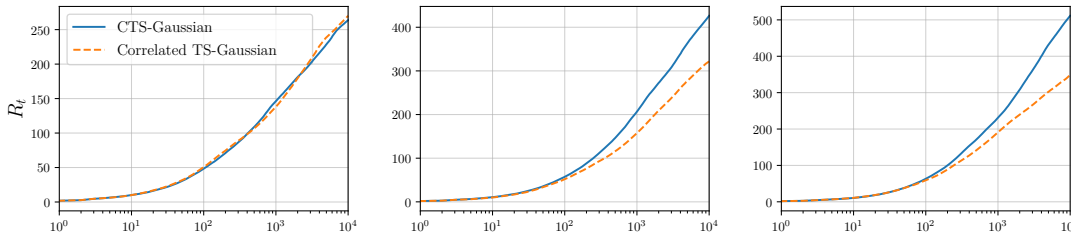

Figure 1: Comparison with correlated prior sampling for the $K_{4,4}$ matching problem, with $c = 0, 0.5, 1$.

We want to thank the reviewers for their useful and positive feedbacks. We answer their comments/questions in the following paragraphs, that will be incorporated in the paper.

Correlated vs independent prior We briefly discussed the use of a correlated prior in future work and also in footnote 3, mentioning that the policy would perform better than using an independent prior. We ran additional empirical comparisons to assess this, plotting the results in Figure 1. As expected, the correlated prior policy is better (when outcomes are correlated). This motivates the theoretical study of such policy for future work. However, some new challenges are raised, as we see in the following paragraph.

Importance of using a factorized prior in our analysis A factorized prior allows us to bound the filtered regret against the event $\mathfrak{S}_t(Z) \wedge \mathfrak{T}_t(Z)$ (see p. 13, beginning of step 4). More precisely, we count the number of rounds needed for $\mathfrak{S}_t(Z) \wedge \neg\mathfrak{T}_t(Z)$ to occur for the q-th time. Under such event, the arms of Z are all observed, i.e., the corresponding priors are updated and counters are thus lower bounded by q. The importance of having a factorize prior resides in bounding the expected number of rounds for $\mathfrak{S}_t(Z) \wedge \neg\mathfrak{T}_t(Z)$ to occur. Indeed, this is (essentially) equal to $\mathbb{E}[1/\mathbb{P}[\neg\mathfrak{T}_t(Z)|\mathfrak{S}_t(Z)]]$. Since $\mathfrak{S}_t(Z)$ and $\mathfrak{T}_t(Z)$ are independent, this is further equal to $\mathbb{E}[1/\mathbb{P}[\neg\mathfrak{T}_t(Z)]]$, allowing us to conduct our analysis by showing that this last quantity is exponentially decreasing in q (so summable over q). To the best of our knowledge, it is unknown how to get such a bound when $\mathfrak{S}_t(Z)$ and $\mathfrak{T}_t(Z)$ are not independent.

Technical contributions for cts-beta Although the gain in the upper bound for the beta prior might be considered as marginal, we want to stress on the asymptotic (quasi) optimality of the new bound (considering that the $\log^2(m)$ factor is negligible compared to n). We also fixed some technical issues in the proof of the previous bound; as a consequence, we believe that, although our work on CTS-BETA might appear somewhat incremental over Wang and Chen [2018], it brings essential clarifications to the literature.

Technical contributions for cts-gaussian Although the CTS-GAUSSIAN policies[1] are natural and essentially not new, our contribution is in their analyses, that are non-incremental. In particular, the stochastic dominance method is completely new, and allows us to convert correlated outcomes into independent Gaussian ones. Independence is crucial to be able to factorize the expectation $\mathbb{E}[1/\mathbb{P}[\neg\mathfrak{T}_t(Z)]]$. To the best of our knowledge, asymptotic quasi optimal analysis only exists for a restrictive class of action spaces (matroid) or outcomes distributions (independent). Dealing with both a general action space and outcome distribution is challenging, and represents our main contribution.

Motivation for sub-Gaussian outcomes In the same way as boundedness generalizes to sub-Gaussianity in 1d, we have that if $\mathbf{X}$ is a.s. in a compact $\mathcal{K}$, it is $\mathbf{C}$-sub-Gaussian, with $\mathbf{C}$ built from the John's ellipsoid of $\mathcal{K}$. In this case, D_i is computed with a linear maximization over $\mathcal{A}$ (see footnote 4). In particular, $\mathcal{K} = B_{\ell_\infty}(0,1)$ gives $D_i = m$, and $\mathcal{K} = B_{\ell_2}(0,1)$ gives $D_i = 1$. We can also use other structures on the outcomes to have D_i, such as negative dependence (as in our shortest path experiments).

Comparison to previous work We will add a comparison to Thompson Sampling for the MNL Bandit in the revised version, mentioning that the use of a common posterior distribution boosts the probability that the samples are all optimistic, and can thus greatly improve the constant term in our bound. However, as we saw, obtaining a quasi optimal gap dependent bound for such correlated sampling is an open question.

[1] $\beta > 1$ is an artefact of the analysis and can in practice be taken equal to 1 (as we did in our experiments).