UNIVERSITY OF CAMBRIDGE

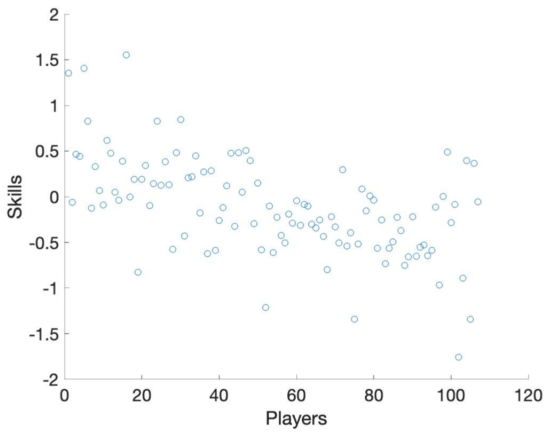
Probabilistic Ranking

5568D

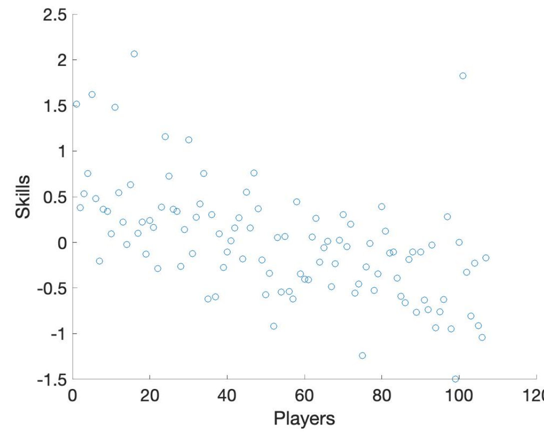
20th November 2019

1 Task A

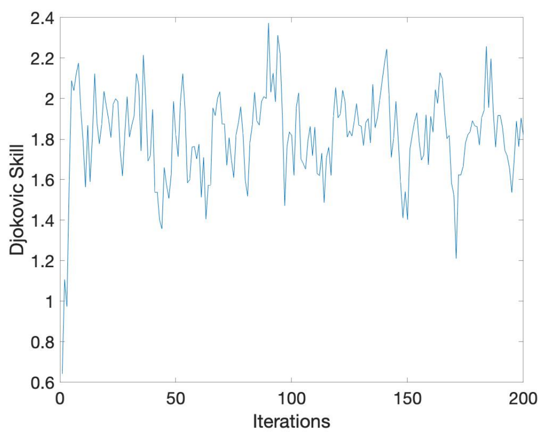
The first method that it is investigated is **Gibbs Sampling**. The skills of the players are jointly sampled by the vector $\mathbf{w}^i$ at stage i . So the new samples are produced based on $\mathbf{w}^i$. Therefore a Markov Chain is formed with a transition probability $q(\mathbf{w}^{i+1}|\mathbf{w}^i)$. It is hoped that as we continue to draw samples from the Markov Chain it will reach a stationary distribution which should equal the true distribution of the skills of the players $p(\mathbf{w})$[Bar12]. The results of applying **Gibbs Sampling** to matrices $\mathbf{W}$ and $\mathbf{G}$ are shown in the figure below.



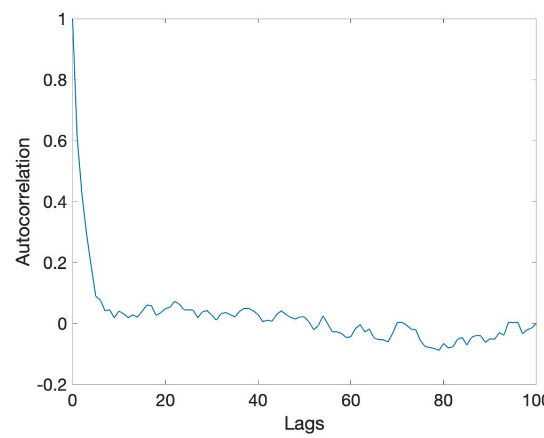
(a) Players skills for 500 iteration of Gibbs Sampling



(b) Players skills for 1100 iteration of Gibbs Sampling



(c) Djokovic Skill sampled for 200 iterations

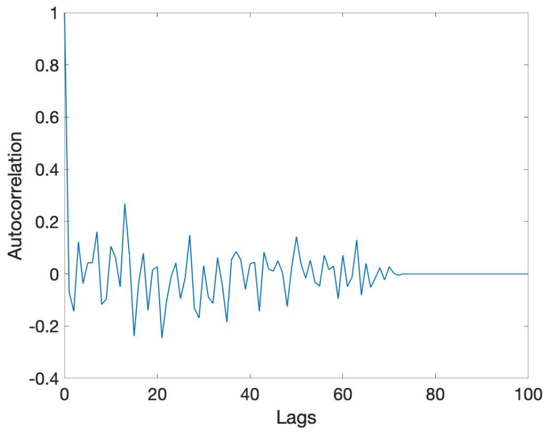


(d) Auto covariance coefficients for Djokovic with 100 lag for 1100 iterations

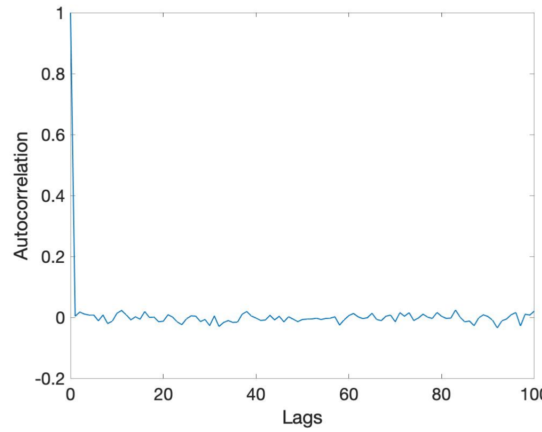
Figure 1

It is observed that in *Figures 1a and 1b* the distribution of the shape is looking towards the true distribution as players such as Nadal, Federer and Djokovic are having a large skill. From *Figure 1c* it is observed that the skill of Djokovic reaches a stationary distribution after about 30 iterations. As this is a relatively small number of iterations it is not necessary to remove the *burn in* samples. Because **Gibbs Sampling** is based on Markov Chain there would always be dependencies between different samples. From *Figure 1d*, it is observed that it takes about 15 before the auto-correlation decreases to zero. From then onward, some oscillation is observed around zero which shows that there haven't been enough samples so that there would be a smooth line at zero. If the thinning is added the result can be seen in *Figure 2a*.

Because only the 15th sample is taken it can be seen that the auto-correlation time has decreased considerably but the oscillation around the zero value has increased. Therefore it is required to increase the number of iterations. In *Figure 2b* it can be seen that if the number of iterations are increase to 10000 the variance from zero decreases.



(a) Auto-correlation with thinning for 1100 iterations



(b) Auto-correlation with thinning for 10000 iterations

Figure 2

2 Task B

As discussed previously in **Task A** it is desired for **Gibbs Sampling** that the sample distribution mean and variance will be steady for a large number of iterations. However, because **Gibbs Sampling** is converging to the distribution itself it is difficult to find the exact moment. The next method investigated to find the distribution of the skills of the players $p(\mathbf{w})$ is called **Message Passing**. In **Message Passing** it is assumed or approximated that all the marginal all Gaussians. Therefore, only the mean and variance (also precision and natural means) are used as a message. Based on a factor graph a recursive model can be developed in which the mean and variance for the marginal skill at iteration τ, $q(\mathbf{w})^\tau$, are used as messages to update the marginal performances, $q(\mathbf{t})^\tau)$, which it will respectively be used to pass on messages to the marginal skills and update it, $q(\mathbf{w})^{\tau+1}$. Therefore it is observed that convergence will be reached if:

$$q(\mathbf{w})^{\tau+1} = q(\mathbf{w})^\tau \tag{1}$$

Applying the **Message Passing** algorithm to **W** and **G** for 30 iterations results in the following mean and variance values for 3 players: Novak Djokovic, Victor Hanescu and Santiago Gonzalez:

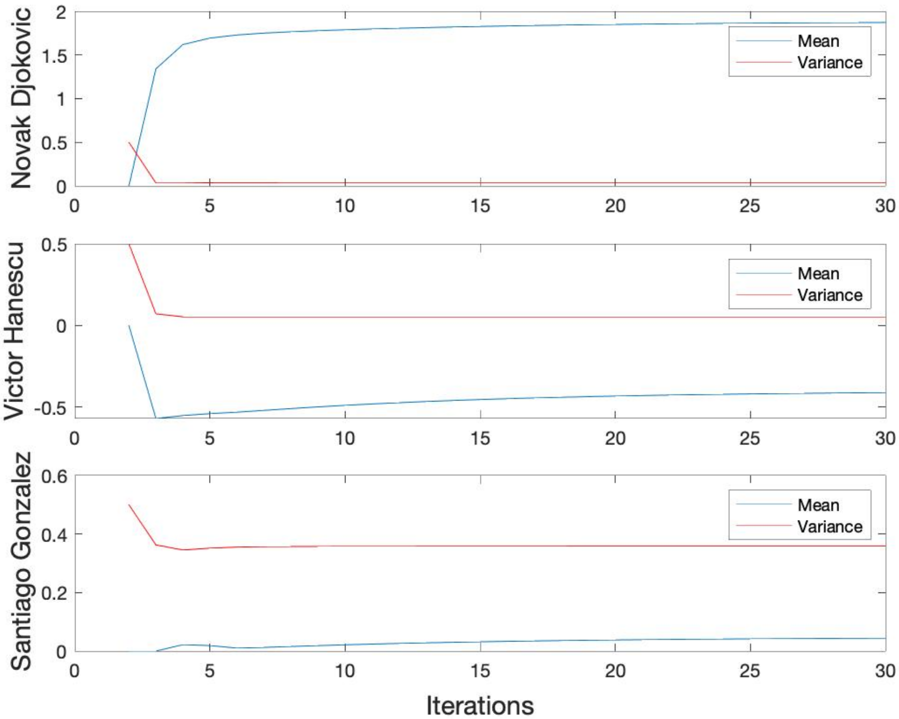


Figure 3: Mean and Variance values for the three values for 30 iterations

From *Figure 3* it is observed that different player will reach convergence for a different number of iterations. Therefore a good convergence condition that could be implemented in the algorithm is:

$$||q(\mathbf{w})^{\tau}||_2^2 - ||q(\mathbf{w})^{\tau+1}||_2^2 < \epsilon \tag{2}$$
$$\text{where} \quad \epsilon \approx 10^{-5} \tag{3}$$

3 Task C

Now that the skills distribution have been defined it is time to look at the probability that player 1 will defeat player 2. The first four ATP players are considered:

Player	Rank
Novak Djokovic	1
Rafael Nadal	2
Roger Federer	3
Andy Murray	4

Table 1: First four of ATP Rankings 2011

To find the probability that a player skill is bigger than another player it is required to look at the difference between the skills. In the following equations it is described how to calculate this probability:

$$p(w_1 > w_2) = p(w_1 - w_2 > 0) \tag{4}$$
$$\text{Therefore let} \quad z = w_1 - w_2 \quad \Longrightarrow z \sim \mathcal{N}(z; \mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) \tag{5}$$
$$p(z > 0) = \int_0^{\infty} \mathcal{N}(z; \mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) = 1 - \int_{-\infty}^{0} \mathcal{N}(z; \mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) = 1 - \phi(z) \tag{6}$$

By applying the convergence condition above to the algorithm and applying the equation above to determine the probabilities the table below can be displayed:

		Player 2			
		Novak Djokovic	Rafael Nadal	Roger Federer	Andy Murray
Player 1	Novak Djokovic	NA	0.9407	0.9092	0.9852
	Rafael Nadal	0.0593	NA	0.4250	0.7634
	Roger Federer	0.0908	0.5750	NA	0.8096
	Andy Murray	0.0148	0.2366	0.1904	NA

Table 2: Probabilities that skill of Player 1 is greater than the skill of Player 2

Moving on from the skills,it is time to look at the matches. The probability of a match outcome(y) in favour of player one is given by:

$$p(y = 1) = \iint \mathcal{N}(w_1; \mu_1, \sigma_1^2)\mathcal{N}(w_2; \mu_2, \sigma_2^2)\phi(y(w_1 - w_2)) = \phi(\frac{y(\mu_1 - \mu_2)}{\sqrt{1 + \sigma_1^2 + \sigma_2^2}}) \tag{7}$$

Therefore a similar table to *Table 2* can be displayed.

		Player 2			
		Novak Djokovic	Rafael Nadal	Roger Federer	Andy Murray
Player 1	Novak Djokovic	NA	0.6558	0.6380	0.7192
	Rafael Nadal	0.3442	NA	0.4811	0.5720
	Roger Federer	0.3620	0.5189	NA	0.5902
	Andy Murray	0.2808	0.4280	0.4098	NA

Table 3: Probability that Player 1 will win against Player 2

It is observed from *Tables 2 and 3* that the values of the probability are correlated. However it is observed that in *Table '3* the probabilities are closer to 0.5. This due to the fact that noise is introduced into the modelling of the game performance. If the variance of the noise is to increase further it is expected that all the probabilities in the *Table 3* to be 0.5. On the other hand, if the noise is to be removed from our model it would be expected that *Table 3* values would be equal to *Table 2* values.

4 Task D

The skills of Novak Djokovic and Rafael Nadal are compared in three different ways. The first one is by approximating marginal skills by a Gaussian. Firstly, the means and variances are found from the samples:

$$\mu_i = \text{mean}(\mathbf{w}_i) \quad \text{and} \quad \sigma_i^2 = \text{var}(\mathbf{w}_i) \tag{8}$$

Using these values, the same method applied in *Table 2* is used. The results are summarised below:

		Player 2	
		Novak Djokovic	Rafael Nadal
Player 1	Novak Djokovic	NA	0.9218
	Rafael Nadal	0.0782	NA

Table 4: Probability that the skill of Player 1 is bigger than Player 2 by computing the marginal skills using a Gaussian

The second method is by approximating the jointly skills by Gaussian. Therefore the mean and variance can be found by:

$$\mu_1 = \text{mean}(\mathbf{w}_1 > \mathbf{w}_2) \quad \text{and} \quad \sigma_1^2 = \text{var}(\mathbf{w}_1 > \mathbf{w}_2) \tag{9}$$

And again by applying the same method as in *Table 2* the following results are displayed:

		Player 2	
		Novak Djokovic	Rafael Nadal
Player 1	Novak Djokovic	NA	0.9977
	Rafael Nadal	0.0023	NA

Table 5: Probability that the skill of Player 1 is bigger than Player 2 by computing the jointly skills using a Gaussian

The 3rd method studied is by simply looking at the samples directly which is summarised in *Equation 9* . The results are:

		Player 2	
		Novak Djokovic	Rafael Nadal
Player 1	Novak Djokovic	NA	0.9474
	Rafael Nadal	0.0526	NA

Table 6: Probability that the skill of Player 1 is bigger than Player 2 by computing the jointly skills directly

The probabilities in *Table 5 and 6* are highly depend on the number of iterations are they are computed by binary counting. On the other hand, in *Table 4* the mean and variance of the skills of each player are found. Therefore the first method is preferred. The results of all the 4 players skill probabilities using the first method is summarised below.

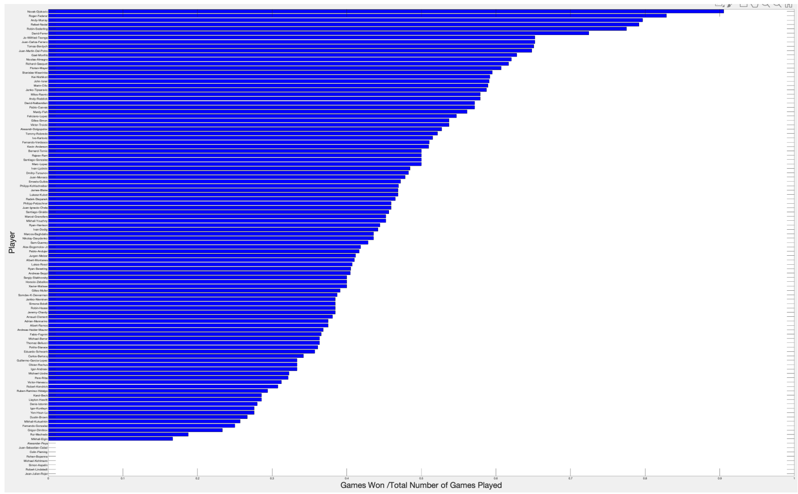
		Player 2			
		Novak Djokovic	Rafael Nadal	Roger Federer	Andy Murray
Player 1	Novak Djokovic	NA	0.9218	0.8864	0.9754
	Rafael Nadal	0.0782	NA	0.4327	0.7420
	Roger Federer	0.1136	0.5673	NA	0.7844
	Andy Murray	0.0246	0.2580	0.2156	NA

Table 7: Probability that Player 1 will win against Player 2 using marginal skills

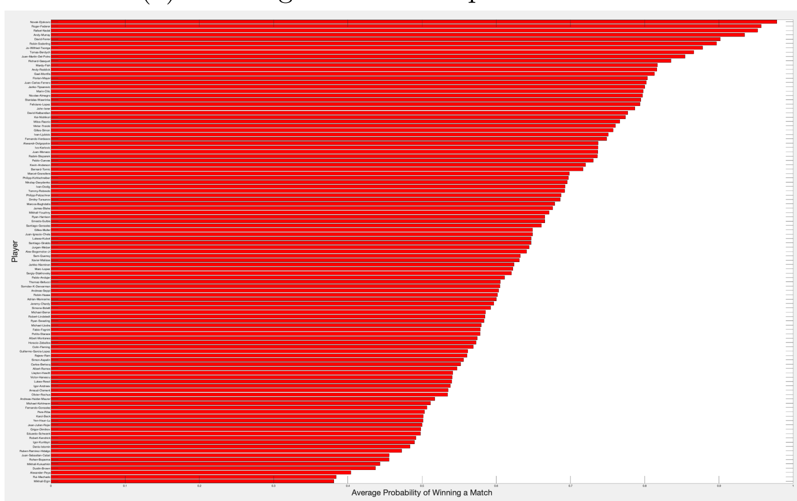
Comparing *Table 2 and 7* it can be observed they both methods generally agree with each other. In the **Gibbs Sampling** method shows that if Player 1 has a generally a bit of lower probability to have better skill than Player 2 compared to **Message Passing** method.

5 Task E

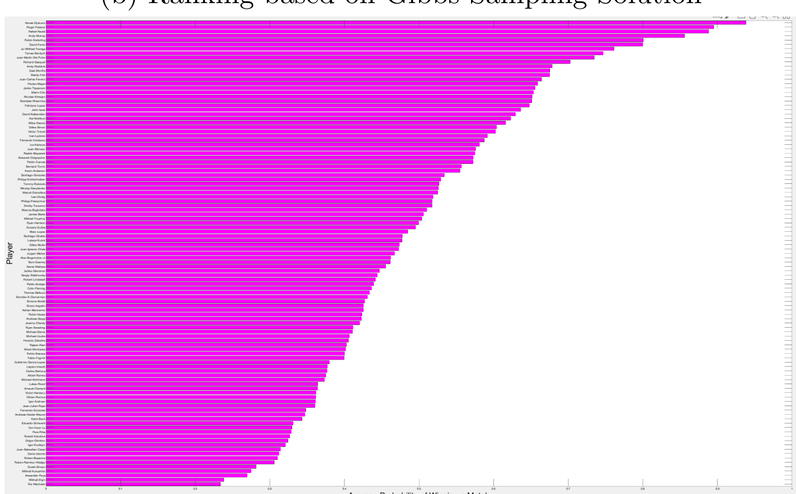
A new ranking system has been implement using **Empirical Method**, **Gibbs Sampling** and **Message Passing**. The results are displayed below.



(a) Ranking based on Empirical Solution



(b) Ranking based on Gibbs Sampling Solution



(c) Ranking based on Message Passing Solution

Figure 4

It is observed that compared with *Figure 4a*, *Figure 4b and 4c* have given a rank to the players who do not have won any game. This is because they have played with a strong opponent and therefore could have a higher skill than players who have won against low skill players. It is also observed that both **Gibbs Sampling** and **Message Passing** provide similar solutions. As discussed, the difference between them remain the time it takes to compute the skill distribution.

WORDCOUNT = 1026

References

[Bar12] David Barber. *Bayesian reasoning and machine learning / David Barber.* eng. Cambridge ; New York: Cambridge University Press, 2012. ISBN: 0521518148.