Dear reviewers, first of all we would like to thank you for taking the time to review our paper during those challenging times! Answers to your questions are in place.

Exponential family of stochastic matrices: We sketch two examples illustrating that one-parameter exponential families of stochastic matrices generalize all the applications of exponential families of discrete probability distributions.

1. Take example 1 from the paper. As discussed in the paper this generalizes the Bernoulli exponential family. In the i.i.d. bandit model one would have K Bernoulli processes, $\mathrm{Ber}(p_{\theta_1}), \ldots, \mathrm{Ber}(p_{\theta_K})$. In the Markovian bandit model under consideration one has K Markov processes specified by K stochastic matrices $P_{\theta_1}, \ldots, P_{\theta_K}$, and each row of them is specified by a coin flip as discussed in example 1. Those K Markovian processes form a generalization of the K Bernoulli processes.
2. In the same spirit as before, we will sketch an approximation for the generalization of the Poisson exponential family, which is useful for count data. Due to the finite state space, approximate the $\mathrm{Pois}(\lambda)$ distribution, with $\mathrm{Bin}(n, \lambda/n)$, and n large enough so that the two distributions are ϵ-close with respect to the total variation distance. Now our state space is $S = \{0, \ldots, n\}$. For the generator stochastic matrix we pick $n+1$ row distributions $\mathrm{Bin}(n, \lambda_0/n), \ldots, \mathrm{Bin}(n, \lambda_n/n)$ which are approximately $\mathrm{Pois}(\lambda_0), \ldots, \mathrm{Pois}(\lambda_n)$, and for the Markovian bandit model we tilt the generator stochastic matrix by parameters $\theta_1, \ldots, \theta_K$, to obtain K Markovian processes, each of them giving rewards and transitioning according to approximate Poisson distributions. In the i.i.d. case we would have just a single $\mathrm{Pois}(\lambda)$ distribution, and we would produce K tilts which would correspond to the distributions of the K arms.

The problem with generalizing even further to countably infinite state spaces, or even continuous state spaces is the peculiar behavior of eigenvalues and eigenfunctions (which may not even exist) on infinite dimensional spaces.

Round-robin scheme: As discussed in the paper the round-robin idea dates back to Lai and Robbins, although forgotten nowadays. In this paper:

1. We use different statistics to make the scheme computable in the case of multiple-plays and Markovian rewards (note that coming up with a computable algorithm in the case of multiple-plays is the motivation of [18] Komiyama, Honda, Nakagawa, where they study Thompson sampling for multiple plays and i.i.d. Bernoulli rewards).
2. We provide a finite-time analysis (as opposed to asymptotics in prior work).
3. Through experimental results we bring to the attention of the research community that this type of scheme is computationally more efficient, and equally as effective as the status quo of calculating UCB or KL-UCB indices for each arm at each round. In particular for KL-UCB type of indices the computational improvement can be quantified as $O(K + \log 1/\epsilon)$ (this paper) vs $O(K \log 1/\epsilon)$ (KL-UCB paper) cost per round, where K is the number of arms, and ϵ is the quality of the KL divergence inverses that we're interested in.

Reviewer #3 we don't think that your suggestion as an alternative to the round-robin scheme works. For instance take $B = T/2$, where T is the time horizon. Then you're claiming that by calculating UCB scores just twice, the regret might only get double, which is clearly wrong. Additionally, the generator stochastic matrix P and the function f need not be know for the algorithm. All need to be known is what declared as parameters in the preamble of the algorithm, so in particular the only thing need to be known from the family is the KL divergence rate function. Finally, we could even eliminate the presence of δ, but we decided to keep it as knob to tune the algorithm. In particular for the experiments we tuned δ by playing around with several values in the range $(0, 1/K)$.

Maximal inequality: The workhorse behind a maximal inequality is typically some variant of Doob's martingale inequality. For our paper we have reviewed the literature on martingale methods to derive deviation inequalities for Markov chains or more general dependent processes [8, 15, 19, 24, 25, 26, 27, 30, 33, 34], and to the best of our knowledge none of them seems capable to deliver a maximal inequality, due to the fact that they don't use the exponential martingale (lemma 1), but instead they use Doob's or Dynkin's martingale. For your interest, A. Kontorovich is one of the authors of [15] and also an AC, so maybe you could consult him for an extra opinion?