Supplementary Material: Efficient Sampling for Bipartite Matching Problems

1 Proof of Proposition 1

Proposition 1. *For any reference permutation σ and any choice of matching probabilities that satisfy $\sum_{v_j \in V \setminus \pi_{1:t-1}} p(v_j | u_{\sigma(t)}, \pi_{1:t-1}) = 1$, the distribution given by: $Q(\pi|\sigma) = \prod_{t=1}^{N} p(v_{\pi(\sigma(t))} | u_{\sigma(t)}, \pi_{1:t-1})$ is a valid probability distribution over assignments.*

Proof. We prove this by induction, the proposition holds for $N = 2$ ($N = 1$ case is trivial) since:

$$\sum_{\pi} Q(\pi|\sigma) = p(v_1 | u_{\sigma(1)}, \pi_{1:0}) \times 1 + p(v_2 | u_{\sigma(1)}, \pi_{1:0}) \times 1 = 1 \tag{1}$$

Now assuming that the proposition holds for some $N \geq 1$ we need to show that it holds for $N + 1$. Considering $N + 1$ possible matches for $u_{\sigma(1)}$ the summation can be factorized as:

$$\sum_{\pi} Q(\pi|\sigma) = \sum_{i=1}^{N+1} p(v_i | u_{\sigma(1)}, \pi_{1:0}) \left[\sum_{\pi' \in \Omega_i} \prod_{t=2}^{N+1} p(v_{\pi'(\sigma(t))} | u_{\sigma(t)}, \pi'_{1:t-1}) \right] \tag{2}$$

where Ω_i is the set of permutations where $u_{\sigma(1)}$ is matched with v_i and $\prod_{t=2}^{N+1} p(v_{\pi'(\sigma(t))} | u_{\sigma(t)}, \pi'_{1:t-1})$ is the probability of $\pi' \in \Omega_i$. Note that Ω_i has $N!$ assignments of N items and all the assignment probabilities satisfy the Theorem's conditions, therefore from our assumption we have that:

$$\sum_{\pi' \in \Omega_i} \prod_{t=2}^{N+1} p(v_{\pi'(\sigma(t))} | u_{\sigma(t)}, \pi'_{1:t-1}) = 1, \ \forall i \tag{3}$$

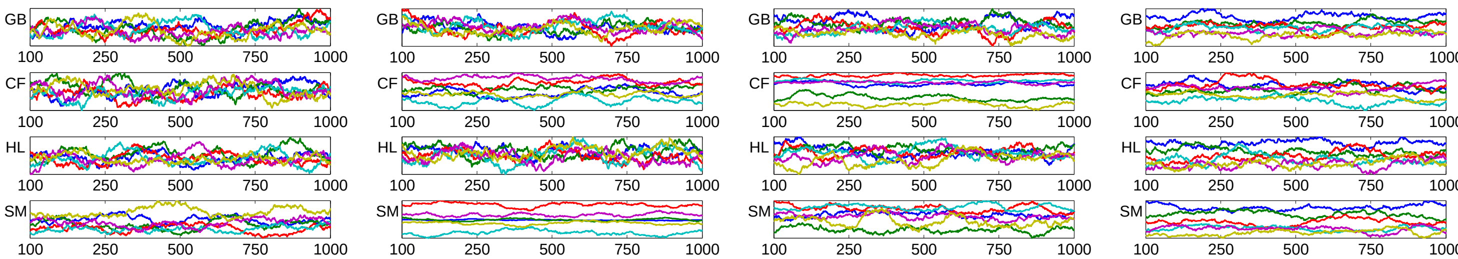
and it follows that:

$$\sum_{\pi} Q(\pi|\sigma) = \sum_{i=1}^{N+1} p(v_i | u_{\sigma(1)}, \pi_{1:0}) = 1 \tag{4}$$

□

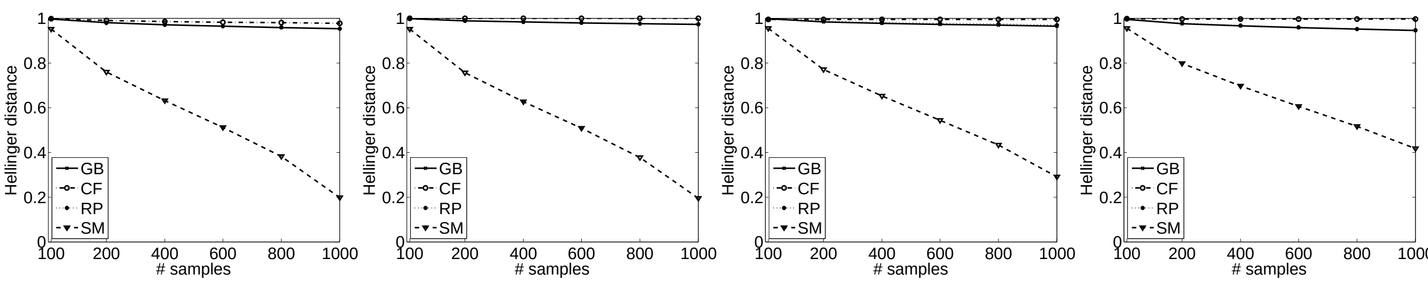
2 Learning To Rank

The trace plots for the four methods for one query with $N = 25$ (plots for other queries and N look similar) are shown in Figure 1. The plots do not show any trending patterns, indicating that the chains are mixing. Figure 2 shows the average Hellinger distances versus the number of samples for each of the four methods with $N = 25$. From the figures it is seen that SM consistently improves the approximation of P as more samples are generated whereas the other samplers are unable to make significant progress throughout the sampling.



(a) 25 documents, $c = 20$ (b) 25 documents, $c = 40$ (c) 25 documents, $c = 60$ (d) 25 documents, $c = 80$

Figure 1: Learning to Rank: moving average (lag 100) trace plots for 5 randomly selected documents from $N = 25$.

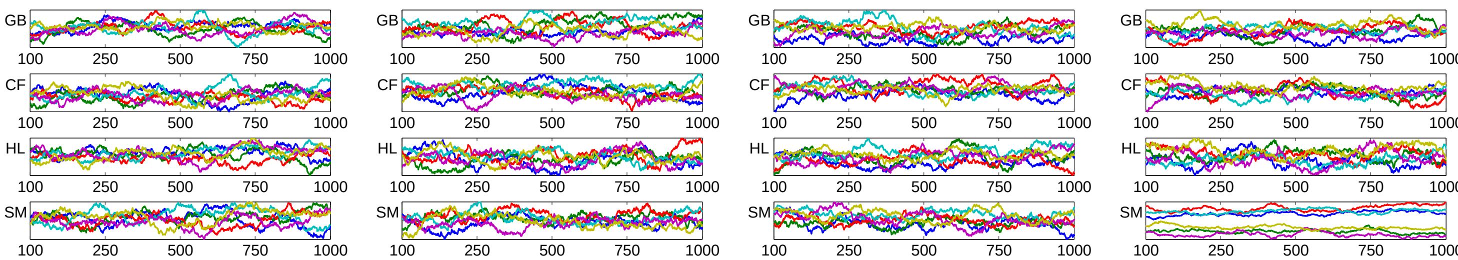


(a) 25 documents, $c = 20$ (b) 25 documents, $c = 40$ (c) 25 documents, $c = 60$ (d) 25 documents, $c = 80$

Figure 2: Learning to Rank: average Hellinger distances versus the number of samples for $N = 25$.

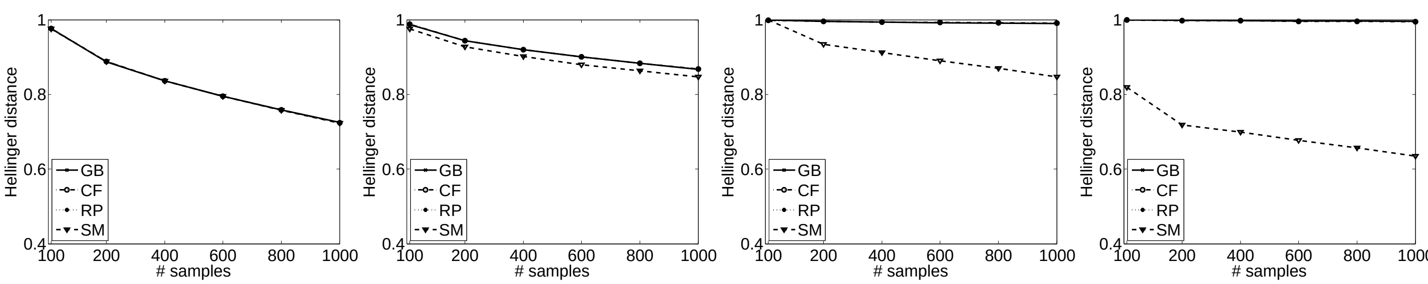
3 Image Matching

The trace plots for the four methods for one image pair with $N = 25$ (plots for other image pairs and N look similar) are shown in Figure 3. Similarly to learning to rank, the plots do not show any trending patterns, indicating that the chains are mixing. Figure 4 also shows the average Hellinger distances versus the number of samples for each of the four methods with $N = 25$. From the figures it is seen that for sharper distributions with several well defined modes ($c \geq 0.6$) SM is able to consistently improve the approximation as more samples are generated whereas the other samplers are again unable to make significant progress.



(a) 25 points, $c = 0.25$ (b) 25 points, $c = 0.4$ (c) 25 points, $c = 0.6$ (d) 25 points, $c = 0.8$

Figure 3: Image Matching: moving average (lag 100) trace plots for 5 randomly selected points from $N = 25$.



(a) 25 points, $c = 0.2$ (b) 25 points, $c = 0.4$ (c) 25 points, $c = 0.6$ (d) 25 points, $c = 0.8$

Figure 4: Image Matching: average Hellinger distances versus the number of samples for $N = 25$.