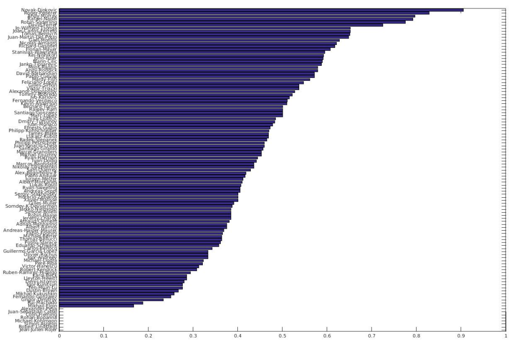
1 Ranking based on empirical ratio

A ranking of tennis players is plotted based on the empirical ratio of number of wins to total number of games played. This is not a good way to estimate player skills because this ranking ignores which players each player played against. A player can win many games against players with worse performance and be highly ranked simply because of the number of games this player won. Another player might be defeated by a strong player without having any other opportunity to play against other less stronger players. This player is ranked in the bottom and the ranking will not reflect his/her true skill.



(a)

Figure 1: Ranking of players based on the empirical ratio of number of wins to total number of games played.

2 Implementation for the Gibbs sampling for the ranking model

The code below is part of the implementation of the Gibbs sampler which samples from the conditional distributions needed to rank players as discussed in lecture 6 and 7.

```
% Second, jointly sample skills given the performance differences
 m = nan(M,1);  % container for the mean of the conditional skill distribution given the t_g samples
 for p = 1:M
   m(p) = t'*(((p-G(:,1))==0)-((p-G(:,2))==0));
 end
 iS = zeros(M,M); % container for the sum of precision matrices contributed
                  % by all the games (likelihood terms)
 for g=1:N
   iS(G(g,1),G(g,1)) = iS(G(g,1),G(g,1))+1;
   iS(G(g,2),G(g,2)) = iS(G(g,2),G(g,2))+1;
   iS(G(g,1),G(g,2)) = iS(G(g,1),G(g,2))-1;
   iS(G(g,2),G(g,1)) = iS(G(g,2),G(g,1))-1;
 end
```

3 Samples after 100 iterations

After 100 iterations of the Gibbs sampler, the autocorrelation of the samples is plotted. Correlation between these samples is noisy and this is because the first few samples do not accurately represent the distribution. By looking at the samples for three players, we notice that samples start moving around a mean with a constant variance after about 20 iterations. We can conclude that the Gibbs sampler is able to move around the posterior distribution after some iterations. It is necessary to discard some samples generated at the beginning. After running 1000 (and 10000 iterations), the samples autocorrelation is less random and we notice that the peak in correlation drops after about 10 lags. For samples to be roughly independent, we can retain every tenth sample generated by the Gibbs sampler.

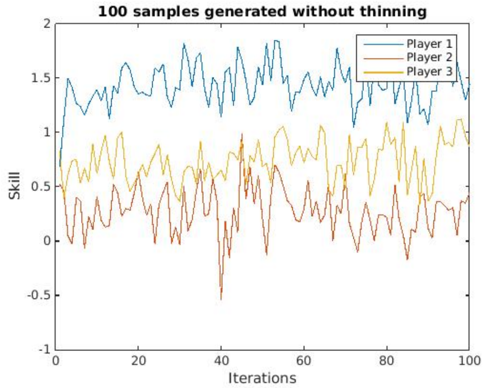
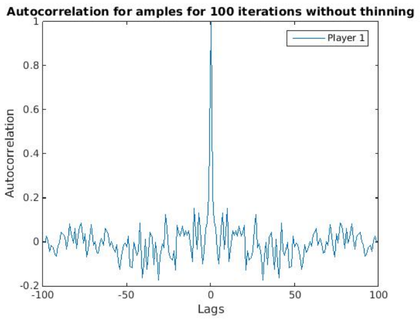
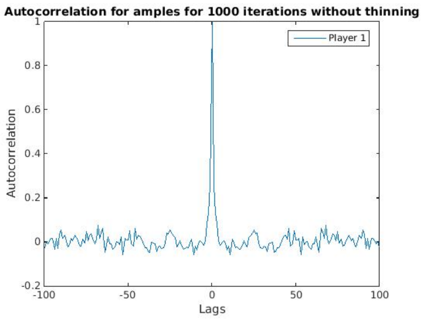


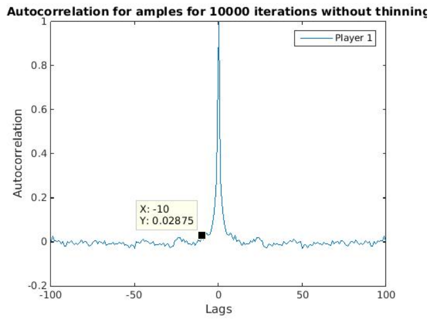
Figure 2: Samples for three players (Rafael-Nadal, Juan-Monaco, Juan-Martin-Del-Potro) after running the Gibbs sampler for 100 iterations.



(a)



(b)

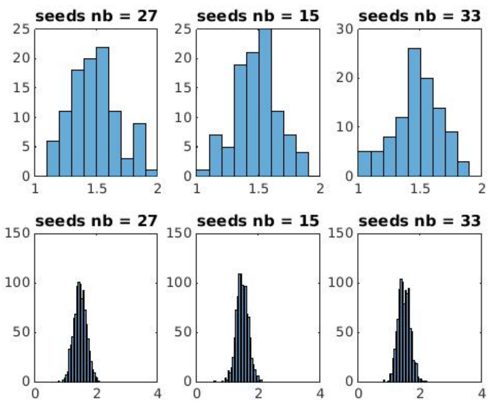


(c)

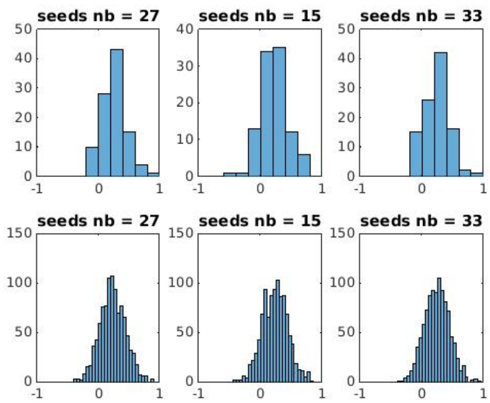
Figure 3: Autocorrelations between samples (100 lags) for (a) 100 samples, (b) 1000 samples and (c) 10000 samples generated by the Gibbs sampler for Rafael-Nadal. Note that in these three cases, no thinning or burn in were applied.

4 Convergence of the Gibbs Sampler

Convergence of a Markov Chain is when values of the Markov chain provide samples from the posterior, i.e. the sample values have the same distribution as if they were sampled from the true posterior joint distribution. The Gibbs Sampler converges when "the variation distance between the law of the Markov chain and the true posterior is appropriately small" [1]. To determine whether the Gibbs sampler seem to converge, the sampler is ran for different pseudo random number seeds. We reach the same distribution form runs with different pseudo random number seeds, especially after running the sampler for 1000 iterations. The distributions will look alike even more for a bigger number of iterations and especially after thinning (by saving every tenth sample only). Based on these observations, Gibbs sampling does converge to the true distribution.



(a)



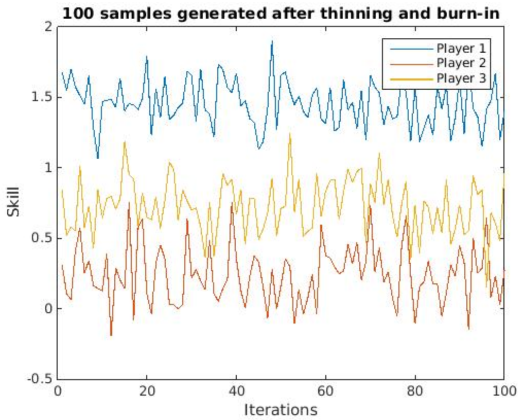
(b)

Figure 4: Distributions of samples for different pseudo random number seeds. In each group of plots, the two plots on the left are distributions with samples for 27 pseudo random number seeds; the samples for the distribution at the top are drawn after 1000 iterations retaining every tenth sample only, and at the bottom after 10000 iterations retaining every tenth sample only. For the remaining plots, the ones at the top are drawn from 100 samples without thinning and the ones at the bottom from 1000 samples. (a) is for Rafael-Nadal and (b) is for Juan-Monaco.

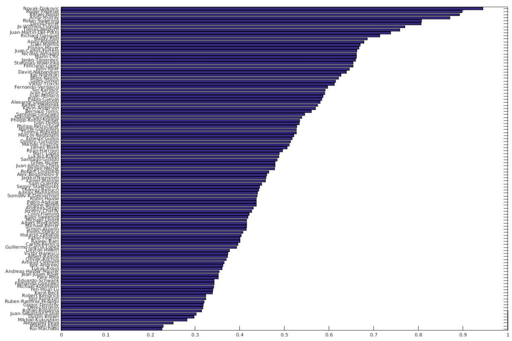
[1] Rosenthal, J. S. (1995). Rates of convergence for Gibbs sampling for variance component models. The Annals of Statistics, Vol. 23, No. 3 (Jun., 1995), pp. 740-761

5 e. Average probability that each player will win against other players

The ranking of players based on the average probability that each player will win against another using 100 samples generated by the Gibbs sampler with samples retained after every 10th iteration is very different from the one obtained in part (a) (the Spearman rank correlation is 0.50). For some top players, the probability of winning using samples from the posterior distribution is higher then their probability in part a. In addition, bottom players do not have a 0 probability of winning as in part (a). This is because the new ranking system takes into consideration the skills of the payers whom a player wins against. A player who wins once against a top player will be ranked higher than a player who wins once against a lower skill player for instance. A player who is defeated by a top player might be ranked higher than a player defeated by someone at the bottom. The ranking is computed based on samples drawn from the exact posterior distribution of the skills.
The ranking is computed using the likelihood $P(y|w_1, w_2) = \phi(y(w_1 - w_2))$.



(a)



(b)

Figure 5: (a) Samples for the skills of three players after running 1000 iterations and retaining the tenth sample only (the first 20 samples from the 1000 iterations were discarded). (b) Ranking of players based on the average probability that each player will win against another using 100 samples generated by the Gibbs sampler with samples retained after every 10th iteration

6 Probability for winning for top players (Gibbs Sampling)

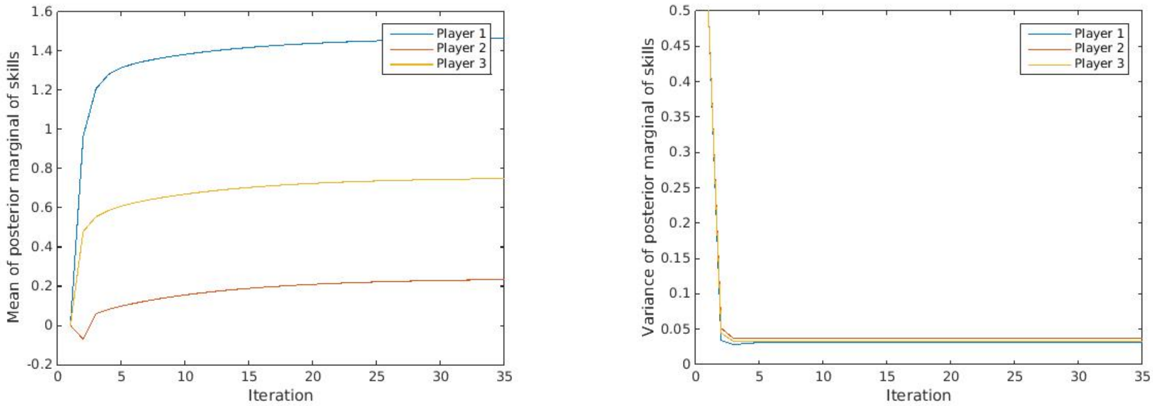
The probability of a player 1 to win against a player 2 is given by the likelihood $P(y|w_1, w_2) = \phi(y(w_1 - w_2))$. The results for the top 4 players are summarized in Table 1.

Winner ↓	Novak-Djokovic	Roger-Federer	Rafael-Nadal	Andy-Murray
Novak-Djokovic	-	0.6528	0.6681	0.7086
Roger-Federer	0.3472	-	0.5164	0.5648
Rafael-Nadal	0.3319	0.4836	-	0.5486
Andy-Murray	0.2914	0.4352	0.4514	-

Table 1: Probability that each of the 4 top players according to the ATP rankings would have of winning against each other. Each row includes the probability that the player in the row wins against the player in each of the columns.

7 Inference with Expectation Propagation

Message Passing and Expectation Propagation are used to approximate inference in the model. To examine the number of iterations necessary for convergence, a threshold is set for the difference of values between iterations for the mean and the variance of performance distribution. The threshold I set is 0.001. In this case, 28 iterations were sufficient to reach this threshold.



(a)

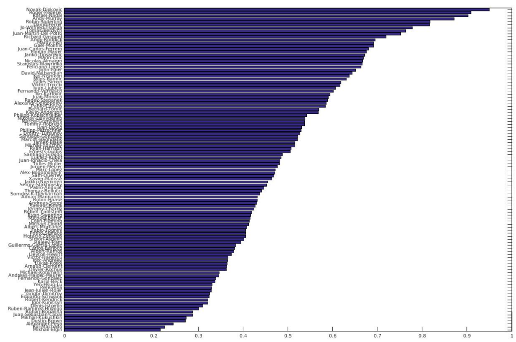
(b)

Figure 6: (a) Mean and (b) variance of performance distributions for three players after 35 iterations of EP.

8 Ranking based on message passing and EP

The ranking based on message passing and EP is obtained using the prior over y as given on page 11 of the lecture handout Introduction to Ranking:

$$p(y) = \phi(\frac{y(\mu_1 - \mu_2)}{\sqrt{1 + \sigma_1^2 + \sigma_2^2}}) \tag{1}$$



(a)

Figure 7: Ranking of players based on the marginal distributions of skills calculated using message passing and EP.

The rankings in e and h are very similar (the Spearman rank correlation is 0.76 which is pretty high). For all players, their average probability of winning obtained via inference with EP is very similar to the one obtained with samples from the Gibbs sampler (difference of magnitude of 0.01 for a player between both methods). Top players have the same rankings in e and h. Rankings for middle and bottom players have slightly different rankings and this is mainly due to small differences in very close probabilities. In table 2, Rui and Mikhail are ranked 106th and 107th in e and 207th and 106th respectively in h because the difference between their probability of winning is very small (in both rankings).

Compared to the ranking in a, both rankings and probabilities of winning of players are very different to the ones in e and h (spearman rank correlation is 0.5), and as explained previously, this is mainly because the ranking based on the empirical ratio does not take into account skills of player (and who won against who).

Gibbs sampling provides samples from the exact inference (as long as retained samples are spaced enough to reduce correlation between samples). The EP method provides an approximation that is very close to the exact inference obtained with Gibbs sampling. However, EP is much faster (35 iterations) than Gibbs sampling (at least 1000 iterations) and is best to use to provide a ranking. The ranking based on skills marginals obtained though message passing/EP is best mainly due to the computational performance.

Player	Gibbs Sampling Ranking	Gibbs Sampling Average Prob of Winning	EP Ranking	EP Prob of Winning
Alexander-Peya	105	0.2495	105	0.2621
Mikhail-Elgin	106	0.2294	107	0.2319
Rui-Machado	107	0.2245	106	0.2367

Table 2: Ranking and average probability for winning against other players of bottom players using samples from the Gibbs sampler vs inference with message passing and EP.

9 Probability for winning for top players (Message Passing and EP)

Winner ↓	Novak-Djokovic	Roger-Federer	Rafael-Nadal	Andy-Murray
Novak-Djokovic	-	0.6380	0.6554	0.7198
Roger-Federer	0.3620	-	0.5184	0.5908
Rafael-Nadal	0.3446	0.4816	-	0.5731
Andy-Murray	0.2802	0.4092	0.4269	-

Table 3: Probability that each of the 4 top players would have of winning against each other based on skills marginals computed with message passing and EP ranking. Each row includes the probability that the player in the row wins against the player in each of the columns.

The probabilities of winning for the 4 top players (against each other) are very similar for both methods, Gibbs Sampling and message passing/EP. This is consistent with the findings in h regarding EP/message passing providing a very good approximation to the inference of the skills posteriors.

10 Probability that a player has a higher skill

For the Gibbs Sampler, the probability that a player has a higher skill is given with:

$$\begin{aligned} p(w_i > w_j) = E[g(w_i, w_j)] \\ g(w_i, w_j) = 1 \qquad for \qquad w_i > w_j, \\ g(w_i, w_j) = 0 \quad otherwise \end{aligned} \tag{2}$$

Higher skill ↓	Novak-Djokovic	Roger-Federer	Rafael-Nadal	Andy-Murray
Novak-Djokovic	-	0.9152	0.9621	0.9850
Roger-Federer	0.0848	-	0.5858	0.8084
Rafael-Nadal	0.0379	0.4142	-	0.7804
Andy-Murray	0.0150	0.1916	0.2196	-

Table 4: Probability that each of the 4 top players has higher skills than the other based on the samples computed with the Gibbs Sampler. Each row includes the probability that the player in the row has a higher skill than the player in each of the columns.

Using the results obtained with EP, the probability that a skill of a player i is bigger than the skill of player j is:

$$\begin{aligned} p(w_i) = N(w_i; \mu_i, \sigma_i); p(w_j) = N(w_j; \mu_j, \sigma_j) \\ p(w_i > w_j) = p(w_w - w_j > 0) = 1 - p(w_i - w_j \leq 0) \\ z_i = w_i - w_j; p(z) = N(z; \mu_i - \mu_j, \sigma_i^2 + \sigma_j^2) \\ p(z > 0) = 1 - p(z \leq 0) = 1 - \phi(\frac{\mu_z}{\sigma_z}) = 1 - \phi(\frac{\mu_j - \mu_i}{\sqrt{\sigma_i^2 + \sigma_j^2}}) \end{aligned} \tag{3}$$

Higher skill ↓	Novak-Djokovic	Roger-Federer	Rafael-Nadal	Andy-Murray
Novak-Djokovic	-	0.9089	0.9399	0.9853
Roger-Federer	0.0911	-	0.5729	0.8108
Rafael-Nadal	0.0601	0.4271	-	0.7664
Andy-Murray	0.0147	0.1892	0.2336	-

Table 5: Probability that each of the 4 top players has higher skills than the other based on marginal skills computed with message passing and EP.

The results obtained with message passing/EP are very similar to the ones obtained with Gibbs sampling (which is consistent with observations in h and i). However, for the top 2 players, their probability of having higher skill than the top 3rd and 4th are much higher that their probability of winning against those players. The expression for probability of winning $p(y) = \phi(\frac{y(\mu_1 - \mu_2)}{\sqrt{1+\sigma_1^2+\sigma_2^2}})$ in EP ranking is the same as for $p(w_i > w_j) = \phi(\frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2+\sigma_j^2}})$ except that there is a 1 added in the square root of the dominator in the probit function which has a smoothing effect resulting in lower probabilities for a top player to win against a lower skill player. This is further explained by the fact that while a top player has higher skill than a less skillful player, the probability of winning for that player is lower that the probability of having higher skill because there are more factors that affect the outcome of a game.