Reviewer #1: Thank you for the insightful comments. We are encouraged you find our paper a significant contribution.

"*many hyperparameters...* " Our results hold even if the parameters $i_{\max}, c, b$ are set above the values suggested, or if the parameters η_0, C' are set below the values suggested in Theorem A.7. In the simulations, we found a fixed batch size of $b = 64$ and a step size of $\eta_t = \frac{\eta_0}{c+t}$ (without the resetting step) attains a marginal accuracy equal to that attained by the Pólya-Gamma method specialized to logistic regression. In this case, there are just two hyperparameters (η_0, c) to tune.

"*f_t are not random variables"* In some applications, one observes iid random variables and defines functions f_t based on them, which are then iid random functions. For example, for logistic regression, an iid observation (u_t, y_t) leads to iid random functions $f_t(\theta) = -\log(1/(1 + e^{-y_t u_t^\top \theta}))$ (lines 179–186). However, our main results (Theorems 2.1 and 2.2) do not require the f_t's to be iid random.

"*specialized method already exists"* The specialized method (Pólya-Gamma sampler) has running time at each epoch that scales linearly with t, while our algorithm scales as $\text{polylog}(t)$. Moreover, it is unknown if it attains TV-error ε in time polynomial in $\frac{1}{\varepsilon}$, t, d, and other problem parameters.

"*authors could comment on other possible scenarios"* We focused on logistic regression as it is one of the most common applications. Our method also applies to other log-concave distributions (exponential, Laplace, Dirichlet, gamma, beta, chi-squared, etc.). Another potential application is for online inference for Gaussian processes (for example, an online version of [Filippone and Engler, ICML 2015]); we will add a discussion on this to the paper.

Reviewer #2: Thank you for the valuable feedback. We are encouraged you find our paper interesting and substantial.

"*main motivation ... estimate integrals"* We agree that Bayesian inference is an important motivation of our work. However, there are also interesting applications that make use of sampling but not integration, such as reinforcement learning (via Thompson sampling), or online optimization. We describe these applications in lines 44–50 of our paper.

"*Do you have a CLT for your chain?"* We do not have a CLT for our chain, but believe it is possible to use our results to prove a CLT. We would do this by modifying our Lemma A.3 to show that the samples generated by our algorithm stay within a Euclidean ball of radius $r > 0$ with probability roughly $1 - e^{-r}$, and then applying our bound for convergence in TV distance. The rate would not be $\sqrt{i}$ (we use i to represent the number of Markov chain steps). Although MALA can attain a $\sqrt{i}$ rate, it has the serious drawback in the online setting that the number of gradients to compute at each step scales linearly with the number of component functions t, while our algorithm scales poly-logarithmically with t.

"*experimental comparison...* " At your suggestion, we performed initial experiments comparing the full Laplace approximation to SAGA-LD on online logistic regression, and found they attain marginal accuracy within 0.003 of each other. We note that the full Laplace approximation currently requires optimizing a sum of t functions, which has runtime that scales linearly with t at each epoch, while our method only scales as $\text{polylog}(t)$. It is an interesting open problem to compute the Laplace approximation with runtime scaling as $\text{polylog}(t)$. In previous experiments, we found the marginal accuracy for our SAGA-LD algorithm to be 0.921, for *online* Laplace to be 0.571, and for SGLD to be 0.442 (Figure 1 in the appendix). As part of our reorganization, we will include a section on experiments in the main body of the paper.

"*genuinely interested in the results... massive block of text"* We are very sorry we caused you discomfort in reading the paper. We will ensure the final version is self-contained and friendly to the reader. Specifically, we will (1) move the offline sampling results (Section 2.3) to the appendix, (2) streamline the related work section, (3) shorten the proof overview, and (4) use the freed-up space to include more examples and intuition, and give full statements of theorems.

"*contribution seems both interesting and substantial, but I do not think NIPS is the right outlet."* We are glad that you find the result interesting, and agree that the format of the paper should be improved for a NeurIPS audience. We will do this using the steps above, and welcome any other suggestions you may have. We hope you will consider increasing your score and supporting the paper.

Reviewer #3: Thank you for the helpful suggestions to improve the flow of the paper, and for the encouraging review.

"*not clear why Section 2.3 exists"* The offline sampling problem is well studied by the ML community. (See e.g. [DRW+16, CFM+18] in our paper's references.) We show our algorithm can succeed in this setting under weaker conditions than in the literature, for a class of weakly convex log-densities under a cold start X_0. If the reviewers find it less interesting than the online problem, we will move Section 2.3 to the appendix to better focus on the online problem.

"*not clear why t should start from 1"* We will explain that f_0 can be thought of as a prior before stating Problem 1.1.

"*title of Section 3 is misleading"* We will change the title of Section 3 to "Algorithm for online sampling."

"*The paper should be reorganized to improve clarity...* " We will reorganize the paper's main body in the order you suggested, shorten the proof overview (moving the more technical parts to the appendix), and make sure all terms are explained before stating the main results. We hope you will consider increasing your score and supporting the paper.