We thank the reviewers for their constructive and positive comments. They will improve the quality of the paper.

About motivation and potential practical applications (Reviewers 1 and 3). The study of the generalization of this learning problem from real-valued distributions (i.i.d. setup) to Markov chains is interesting in itself from a theoretical perspective: In contrast to our studied problem where various regimes appear as the budget varies, in the i.i.d. case only a single regime exists. Markov chains have been successfully used for modeling a broad range of practical problems, and their success makes "active learning in Markov chains" relevant in practice. Furthermore, there are practical applications in reinforcement learning (RL) and in rested Markov bandits, for which our results could prove beneficial. As an instance in RL, we mention the problem of "active exploration in MDPs" (see [28]), where the task is to estimate the transition kernel of an unknown MDP uniformly well over state-action space, using a budget of n samples. For the case of ergodic MDPs, each policy in the MDP defines an ergodic chain, and hence, the leaning task becomes actively learning multiple Markov chains (we also note that compared to the setup in the present paper, active learning in MDPs poses more challenges, as one has to consider a subset of all policies due to overlap among them. However, we believe that our contribution could be beneficial for researchers in the RL community studying problems related to active learning and exploration in MDPs). We may also refer to applications falling in the framework of rested Markov bandits, for example channel allocation in wireless communication systems where a given channel's state follows a Markov chain (e.g., Gilbert-Elliot channel model). Active learning in Markov chains is a relevant problem for such applications, and we believe our contributions could serve as a technical tool for these applications. We agree to strengthen the motivation of studying this problem and to widen the scope of the paper in view of this discussion.

About the use of empirical stationary distribution in the loss function (Reviewer 1). The intention for using the term "less meaningful" is partly illustrated in the paper (lines 207–218). We provide further detailed explanation below, and agree to rewrite the corresponding part in Section 2.3, in view of the following discussion, so as to further clarify the motivation of using $\hat{\pi}_{k,n}$. We aim to derive performance guarantees on the algorithm's loss that hold with high probability (for $1-\delta$ portions of the sample paths of the algorithm for a given δ), as opposed to those holding only in expectation. To this end, the loss L_n (which uses $\hat{\pi}_{k,n}$) is more natural and meaningful than L''_n (which uses π_k; see line 189) as L_n penalizes the algorithm's performance by the relative visit counts of various states in a given sample path (through $\hat{\pi}_{k,n}$), and not by the expected value of these. This matters a lot in the small-budget regime ($n < n_{\text{cutoff}}$), where $\hat{\pi}_{k,n}$ could differ significantly from π_k — Otherwise when $n \geq n_{\text{cutoff}}$, $\hat{\pi}_{k,n}$ is well concentrated around π_k with high probability. Reiterating the discussion in Section 2.3, let us consider the small-budget regime, and some state x where $\pi_k(x)$ is not small. In the case of L_n, using $\hat{\pi}_{k,n}$ we penalize the performance by the mismatch between $\widehat{P}_{k,n}(x,\cdot)$ and $P_k(x,\cdot)$, weighted proportionally to the number of rounds the algorithm has actually visited x. In contrast, in the case of L''_n, weighting the mismatch proportionally to $\pi_k(x)$ does not seem reasonable since in a given sample path, the algorithm might not have visited x enough even though $\pi_k(x)$ is not small.

Minor comments. About chains with $\sum_x G_k(x) = 0$ (Reviewer 2): There exists ergodic chains with $\sum_x G_k(x) = 0$. The definition of the Gini index implies that such chains are necessarily deterministic (or degenerate), i.e. their transition matrices belong to $\{0,1\}^{S\times S}$. One example is a deterministic cycle with S nodes. So by assuming $\sum_x G_k(x) > 0$, the analysis of Theorem 2 indeed excludes degenerate ergodic chains (satisfying $\sum_x G_k(x) = 0$). In other words, the theorem is valid for *almost* all ergodic chains. We note however that the assertion of Theorem 1 still holds even if $\sum_x G_k(x) = 0$. We will provide a footnote in page 7 to clarify this.

About estimator for empirical stationary distribution (Reviewer 2): This is indeed a nice remark. Our algorithm and proofs do not rely on this fact, and we will include a remark on this in the paper. We also note that we use empirical estimate $\hat{\pi}_{k,n}$ of π_k in L_n as it naturally corresponds to the occupancy of various states according to a given sample path, and hence, its use can be intuitively justified.

About Remark 1 (Reviewer 2): The proof of Theorem 1 uses *entry-wise* concentration of $\widehat{P}_{k,n}$ around P_k, under the event C (which occurs with probability greater than $1-\delta$); the proof does not rely on any *trajectory-wise* concentration. As a result, the theorem is valid even if irreducibility and aperiodicity are dropped. Moreover, the proof does not use the arguments in the proof of Theorem 2, which require $\sum_x G_k(x) > 0$. Hence, Theorem 1 is valid even for deterministic ergodic chains for which $\sum_x G_k(x) = 0$. We agree to make Remark 1 more precise in view of this discussion.

About sketch proof of Lemma 1 (Reviewer 3): We explain the second step of the proof with more details.

Typos (all reviewers): We will fix typos. Thanks a lot for constructive comments!