_{i,j}(x_{i,j}) = \beta'_{i,j}(x_{i,j}; x_{<i}) = \beta_{i,j} \cdot \mathbb{1}_{\{x_{i,j}\in\mathcal{X}_{i,j}\}}$.

- $\gamma_i(x_i y_i) = \gamma_i(x_i y_i; x_{<i}) = \left(\sum_{j \in \mathcal{S}_i \cap 1_{y_i}} \frac{\omega_{i,j} \cdot \beta'_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right) / \left(\sum_{j \in \mathcal{S}_i} \omega_{i,j} \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j})\right) - 1.$

Definition 6.3 (Events)**.** *The event B_i is defined over $Q_{X,J}$ by B_i: $J \in \mathcal{G}_i(X_{\leq i})$.*

The following events are defined over $P = P_{\cdot,X}$:

- G_i: $|\mathcal{S}_i(X_{<i})| \geq 0.9n$.
- T_i: $|\gamma_i(X_i X_i; X_{<i})| \leq 1/2$.
- T'_i: $\widetilde{P}[T_i \mid X_{<i}] \geq 1 - 1/n$.
- $A_i = G_i \wedge T_i \wedge T'_i$.
- $\widetilde{B}_i$: $\mathrm{Bern}(Q[B_i \mid X_{<i}, B_{<i}]) = 1$.
  (i.e., a coin that takes one with probability $Q[B_i \mid X_{<i}, B_{<i}]$ is flipped and its outcome is one).
- $C_i = A_i \wedge \widetilde{B}_i$.

A few words about these definitions are in order. For $i \in [m]$, the set $\mathcal{G}_i(x_i)$ is basically the set of all columns $j \in [n]$ that are "good" for all rows $s \in [i]$ (in a sense that all values of $\tilde{\delta}_{s,j}$, $\beta_{s,j}$, $\omega_{s,j}$ are bounded as we would like), and the set $\mathcal{S}_i$ is the set of all (potential) "good" columns with respect to the history $x_{<i}$ (i.e., $\tilde{\delta}_{s,j}$, $\omega_{s,j}$ are bounded for all $s \in [i]$, but $\beta_{s,j}$ are only bounded for $s \in [i-1]$). A_i is the event (over $\widetilde{P}$) that we have large number of potential good columns for the row i (described by the event G_i), and that $|\gamma_i|$, the term that will appear in the analysis, is promised to be small (described by the event T_i). B_i is the event (over Q) that J is "good" for all rows in $[i]$.

The proof of Theorem 4.3 follows by the following two lemmatas and Lemma 3.5.

Lemma 6.4 (Bounding KL-divergence of conditional distributions)**.** *Let $P, \widetilde{P}, Q, W, \mathcal{E}, Y, \delta, d$ as defined in Theorem 4.3, and let $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{T_i\}_{i=1}^m$ be the events defined in Definition 6.3. Assuming that $\widetilde{P}[T_1 \wedge \ldots \wedge T_n] \geq 1/2$, then for every $i \in [m]$ it holds that*

$$D(\widetilde{P}_{X_i|A_{\leq i}} || Q_{X_i|B_{\leq i}} \mid \widetilde{P}_{X_{<i}|C_{\leq i}}) \leq \frac{c}{\delta n}(d_i + 1) \cdot \frac{1}{\widetilde{P}[C_{\leq i}]}$$

for some universal constant $c > 0$, and $d_i = D(\widetilde{P}_{X_i Y_i} || P_{X_i Y_i} | \widetilde{P}_{X_{<i}})$.

Lemma 6.5 (Bounding probability of bad events under $\widetilde{P}$)**.** *Let $P, \widetilde{P}, Q, W, \mathcal{E}, Y, \delta, d$ as defined in Theorem 4.3, and let $\{C_i\}_{i=1}^m$ be the events defined in Definition 6.3. Then there exists a universal constant $c > 0$ such that if $n \geq c \cdot m/\delta$ and $d \leq \delta n / c$, then*

$$\widetilde{P}[C_1 \wedge \ldots \wedge C_m] \geq 1 - c \cdot (d+1)/\delta n.$$

Proving Theorem 4.3.

Proof of Theorem 4.3. We start by setting the constant of Theorem 4.3 to $c = 4 \cdot \max\{c_1, c_2 + 1\}$ where c_1 is the constant from Lemma 6.4 and c_2 is the constant from Lemma 6.5. By Lemma 6.5 it holds that

$$\widetilde{P}[C_1 \wedge \ldots \wedge C_m] \geq 1 - (c_2 + 1) \cdot (d+1)/\delta n \tag{49}$$

$$\geq 1/2, \tag{50}$$

the last inequality holds by the assumption on n and d. In particular, it holds that

$$\widetilde{P}[T_1 \wedge \ldots \wedge T_m] \geq 1/2 \tag{51}$$

Therefore, by (51) and Lemma 6.4 it holds that

$$\begin{aligned} D(\widetilde{P}_{X_i|A_{\leq i}} || Q_{X_i|B_{\leq i}} \mid \widetilde{P}_{X_{<i}|C_{\leq i}}) &\leq \frac{c_1}{\delta n}(d_i+1) \cdot \frac{1}{\widetilde{P}[C_{\leq i}]} \\ &\leq \frac{c_1}{\delta n}(d_i+1) \cdot \frac{1}{\widetilde{P}[C_1 \wedge \ldots \wedge C_m]} \\ &\leq \frac{c}{\delta n}(d_i+1), \end{aligned} \tag{52}$$

the last inequality holds by Equation (50). The proof now holds by Equations (49) and (52) and Lemma 3.5. □

In addition, the proof of Lemma 4.5 now follows by Lemma 6.5.

Proving Lemma 4.5

Corollary 6.6 (Restatement of Lemma 4.5)**.** *Let $P, \widetilde{P}, Q, W, \mathcal{E}, \delta, d$ be as in Theorem 4.3, let c be the constant from Lemma 6.5, let $t > 0$ and let*

$$p_t := \Pr_{x \sim \widetilde{P}_X \,;\, j \sim Q_{J|X=x}}\left[\exists i \in [m] : P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] < \frac{P[W]}{t}\right]$$

Assuming $n \geq c \cdot m/\delta$ and $d \leq \delta n/c$, then

$$p_t \leq \frac{2m}{t} + \frac{c(d+1)}{\delta n}.$$

Proof. Let $\mathcal{G}_m, \tilde{\delta}_{i,j}, \beta_{i,j}$ be according to Notation 6.2. Observe that for any fixing of $x \in \mathrm{Supp}(\widetilde{P}_X)$ and any $j \in \mathcal{G}_m(x)$, the following holds for all $i \in [m]$:

$$\begin{aligned} P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}] &= \frac{\widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}]}{P[E_{i,j} \mid X_{<i} = x_{<i}]} \cdot \frac{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j})}{P_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j})} \cdot P[W \mid X_{<i} = x_{<i}] \\ &= \frac{\tilde{\delta}_{i,j}(x_{<i})}{\delta_{i,j}} \cdot \beta_{i,j}(x_{i,j}; x_{<i}) \cdot P[W \mid X_{<i} = x_{<i}] \\ &\geq P[W \mid X_{<i} = x_{<i}]/2, \end{aligned} \tag{53}$$

where second equality holds since

$$\begin{aligned} P_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j}) &= \frac{P[E_{i,j} \mid X_{\leq i} = x_{\leq i}] \cdot P[X_{i,j} = x_{i,j} \mid X_{<i} = x_{<i}]}{P[E_{i,j}]} \\ &= P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j}), \end{aligned}$$

(recall that by assumption, $P[E_{i,j} \mid X_{\leq i} = x_{\leq i}] = P[E_{i,j}]$ for any fixing of $x_{\leq i}$), and the inequality holds since $j \in \mathcal{G}_m(x)$. Let $\{\widetilde{B}_i\}, \{C_i\}$ be the events from Definition 6.3. We deduce that

$$\begin{aligned}\Pr_{\substack{x\sim\widetilde{P}_X \\ j\sim Q_{J|X=x}}}\left[\exists i \in [m] : \frac{P[W \mid (X_{<i}, X_{i,j}) = (x_{<i}, x_{i,j}), E_{i,j}]}{P[W \mid X_{<i} = x_{<i}]} < \frac{1}{2}\right] &\leq \Pr_{\substack{x\sim\widetilde{P}_X \\ j\sim Q_{J|X=x}}}[j \in \mathcal{G}_m(x)] \\ &\leq \widetilde{P}[\widetilde{B}_1 \wedge \ldots \wedge \widetilde{B}_m] \\ &\leq \frac{c(d+1)}{\delta n}, \end{aligned} \tag{54}$$

where the first inequality holds by Equation (53) and the last one holds by Lemma 6.5. In addition, by Fact 2.20 along with Markov's inequality and a union bound it holds that

$$\Pr_{x\sim\widetilde{P}_X}\left[\exists i \in [m] : P[W \mid X_{<i} = x_{<i}] < \frac{2P[W]}{t}\right] < \frac{2m}{t}. \tag{55}$$

The proof now follows by Equations (54) and (55) □

6.3 Bounding KL-Divergence of the Conditional Distributions

In this section we prove Lemma 6.4.

Proof of Lemma 6.4. We start by noting that for any $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{X_{<i}|C_{\leq i}}) \subseteq \mathrm{Supp}(\widetilde{P}_{X_{<i}|S_{\leq i},T'_{\leq i},T_{\leq i}})$, the following assertions hold.

Assertion 6.7.

1. $\left(\widetilde{P}[T_{\leq i} \mid X_{<i} = x_{<i}] > 1 - \frac{1}{n}\right)$ *(holds by the event T'_i and $T_{\leq i-1}$).*
2. $(|\mathcal{S}_i| \geq 0.9n)$ *(holds by the event G_i).*
3. $\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},A_{\leq i}} \equiv \widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}$ *(holds since G_i, T'_i and $\widetilde{B}_i$ are just random functions of $X_{<i}$).*
4. $Q_{X_iY_i|X_{<i}=x_{<i},B_{\leq i}} \equiv Q_{X_iY_i|X_{<i}=x_{<i},J\in\mathcal{G}_i(X_i)}$.
5. *For all $x_iy_i \in \mathrm{Supp}(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}})$ it holds that $|\gamma_i(x_iy_i)| \leq 1/2$.*
6. *For all $x_iy_i \in \mathrm{Supp}(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}})$ it holds that $\gamma_i(x_iy_i) \leq 2/\delta$.*

Note that Assertion 6 holds since for any $j \in \mathcal{S}_i$ and any x_i it holds that: $\tilde{\delta}_{i,j} \geq 0.9\delta$, $\omega_{i,j} \in 1 \pm 0.1$, $\beta'_{i,j}(x_i) \leq 1.1$ and $P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j}) \geq 0.9$. Therefore,

$$\begin{aligned}\gamma_i &\leq \left(\sum_{j\in\mathcal{S}_i} \frac{\omega_{i,j} \cdot \beta'_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right) / \left(\sum_{j\in\mathcal{S}_i} \omega_{i,j} \cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j})\right) \\ &\leq \left(\frac{1.1 \cdot 1.1}{0.9\delta} \cdot |\mathcal{S}_i|\right) / (0.9 \cdot 0.9 \cdot |\mathcal{S}_i|) \leq 2/\delta \end{aligned}$$

Our goal now is to show that for any fixing of $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{X_{<i}|C_{\leq i}})$ is holds that

$$D(\widetilde{P}_{X_i|X_{<i}=x_{<i},T_{\leq i}}||Q_{X_i|X_{<i}=x_{<i},J\in\mathcal{G}_i(X_i)}) \leq \frac{c}{\delta n}\cdot\Big(D(\widetilde{P}_{X_i|X_{<i}=x_{<i}}||P_{X_i|X_{<i}=x_{<i}})+1\Big), \tag{56}$$

for some constant $c>0$. The proof then follow by Equation (56) since

$$\begin{aligned}
D(\widetilde{P}_{X_i|A_{\leq i}}||Q_{X_i|B_{\leq i}} \mid \widetilde{P}_{X_{<i}|C_{\leq i}}) &= \mathrm{E}_{x_{<i}\sim\widetilde{P}_{X_{<i}|C_{\leq i}}}\Big[D(\widetilde{P}_{X_i|X_{<i}=x_{<i},A_{\leq i}}||Q_{X_i|X_{<i}=x_{<i},B_{\leq i}})\Big] \\
&= \mathrm{E}_{x_{<i}\sim\widetilde{P}_{X_{<i}|C_{\leq i}}}\Big[D(\widetilde{P}_{X_i|X_{<i}=x_{<i},T_{\leq i}}||Q_{X_i|X_{<i}=x_{<i},J\in\mathcal{G}_i(X_i)})\Big] \\
&\leq \frac{c}{\delta n}\cdot\Big(\mathrm{E}_{x_{<i}\sim\widetilde{P}_{X_{<i}|C_{\leq i}}}\Big[D(\widetilde{P}_{X_i|X_{<i}=x_{<i}}||P_{X_i|X_{<i}=x_{<i}})\Big]+1\Big) \\
&\leq \frac{c}{\delta n}(d_i+1)\cdot\frac{1}{\widetilde{P}[C_{\leq i}]},
\end{aligned}$$

where the second equality holds by Properties 3 and 4 in 6.7, and the last inequality holds by Fact 2.5. We now focus on proving Equation (56) in a similar spirit to the proof given in Section 6.1.

In the following, fix $i\in[m]$ and $x_{<i}\in\mathrm{Supp}(\widetilde{P}_{X_{<i}|C_{\leq i}})$. By data-processing of KL-divergence (Fact 2.3(5)), it holds that

$$D(\widetilde{P}_{X_i|X_{<i}=x_{<i},T_{\leq i}}||Q_{X_i|X_{<i}=x_{<i},J\in\mathcal{G}_i(X_i)}) \leq D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}||Q'_{X_iY_i|X_{<i}=x_{<i}}), \tag{57}$$

where

$$\begin{aligned}
Q'_{X_iY_i|X_{<i}=x_{<i}} &= \widetilde{P}_{X_iY_i|X_{<i}=x_{<i},X_i,Y_{i,J}=1}\circ Q_{J,X_i|X_{<i}=x_{<i},J\in\mathcal{G}_i(X_i)} \\
&\equiv \widetilde{P}_{X_iY_i|X_{<i}=x_{<i},X_{i,J},Y_{i,J}=1}\circ Q_{J,X_{i,J}|X_{<i}=x_{<i},J\in\mathcal{S}_i,X_{i,J}\in\mathcal{X}_{i,J}}
\end{aligned}$$

Similar calculation to the one in Equation (34) yields that for any fixing of $x_iy_i\in\mathrm{Supp}(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}})$ it holds that

$$\begin{aligned}
&Q'_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i) \\
&= \sum_{j\in\mathcal{G}_i(x_i)\bigcap 1_{y_i}} Q_{J|X_{<i}=x_{<i},J\in\mathcal{G}_i(x_i)}(j)\cdot\frac{\beta_{i,j}(x_{i,j})\cdot\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)}{\tilde{\delta}_{i,j}}
\end{aligned} \tag{58}$$

In addition, for any $j\in\mathcal{G}_i(x_i)$ it holds that

$$\begin{aligned}
Q_{J|X_{<i}=x_{<i},J\in\mathcal{G}_i(x_i)}(j) &= \frac{Q_{J|X_{<i}=x_{<i},J\in\mathcal{S}_i}(j)\cdot\mathbb{1}_{\{x_{i,j}\in\mathcal{X}_{i,j}\}}}{Q[X_{i,j}\in\mathcal{X}_{i,j}\mid X_{<i}=x_{<i},J=j]} \\
&= \frac{Q_{J|X_{<i}=x_{<i},J\in\mathcal{S}_i}(j)\cdot\mathbb{1}_{\{x_{i,j}\in\mathcal{X}_{i,j}\}}}{P[X_{i,j}\in\mathcal{X}_{i,j}\mid X_{<i}=x_{<i}]} \\
&= \frac{\omega_{i,j}\cdot\mathbb{1}_{\{x_{i,j}\in\mathcal{X}_i\}}}{\sum_{t\in\mathcal{S}_i}\omega_{i,t}\cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j})},
\end{aligned} \tag{59}$$

where the first equality holds since $\mathcal{G}_i(x_i)=\{j\in[n]\colon j\in\mathcal{S}_i\wedge x_{i,j}\in\mathcal{X}_{i,j}\}$ and the last equality holds by Equation (37). Therefore, by combining Equations (58) and (59) we now can write

$$Q'_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i) = \frac{\sum_{j\in\mathcal{S}_i\bigcap 1_{y_i}}\frac{\omega_{i,j}\cdot\beta'_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}}{\sum_{j\in\mathcal{S}_i}\omega_{i,j}\cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j})}\cdot\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i) \tag{60}$$

Using Equations (57) and (60), we deduce that

$$
\begin{aligned}
&D(\widetilde{P}_{X_i|X_{<i}=x_{<i},T_{\leq i}}||Q_{X_i|X_{<i}=x_{<i},J\in\mathcal{G}_i(X_i)}) \qquad (61)\\
&\leq D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}||Q'_{X_iY_i|X_{<i}=x_{<i}})\\
&= \mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[\log\frac{\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}(x_iy_i)}{Q'_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)}\right]\\
&\leq \mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[\log\frac{\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)/\widetilde{P}[T_{\leq i}\mid X_{<i}=x_{<i}]}{Q'_{X_iY_i|X_{<i}=x_{<i}}(x_iy_i)}\right]\\
&\leq \mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[\log\frac{\sum_{j\in\mathcal{S}_i}\omega_{i,j}\cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j})}{\sum_{j\in\mathcal{S}_i\bigcap 1_{y_i}}\frac{\omega_{i,j}\cdot\beta'_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}}\right]+2/n,\\
&= \mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[-\log(1+\gamma_i(x_iy_i))\right]+2/n\\
&\leq \mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[-\gamma_i(x_iy_i)+\gamma_i(x_iy_i)^2\right]+2/n\\
&\leq -\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}[\gamma_i(x_iy_i)]/\widetilde{P}[T_{\leq i}\mid X_{<i}=x_{<i}]+\frac{2}{\delta n}+\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[\gamma_i(x_iy_i)^2\right]+2/n
\end{aligned}
$$

The third inequality holds by Equation (60) and by Assertion 6.7(1) which yields that $\log\frac{1}{\widetilde{P}[T_{\leq i}|X_{<i}=x_{<i}]}\leq 2/n$. The one before last inequality holds by the inequality $-\log(1+x)\leq -x+x^2$ for $|x|\leq 1/2$ (recall Assertion 6.7(5)). The last inequality holds since for any random variable $X\leq M$ and any event T it holds that $\mathrm{E}[-X\mid T]=\frac{-\mathrm{E}[X]+\mathrm{E}[X|\overline{T}]\cdot\Pr[\overline{T}]}{\Pr[T]}\leq -\mathrm{E}[X]/\Pr[T]+M\cdot\Pr[\overline{T}]$ (recall Assertions 6.7(1,6)). Note that

$$
\begin{aligned}
&\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}\left[\sum_{j\in\mathcal{S}_i\bigcap 1_{y_i}}\frac{\omega_{i,j}\cdot\beta'_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right] \qquad (62)\\
&=\sum_{j\in\mathcal{S}_i}\mathrm{E}_{x_{i,j}y_{i,j}\sim\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}}\left[y_{i,j}\cdot\frac{\omega_{i,j}\cdot\beta'_{i,j}(x_{i,j})}{\tilde{\delta}_{i,j}}\right]\\
&=\sum_{j\in\mathcal{S}_i}\omega_{i,j}\cdot\mathrm{E}_{x_{i,j}\sim\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}}\left[\beta'_{i,j}(x_{i,j})\right]\\
&=\sum_{j\in\mathcal{S}_i}\omega_{i,j}\cdot\mathrm{E}_{x_{i,j}\sim\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}}\left[\frac{P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})}{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}(x_{i,j})}\cdot\mathbb{1}_{\{x_{i,j}\in\mathcal{X}_{i,j}\}}\right]\\
&=\sum_{j=1}^{n}\omega_{i,j}\cdot P_{X_{i,j}|X_{<i}=x_{<i}}\left(\mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1})\bigcap\mathcal{X}_{i,j}\right)=\sum_{j=1}^{n}\omega_{i,j}\cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j}),
\end{aligned}
$$

where the second equality holds since $y_{i,j}\in\{0,1\}$ and since for all $j\in\mathcal{S}_i$ it holds that $\widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}}(1)=\tilde{\delta}_{i,j}>0$, and the last equality holds since $\mathcal{X}_{i,j}\subseteq\mathrm{Supp}(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1})$. Therefore, we deduce from Equation (62) that

$$\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}}[\gamma_i(x_iy_i)]=0 \tag{63}$$

Therefore, by Equations (61) and (63), in order to prove Equation (56) it is left to show that

$$\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[\gamma_i(x_iy_i)^2\right]\leq O\left(\frac{1}{\delta n}\right)\cdot\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||P_{X_iY_i|X_{<i}=x_{<i}})+1\right). \tag{64}$$

Let Γ be the value of $\gamma_i(x_iy_i)$ when x_iy_i is drawn from either $\widetilde{P}'=\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}$ or $\widetilde{P}^{\Pi}=\prod_{j=1}^{n}\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}}$. We now prove that there exists constants $K_1,K_2>0$ such that for every $\gamma\in[0,1]$ it holds that

$$\widetilde{P}^{\Pi}[|\Gamma|\geq\gamma]\leq K_2\cdot\exp\left(-\frac{\gamma^2}{K_1\cdot\sigma^2}\right), \tag{65}$$

The proof of Equation (64) then follows since

$$\begin{aligned}\mathrm{E}_{x_iy_i\sim\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}}\left[\gamma_i(x_iy_i)^2\right]=\mathrm{E}_{\widetilde{P}'}\left[\Gamma^2\right]&\leq\frac{K_3}{\delta n}\cdot\left(D(\widetilde{P}'||\widetilde{P}^{\Pi})\right) \qquad (66)\\ &=\frac{K_3}{\delta n}\cdot\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}||\prod_{j=1}^{n}\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}})+1\right)\\ &\leq\frac{K_3}{\delta n}\cdot\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i},T_{\leq i}}||P_{X_iY_i|X_{<i}=x_{<i}})+1\right)\\ &\leq\frac{K_3}{\delta n}\cdot\left(\frac{1}{\widetilde{P}[T_{\leq j}]}\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||P_{X_iY_i|X_{<i}=x_{<i}})+1/e+1\right)+1\right)\\ &\leq\frac{5K_3}{\delta n}\cdot\left(D(\widetilde{P}_{X_iY_i|X_{<i}=x_{<i}}||P_{X_iY_i|X_{<i}=x_{<i}})+1\right)\end{aligned}$$

where the first inequality holds by Proposition 2.9 and the fact that $|\Gamma|\leq 1$ under $\widetilde{P}'$, the second inequality holds by chain rule of KL-divergence when the right-hand side distribution is product (Fact 2.3(3)), the one before last inequality holds by Fact 2.6, and the last one holds since $\widetilde{P}[T_{\leq j}]\geq 1/2$.

We now prove Equation (65). In the following, given x_iy_i which are drawn from either $\widetilde{P}^{\Pi}$ or $\widetilde{P}^{\Pi'}=\prod_{j=1}^{n}\widetilde{P}_{Y_{i,j}|X_{<i}=x_{<i}}\cdot\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1}$, we define the random variables L_j,Z_j,L and Z (in addition to Γ), where L_j is the value of $\omega_j\cdot\beta'_j(x_{i,j})$, $L=\sum_{j=1}^{n}L_j$, $Z_j=\begin{cases}L_j/\tilde{\delta}_j & y_{i,j}=1\\ 0 & y_{i,j}=0\end{cases}$ and $Z=\sum_{j=1}^{n}Z_j$, letting $\omega_j=\omega_{i,j}$, $\beta_j(\cdot)=\beta_{i,j}(\cdot)$ and $\tilde{\delta}_j=\tilde{\delta}_{i,j}$. Note that by definition, $Z=(1+\Gamma)\mu$ for $\mu=\sum_{j\in\mathcal{S}_i}\omega_j\cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j})$ (follows from Equation (62)). Moreover, by the definition of $\mathcal{S}_i$ and the fact that $|\mathcal{S}_i|\geq 0.9n$ (Assertion 6.7(2)), it holds that $|L_j|\leq 2$, $\tilde{\delta}_j\geq 0.9\delta$ and $\mu\geq n/2$. Hence,

$$\widetilde{P}^{\Pi}[|\Gamma|\geq\gamma]=\widetilde{P}^{\Pi'}[|\Gamma|\geq\gamma]\leq 4\exp\left(-\frac{\delta n\gamma^2}{100}\right), \tag{67}$$

where the equality holds since Γ has the same distribution under $\widetilde{P}^{\Pi}$ and under $\widetilde{P}^{\Pi'}$, and the inequality holds by Fact 2.23 since under $\widetilde{P}^{\Pi'}$ the L_j's are independent random variables with $\mathrm{E}_{\widetilde{P}^{\Pi'}}[L] = \mu$ and for all $j \in \mathcal{S}_i$ we have $Z_j = (L_j/\tilde{\delta}_j) \cdot \mathrm{Bern}(\tilde{\delta}_j)$. This proves Equation (65) and concludes the proof.

□

6.4 Bounding the Probability of Bad Events Under $\widetilde{P}$

In this section we prove Lemma 6.5. We start with few more notations (in addition to the ones given in Notation 6.2), then we prove facts about the distribution $\widetilde{P}$ (Section 6.4.1), and in Section 6.4.2 we present the proof of Lemma 6.5.

Notation 6.8 (Additional notations)**.**

- $\alpha_{i,j} = \alpha_{i,j}(x_{i,j}; x_{<i}) = \frac{P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})}{\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})} - 1.$
- $\rho_{i,j} = \rho_{i,j}(x_{<i}) = \frac{\widetilde{P}[E_{i,j}|X_{<i}=x_{<i}]}{P[E_{i,j}|X_{<i}=x_{<i}]} - 1 = \frac{\tilde{\delta}_{i,j}}{\delta_{i,j}} - 1.$
- $\tau_{i,j} = \tau_{i,j}(x_i; x_{<i}) = \frac{\widetilde{P}[E_{i,j}|X_{\le i}=x_{\le i}]}{P[E_{i,j}|X_{\le i}=x_{\le i}]} - 1 = \frac{\widetilde{P}[E_{i,j}|X_{\le i}=x_{\le i}]}{\delta_{i,j}} - 1.$
- $\xi_{i,j} = \xi_{i,j}(x_{i,j}; x_{<i}) = \frac{\widetilde{P}[E_{i,j}|X_{<i}=x_{<i},X_{i,j}=x_{i,j}]}{P[E_{i,j}|X_{<i}=x_{<i},X_{i,j}=x_{i,j}]} - 1 = \frac{\widetilde{P}[E_{i,j}|X_{<i}=x_{<i},X_{i,j}=x_{i,j}]}{\delta_{i,j}} - 1.$
- $U_{i,j} = U_{i,j}(x_{i,j}; x_{<i}) = \prod_{s=1}^{i} \frac{\widetilde{P}[E_{s,j}|X_{<s}=x_{<s},X_{s,j}=x_{s,j}]}{\widetilde{P}[E_{s,j}|X_{<s}=x_{<s}]} = U_{i-1,j} \cdot \frac{1+\xi_{i,j}}{1+\rho_{i,j}}.$
- $V_{i,j} = V_{i,j}(x_i; x_{<i}) = \prod_{s=1}^{i} \frac{\widetilde{P}[E_{s,j}|X_{\le s}=x_{\le s}]}{\widetilde{P}[E_{s,j}|X_{<s}=x_{<s}]} = V_{i-1,j} \cdot \frac{1+\tau_{i,j}}{1+\rho_{i,j}}.$
- $R_{i,j} = R_{i,j}(x_{<i}) = \frac{n \cdot \omega'_{i,j}}{\sum_{t=1}^{n} \omega'_{i,t}} = n \cdot \left(\prod_{s=1}^{i-1} \frac{P_{X_{s,j}|X_{<s}=x_{<s}}(x_{s,j})}{\widetilde{P}_{X_{s,j}|X_{<s}=x_{<s}}(x_{s,j})} \right) / \left(\sum_{t=1}^{n} \prod_{s=1}^{i-1} \frac{P_{X_{s,t}|X_{<s}=x_{<s}}(x_{s,t})}{\widetilde{P}_{X_{s,t}|X_{<s}=x_{<s}}(x_{s,t})} \right).$

  where in all definitions, recall that $\delta_{i,j} = P[E_{i,j} \mid X_{\le i}]$ for any fixing of $X_{\le i}$.

6.4.1 Facts about $\widetilde{P}$

Fact 6.9. *For all $r \in \{\rho, \tau, \xi\}$ it holds that*

1. $\mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{m} \sum_{j=1}^{n} \min\{|r_{i,j}|, r_{i,j}^2\}\right] \le \frac{4d}{\delta}.$
2. *For all $\lambda > 0$:* $\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \text{ s.t. } |r_{i,j}| \ge \lambda\}|] \le \frac{4d}{\delta \cdot \min\{\lambda, \lambda^2\}}.$

Proof. Assuming Item 1 holds, then Item 2 holds since

$$\begin{aligned}
\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \text{ s.t. } |r_{i,j}| \ge \lambda\}|] &\le \frac{1}{\min\{\lambda, \lambda^2\}} \cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{m} \sum_{j=1}^{n} \min\{|r_{i,j}|, r_{i,j}^2\} \cdot \mathbb{1}_{\{\min\{|r_{i,j}|, r_{i,j}^2\} \ge \min\{\lambda, \lambda^2\}\}}\right] \\
&\le \frac{1}{\min\{\lambda, \lambda^2\}} \cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{m} \sum_{j=1}^{n} \min\{|r_{i,j}|, r_{i,j}^2\}\right]
\end{aligned}$$

Item 1 for $r=\rho$ holds since

$$\begin{aligned}
d &= \sum_{i=1}^{m} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_iY_i|X_{<i}}||P_{X_iY_i|X_{<i}})\right] \geq \sum_{i=1}^{m} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{Y_i|X_{<i}}||P_{Y_i|X_{<i}})\right] \\
&\geq \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{Y_{i,j}|X_{<i}}||P_{Y_{i,j}|X_{<i}})\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}[D((1+\rho_{i,j})\delta_{i,j}||\delta_{i,j})] \\
&\geq \sum_{i=1}^{m}\sum_{j=1}^{n} \delta_{i,j}\cdot \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[\min\{|\rho_{i,j}|,\rho_{i,j}^2\}\right]/4 \geq \delta\cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\min\{|\rho_{i,j}|,\rho_{i,j}^2\}\right]/4,
\end{aligned}$$

where the first inequality holds by data processing of KL divergence (Fact 2.3(5)), the second one holds by chain rule of KL-divergence when the right-hand side distribution is product (Fact 2.3(3)), and the one before last inequality holds by Fact 2.8.

For $r=\tau$, Item 1 holds since

$$\begin{aligned}
d &= \sum_{i=1}^{m} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_iY_i|X_{<i}}||P_{X_iY_i|X_{<i}})\right] \geq \sum_{i=1}^{m} \mathrm{E}_{\widetilde{P}_{X_{\leq i}}}\left[D(\widetilde{P}_{Y_i|X_{\leq i}}||P_{Y_i|X_{\leq i}})\right] \\
&\geq \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{\leq i}}}\left[D(\widetilde{P}_{Y_{i,j}|X_{\leq i}}||P_{Y_{i,j}|X_{\leq i}})\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{\leq i}}}[D((1+\tau_{i,j})\delta_{i,j}||\delta_{i,j})] \\
&\geq \sum_{i=1}^{m}\sum_{j=1}^{n} \delta_{i,j}\cdot \mathrm{E}_{\widetilde{P}_{X_{\leq i}}}\left[\min\{|\tau_{i,j}|,\tau_{i,j}^2\}\right]/4 \geq \delta\cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\min\{|\tau_{i,j}|,\tau_{i,j}^2\}\right]/4,
\end{aligned}$$

where the first inequality holds by chain rule (Fact 2.3(3)) and the second one holds by chain rule when the right-hand side distribution is product (Fact 2.3(3)). For $r=\xi$, Item 1 holds since

$$\begin{aligned}
d &= \sum_{i=1}^{m} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_iY_i|X_{<i}}||P_{X_iY_i|X_{<i}})\right] \geq \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}}||P_{X_{i,j}Y_{i,j}|X_{<i}})\right] \\
&\geq \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i},X_{i,j}}}\left[(\widetilde{P}_{Y_{i,j}|X_{<i},X_{i,j}}||P_{Y_{i,j}|X_{<i},X_{i,j}})\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i},X_{i,j}}}[D((1+\xi_{i,j})\delta_{i,j}||\delta_{i,j})] \\
&\geq \sum_{i=1}^{m}\sum_{j=1}^{n} \delta_{i,j}\cdot \mathrm{E}_{\widetilde{P}_{X_{<i},X_{i,j}}}\left[\min\{|\xi_{i,j}|,\xi_{i,j}^2\}\right]/4 \geq \delta\cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\min\{|\xi_{i,j}|,\xi_{i,j}^2\}\right]/4,
\end{aligned}$$

where the first inequality holds by chain rule when the right-hand side distribution is product (Fact 2.3(3)) and the second one holds by standard chain-rule ok KL-divergence (Fact 2.3(3)).

□

Fact 6.10. *For all $L\in\{U,V\}$ it holds that*

1. *For any $j\in[n]$: the sequence $\{L_{i,j}\}_{i=1}^m$ is a martingale with respect to $\{X_i\}_{i=1}^m$ which are drawn from $\widetilde{P}$ (recall Definition 2.17).*
2. *For any $\lambda\in(0,\frac{1}{4})$: $\mathrm{E}_{\widetilde{P}}[|\{j\in[n]\colon \exists i\in[m] \text{ s.t. } |L_{i,j}-1|\geq\lambda\}|] \leq \frac{c\cdot d}{\delta\lambda^2}$, for some universal constant $c>0$.*

Proof. Note that for any fixing of $x_{<i}$ it holds that

$$\mathrm{E}_{\widetilde{P}_{X_i|X_{<i}=x_{<i}}}[U_{i,j}] = U_{i-1,j} \cdot \frac{\mathrm{E}_{\widetilde{P}_{X_i|X_{<i}=x_{<i}}}\left[\widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}, X_{i,j} = x_{i,j}]\right]}{\widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}]} = U_{i-1,j}$$

and

$$\mathrm{E}_{\widetilde{P}_{X_i|X_{<i}=x_{<i}}}[V_{i,j}] = V_{i-1,j} \cdot \frac{\mathrm{E}_{\widetilde{P}_{X_i|X_{<i}=x_{<i}}}\left[\widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}, X_i = x_i]\right]}{\widetilde{P}[E_{i,j} \mid X_{<i} = x_{<i}]} = V_{i-1,j}$$

This proves Item 1. By Proposition 2.19, there exists a constant $c' > 0$ such that for any $j \in [n]$ and $\lambda \in (0, \frac{1}{4})$ it holds that

$$\widetilde{P}[\exists i \in [m] \text{ s.t. } |U_{i,j} - 1| \geq \lambda] \leq \frac{c' \cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{n}\left(\min\{|\rho_{i,j}|, \rho_{i,j}^2\} + \min\{|\xi_{i,j}|, \xi_{i,j}^2\}\right)\right]}{\lambda^2}$$

and that

$$\widetilde{P}[\exists i \in [m] \text{ s.t. } |V_{i,j} - 1| \geq \lambda] \leq \frac{c' \cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{i=1}^{n}\left(\min\{|\rho_{i,j}|, \rho_{i,j}^2\} + \min\{|\tau_{i,j}|, \tau_{i,j}^2\}\right)\right]}{\lambda^2}$$

The proof of Item 2 now follows from the bounds in Fact 6.9(1).

□

Fact 6.11. *For every $\lambda > 0$ it holds that*

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \text{ s.t. } |R_{i,j} - 1| > \lambda\}|] \leq \frac{16 \cdot d}{\min\{\lambda, \lambda^2\}}.$$

Proof. We prove that for every $i \in [m]$ it holds that

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon |R_{i,j} - 1| > \lambda\}|] \leq \frac{16 \cdot d_i}{\min\{\lambda, \lambda^2\}}, \tag{68}$$

The proof of the fact then follows since

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \text{ s.t. } |R_{i,j} - 1| > \lambda\}|] \leq \sum_{i=1}^{m} \mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon |R_{i,j} - 1| > \lambda\}|]$$

In the following, let

$$Q' = Q'_X = \prod_{i=1}^{m} P_{X_{i,J}|X_{<i,J}} \widetilde{P}_{X_{i,-J}|X_{<i},X_{i,J}} \circ Q'_J,$$

where $Q'_J = U_{[n]}$. Applying Equation (37) on Q' (note that Q' is a special case of a skewed distribution Q when choosing events $\{E_{i,j}\}$ with $P[E_{i,j}] = 1$ for all i, j), we obtain for any $i \in [m]$, $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{X_{<i}})$ and any $j \in [n]$:

$$Q'_{J|X_{<i}=x_{<i}}(j) = \left(\prod_{s=1}^{i-1} \frac{P_{X_{s,j}|X_{<s}=x_{<s}}(x_{s,j})}{\widetilde{P}_{X_{s,j}|X_{<s}=x_{<s}}(x_{s,j})}\right) \Big/ \left(\sum_{t=1}^{n} \prod_{s=1}^{i-1} \frac{P_{X_{s,t}|X_{<s}=x_{<s}}(x_{s,t})}{\widetilde{P}_{X_{s,t}|X_{<s}=x_{<s}}(x_{s,t})}\right) = \frac{R_{i,j}}{n} \tag{69}$$

Moreover, as proven in [CP15], by letting $\widetilde{P}_J = U_{[n]}$ (i.e., the uniform distribution over $[n]$), it holds that

$$\mathrm{E}_{\widetilde{P}_J \widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_i|X_{<i}}||Q'_{X_i|J,X_{<i}})\right] = \frac{1}{n}\cdot \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[\sum_{j=1}^{n} D(\widetilde{P}_{X_i|X_{<i}}||Q'_{X_i|J=j,X_{<i}})\right]$$
$$= \frac{1}{n}\cdot \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[\sum_{j=1}^{n} D(\widetilde{P}_{X_{i,j}|X_{<i}}||P_{X_{i,j}|X_{<i}})\right] \leq \frac{1}{n}\cdot \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_i|X_{<i}}||P_{X_i|X_{<i}})\right] \leq \frac{d_i}{n},$$

where inequality holds by chain-rule of KL-divergence where the right-hand side is product. The above yields that

$$\mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(U_{[n]}||Q'_{J|X_{<i}})\right] \leq \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(U_{[n]}\widetilde{P}_{X_i|X_{<i}}||Q'_J Q'_{X_i|J,X_{<i}})\right] \tag{70}$$
$$= \mathrm{E}_{\widetilde{P}_J \widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_i|X_{<i}}||Q'_{X_i|J,X_{<i}})\right] \leq \frac{d_i}{n},$$

where the first inequality holds by data-processing of KL-divergence, and the equality holds by chain-rule of KL-divergence along with the fact that $\widetilde{P}_J \equiv Q'_J \equiv U_{[n]}$. In the following, fix $i \in [m]$ and let $\mathcal{B}_i^+ = \mathcal{B}_i^+(x_{<i}) = \{j \in [n] \colon Q'_{J|X_{<i}=x_{<i}}(j) > (1+\lambda)/n\}$ and let $\mathcal{B}_i^- = \mathcal{B}_i^-(x_{<i}) = \{j \in [n] \colon Q'_{J|X_{<i}=x_{<i}}(j) < (1-\lambda)/n\}$. By Equation (70) along with data-processing of KL-divergence, it holds that

$$\mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\frac{|\mathcal{B}_i^+|}{n}||(1+\lambda)\frac{|\mathcal{B}_i^+|}{n})\right] \leq \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(U_{[n]}(\mathcal{B}_i^+)||Q'_{J|X_{<i}}(\mathcal{B}_i^+))\right] \leq d_i/n$$

and by Fact 2.8 we deduce that $\mathrm{E}_{\widetilde{P}}\left[\left|\mathcal{B}_i^+\right|\right] \leq \frac{8\cdot d_i}{\min\{|\lambda|,\lambda^2\}}$. Similarly it holds that $\mathrm{E}_{\widetilde{P}}\left[\left|\mathcal{B}_i^-\right|\right] \leq \frac{8\cdot d}{\min\{|\lambda|,\lambda^2\}}$. The proof of Equation (68) now follows since for any $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{X_{<i}})$ and any $j \notin \mathcal{B}_i^+(x_{<i}) \bigcup \mathcal{B}_i^-(x_{<i})$ it holds that $\frac{R_{i,j}(x_{<i})}{n} = Q'_{J|X_{<i}=x_{<i}}(j) \in (1\pm\lambda)/n$ (the equality holds by Equation (69)). □

Fact 6.12. *For all $\lambda \in (0, \frac{1}{4})$ it holds that*

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \ s.t. \ |\omega_{i,j} - 1| \geq \lambda\}|] \leq \frac{c\cdot d}{\delta \cdot \min\{\lambda, \lambda^2\}},$$

for some universal constant $c > 0$.

Proof. Note that

$$\omega_{i,j} = \frac{R_{i,j}\cdot V_{i-1,j}}{U_{i-1,j}}$$

Therefore, we deduce that

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \text{ s.t. } |\omega_{i,j} - 1| \geq \lambda\}|]$$
$$\leq \mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \text{ s.t. } (|U_{i-1,j} - 1| > \lambda/10) \vee (|V_{i-1,j} - 1| > \lambda/10) \vee (|R_{i,j} - 1| > \lambda/10)\}|]$$
$$\leq 100(c_1 + c_2 + c_3)\cdot d/\delta,$$

where c_1, c_2 and c_3 are the constants from Fact 6.9(2), Fact 6.11 and Fact 6.10(2), respectively. □

Fact 6.13. *For every $\lambda \in (0, \frac{1}{2})$ it holds that*

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \ s.t. \ \alpha_{i,j} > \lambda\}|] \leq \frac{4 \cdot d}{\lambda^2},$$

for some constant $c > 0$.

Proof. We prove that for every $i \in [m]$ it holds that

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \alpha_{i,j} > \lambda\}|] \leq \frac{4 \cdot d_i}{\min\{\lambda, \lambda^2\}}, \tag{71}$$

The proof of the fact then follows since

$$\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \text{ s.t. } \alpha_{i,j} > \lambda\}|] \leq \sum_{i=1}^{m} \mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \alpha_{i,j} > \lambda\}|]$$

In the following, fix $i \in [m]$ and compute

$$\begin{aligned}
d_i &\geq \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_i|X_{<i}} || P_{X_i|X_{<i}})\right] \geq \sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_{i,j}|X_{<i}} || P_{X_{i,j}|X_{<i}})\right] \\
&\geq \sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_{i,j}|X_{<i}}[\alpha_{i,j} > \lambda] || P_{X_{i,j}|X_{<i}}[\alpha_{i,j} > \lambda])\right] \\
&\geq \sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[D(\widetilde{P}_{X_{i,j}|X_{<i}}[\alpha_{i,j} > \lambda] || (1+\lambda) \cdot \widetilde{P}_{X_{i,j}|X_{<i}}[\alpha_{i,j} > \lambda])\right] \\
&\geq \sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[\frac{1}{2} \cdot \left(\frac{\lambda}{1+\lambda}\right)^2 \cdot (1+\lambda) \cdot \widetilde{P}_{X_{i,j}|X_{<i}}[\alpha_{i,j} > \lambda]\right] \\
&\geq \frac{\lambda^2}{4} \cdot \sum_{j=1}^{n} \mathrm{E}_{\widetilde{P}_{X_{<i}}}\left[\widetilde{P}_{X_{i,j}|X_{<i}}[\alpha_{i,j} > \lambda]\right] = \frac{\lambda^2}{4} \cdot \mathrm{E}_{\widetilde{P}}\left[\sum_{j=1}^{n} \mathbb{1}_{\{\alpha_{i,j} > \lambda\}}\right] \\
&= \frac{\lambda^2}{4} \cdot \mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \alpha_{i,j} > \lambda\}|].
\end{aligned}$$

Which concludes the proof of Equation (71). The second inequality holds by data-processing of KL-divergence when the right-hand side distribution is product. The third inequality holds by data-processing of KL-divergence. The fourth inequality holds since for any $x_{i,j}$ with $\alpha_{i,j}(x_{i,j}) > \lambda$, it holds that $P_{X_{i,j}|X_{<i}}(x_{i,j}) \geq (1+\lambda)\widetilde{P}_{X_{i,j}|X_{<i}}(x_{i,j})$. The fifth inequality holds by Fact 2.8(1). □

Fact 6.14. *There exist constants $c, c' > 0$ such that for all $\lambda > 0$ it holds that*

1. $\mathrm{E}_{\widetilde{P}}[|\{j \in [n] \colon \exists i \in [m] \ s.t. \ \beta_{i,j} \geq 1 + \lambda\}|] \leq \frac{c \cdot d}{\delta}$.
2. $\mathrm{E}_{x \sim \widetilde{P}_X}\left[\sum_{i=1}^{m} \sum_{j=1}^{n} P_{X_{i,j}|X_{<i}=x_{<i}}(\overline{\mathcal{X}_{i,j}}) \cdot \mathbb{1}_{\{\rho_{i,j} \geq -0.5\}}\right] \leq \frac{c \cdot d}{\delta}$.
3. $\mathrm{E}_{x \sim \widetilde{P}_X}\left[\left|\{j \in [n] \colon \exists i \in [m] \ s.t. \ P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j}) < 0.9\}\right|\right] \leq \frac{c' \cdot d}{\delta}$.

Proof. Note that by definition, $\beta_{i,j} = \frac{1+\alpha_{i,j}}{1+\xi_{i,j}}$. Therefore, $\beta_{i,j} \geq 1+\lambda \implies (\alpha_{i,j} > 0.01) \vee (|\xi_{i,j}| > 0.01)$. The proof of Item 1 then follows by Fact 6.9(2) and Fact 6.13. Moreover, note that the proof of Item 3 follows by Item 2 and Fact 6.9(2) (for $r = \rho$ and $\lambda = 1/2$). Therefore, it is left to prove Item 2. Note that

$$
\begin{aligned}
d &\geq \mathrm{E}_{x\sim\widetilde{P}_X}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} D(\widetilde{P}_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}} || P_{X_{i,j}Y_{i,j}|X_{<i}=x_{<i}})\right] \qquad (72)\\
&\geq \mathrm{E}_{x\sim\widetilde{P}_X}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} \widetilde{P}_{Y_{i,j}|X_{<i}}(1)\cdot D(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1} || P_{X_{i,j}|X_{<i}=x_{<i},Y_{i,j}=1})\right]\\
&= \mathrm{E}_{x\sim\widetilde{P}_X}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} (1+\rho_{i,j})\cdot\delta_{i,j}\cdot D(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}} || P_{X_{i,j}|X_{<i}=x_{<i}})\right]\\
&\geq \mathrm{E}_{x\sim\widetilde{P}_X}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} (1+\rho_{i,j})\cdot\delta_{i,j}\cdot D(\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(\overline{\mathcal{X}_{i,j}}) || P_{X_{i,j}|X_{<i}=x_{<i}}(\overline{\mathcal{X}_{i,j}}))\right]\\
&\geq \mathrm{E}_{x\sim\widetilde{P}_X}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} (1+\rho_{i,j})\cdot\delta_{i,j}\cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\overline{\mathcal{X}_{i,j}})\right]/400,\\
&\geq \mathrm{E}_{x\sim\widetilde{P}_X}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} \delta\cdot P_{X_{i,j}|X_{<i}=x_{<i}}(\overline{\mathcal{X}_{i,j}})\cdot \mathbb{1}_{\{\rho_{i,j}\geq -0.5\}}\right]/800,
\end{aligned}
$$

which concludes the proof. Note that the one before last inequality holds by Fact 2.8 since (recall that)

$$\mathcal{X}_{i,j} = \{x_{i,j} \in \mathrm{Supp}(P_{X_{i,j}|X_{<i}=x_{<i}})\colon P_{X_{i,j}|X_{<i}=x_{<i}}(x_{i,j})/\widetilde{P}_{X_{i,j}|X_{<i}=x_{<i},E_{i,j}}(x_{i,j}) \leq 1.1\}$$

and the equality holds since for any $x_{i,j}$ it holds that

$$
\begin{aligned}
P[X_{i,j} = x_{i,j} \mid X_{<i} = x_{<i}, Y_{i,j} = 1] &= \frac{P[E_{i,j} \mid X_{\leq i} = x_{\leq i}]\cdot P[X_{i,j} = x_{i,j} \mid X_{<i} = x_{<i}]}{P[E_{i,j}]}\\
&= P[X_{i,j} = x_{i,j} \mid X_{<i} = x_{<i}],
\end{aligned}
$$

(recall that by assumption, $P[E_{i,j} \mid X_{\leq i} = x_{\leq i}] = P[E_{i,j}]$ for any fixing of $x_{\leq i}$). □

6.4.2 Proving Lemma 6.5

We now ready to prove Lemma 6.5, restated for convenience below.

Lemma 6.15 (Restatement of Lemma 6.5)**.** *Let $P, \widetilde{P}, Q, W, \mathcal{E}, Y, \delta, d$ as defined in Theorem 4.3, and let $\{C_i\}_{i=1}^m$ be the events defined in Definition 6.3. Then there exists a universal constant $c > 0$ such that if $n \geq c\cdot m/\delta$ and $d \leq \delta n/c$, then*

$$\widetilde{P}[C_1 \wedge \ldots \wedge C_m] \geq 1 - c\cdot(d+1)/\delta n.$$

Proof. The proof is divided into three parts. We prove that

1. $\widetilde{P}[G_m] = \widetilde{P}[G_1 \wedge \ldots \wedge G_m] \geq 1 - c \cdot d/\delta n$.
2. $\widetilde{P}[\widetilde{B}_1 \wedge \ldots \wedge \widetilde{B}_m \mid G_m] \geq 1 - c' \cdot d/\delta n$.
3. $\sum_{i=1}^{m} \widetilde{P}[T_i \wedge T_i' \mid G_i] \geq 1 - 2/n$.

Proving Part 1 Note that

$$\widetilde{P}[G_m] = \widetilde{P}[|\mathcal{S}_m| \geq 0.9n] \geq 1 - \widetilde{P}[|\mathcal{B}| > 0.1n]$$

where $\mathcal{B} = \mathcal{B}(x) = \biguplus_{i=1}^{n}(\mathcal{B}_i \setminus \mathcal{B}_{i-1})$, letting $\mathcal{B}_0 = \emptyset$ and

$$\mathcal{B}_i = \mathcal{B}_i(x) = \{j \in [n] \colon (|\rho_{i,j}| > 0.1) \vee (|\omega_{i,j} - 1| > 0.1) \vee (\beta_{i,j} > 1.1) \vee P_{X_{i,j}|X_{<i}=x_{<i}}(\mathcal{X}_{i,j}) < 0.9\}$$

for $i \in [m]$. By Fact 6.9(2), Fact 6.12 and Fact 6.14(1,3) it holds that

$$\mathrm{E}_{\widetilde{P}}[|\mathcal{B}|] \leq c \cdot d/\delta \tag{73}$$

for some universal constant $c > 0$. Therefore, by Markov inequality we deduce that

$$\widetilde{P}[|\mathcal{B}| > 0.1n] \leq \frac{10c \cdot d}{\delta n}. \tag{74}$$

which ends the proof of Part 1.

Proving Part 2 By definition of $\widetilde{B}_i$ it holds that

$$\begin{aligned}
\widetilde{P}[\neg\widetilde{B}_i \mid G_m] &= \mathrm{E}_{x_{<i} \sim \widetilde{P}_{X_{<i}|G_m}}[Q[\neg B_i \mid B_{<i}, X_{<i} = x_{<i}]] \qquad (75)\\
&= \mathrm{E}_{x_{<i} \sim \widetilde{P}_{X_{<i}|G_m}}[Q[J \notin \mathcal{G}_i(X_i) \mid J \in \mathcal{G}_{i-1}, X_{<i} = x_{<i}]]\\
&= \mathrm{E}_{x_{<i} \sim \widetilde{P}_{X_{<i}|G_m}}[Q[(J \notin \mathcal{J}_i) \vee (X_{i,J} \notin \mathcal{X}_{i,J}) \mid J \in \mathcal{G}_{i-1}, X_{<i} = x_{<i}]]\\
&= \mathrm{E}_{x_{<i} \sim \widetilde{P}_{X_{<i}|G_m}}[Q[J \notin \mathcal{J}_i \mid J \in \mathcal{G}_{i-1}, X_{<i} = x_{<i}] + Q[X_{i,J} \notin \mathcal{X}_{i,J} \mid J \in \mathcal{S}_i, X_{<i} = x_{<i}]]
\end{aligned}$$

where in the last equality recall that $\mathcal{S}_i = \mathcal{G}_{i-1} \bigcap \mathcal{J}_i$. In the following, fix $x_{<i} \in \mathrm{Supp}(\widetilde{P}_{X_{<i}|G_m})$. We first bound the left-hand side term with respect to $x_{<i}$. Note that by definition, for all $j \in \mathcal{G}_{i-1}$ it holds that $\omega_{i-1,j}, \beta_{i-1,j}, (1 + \rho_{i-1,j}) \in 1 \pm 0.1$ which yields that

$$\omega_{i,j} = \omega_{i-1,j} \cdot \beta_{i-1,j} \cdot \frac{1 + \tau_{i-1,j}}{1 + \rho_{i-1,j}} \leq 2(1 + \tau_{i-1,j})$$

Moreover, by the event G_m it holds that $|\mathcal{S}_i| \geq 0.9n$ and note that by definition of $\mathcal{S}_i$ it holds that $\mathcal{S}_i \subseteq \mathcal{G}_{i-1}$ and that $\omega_{i,j} \geq 0.9$ for all $j \in \mathcal{S}_i$. We deduce that

$$
\begin{aligned}
&Q[J \notin \mathcal{J}_i \mid J \in \mathcal{G}_{i-1}, X_{<i} = x_{<i}] \\
&= \frac{\sum_{j \in \mathcal{G}_{i-1} \setminus \mathcal{J}_i} \omega_{i,j}}{\sum_{j \in \mathcal{G}_{i-1}} \omega_{i,j}} \leq \frac{\sum_{j \in \mathcal{G}_{i-1} \setminus \mathcal{J}_i} \omega_{i,j}}{\sum_{j \in \mathcal{S}_i} \omega_{i,j}} \leq \frac{2 \cdot \sum_{j \in \mathcal{G}_{i-1} \setminus \mathcal{J}_i} (1 + \tau_{i,j})}{0.9 \cdot |\mathcal{S}_i|} \\
&\leq \frac{3}{n} \cdot \sum_{j \in \mathcal{G}_{i-1} \setminus \mathcal{J}_i} (1 + \tau_{i,j}) \leq \frac{6}{n} \cdot \left(|\mathcal{G}_{i-1} \setminus \mathcal{J}_i| + \sum_{j \in \mathcal{G}_{i-1} \setminus \mathcal{J}_i} \tau_{i,j} \cdot \mathbb{1}_{\{\tau_{i,j} > 1\}} \right) \\
&\leq \frac{6}{n} \cdot \left(|\mathcal{B}_i| + \sum_{j=1}^{n} \min\{|\tau_{i,j}|, \tau_{i,j}^2\} \right) \qquad (76)
\end{aligned}
$$

We now bound the right-hand side term in Equation (75) with respect to $x_{<i}$. Compute

$$
\begin{aligned}
&Q[X_{i,J} \notin \mathcal{X}_{i,J} \mid J \in \mathcal{S}_i, X_{<i} = x_{<i}] \\
&\leq \frac{2}{|\mathcal{S}_i|} \cdot \sum_{j \in \mathcal{S}_i} Q[X_{i,j} \notin \mathcal{X}_{i,j} \mid J = j, X_{<i} = x_{<i}] \\
&= \frac{2}{|\mathcal{S}_i|} \cdot \sum_{j \in \mathcal{S}_i} P_{X_{i,j} | X_{<i} = x_{<i}}(\neg \mathcal{X}_{i,j}) = \frac{2}{|\mathcal{S}_i|} \cdot \sum_{j \in \mathcal{S}_i} P_{X_{i,j} | X_{<i} = x_{<i}}(\neg \mathcal{X}_{i,j}) \cdot \mathbb{1}_{\{\rho_{i,j} > -0.5\}} \\
&\leq \frac{4}{n} \cdot \sum_{j=1}^{n} P_{X_{i,j} | X_{<i} = x_{<i}}(\neg \mathcal{X}_{i,j}) \cdot \mathbb{1}_{\{\rho_{i,j} > -0.5\}} \qquad (77)
\end{aligned}
$$

The first inequality holds since given $X_{<i}$ and given $J \in \mathcal{S}_i$, then by definition J is distributed (almost) uniformly over $\mathcal{S}_i$ (i.e., has high min entropy). The last equality holds since, by definition, for all $j \in \mathcal{S}_i$ it holds that $\rho_{i,j} > -0.5$. The last inequality holds since the event G_m implies that $|\mathcal{S}_i| \geq 0.9n$. By combining Equations (75) to (77) we deduce that

$$
\begin{aligned}
\sum_{i=1}^{m} \widetilde{P}[\neg \widetilde{B}_i \mid G_m] &\leq \mathrm{E}_{x \sim \widetilde{P}_{X|G_m}} \left[\sum_{i=1}^{m} Q[J \notin \mathcal{J}_i \mid J \in \mathcal{G}_{i-1}, X_{<i} = x_{<i}] + Q[X_{i,J} \notin \mathcal{X}_{i,J} \mid J \in \mathcal{G}_i, X_{<i} = x_{<i}] \right] \\
&\leq \frac{6}{n} \cdot \mathrm{E}_{x \sim \widetilde{P}_{X|G_m}} \left[\sum_{i=1}^{n} \left(|\mathcal{B}_i| + \sum_{j=1}^{n} \min\{|\tau_{i,j}|, \tau_{i,j}^2\} + \sum_{j=1}^{n} P_{X_{i,j} | X_{<i} = x_{<i}}(\neg \mathcal{X}_{i,j}) \cdot \mathbb{1}_{\{\rho_{i,j} > -0.5\}} \right) \right]
\end{aligned}
$$

and the proof of Part 2 follows by Part 1, Fact 6.9(1), Fact 6.14(2) and Equation (73) (recall that $|\mathcal{B}| = \sum_{i=1}^{n} |\mathcal{B}_i|$).

Proving Part 3 Assume (towards a contradiction) that $\exists i \in [m]$ with $\widetilde{P}[T_i \mid G_i] \geq \frac{1}{n^3} \geq \frac{2}{\delta n^4}$ (recall that $n \geq c \cdot m/\delta$ for a large constant c of our choice) and let $\widetilde{P}'_{X_{\leq i} Y_i} = \widetilde{P}_{X_{<i}} \prod_{j=1}^{n} \widetilde{P}_{X_{i,j} Y_{i,j} | X_{<j}}$ (namely, $\widetilde{P}'$ behaves as $\widetilde{P}$ in the first $i-1$ rows, and in row i it becomes the product of the marginals of $\widetilde{P}$ given $X_{<i}$). It holds that

$$
\begin{aligned}
d \geq D(\widetilde{P}_{X_{\leq i}Y_i}||P_{X_{\leq i}Y_i}) \geq D(\widetilde{P}_{X_{\leq i}Y_i}||\widetilde{P}'_{X_{\leq i}Y_i}) \geq D(\widetilde{P}[G_i \wedge \neg T_i]||\widetilde{P}'[G_i \wedge \neg T_i]) \\
\geq D\left(\frac{1}{\delta n^4}||\widetilde{P}'[|\gamma_i| > 1/2 \mid G_i]\right) \\
\geq D\left(\frac{1}{\delta n^4}||4 \cdot \exp(-\delta n/400)\right) \geq \frac{\delta n}{500}
\end{aligned} \tag{78}
$$

where the first inequality holds by chain rule and data processing of KL-divergence (recall that $d = \sum_{i=1}^{m} D(\widetilde{P}_{X_iY_i}||P_{X_iY_i}|\widetilde{P}_{X_{<i}}))$, the second one holds by the product case of chain rule, the third one holds by data-processing (indicator to the event $G_i \wedge \neg T_i$) and the fourth one holds by assumption (recall that $\widetilde{P}[G_i] \geq 1 - O(d/\delta n) \geq 1/2$). The one before last inequality holds by Equation (67) (under product, when G_i occurs, there is a strong concentration), and last inequality holds since n is large enough. This contradicts the assumption on d (by setting the constant there to be larger than 500). Therefore, we deduce that for all $i \in [m]$:

$$
\widetilde{P}[\neg T_i \mid G_i] \leq 1/n^3 \tag{79}
$$

Moreover, by definition of T'_i (recall that T'_i is the event that $\widetilde{P}[T_i \mid X_{<i}] \geq 1 - 1/n$), it holds that

$$
\widetilde{P}[\neg T'_i \mid G_i] \leq \frac{\widetilde{P}[\neg T_i \mid G_i]}{\widetilde{P}[\neg T_i \mid \neg T'_i \wedge G_i]} \leq \left(1/n^3\right)/(1/n) = 1/n^2. \tag{80}
$$

The proof now immediately follows by Equations (79) and (80). □

7 Lower Bound

In this section we formally state and prove Theorem 1.5, showing that Theorem 5.1 is tight for partially prefix-simulatable interactive arguments. In Section 7.1 we start by showing how random termination helps to beat [BIN97]'s counterexample, and in Section 7.2 we restate and prove Theorem 1.5 using a variant of [BIN97]'s protocol.

7.1 Random Termination Beats Counterexample of [BIN97]

In this section we exemplify the power of random termination, showing that the counterexample of [BIN97] does not apply to random-terminating verifiers. We do so by presenting [BIN97]'s counterexample against n repetitions and see how random termination helps in this case. The protocol is described below.

Protocol 7.1 ([BIN97]'s Protocol $\pi = (\mathrm{P}, \mathrm{V})$)**.**

Common input: Security parameter 1^κ *and public key* pk *.*

Prover's private input: Secret key sk*.*

Operation:

1. *Round* 1*:*
   (a) V *uniformly samples* $b \leftarrow \{0,1\}$ *and* $r \leftarrow \{0,1\}^\kappa$*, and sends* $B = \mathrm{Enc}_{pk}(b, r)$ *to* P*.*

(b) P *computes* $(b, r) = \mathrm{Dec}_{sk}(B)$ *and for any* $i \in [n-1]$*, it uniformly samples* $b'_i \in \{0,1\}$ *and* $r'_i \in \{0,1\}^\kappa$ *conditioned on* $b = \oplus_{i=1}^{n-1} b'_i$*. Then it computes* $C_i = \mathrm{Enc}_{pk}(b'_i, r'_i)$*, and sends* $(C_1, \ldots, C_{n-1})$ *to* V*.*

2. *Round* 2*:*

   (a) V *sends* (b, r) *to* P*.*

   (b) P *sends* $\big((b'_1, r'_1), \ldots, (b'_{n-1}, r'_{n-1})\big)$ *to* V*.*

3. *At the end:* V *accepts iff* $b = \oplus_{i=1}^{n-1} b'_i$*, and for any* $i \in [n-1]$*:* $C_i = \mathrm{Enc}_{pk}(b'_i, r'_i)$ *and* $B \neq C_i$*.*

Intuitively, assuming the cryptosystem is CCA2-secure, if a single instance of the protocol is run, then a prover without access to sk can only convince the honest verifier with probability $1/2$, since it must commit itself to a guess $\oplus_{i=1}^{n-1} b'_i$ of b before receiving (b, r). On the other hand, if n instances of the protocol are run in parallel, then a cheating prover can send the tuple $(C_1, \ldots, C_{n-1}) = (B_1, \ldots, B_{i-1}, B_{i+1}, \ldots, B_n)$ to V_i and then either all verifier instances accept or all verifier instances fail, the first event occurring with probability at least $1/2$.

Let's look now on a n instances that run in parallel of the protocol $\pi = (\mathrm{P}, \widetilde{\mathrm{V}})$, where $\widetilde{\mathrm{V}}$ is the random-terminating variant of V (note that this protocol has only two rounds, and therefore, a random terminating bit takes one with probability $1/2$). First, we expect that $\approx n/2$ of the verifiers abort at the first round, and with high probability at least $n/4$ of the verifiers remain active (assume that n is large enough). For a cheating prover, aborting at the first round is not an issue since it can completely simulate the aborted verifiers. However, even if a single verifier V_i aborts at the second round, then the attack presented above completely fail since the prover has no way to reveal (b_i, r_i), needed for the other verifiers. Note that the attack do succeed in case non of the verifiers abort at the second round, but the probability of this to happen is at most $2^{-n/4}$.

7.2 Proving Theorem 1.5

We now restate and prove Theorem 1.5.

Theorem 7.2 (lower bound, retstment of Theorem 1.5.)**.** *Assume the existence of CCA2-secure public-key cryptosystem. Then for every* $m = m(\kappa) \in [2, \mathrm{poly}(\kappa)]$ *and* $\varepsilon = \varepsilon(\kappa) \in [1/\mathrm{poly}(\kappa), 1/3]$ *and* $n = n(\kappa) \in [m/\varepsilon, \mathrm{poly}(\kappa)]$*, there exists an* m*-round interactive argument* (P, V) *with soundness error* $1-\varepsilon$ *such that* $(\mathrm{P}^n, \widetilde{\mathrm{V}}^n)$ *has soundness error of at least* $(1-\varepsilon)^{c \cdot n/m}$ *for some universal constant* $c > 0$*, where* $\widetilde{\mathrm{V}}$ *is the* $1/m$*-random-terminating variant of* V *(according to Definition 2.11) and* $(\mathrm{P}^n, \widetilde{\mathrm{V}}^n)$ *denotes the* n*-parallel repetition of* $(\mathrm{P}, \widetilde{\mathrm{V}})$*.*[7]

Fix large enough κ and fix m, ε, n as in the theorem statements, and let $CS = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ be a CCA2-secure public-key cryptosystem. Consider the following m-round variant (P, V) of [BIN97]'s protocol:

Protocol 7.3 (The counterexample protocol $\pi = (\mathrm{P}, \mathrm{V})$)**.**

Common input: Security parameter 1^κ *and public key* pk *.*

[7] Assuming the existence of collision-free family of hash functions and CCA2-secure cryptosystem with respect to superpolynomial adversaries, one can adopt the techniques used in [PW12] for constructing a single protocol (P, V) such that for any polynomial bounded n, $(\mathrm{P}^n, \widetilde{\mathrm{V}}^n)$ has soundness error of at least $(1-\varepsilon)^{c \cdot n/m}$. This, however, is beyond the scope of this paper.

Prover's private input: Secret key sk.

Operation:

1. *Round* 1*:*
   (a) V *flips a coin that takes one with probability* $1-3\varepsilon$ *and zero otherwise.*
   If the coin outcome is one, V *sends* $\perp$ *to* P*, accepts and the protocol terminates.*
   Else, V *uniformly samples* $b \leftarrow \{0,1\}$ *and* $r \leftarrow \{0,1\}^\kappa$*, and sends* $B = \mathrm{Enc}_{pk}(b,r)$ *to* P*.*
   (b) P *computes* $(b,r) = \mathrm{Dec}_{sk}(B)$ *and for any* $i \in [n-1]$*, it uniformly samples* $b'_i \in \{0,1\}$ *and* $r'_i \in \{0,1\}^\kappa$ *conditioned on* $b = \oplus_{i=1}^{n-1} b'_i$*. Then it computes* $C_i = \mathrm{Enc}_{pk}(b'_i, r'_i)$*, and sends* $(C_1, \ldots, C_{n-1})$ *to* V*.*
2. *Round* 2*:*
   (a) V *sends* (b,r) *to* P*.*
   (b) P *sends* $\big((b'_1, r'_1), \ldots, (b'_{n-1}, r'_{n-1})\big)$ *to* V*.*
3. *Rounds* 3 *to* m*: parties exchange dummy messages.*
4. *At the end:* V *accepts iff* $b = \oplus_{i=1}^{n-1} b'_i$*, and for every* $i \in [n-1]$*:* $C_i = \mathrm{Enc}_{pk}(b'_i, r'_i)$ *and* $B \neq C_i$*.*

Namely, Protocol 7.3 first transforms [BIN97]'s two-rounds protocol, of soundness error $1/2 + \mathrm{neg}(\kappa)$, into an m-round protocol with soundness error $1-\varepsilon$, by flipping a coin at Step 1a (for increasing the soundness error) and adding dummy rounds at the end for increasing the number of rounds (Step 3).[8]

We first note that soundness error of π is indeed low.

Claim 7.4. *The soundness error of* $\pi(1^\kappa)$ *is at most* $1-\varepsilon$*.*

Proof. Let P^* be some efficient cheating prover and let T be the event over a random execution of $(\mathrm{P}^*, \mathrm{V})$ that the outcome of the $(1-3\varepsilon, 3\varepsilon)$ bit (flipped by V at Step 1a) is 0 (i.e., V does not abort). Conditioned on T, P^* must commit itself to a guess $\oplus_{i=1}^{n-1} b'_i$ before receiving (b,r). Since the encryption scheme is CCA2-secure (which implies non-malleability), we obtain that

$$\Pr_{(pk,sk)\leftarrow \mathrm{Gen}(1^\kappa)}[(\mathrm{P}^*, \mathrm{V})(1^\kappa, pk) = 1 \mid T] \leq 1/2 + \mathrm{neg}(\kappa),$$

and hence

$$\begin{aligned}\Pr_{(pk,sk)\leftarrow \mathrm{Gen}(1^\kappa)}[(\mathrm{P}^*, \mathrm{V})(1^\kappa) = 1] &\leq \Pr[\neg T] + \Pr[T] \cdot \Pr_{(pk,sk)\leftarrow \mathrm{Gen}(1^\kappa)}[(\mathrm{P}^*, \mathrm{V})(1^\kappa, pk) = 1 \mid T] \\ &\leq 1 - 3\varepsilon + 3\varepsilon \cdot (1/2 + \mathrm{neg}(\kappa)) \\ &\leq 1 - \varepsilon.\end{aligned}$$

□

So it is left to show that the soundness error of the n parallel repetition of the random terminating variant of π is high. Let $\widetilde{\mathrm{V}}$ and $(\mathrm{P}^n, \widetilde{\mathrm{V}}^n)$ be as in the theorem statement with respect to (P, V) (Protocol 7.3) and assume without loss of generality that $\widetilde{\mathrm{V}}$ sends $\perp$ to the prover right after flipping a termination coin with outcome one. Consider the following cheating prover P^*:

[8] As in [BIN97; PW12], the soundness error holds with respect to a prover without access to sk.

Algorithm 7.5 (Cheating prover P^{n*})**.**

Input: Security parameter 1^κ*.*

Operation:

1. *Upon receiving a* n*-tuple* $(a_1, \ldots, a_n)$ *from* $\widetilde{\mathrm{V}}^n = (\widetilde{\mathrm{V}}_1, \ldots, \widetilde{\mathrm{V}}_n)$*, let* $\mathcal{S} = \{i \in [n] \colon a_i \neq \perp\}$ *(the set of active verifiers) and for* $i \notin \mathcal{S}$ *sample uniformly* $b_i \leftarrow \{0,1\}$ *and* $r_i \leftarrow \{0,1\}^\kappa$*. Then for any* $i \in \mathcal{S}$ *send* $(a'_1, \ldots, a'_{i-1}, a'_{i+1}, \ldots, a'_n)$ *to* $\widetilde{\mathrm{V}}_i$*, where* $a'_j = \begin{cases} a_j & j \in \mathcal{S} \\ \mathrm{Enc}_{pk}(b_j, r_j) & o.w \end{cases}$.
2. *If at least one verifier in* $\mathcal{S}$ *sends* $\perp$ *(after aborting at the second round), fail and abort. Otherwise, upon receiving* (b_i, r_i) *for all* $i \in \mathcal{S}$*, send the tuple* $((b_1, r_1), \ldots, (b_{i-1}, r_{i-1}), (b_{i+1}, r_{i+1}), \ldots, (b_n, r_n))$ *to* $\widetilde{\mathrm{V}}_i$*.*

Namely, P^{n*} performs [BIN97]'s attack on the verifiers that remain active after the first round. The attack, however, can only be performed if none of these active verifiers abort in the second round. Yet, we show that the probability for this to happen is high enough. The following claim conclude the proof of Theorem 7.2.

Claim 7.6. *Let* ε, m, n *as in the theorem statement, let* (P, V) *be Protocol 7.3 and let* P^{n*} *be the cheating prover described in Algorithm 7.5 (with respect to* n*). Then*

$$\Pr_{(pk,sk)\leftarrow\mathrm{Gen}(1^\kappa)}\left[(\mathrm{P}^{n*}, \widetilde{\mathrm{V}}^n)(1^\kappa, pk) = 1\right] \geq (1-\varepsilon)^{14\cdot n/m}.$$

Proof. Fix pk and let L be the random variable that denotes the value of $|\mathcal{S}|$ (the number of active verifiers after the first round) in a random execution of $(\mathrm{P}^{n*}, \widetilde{\mathrm{V}}^n)(1^\kappa, pk)$. Note that each verifier aborts with probability greater than $1-3\varepsilon$ at the first round (it can abort by the $(1-3\varepsilon, 3\varepsilon)$ coin or by the $(1/m, 1-1/m)$ random-terminating coin). Therefore, $\mathrm{E}[L] \leq 3\varepsilon n$ and we obtain by Markov's inequality that $\Pr[L \leq 6\varepsilon n] \geq 1/2$. Let G be the event that none of the verifiers abort at the second round. Note that

$$\begin{aligned} \Pr[G] &\geq \Pr[L \leq 6\varepsilon n] \cdot \Pr[G | L \leq 6\varepsilon n] \\ &\geq 1/2 \cdot (1-1/m)^{6\varepsilon n} \\ &\geq 1/2 \cdot \exp(-12\varepsilon n/m). \end{aligned} \tag{81}$$

The second inequality holds since $1 - x \geq e^{-2x}$ for $x \in [0, 1/2]$. In addition, observe that

$$\begin{aligned} \Pr\left[(\mathrm{P}^{n*}, \widetilde{\mathrm{V}}^n)(1^\kappa, pk) = 1 \mid G\right] &\geq \Pr_{(b_1,\ldots,b_n)\leftarrow\{0,1\}^n}[\oplus_{i=1}^n b_i = 0] - \mathrm{neg}(\kappa) \\ &= 1/2 - \mathrm{neg}(\kappa) \end{aligned} \tag{82}$$

and we conclude by Equations (81) and (82) that

$$\begin{aligned} \Pr\left[(\mathrm{P}^{n*}, \widetilde{\mathrm{V}}^n)(1^\kappa, pk) = 1\right] &\geq \Pr[G] \cdot \Pr\left[(\mathrm{P}^{n*}, \widetilde{\mathrm{V}}^n)(1^\kappa, pk) = 1 \mid G\right] \\ &\geq 1/2 \cdot \exp(-12\varepsilon n/m) \cdot (1/2 - \mathrm{neg}(\kappa)) \\ &\geq \exp(-14\varepsilon n/m) \\ &\geq (1-\varepsilon)^{14n/m}. \end{aligned}$$

The penultimate inequality holds since we assumed that $n \geq m/\varepsilon$, and the last one since $1 + x \leq e^x$ for any $x \in \mathbb{R}$. □

Putting it together.

Proof of Theorem 7.2. Immediate by Claim 7.6. □

Acknowledgment

We thank Chris Brzuska, Or Ordentlich and Yury Polyanskiy for very useful discussions.

References

[BHT19a] I. Berman, I. Haitner, and E. Tsfadia, "A tight parallel-repetition theorem for random-terminating interactive arguments," Electronic Colloquium on Computational Complexity, Tech. Rep. TR19-049, 2019.

[BHT19b] ——, "A tight parallel-repetition theorem for random-terminating interactive arguments," IACR Cryptol. ePrint Arch., Tech. Rep. 2019/393, 2019.

[BIN97] M. Bellare, R. Impagliazzo, and M. Naor, "Does parallel repetition lower the error in computationally sound protocols?" In *38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997*, 1997, pp. 374–383 (cit. on pp. 1, 7–9, 45–48).

[BV14] Z. Brakerski and V. Vaikuntanathan, "Efficient fully homomorphic encryption from (standard) lwe," *Journal of the ACM*, vol. 43, no. 2, pp. 831–871, 2014 (cit. on p. 8).

[CHS05] R. Canetti, S. Halevi, and M. Steiner, "Hardness amplification of weakly verifiable puzzles," in *Theory of Cryptography, Second Theory of Cryptography Conference, TCC 2005*, 2005, pp. 17–33 (cit. on p. 7).

[CL02] F. Chung and L. Lu, "Connected components in random graphs with given expected degree sequences," 2002. [Online]. Available: https://link.springer.com/content/pdf/10.1007/PL00012580.pdf (cit. on pp. 1, 15).

[CL10] K. Chung and F. Liu, "Parallel repetition theorems for interactive arguments," in *Theory of Cryptography, Sixth Theory of Cryptography Conference, TCC 2010*, 2010, pp. 19–36 (cit. on pp. 6–8).

[CP11] K.-M. Chung and R. Pass, "The randomness complexity of parallel repetition," in *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, 2011, pp. 658–667 (cit. on p. 8).

[CP15] K. Chung and R. Pass, "Tight parallel repetition theorems for public-coin arguments using kl-divergence," in *Theory of Cryptography, 11th Theory of Cryptography Conference, TCC 2015*, 2015, pp. 229–246 (cit. on pp. 1, 2, 4, 7, 40).

[Das11] A. DasGupta, *Probability for Statistics and Machine Learning. Chapter 14: Discrete Time Martingales and Concentration Inequalities*. 2011. [Online]. Available: https://www.researchgate.net/publication/226263860_Discrete_Time_Martingales_and_Concentration_Inequalities (cit. on p. 55).

[DJMW12] Y. Dodis, A. Jain, T. Moran, and D. Wichs, "Counterexamples to hardness amplification beyond negligible," in *Theory of Cryptography, 8th Theory of Cryptography Conference, TCC 2012*, 2012, pp. 476–493 (cit. on p. 2).

[DP98] I. B. Damgärd and B. Pfitzmann, "Sequential iteration arguments and an efficientzero-knowledge argument for NP," in *Annual International Colloquium on Automata, Languages and Programming (ICALP)*, 1998, pp. 772–783 (cit. on p. 1).

[DS14] I. Dinur and D. Steurer, "Analytical approach to parallel repetition," in *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, 2014, pp. 624–633 (cit. on p. 8).

[Duc16] J. Duchi, "Lecture notes for statistics 311/electrical engineering 377," 2016. [Online]. Available: https://stanford.edu/class/stats311/Lectures/full_notes.pdf (cit. on p. 54).

[DV83] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain markov process expectations for large time. iv," *Communications on Pure and Applied Mathematics*, vol. 36, no. 2, pp. 183–212, 1983 (cit. on p. 12).

[Fei91] U. Feige, "On the success probability of the two provers in one-round proof systems," in *Proceedings of the Sixth Annual Structure in Complexity Theory Conference, Chicago, Illinois, USA, June 30 - July 3, 1991*, 1991, pp. 116–123 (cit. on p. 8).

[FRS90] L. Fortnow, J. Rompel, and M. Sipser, "Errata for on the power of multi-prover interactive protocols," in *Proceedings: Fifth Annual Structure in Complexity Theory Conference, Universitat Politècnica de Catalunya, Barcelona, Spain, July 8-11, 1990*, 1990, pp. 318–319 (cit. on p. 8).

[FV02] U. Feige and O. Verbitsky, "Error reduction by parallel repetition - A negative result," *Combinatorica*, vol. 22, no. 4, pp. 461–478, 2002 (cit. on p. 8).

[Gol99] O. Goldreich, *Modern Cryptography, Probabilistic Proofs and Pseudorandomness*. Springer, 1999 (cit. on p. 1).

[Hai13] I. Haitner, "A parallel repetition theorem for any interactive argument," *SIAM J. Comput.*, vol. 42, no. 6, pp. 2487–2501, 2013. DOI: 10.1137/100810630. [Online]. Available: https://doi.org/10.1137/100810630 (cit. on pp. 1, 2, 4, 6, 7, 13, 14).

[Hol09] T. Holenstein, "Parallel repetition: Simplification and the no-signaling case," *Theory of Computing*, vol. 5, no. 1, pp. 141–172, 2009 (cit. on pp. 1, 8).

[HPWP10] J. Håstad, R. Pass, D. Wikström, and K. Pietrzak, "An efficient parallel repetition theorem," in *Theory of Cryptography, Sixth Theory of Cryptography Conference, TCC 2010*, 2010, pp. 1–18 (cit. on pp. 1, 2, 4, 6, 7, 13).

[Mos14] D. Moshkovitz, "Parallel repetition from fortification," in *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, 2014, pp. 414–423 (cit. on p. 8).

[Mul] W. Mulzer, *Chernoff bounds*. [Online]. Available: https://page.mi.fu-berlin.de/mulzer/notes/misc/chernoff.pdf (cit. on p. 12).

[Pat90] J. Patarin, "Pseudorandom permutations based on the DES scheme," in *EUROCODE '90, International Symposium on Coding Theory and Applications, Udine, Italy, November 5-9, 1990, Proceedings*, 1990, pp. 193–204. DOI: 10.1007/3-540-54303-1_131. [Online]. Available: https://doi.org/10.1007/3-540-54303-1_131 (cit. on p. 16).

[PV12] R. Pass and M. Venkitasubramaniam, "A parallel repetition theorem for constant-round arthur-merlin proofs," *TOCT*, vol. 4, no. 4, 10:1–10:22, 2012 (cit. on p. 7).

[PW12] K. Pietrzak and D. Wikström, "Parallel repetition of computationally sound protocols revisited," *Journal of Cryptology*, vol. 25, no. 1, pp. 116–135, 2012 (cit. on pp. 1, 8, 46, 47).

[PW17] Y. Polyanskiyi and Y. Wu, "Lecture notes on information theroy," 2017. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf (cit. on p. 53).

[Rao11] A. Rao, "Parallel repetition in projection games and a concentration bound," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1871–1891, 2011 (cit. on pp. 1, 8).

[Raz98] R. Raz, "A parallel repetition theorem," *SIAM J. Comput.*, vol. 27, no. 3, pp. 763–803, 1998 (cit. on pp. 1, 8).

[Ver10] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *ArXiv e-prints*, Nov. 2010. arXiv: 1011.3027 [math.PR]. [Online]. Available: https://arxiv.org/abs/1011.3027 (cit. on p. 54).

8 Missing Proofs

8.1 Proof of Proposition 3.3

Proposition 8.1 (Restatement of Proposition 3.3)**.** *Let P and Q be two distributions over $\mathcal{U}$ with $D^{\alpha}(P||Q) < \beta$. Then for every event E over $\mathcal{U}$, it holds that $Q[E] < 2 \cdot \max\{\alpha + P[E], 4\beta\}$.*

Proof. We assume that $\max\{\alpha + P[E], 4\beta\} \leq 1/2$, as otherwise the proof holds trivially. The definition of smooth KL-divergence yields the existence of randomized function F_P, F_Q satisfying

a. $\Pr_{x\sim P}[F_P(x) \neq x] \leq \alpha$,

b. $D(F_P(P)||F_Q(Q)) < \beta$, and

c. $\forall x \in \mathcal{U}$: $\mathrm{Supp}(F_P(x)) \cap \mathcal{U} \subseteq \{x\}$ and $\mathrm{Supp}(F_Q(x)) \cap \mathcal{U} \subseteq \{x\}$.

Let $E' = E \cup (\mathrm{Supp}(F_P(\mathcal{U})) \cup \mathrm{Supp}(F_Q(\mathcal{U})) \setminus \mathcal{U})$. By Item a and data processing of (standard) KL-divergence,

$$D\left(\mathbb{1}_{\{F_P(P)\in E'\}}||\mathbb{1}_{\{F_Q(Q)\in E'\}}\right) < \beta \tag{83}$$

By Items b and c,

$$\mathcal{F}_P(P)[E'] \leq \Pr_{x\sim P}[F_P(x) \neq x] + P[E] \leq \alpha + P[E] \tag{84}$$

Assume toward a contradiction that $\mathcal{F}_Q(Q)[E'] \geq 2\cdot\max\{\alpha+P[E], 4\beta\}$, then by the above equations

$$D(\alpha + P[E]||2 \cdot \max\{\alpha + P[E], 4\beta\}) < \beta \tag{85}$$

If $\alpha + P[E] > 4\beta$, then Equation (85) yields that $D(\alpha + P[E]||2(\alpha + P[E])) < \beta$. Otherwise, Equation (85) yields that $D(4\beta||8\beta) < \beta$. In both cases we get a contradiction to Fact 2.8(1). Since by Item c it holds that $Q[E] \leq \mathcal{F}_Q(Q)[E']$, we conclude that $Q[E] < 2 \cdot \max\{\alpha + P[E], 4\beta\}$. □

8.2 Proof of Proposition 3.4

Proposition 8.2 (Restatement of Proposition 3.4)**.** *Let P and Q be two distributions over a universe $\mathcal{U}$, let $\alpha \in [0,1]$ and let H be a randomized function over $\mathcal{U}$. Then $D^{\alpha}(H(P)||H(Q)) \leq D^{\alpha}(P||Q)$.*

Proof. Let (F_P, F_Q) be a pair of functions such that

1. $\Pr_{x\sim P}[F_P(x) \neq x] \leq \alpha$, and
2. $\forall x \in \mathcal{U}$: $\mathrm{Supp}(F_P(x)) \cap \mathcal{U} \subseteq \{x\}$ and $\mathrm{Supp}(F_Q(x)) \cap \mathcal{U} \subseteq \{x\}$.

We assume without loss of generality that for both $T \in \{P, Q\}$:

$$\forall x \in \mathcal{U} : \ \mathrm{Supp}(F_T(x)) \cap \mathrm{Supp}(H(x)) \subseteq \{x\}. \tag{86}$$

Indeed, since $F_T(x) \neq x$ implies $F_T(x) \notin \mathcal{U}$, one can add a fixed prefix to the value of $F_T(x)$ when $F_T(x) \neq x$ (same prefix for both $T \in \{P, Q\}$) such that Equation (86) holds. Such a change neither effect the properties of F_P and F_Q stated above, nor the value of $D(F_P(P)||F_Q(Q))$.

For $T \in \{P, Q\}$, let $G_T(y)$ be the randomized function defined by the following process:

a. Sample $x \sim T_{X|H(X)=y}$.

b. Sample $z \sim F_T(x)$.

c. If $z = x$, output y.

   Else, output z.

By construction and Equation (86), for both $T \in \{P, Q\}$:

$$\forall y \in H(\mathcal{U}) : \mathrm{Supp}(G_T(y)) \cap H(\mathcal{U}) \subseteq \{y\}. \tag{87}$$

Let $Y_T = H(T)$ and let X_T be the value of x in a random execution of $G_T(Y_T)$. It is clear that $X_T \sim T$. We note that

$$\begin{aligned} \Pr[G_P(Y_P) \neq Y_P] &= \Pr[F_P(X_P) \neq X_P] \\ &= Pr_{x\sim P}[F_P(x) \neq x] \\ &\leq \alpha. \end{aligned} \tag{88}$$

The inequality is by the assumption about F_P.

Consider the randomized function $K(z)$ that outputs $H(z)$ if $z \in \mathcal{U}$, and otherwise outputs z. It holds that

$$\begin{aligned} \Pr[K(F_T(T)) = z] &= \Pr[F_T(T) \in \mathcal{U}] \cdot \Pr[H(F_T(T)) = z | F_T(T) \in \mathcal{U}] \\ &\quad + \Pr[F_T(T) \notin \mathcal{U}] \cdot \Pr[F_T(T) = z | F_T(T) \notin \mathcal{U}] \\ &= \Pr[F_T(T) = T] \cdot \Pr[H(T) = z | F_T(T) = T] \\ &\quad + \Pr[F_T(T) \neq T] \cdot \Pr[F_T(T) = z | F_T(T) \neq T], \end{aligned}$$

where the second inequality follows from the second property of (F_P, F_Q); namely, $F_T(T) \in \mathcal{U} \iff F_T(T) = T$. Similarly,

$$\begin{aligned}\Pr[G_T(H(T)) = z] &= \Pr[F_T(X_T) = X_T] \cdot \Pr[H(X_T) = z | F_T(X_T) = X_T] \\ &\quad + \Pr[F_T(X_T) = X_T] \cdot \Pr[F_T(X_T) = z | F_T(X_T) \neq X_T]. \\ &= \Pr[F_T(T) = T] \cdot \Pr[H(T) = z | F_T(T) = T] \\ &\quad + \Pr[F_T(T) \neq T] \cdot \Pr[F_T(T) = z | F_T(T) \neq T],\end{aligned}$$

where the second inequality holds since $X_T \sim T$. Hence, we have $G_T(H(T)) \equiv K(F_T(T))$. Thus, the data-processing inequality for (standard) KL-divergence implies that

$$\begin{aligned}D(F_P(P)||F_Q(Q)) &\geq D(K(F_P(P))||K(F_Q(Q))) \\ &= D(G_P(H(P))||G_Q(H(Q))).\end{aligned} \tag{89}$$

The proof now follows by Properties (87), (88), (89) of G_P and G_Q. □

8.3 Proof of Proposition 2.9

Proposition 8.3 (Restatement of Proposition 2.9)**.** *Let X be a random variable drawn form either P or Q. Assume that $\Pr_P[|X| \leq 1] = 1$ (i.e., if X is drawn from P then $|X| \leq 1$ almost surely) and that there exist $\varepsilon, \sigma^2, K_1, K_2 > 0$ such that $\Pr_Q[|X| \leq 1] \geq 1 - \varepsilon$ and*

$$\Pr_Q[|X| \geq t] \leq K_2 \cdot \exp\left(-\frac{t^2}{K_1 \sigma^2}\right) \quad \text{for all } 0 \leq t \leq 1.$$

Then, there exists $K_3 = K_3(K_1, K_2, \varepsilon) > 0$ such that

$$\mathrm{E}_P[X^2] \leq K_3 \cdot \sigma^2 \cdot (D(P||Q) + 1).$$

Note that for $\sigma \geq 1$, the statement is trivial, and thus not interesting. We would use this proposition when $\sigma \ll 1$.

Proof. Assume that $\sigma^2 \leq 1$ and that $D(P||Q) < \infty$, since otherwise the statement is trivial. We use the following two fundamental theorems. The first theorem gives a variational characterization for divergence that is useful for bounding expected values of random variables.

Theorem 8.4 (Donsker-Varadhan; cf. [PW17, Theorem 3.5])**.** *Let P and Q be probability measures on $\mathcal{X}$ and let $\mathcal{C}$ denote the set of functions $f \colon \mathcal{X} \to \mathbb{R}$ such that $\mathrm{E}_Q[\exp(f(X))] < \infty$. If $D(P||Q) < \infty$, then*

$$D(P||Q) = \sup_{f \in \mathcal{C}} \mathrm{E}_P[f(X)] - \log \mathrm{E}_Q[\exp(f(X))].$$

In particular, for every $f \in \mathcal{C}$, it holds that

$$\mathrm{E}_P[f(X)] \leq \log \mathrm{E}_Q[\exp(f(X))] + D(P||Q).$$

The second theorem is the super-exponential moment characterization condition for sub-Gaussianity.

Theorem 8.5 (Sub-Gaussian characterization; cf. [Duc16, Theorem 3.10][9])**.** *Let X be a random variable and $\sigma^2 > 0$ be a constant. Assume that there exist $K_1', K_2' > 0$ such that*

$$\Pr[|X| \geq t] \leq K_2' \cdot \exp\left(-\frac{t^2}{K_1'\sigma^2}\right) \quad \textit{for all } t \geq 0.$$

Then, there exists $K_3' = K_3'(K_1', K_2')$ such that

$$\mathrm{E}\left[\exp\left(\frac{X^2}{K_3'\sigma^2}\right)\right] \leq e.$$

We would like to apply the above theorems to derive the proof. However, under the Q distribution X is not sub-Gaussian, since its concentration bound apply only for $0 \leq t \leq 1$. Instead, we let $\mathcal{W} = [0, 1]$, $K_2' = K_2/(1-\varepsilon)$ and observe that

$$\Pr_Q[|X| \geq t \mid |X| \in \mathcal{W}] \leq K_2' \cdot \exp\left(-\frac{t^2}{K_1\sigma^2}\right) \quad \text{for all } t \geq 0.$$

Indeed, for $t > 1$ this inequality holds trivially. For $0 \leq t \leq 1$, it holds that

$$\begin{aligned}
\Pr_Q[|X| \geq t \mid |X| \in \mathcal{W}] &\leq \frac{\Pr_Q[|X| \geq t]}{\Pr_Q[|X| \in \mathcal{W}]} \\
&\leq \frac{\Pr_Q[|X| \geq t]}{1-\varepsilon} \\
&\leq K_2' \cdot \exp\left(-\frac{t^2}{K_1\sigma^2}\right),
\end{aligned}$$

where the second inequality follows from the assumption of the proposition and since $\sigma^2 \leq 1$, and the third inequality again follows from the assumption of the proposition.

Let $K_3 = K_3'(K_1, K_2')$ from the statement of Theorem 8.5. Furthermore, note that $D(P_X||Q_{X|(|X|\in\mathcal{W})}) < \infty$, since $D(P_X||Q_X) < \infty$ and $|X| \in \mathcal{W}$ under P almost surely. Using Theorems 8.4 and 8.5, it follows that

$$\begin{aligned}
\frac{1}{K_2\sigma^2}\mathrm{E}_P[X^2] &\leq \log \mathrm{E}_Q[\exp(X^2/(K_2\sigma^2))||X| \in \mathcal{W}] + D(P_X||Q_{X|(|X|\in\mathcal{W})}) \\
&\leq \log e + D(P_X||Q_{X|(|X|\in\mathcal{W})}).
\end{aligned}$$

Finally, the proposition follows since

$$\begin{aligned}
D(P_X||Q_{X|(|X|\in\mathcal{W})}) &= \mathrm{E}_{x\sim P_X} \log \frac{P_X(x)}{Q_X(x)/\Pr_Q[|X| \in \mathcal{W}]} \\
&= D(P_X||Q_X) + \log(\Pr_Q[|X| \in \mathcal{W}]) \\
&\leq D(P_X||Q_X),
\end{aligned}$$

where in the first equality we again used that $|x| \in \mathcal{W}$ for every $x \in \mathrm{Supp}(P_X)$, so $\Pr_Q[X = x \wedge |X| \in \mathcal{W}] = Q_X(x)$ for any such x. □

[9] While the statement of [Duc16, Theorem 3.10] explicitly take $K_2' = 2$ and require that X's mean is zero, it is easy to see how to modify the proof to work with any constant K_2' and that the proof of this part does not actually use that X has a zero mean. For example, see [Ver10, Lemma 5.5] that uses $K_2' = e$ and does not assume that X has zero mean.

8.4 Proof of Lemma 2.18

Proposition 8.6 (Restatement of Lemma 2.18)**.** *Let $Y_0 = 1, Y_1, \ldots, Y_n$ be a martingale w.r.t $X_0, X_1, \ldots, X_n$ and assume that $Y_i \geq 0$ for all $i \in [n]$. Then for every $\lambda \in (0, \frac{1}{4}]$ it holds that*

$$\Pr[\exists i \in [n] \text{ s.t. } |Y_i - 1| \geq \lambda] \leq \frac{23 \cdot \mathrm{E}\left[\sum_{i=1}^{n} \min\{|R_i|, R_i^2\}\right]}{\lambda^2}$$

for $R_i = \frac{Y_i}{Y_{i-1}} - 1$, letting $R_i = 0$ in case $Y_{i-1} = Y_i = 0$.

We use the following fact.

Fact 8.7 ([Das11, Theorem 14.9])**.** *Let $Y_0 = 0, Y_1, \ldots, Y_n$ be a martingale sequence with respect to $X_0, X_1, \ldots, X_n$, and assume that $\mathrm{E}\left[Y_i^2\right] < \infty$ for all $i \in [n]$. Then for every $\lambda > 0$, it holds that*

$$\Pr\left[\max_{i \in [n]} |Y_i| \geq \lambda\right] \leq \frac{\mathrm{E}\left[\sum_{i=1}^{n} D_i^2\right]}{\lambda^2},$$

for $D_i = Y_i - Y_{i-1}$.

Proof of Lemma 2.18. Let $\mu = \mathrm{E}\left[\sum_{i=1}^{n} \min\{|R_i|, R_i^2\}\right]$ and assume without loss of generality that $\mu \leq 0.1$ (otherwise the proof holds trivially). For $i \in [n]$ let $\Delta_i = \mathrm{E}\left[R_i \cdot \mathbb{1}_{\{R_i > 1\}} \mid X_0, \ldots, X_{i-1}\right]$, let

$$\widehat{R}_i = \begin{cases} R_i \cdot \mathbb{1}_{\{|R_i| \leq 1\}} + \Delta_i & \Delta_i \leq 1 \\ 0 & \textit{otherwise}, \end{cases}$$

and let $\widehat{S}_i = \sum_{j=1}^{i} \widehat{R}_j$. Note that for any $i \in [n]$ and a fixing of $X_0, \ldots, X_{i-1}$ such that $\Delta_i \leq 1$, it holds that

$$\begin{aligned} \mathrm{E}\left[\widehat{S}_i \mid X_0, \ldots, X_{i-1}\right] - \widehat{S}_{i-1} &= \mathrm{E}\left[\widehat{R}_i \mid X_0, \ldots, X_{i-1}\right] \\ &= \mathrm{E}\left[R_i \cdot \mathbb{1}_{\{|R_i| \leq 1\}} + \Delta_i \mid X_0, \ldots, X_{i-1}\right] \\ &= \Pr[|R_i| \leq 1 \mid X_0, \ldots, X_{i-1}] \cdot \mathrm{E}[R_i \mid X_0, \ldots, X_{i-1}, (|R_i| \leq 1)] \\ &\quad + \Pr[R_i > 1 \mid X_0, \ldots, X_{i-1}] \cdot \mathrm{E}[R_i \mid X_0, \ldots, X_{i-1}, (R_i > 1)] \\ &= \mathrm{E}[R_i \mid X_0, \ldots, X_{i-1}] = 0. \end{aligned}$$

The penultimate equality holds since $R_i \geq -1$. By definition, for any fixing of $X_0, \ldots, X_{i-1}$ such that $\Delta_i > 1$, it holds that $\widehat{R}_i = 0$. Hence, $\mathrm{E}\left[\widehat{S}_i \mid X_0, \ldots, X_{i-1}\right] = \widehat{S}_{i-1}$ also for any such fixing. Thus, the sequence $\widehat{S}_1, \ldots, \widehat{S}_n$ is a martingale with respect to $X_1, \ldots, X_n$ for any fixing of X_0.

By Fact 8.7,

$$\forall \beta > 0 : \ \Pr\left[\max_{i \in [n]} |\widehat{S}_i| \geq \beta\right] \leq \frac{\mathrm{E}\left[\sum_{i=1}^{n} \widehat{R}_i^2\right]}{\beta^2} \tag{90}$$

In addition, note that

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{n}\Delta_i\right] &= \sum_{i=1}^{n}\mathrm{E}_{X_0,\ldots,X_{i-1}}\left[\mathrm{E}\left[R_i\cdot\mathbb{1}_{\{R_i>1\}}\mid X_0,\ldots,X_{i-1}\right]\right] \\ &\leq \sum_{i=1}^{n}\mathrm{E}_{X_0,\ldots,X_{i-1}}\left[\mathrm{E}\left[\min\{|R_i|,R_i^2\}\mid X_0,\ldots,X_{i-1}\right]\right] \\ &= \mu.\end{aligned} \tag{91}$$

Therefore,

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{n}\widehat{R}_i^2\right] &\leq \mathrm{E}\left[\sum_{i=1}^{n}\left(R_i\cdot\mathbb{1}_{\{|R_i|\leq 1\}}+\Delta_i\cdot\mathbb{1}_{\{\Delta_i\leq 1\}}\right)^2\right] \\ &\leq 2\cdot\mathrm{E}\left[\sum_{i=1}^{n}R_i^2\cdot\mathbb{1}_{\{|R_i|\leq 1\}}\right]+2\cdot\mathrm{E}\left[\sum_{i=1}^{n}\Delta_i^2\cdot\mathbb{1}_{\{\Delta_i\leq 1\}}\right] \\ &\leq 2\mu+2\cdot\mathrm{E}\left[\sum_{i=1}^{n}\Delta_i\cdot\mathbb{1}_{\{\Delta_i\leq 1\}}\right]\leq 2\mu+2\cdot\mathrm{E}\left[\sum_{i=1}^{n}\Delta_i\right] \\ &\leq 4\mu.\end{aligned} \tag{92}$$

The last inequality holds by Equation (91). Combining Equations (90) and (92) yields that

$$\forall\beta>0:\ \Pr\left[\max_{i\in[n]}|\widehat{S}_i|\geq\beta\right]\leq 4\mu/\beta^2 \tag{93}$$

Let $S_i=\sum_{j=1}^{i}R_j$. Note that for any $i\in[n]$:

$$\mathrm{E}\left[\max_{i\in[n]}\{|S_i-\widehat{S}_i|\}\right]\leq\mathrm{E}\left[\sum_{i=1}^{n}R_i\cdot\mathbb{1}_{\{R_i>1\}}\right]+\mathrm{E}\left[\sum_{i=1}^{n}\Delta_i\right]\leq 2\mu,$$

the last inequality holds by Equation (91). Hence, by Markov inequality

$$\forall\beta>0:\ \Pr\left[\max_{i\in[n]}\{|S_i-\widehat{S}_i|\}\geq\beta\right]\leq 2\mu/\beta \tag{94}$$

A Markov inequality yields that for any $i\in[n]$:

$$\Pr\left[|R_i|>\frac{1}{2}\right]=\Pr\left[\min\{|R_i|,R_i^2\}>\frac{1}{4}\right]\leq 4\cdot\mathrm{E}\left[\min\{|R_i|,R_i^2\}\right] \tag{95}$$

Let E be the event that $|R_i|\leq\frac{1}{2}$ for all $i\in[n]$. By Equation (95) and a union bound:

$$\Pr[E]\geq 1-4\mu \tag{96}$$

Since $Y_i=\prod_{j=1}^{i}(1+R_j)$, then conditioned on E , we can use the inequality $e^{x-x^2}\leq 1+x\leq e^x$ for $x\in(0,\frac{1}{2})$ to deduce that

$$e^{\widehat{S}_i-|S_i-\widehat{S}_i|-\sum_{i=1}^{n}R_i^2}\leq e^{S_i-\sum_{i=1}^{n}R_i^2}\leq Y_i\leq e^{S_i}\leq e^{\widehat{S}_i+|S_i-\widehat{S}_i|} \tag{97}$$

Note that if E happens, and $\max_{i\in[n]}\{|\widehat{S}_i|\} \le \frac{1}{2}\ln\frac{1}{1-\lambda}$, and $\max_{i\in[n]}\{\left|S_i-\widehat{S}_i\right|\} \le \frac{1}{4}\ln\frac{1}{1-\lambda}$ and $\sum_{i=1}^n R_i^2 \le \frac{1}{4}\ln\frac{1}{1-\lambda}$, then for every $i\in[n]$:

$$1-\lambda = e^{-\ln\frac{1}{1-\lambda}} \le Y_i \le e^{\frac{3}{4}\cdot\ln\frac{1}{1-\lambda}} = \frac{1}{(1-\lambda)^{3/4}} \le 1+\lambda, \tag{98}$$

the last inequality holds since $\lambda\in(0,\frac{1}{4}]$. The proof follows since the probability that one of the conditions above does not happen is at most

$$\begin{aligned}
&\Pr[\neg E] + \Pr\left[\sum_{i=1}^n R_i^2 > \frac{1}{4}\ln\frac{1}{1-\lambda} \mid E\right] + \Pr\left[\max_{i\in[n]}\{|S_i|\} > \frac{1}{2}\ln\frac{1}{1-\lambda}\right] + \Pr\left[\max_{i\in[n]}\{\left|S_i-\widehat{S}_i\right|\} > \frac{1}{4}\ln\frac{1}{1-\lambda}\right] \\
&\quad\le 4\mu + \frac{\Pr\left[\sum_{i=1}^n \min\{|R_i|, R_i^2\} > \frac{1}{4}\ln\frac{1}{1-\lambda}\right]}{\Pr[E]} + \frac{16\mu}{\ln^2\frac{1}{1-\lambda}} + \frac{8\mu}{\ln\frac{1}{1-\lambda}} \\
&\quad\le 4\mu + \frac{4\mu}{(1-4\mu)\cdot\ln\frac{1}{1-\lambda}} + \frac{16\mu}{\ln^2\frac{1}{1-\lambda}} + \frac{8\mu}{\ln\frac{1}{1-\lambda}} \\
&\quad\le \frac{\mu}{4\lambda^2} + \frac{2\mu}{\lambda^2} + \frac{16\mu}{\lambda^2} + \frac{4\mu}{\lambda^2} \le \frac{23\mu}{\lambda^2}.
\end{aligned}$$

The first inequality holds by Equations (93), (94) and (96), the second one by Equation (96) and by Markov inequality, and the third one holds since $\mu\le 0.1$, $\lambda\in(0,\frac{1}{4}]$ and since $\ln\frac{1}{1-\lambda}\ge\lambda$ and $\ln^2\frac{1}{1-\lambda}\ge\lambda^2$ for $\lambda\in(0,\frac{1}{4}]$. □

8.5 Proof of Proposition 2.19

Proposition 8.8 (Restatement of Proposition 2.19)**.** *Let $Y_0=1, Y_1,\ldots,Y_n$ be a martingale w.r.t $X_0,X_1,\ldots,X_n$ where $Y_i\ge 0$ for all $i\in[n]$. Let $Z_1,\ldots,Z_n$ and $T_1,\ldots,T_n$ be sequences of random variables satisfying for all $i\in[n]$: (1) $Y_i = Y_{i-1}\cdot(1+Z_i)/(1+T_i)$, and (2) T_i is a deterministic function of $X_0,X_1,\ldots,X_{i-1}$. Then*

$$\Pr[\exists i\in[n] \text{ s.t. } |Y_i-1|\ge\lambda] \le \frac{150\cdot\mathrm{E}\left[\sum_{i=1}^n\left(\min\{|Z_i|,Z_i^2\}+\min\{|T_i|,T_i^2\}\right)\right]}{\lambda^2}$$

Proof. Let $\widetilde{Y}_0,\widetilde{Y}_1,\ldots,\widetilde{Y}_n$ be the random variables such that for all $i\in[n]$, $\widetilde{Y}_i = Y_{j_{\min}-1}$ where $j_{\min}\in[i]$ is the value with $|T_1|,\ldots,|T_{j-1}|\le 0.1, |T_{j_{\max}}|>0.1$, letting $j_{\min}=i+1$ (i.e., $\widetilde{Y}_i=Y_i$) in case $|T_1|,\ldots,|T_i|\le 0.1$. Since T_i is a deterministic function of $X_0,\ldots,X_{i-1}$, then $\mathrm{E}\left[\widetilde{Y}_i \mid X_0,\ldots,X_{i-1}\right] = \widetilde{Y}_{i-1}$ (namely, $\widetilde{Y}_0,\widetilde{Y}_1,\ldots,\widetilde{Y}_n$ are martingale w.r.t $X_0,\ldots,X_n$). Lemma 2.18 yields that

$$\Pr\left[\exists i\in[n] \text{ s.t. } \left|\widetilde{Y}_i-1\right|\ge\lambda\right] \le \frac{23\cdot\mathrm{E}\left[\sum_{i=1}^n\min\{\left|\widetilde{R}_i\right|,\widetilde{R}_i^2\}\right]}{\lambda^2}, \tag{99}$$

where $\widetilde{R}_i = \widetilde{Y}_i/\widetilde{Y}_{i-1}-1$. By definition, for any fixing of $X_0,\ldots,X_n$ with $j_{\min}\in[i]$ it holds that $\widetilde{R}_i=0$, and for any fixing with $j_{\min}=i+1$ it holds that $\widetilde{R}_i=(Z_i-T_i)/(1+T_i)$. In the latter case, it holds that

$$\left|\widetilde{R}_i\right| \le (|Z_i|+|T_i|)/(|1+T_i|) \le (|Z_i|+|T_i|)/0.9 \le 2(|Z_i|+|T_i|),$$

and that

$$\left|\widetilde{R}_i^2\right| \leq \left(2Z_i^2 + 2T_i^2\right)/(1+T_i)^2 \leq \left(2Z_i^2 + 2T_i^2\right)/0.9^2 \leq 3\left(Z_i^2 + T_i^2\right),$$

where in the first inequality we used the fact that $(a+b)^2 \leq 2a^2 + 2b^2$. Overall, we deduce that for all $i \in [n]$:

$$\min\{\left|\widetilde{R}_i\right|, \widetilde{R}_i^2\} \leq 3\min\{|Z_i| + |T_i|, Z_i^2 + T_i^2\} \leq 6\big(\min\{|Z_i|, Z_i^2\} + \min\{|T_i|, T_i^2\}\big), \tag{100}$$

where in the last inequality we use the fact that $\min\{x+y, x^2+y^2\} \leq 2\min\{x, x^2\} + 2\min\{y, y^2\}$ for any $x, y \geq 0$. By Equations (99) and (100) we deduce that

$$\Pr\Big[\exists i \in [n] \text{ s.t. } \left|\widetilde{Y}_i - 1\right| \geq \lambda\Big] \leq \frac{138 \cdot \mathrm{E}\big[\sum_{i=1}^n \big(\min\{|Z_i|, Z_i^2\} + \min\{|T_i|, T_i^2\}\big)\big]}{\lambda^2}, \tag{101}$$

By Markov inequality, for any $i \in [n]$ it holds that

$$\Pr[|T_i| \geq 0.1] = \Pr\big[\min\{|T_i|, T_i^2\} \geq 0.01\big] \leq 100 \cdot \mathrm{E}\big[\min\{|T_i|, T_i^2\}\big]$$

and therefore

$$\begin{aligned}\Pr\Big[\exists i \in [n] \text{ s.t. } \widetilde{Y}_i \neq Y_i\Big] &\leq \Pr[\exists i \in [n] \text{ s.t. } |T_i| \geq 0.1] \\ &\leq 100 \cdot \mathrm{E}\left[\sum_{i=1}^n \min\{|T_i|, T_i^2\}\right] \leq \frac{7 \cdot \mathrm{E}\big[\min\{|T_i|, T_i^2\}\big]}{\lambda^2}\end{aligned} \tag{102}$$

The proof now follows by Equations (101) and (102). □