4F13 Coursework 2 - Probabilistic Ranking

November 2024

1 Task A

Figure 1 shows samples of the skills at each Gibbs iteration for 3 players. We see that at each iteration the skill of each player appears random but perhaps not independent of the previous iteration. This is as expected from Gibbs sampling and is demonstrated by the auto covariance plot in Figure 2.
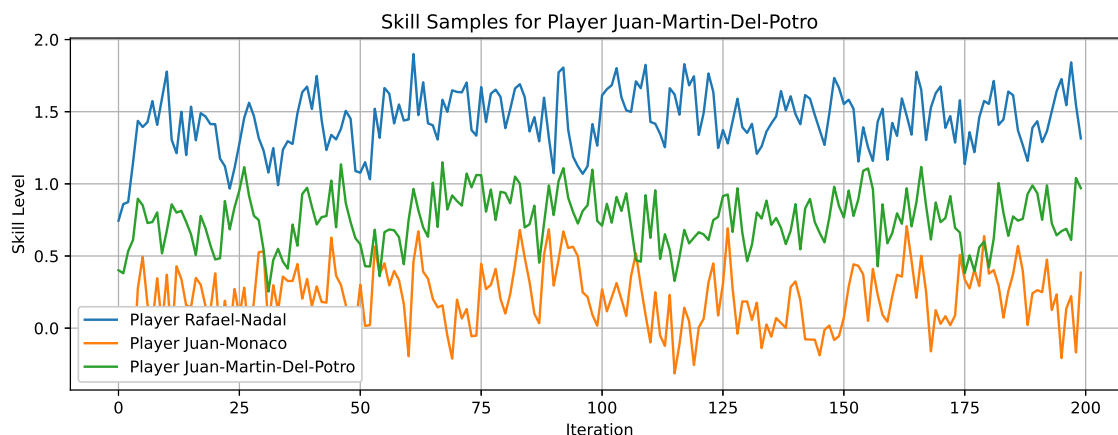

Figure 1: Samples of the skill of 3 players at each Gibbs iteration.

If the samples at each iteration were independent we would expect the auto covariance to be zero for all non zero lags. However, we notice that the auto covariance only converges to zero for lags greater than about 5. This indicates that if we kept only every 5th sample they would be independent, this is the logic behind thinning. We are not running into computational/memory limits in this case so there is no need to thin the samples.
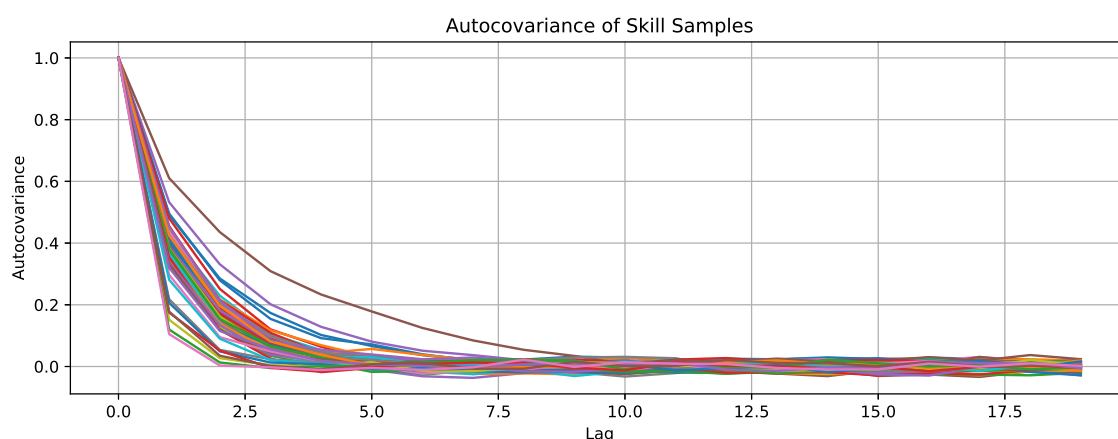

Figure 2: Auto covariance of the skill of all players

The convergence of the skill population is shown in Figure 3. We see that there is no clear burn in time for the population mean but the standard deviation needs about 5 iterations to converge. This makes sense from our auto covariance analysis as we concluded
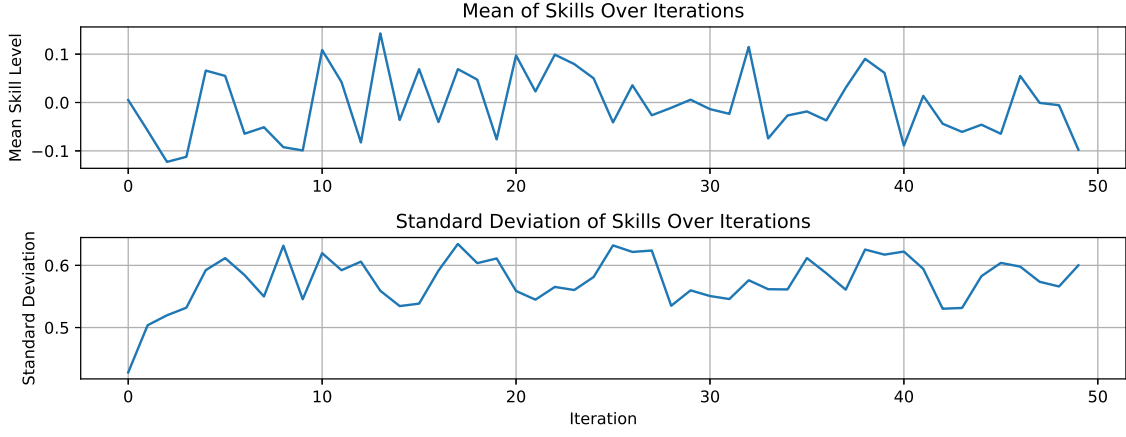
Figure 3: Skill populations mean and standard deviation with Gibbs iteration.

samples 5 iterations apart are roughly independent. For any further we will discard the first 5 Gibbs iterations as burn in time.

To estimate how many iterations we need to run the Gibbs sampler to get skill estimates within a certain tolerance we can analyse the variance of our Monte Carlo skill estimates. Equation 1 calculates the variance of the estimated skill of player i after M Gibbs iterations, the first two equalities are the standard Monte Carlo variance result with N independent samples $w_i^{(n)}$. The final approximate equality is by setting $M = 5N$ as we found that samples are roughly independent every 5 Gibbs iterations.

$$\text{Var}(\bar{w}_i) = \text{Var}(\frac{1}{N}\sum_{n=1}^{N} w_i^{(n)}) = \frac{1}{N}\text{Var}(w_i) \approx \frac{5}{M}\text{Var}(w_i) \tag{1}$$

We take this result and compute the number of Gibbs iterations required to have 95% confidence that the skill estimate is within 1% of the true skill in Equation 2. Here we assume that the marginal skill mean, μ_{w_i}, is of order 1 due to our prior variance. This gives a result of $M \approx 1 \times 10^4$ iterations where we use the median $Var(w_i)$.

$$\frac{2 * \sqrt{\text{Var}(\bar{w}_i)}}{\mu_{w_i}} = 0.01 \implies M \approx 2 \times 10^5 \times \text{Var}(w_i) \tag{2}$$

2   Task B

In the previous section we saw that it took about 10000 iterations for the Gibbs sampled skill estimates to converge within our tolerance. In comparison, Figure 4 shows that the EP skill estimates converge within 10 iterations for the given players.

3   Task C

Using the skill parameters estimated by the EP algorithm we denote the skill of player i as $w_i \sim \mathcal{N}(m_i, p_i^{-1})$, with m_i and p_i being the mean and precision for player i. The distributions of the skill difference, $s_{ij} = w_i - w_j$, and performance difference, $t_{ij} = s_{ij} + \eta$: $\eta \sim \mathcal{N}(0, 1)$, between player i and j are given in Equation 3. These are simple to compute under the assumption that w_i, w_j and η are independent.

$$\begin{aligned} s_{ij} &\sim \mathcal{N}(m_i - m_j, p_i^{-1} + p_j^{-1}) \\ t_{ij} &\sim \mathcal{N}(m_i - m_j, p_i^{-1} + p_j^{-1} + 1) \end{aligned} \tag{3}$$
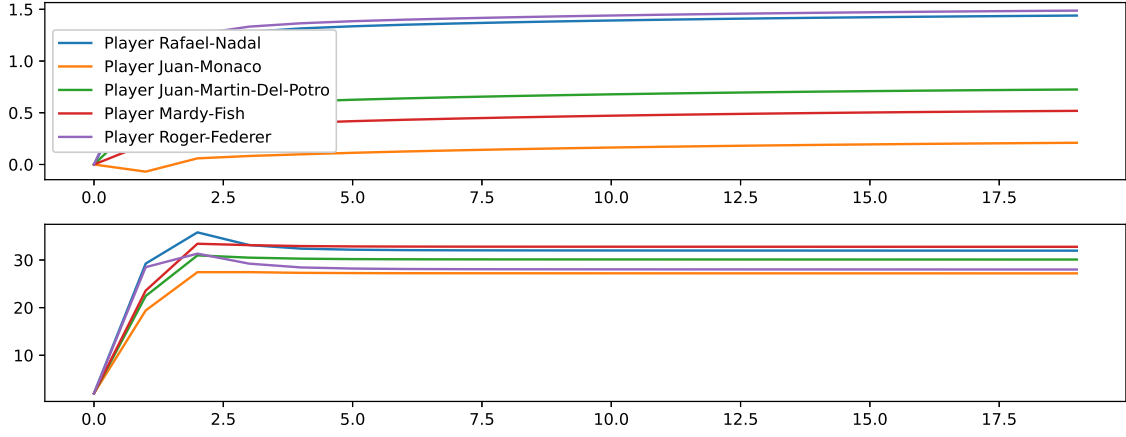
2

Figure 4: EP skill parameter evolution with iteration for 5 players.

$P(s_{ij} > 0)$	Djokovic	Nadal	Federer	Murray	$P(t_{ij} > 0)$	Djokovic	Nadal	Federer	Murray
Djokovic	-	0.94	0.91	0.99	Djokovic	-	0.94	0.91	0.99
Nadal	0.06	-	0.43	0.77	Nadal	0.06	-	0.43	0.77
Federer	0.09	0.57	-	0.81	Federer	0.09	0.57	-	0.81
Murray	0.01	0.23	0.19	-	Murray	0.01	0.23	0.19	-

(a) Probability that row player is more skilled that column player. (b) Probability that row player wins a match against column player.

Table 1: Skill and performance differences between top 4 ATP players based on EP.

We wish to compute the probability that player i has greater skill that player j, $P(s_{ij} > 0)$, and the probability that player i will win a match against player j, $P(t_{ij} > 0)$. We use the identity $P(x > 0) = \phi(\frac{\mu}{\sigma})$ where ϕ is the standard normal c.d.f. and $x \sim \mathcal{N}(\mu, \sigma^2)$. We find these probabilities between the top 4 ranked ATP players in Table 1.

We are always more confident in who is more skilled that who will win a match. This makes sense as $\sigma_t > \sigma_s$, however, intuitively this is clear as we are always less certain about the outcome of a match due to performance variation.

4 Task D

Using our Gibbs samples we would like to estimate the skill difference between Nadal and Federer. We investigate three methods to do this in the following sections.

4.1 Approximate marginal skills with gaussians

We compute the samples mean and variance for each player and model their skills as $w_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Figure 5 shows sample histograms and gaussian approximation. Using the formulas from Task C we find that $P(w_{Nadal} > w_{Federer}) = 0.395$.

4.2 Approximate joint skill distribution with a multivariate gaussian

We can also approximate the joint skill distribution with a gaussian by computing the sample covariance matrix. We find a positive covariance between the two players, $\sigma_{ij}^2 = 0.01$. We visualise the joint distribution and our approximation in Figure 6.
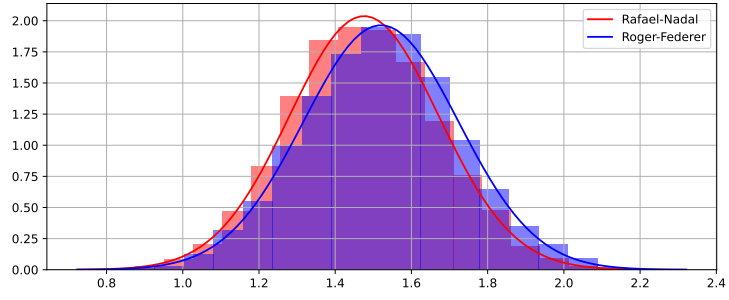
Figure 5: Marginal skill distributions and gaussian approximations for Nadal and Federer.
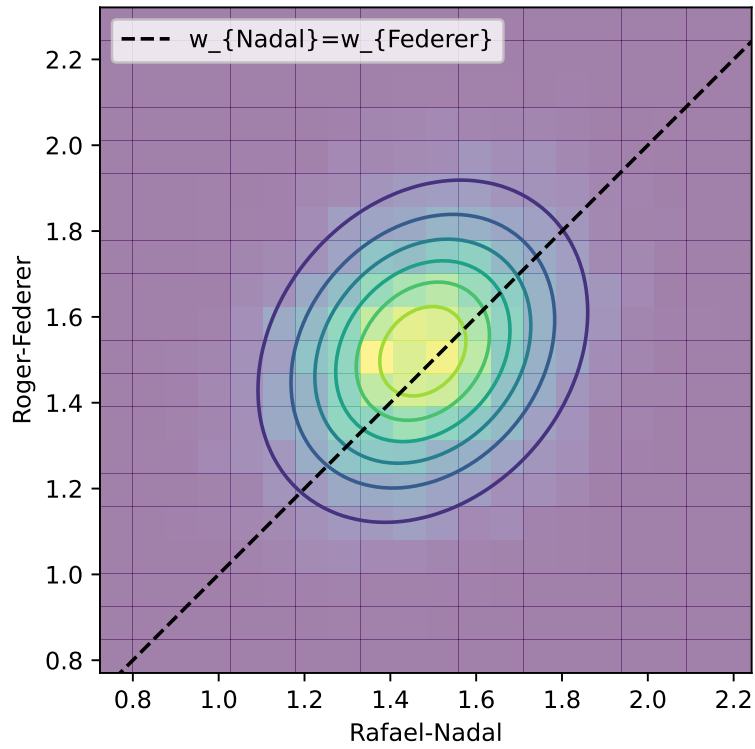

Figure 6: Joint skill distribution (histogram) and gaussian approximation (contours) for Nadal and Federer.

4.3 Use the samples directly

We can also compute a Monte Carlo estimate of $P(w_{Nadal} > w_{Federer})$. This estimate is given in Equation 4 where $(w_i^{(n)}, w_j^{(n)})$ are samples from a Markov chain with stationary distribution $p(w_i, w_j)$. Our Gibbs samples satisfy this condition and we find $P(w_{Nadal} > w_{Federer}) = 0.395$.

$$
\begin{aligned}
\mathrm{P}(w_i > w_j) &= \mathrm{E}(\mathbb{1}(w_i > w_j)) \\
&= \int \mathbb{1}(w_i > w_j) p(w_i, w_j) dw_i dw_j \\
&\approx \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}(w_i^{(n)} > w_j^{(n)})
\end{aligned}
\tag{4}
$$

5 Task E

We have explored methods for skill estimation using both Gibbs sampling and EP algorithms. A simpler model for estimating skill is to assume every player has a fixed win rate r_i. The posterior distribution of r for a player who wins k out of n matches with a uniform prior is the Beta distribution given in Equation 5. We can apply the central limit theorem to approximate this posterior with a gaussian for large n. We interpret this win rate as the skill of the player and plot its mean and 66% error bars in Figure **??** along with those obtains from Gibbs sampling and EP.

$$
r|k, n \sim \mathrm{Beta}(1 + k, 1 + n - k) \approx \mathcal{N}(\frac{k}{n}, \frac{k(n-k)}{n^3})
\tag{5}
$$