Question 1

For the sake of notation, let $x_{i:j}$ denote $(x_i, x_{i+1}, \ldots, x_j)$. Then

$$\sum_{\mathbf{x}} \pi(\mathbf{x})\pi(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x}} \pi(x_1|x_{2:m})\pi(x_{2:m}) \cdot \prod_{k=1}^{m} \pi(y_k|y_{1:k-1}, x_{k+1:m}) \tag{1}$$

We now use the identity

$$\mathbb{P}(X|Y,Z) = \frac{\mathbb{P}(X,Y,Z)}{\mathbb{P}(Y,Z)} = \frac{\mathbb{P}(Z|X,Y)\mathbb{P}(X,Y)}{\mathbb{P}(Z|Y)\mathbb{P}(Y)} = \frac{\mathbb{P}(Z|X,Y)\mathbb{P}(X|Y)}{\mathbb{P}(Z|Y)} \tag{2}$$

Setting $X = y_k$, $Y = y_{1:k-1}$ and $Z = x_{k+1:m}$, this transforms equation (1) to

$$\begin{aligned} &\sum_{\mathbf{x}} \pi(x_1|x_{2:m})\pi(x_{2:m}) \cdot \prod_{k=1}^{m} \frac{\pi(x_{k+1:m}|y_{1:k})\pi(y_k|y_{1:k-1})}{\pi(x_{k+1:m}|y_{1:k-1})} \\ &= \prod_{k=1}^{m} \{(\pi(y_k|y_{1:k-1})\} \sum_{\mathbf{x}} \left\{ \pi(x_1|x_{2:m})\pi(x_{2:m}) \prod_{k=1}^{m-1} \frac{\pi(x_{k+1:m}|y_{1:k})}{\pi(x_{k+1:m}|y_{1:k-1})} \right\} \end{aligned} \tag{3}$$

However we notice that

$$\begin{aligned} &\sum_{\mathbf{x}} \left\{ \pi(x_1|x_{2:m})\pi(x_{2:m}) \prod_{k=1}^{m-1} \frac{\pi(x_{k+1:m}|y_{1:k})}{\pi(x_{k+1:m}|y_{1:k-1})} \right\} \\ &= \sum_{\mathbf{x}} \pi(x_1|x_{2:m})\pi(x_{2:m}) \frac{\pi(x_{2:m}|y_1)}{\pi(x_{2:m})} \cdot \frac{\pi(x_{3:m}|y_{1:2})}{\pi(x_{3:m}|y_1)} \cdots \frac{\pi(x_m|y_{1:m-1})}{\pi(x_m|y_{1:m-2})} \\ &= \sum_{\mathbf{x}} \pi(x_1|x_{2:m})\pi(x_m|y_{1:m-1}) \prod_{k=1}^{m-2} \frac{\pi(x_{k+1:m}|y_{1:k})}{\pi(x_{k+2:m}|y_{1:k})} \end{aligned} \tag{4}$$

Notice that x_1 only appears in the first factor, so by summing over x_1 it disappears. Then notice the only place x_2 occurs is in $\pi(x_{2:m}|y_1)$, so by summing over x_2 this turns into $\pi(x_{3:m}|y_1)$ cancelling with the denominator. Proceeding by induction we claim that the above is equal to 1. We therefore conclude that

$$\sum_{\mathbf{x}} \pi(\mathbf{x})\pi(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^{m} \pi(y_k|y_{1:k-1}) = \pi(\mathbf{y}) \tag{5}$$

and hence π is an invariant measure of this Markov chain.

Question 2

Using the identity in equation (2) we deduce that $\pi(\mu_k|\boldsymbol{\mu}_{-k}, \theta, \boldsymbol{\sigma}^2, \mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\mu}, \theta, \boldsymbol{\sigma}^2)\pi(\mu_k|\boldsymbol{\mu}_{-k}, \theta, \boldsymbol{\sigma}^2) = \pi(\mathbf{y}|\boldsymbol{\mu}, \theta, \boldsymbol{\sigma}^2)\pi(\mu_k|\theta)$. Hence the posterior distribution of μ_k is proportional to

$$\begin{aligned} &\prod_{l=1}^{K}\prod_{t=1}^{T} \left\{ \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{1}{2\sigma_l^2} (y_{lt} - \mu_l)^2 \right) \right\} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2} (\mu_k - \theta)^2 \right) \\ &\propto \prod_{t=1}^{T} \exp\left(-\frac{1}{2\sigma_k^2} (y_{kt} - \mu_k)^2 \right) \exp\left(-\frac{1}{2\sigma_0^2} (\mu_k - \theta)^2 \right) \\ &\propto \exp\left(\left(-\frac{T}{2\sigma_k^2} - \frac{1}{2\sigma_0^2} \right) \mu_k^2 + \left(\frac{1}{\sigma_k^2} \sum_{t=1}^{T} y_{kt} + \frac{\theta}{\sigma_0^2} \right) \mu_k \right) \\ &\propto \exp\left(-\frac{T\sigma_k^{-2} + \sigma_0^{-2}}{2} \left(\mu_k - \frac{\sigma_k^{-2} \sum_{t=1}^{T} y_{kt} + \theta\sigma_0^{-2}}{T\sigma_k^{-2} + \sigma_0^{-2}} \right)^2 \right) \end{aligned} \tag{6}$$

Notice that this is proportional to a normal distribution with the required parameters in the project booklet. An almost identical procedure is used for the second result, since $\pi(\theta|\boldsymbol{\sigma}^2, \boldsymbol{\mu}, \mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \theta)\pi(\theta|\boldsymbol{\mu}, \boldsymbol{\sigma}^2) \propto \pi(\boldsymbol{\mu}|\theta)\pi(\theta)$. Hence taking the product over k and interchanging the parameters

$$
\begin{aligned}
\sigma_k^{-2} &\mapsto \sigma_0^{-2} \\
\sigma_0^{-2} &\mapsto \tau_0^{-2} \\
\mathbf{y}_{kt} &\mapsto \mu_k \\
\theta &\mapsto \mu_0
\end{aligned}
\tag{7}
$$

we derive the second distribution. For the final result, we know that the distribution $\pi(\sigma_k^{-2}|\boldsymbol{\mu}, \theta, \boldsymbol{\sigma}^2_{-k}, \mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\mu}, \theta, \boldsymbol{\sigma}^2)\pi(\sigma_k^{-2}|\boldsymbol{\mu}, \theta, \boldsymbol{\sigma}^2_{-k}) = \pi(\mathbf{y}|\boldsymbol{\mu}, \theta, \boldsymbol{\sigma}^2)\pi(\sigma_k^{-2})$. So the posterior distribution is proportional to

$$
\begin{aligned}
&\prod_{l=1}^{K}\prod_{t=1}^{T}\left\{\frac{1}{\sqrt{2\pi\sigma_l^2}}\exp\left(-\frac{1}{2\sigma_l^2}\left(y_{lt}-\mu_l\right)^2\right)\right\}\cdot\frac{\left(\sigma_k^{-2}\right)^{\alpha_0-1}\exp\left(-\beta_0\sigma_k^{-2}\right)\beta_0^{\alpha_0}}{\Gamma\left(\alpha_0\right)} \\
&\propto \left(\sigma_k^{-2}\right)^{T/2}\exp\left(-\frac{1}{2}\sigma_k^{-2}\sum_{t=1}^{T}\left(y_{kt}-\mu_k\right)^2\right)\left(\sigma_k^{-2}\right)^{\alpha_0-1}\exp\left(-\beta_0\sigma_k^{-2}\right) \\
&= \left(\sigma_k^{-2}\right)^{\alpha_0+T/2-1}\exp\left(-\left(\beta_0+\frac{1}{2}\sum_{t=1}^{T}\left(y_{kt}-\mu_k\right)^2\right)\sigma_k^{-2}\right)
\end{aligned}
\tag{8}
$$

This is proportional to the Gamma distribution with the same parameters given in the project booklet, and hence we're done. The marginal prior distribution for $\pi(\mu_k) = \int_{-\infty}^{+\infty}\pi(\mu_k|\theta)\pi(\theta)d\theta$. We can calculate this as

$$
\begin{aligned}
\pi(\mu_k) &\propto \int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2\sigma_0^2}\left(\theta-\mu_k\right)^2\right)\exp\left(-\frac{1}{2\tau_0^2}\left(\theta-\mu_0\right)^2\right)d\theta \\
&\propto \exp\left(-\frac{1}{2\tau_0^2+2\sigma_0^2}\left(\mu_k-\mu_0\right)^2\right)
\end{aligned}
\tag{9}
$$

From this we conclude that the marginal prior for μ_k is distributed as such

$$
\pi\left(\mu_k\right) \sim \mathrm{N}\left(\mu_0, \tau_0^2+\sigma_0^2\right) \tag{10}
$$

Question 3

The code named Gibbs Sampler, referenced on page 9, was used to sample from the posterior distribution using the described MCMC method.[1] The test data below in Table 1 is a small section of output from the first 5 iterations.

	$\mathbf{x}^0$	$\mathbf{x}^1$	$\mathbf{x}^2$	$\mathbf{x}^3$	$\mathbf{x}^4$	$\mathbf{x}^5$
μ_1	60.0	77.810272	84.941201	84.404838	82.554353	84.320647
μ_2	60.0	53.986773	61.205653	50.919007	47.719512	55.28883
μ_3	60.0	54.46738	52.862724	49.288816	57.539684	47.780734
σ_1^2	100.0	82.554671	18.999169	105.86238	51.541417	27.141017
σ_2^2	100.0	157.20313	65.218544	25.098216	143.12897	79.378385
σ_3^2	100.0	48.066815	32.884554	78.324008	107.22865	159.2443
θ	60.0	66.303817	55.594835	57.2068	60.967041	58.96342

Table 1: Test Data for Gibbs Sampler

The choice of parameters suggest that each team roughly scores 60 points per year, however $\sigma_0^2+\tau_0^2 = 500$, suggesting that there is roughly $\pm 2\sqrt{\sigma_0^2+\tau_0^2} = \pm 45$ uncertainty in the value of any teams mean μ_k. We also note that initially $\mathbb{E}(\sigma_k^2) = 10^{-2}$, and $\mathrm{Var}(\sigma_k^{-2}) = 10$, suggesting that there is a large uncertainty in the data, with a small scope to change this uncertainty.

[1] The function `Gibbs` and the object `Data` will be used implicitly throughout the rest of this project, and wont be included in any subsequent referencing.

We aim to choose the initial state $\mathbf{x}^0$ as close to the prior distribution as possible. Assuming no knowledge of the observed data $\mathbf{y}$, it makes sense to set each of the values of the parameters to their expected value under the prior distribution. Hence $\mu_k^0 = 60$ and $\theta^0 = 60$. Note the value of $\mathbb{E}(\sigma_k^2)$ is negative if we use the extended Gamma function for negative numbers which makes no sense, however using Jensen's inequality we deduce that $\mathbb{E}(\sigma_k^2) \geq 100$, and so it has been set to 100.

Question 4

The posterior means of $(\theta, \mu_k, \sigma_k^2)$ were calculated using $N = 250$ samples by Code 4, referenced on page 11, as well as generating a histogram for $\pi(\theta|\mathbf{y})$ using $N = 10^5$. These are presented below in Table 2 and Figure 1.

k	Estimated μ_k	Empirical μ_k	Estimated σ^2	Empirical σ^2
1	81.496883	83.8	52.714064	22.7
2	52.13558	51.6	89.648396	39.3
3	50.672261	49.6	127.38336	62.8
4	52.508082	51.0	137.94752	76.0
5	48.138956	47.4	27.386959	13.3
6	69.784992	75.2	284.56201	154.2
7	51.63744	49.6	162.13317	96.8
8	48.756523	48.2	38.925028	18.2
9	62.29331	63.2	182.30559	101.2
10	74.386134	76.4	59.714628	21.8
11	49.838218	49.6	24.982545	13.8
12	59.952036	60.8	202.51054	121.7
13	48.286265	46.2	190.7173	78.7
14	51.535571	51.2	45.20259	25.7

Table 2: Estimated Posterior Means of μ_k and σ_k^2

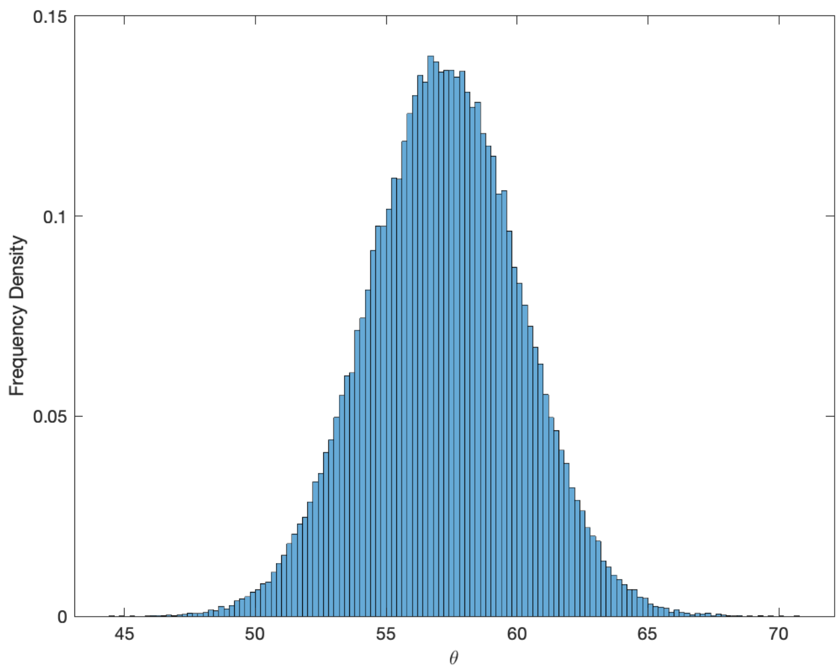

Figure 1: Posterior distribution of θ

The empirical means and variances were calculated from the data $\mathbf{y}$ for comparison. We can see that the expected value of μ_k under the posterior distribution is very close to the average of the data, whereas the variance seems to remain significantly higher than the empirical variance (roughly double). This confirms the predictions in Question 3, that the large uncertainty in the data does not decrease with the observations $\mathbf{y}$.

The posterior mean of θ was calculated as 57.19, slightly less than the starting value 60. The distribution appears to be Gaussian. We can try to calculate this distribution analytically:

$$\begin{aligned}\pi(\theta|\mathbf{y}) &\propto \pi(\mathbf{y}|\theta)\pi(\theta)\\ &= \left(\int_{\boldsymbol{\mu},\boldsymbol{\sigma}^{-2}} \pi(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\sigma}^{-2},\theta)\pi(\boldsymbol{\mu},\boldsymbol{\sigma}^{-2}|\theta)\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\boldsymbol{\sigma}^{-2}\right)\pi(\theta)\\ &= \left(\int_{\boldsymbol{\mu},\boldsymbol{\sigma}^{-2}} \pi(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\sigma}^{-2})\pi(\boldsymbol{\mu}|\theta)\pi(\boldsymbol{\sigma}^{-2})\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\boldsymbol{\sigma}^{-2}\right)\pi(\theta)\end{aligned} \tag{11}$$

Firstly note that we may treat each value of k independently and instead take the product over this at the end. We therefore can simplify the problem to calculating the above for some fixed value of k (for the sake of simplicity drop the subscript). We then calculate the bracketed part above as

$$\begin{aligned}&\int_{-\infty}^{+\infty}\int_{0}^{\infty}\left(\sigma^{-2}\right)^{T/2+\alpha_0-1}\exp\left(-\frac{\sigma^{-2}}{2}\left(\sum_{i=1}^{T}(y_{ki}-\mu)^2\right)-\frac{1}{2\sigma_0^2}(\mu-\theta)^2-\beta_0\sigma^{-2}\right)\mathrm{d}\sigma^{-2}\,\mathrm{d}\mu\\ &\propto\int_{-\infty}^{\infty}\left(\frac{1}{\frac{1}{2}\sum_{i=1}^{T}(y_{ki}-\mu)^2+\beta_0}\right)^{T/2+\alpha_0}\exp\left(-\frac{1}{2\sigma_0^2}(\mu-\theta)^2\right)\mathrm{d}\mu\end{aligned} \tag{12}$$

The values of α_0 and β_0 can be ignored since they're comparably small. Modelling this function gives a very similar (though not exact) Gaussian distribution for θ. Assuming it is, then by taking the product of $K+1$ Gaussian distributions it will indeed give a Gaussian, and thus we conclude that the posterior distribution of θ is approximately (but not exactly) Gaussian.

Question 5

Code 5, referenced on page 12, was used to estimate $\mathbb{P}(\mu_k > \theta|\mathbf{y})$ by finding the average of the indicator function $\mathbb{1}_{\mu_k>\theta}$ over the posterior distribution. The results below in Table 3 were produced when $N = 1000$.

k	$\mathbb{P}(\mu_k > \theta\|\mathbf{y})$
1	1.000
2	0.104
3	0.100
4	0.152
5	0.004
6	0.932
7	0.120
8	0.004
9	0.840
10	0.996
11	0.036
12	0.728
13	0.056
14	0.076

Table 3: Estimated values of $\mathbb{P}(\mu_k > \theta|\mathbf{y})$

Notice how all of these probabilities are either very large or very small, there is not much uncertainty if a team is above or below average. This suggests that the variance in μ_k will be small, since then μ_k will either lie on one side or the other of θ.

Question 6

The simulations in Questions 4-5 were repeated 500 times with $N = 250$ using Code 5, and the sample variances were calculated and tabulated below in Table 4.

k	S.V in μ_k	S.V in $\mathbb{P}(\mu_k > \theta \mid \mathbf{y})$
1	0.31932168	1.9852249e-5
2	0.050576302	4.5196287e-4
3	0.073926224	3.4977321e-4
4	0.079487089	5.1195742e-4
5	0.025197288	3.1824032e-5
6	0.31273951	2.2441523e-4
7	0.11360919	5.2450315e-4
8	0.030712084	6.5253322e-5
9	0.098515333	6.3724319e-4
10	0.081711728	1.0228200e-5
11	0.020928080	9.9218212e-5
12	0.10951773	8.4340066e-4
13	0.11541775	2.3513036e-4
14	0.037308591	3.0127455e-4

Table 4: Sample Variances in μ_k and $\mathbb{P}(\mu_k > \theta|\mathbf{y})$

Notice the sample variances in μ_k are indeed small, which causes the variance in $\mathbb{P}(\mu_k > \theta|\mathbf{y})$ to be small, since μ_k will likely lie to one side or the other of θ (unless the two are very close). The sample variances of $\mathbb{P}(\mu_k > \theta|\mathbf{y})$ is far smaller than those of μ_k, and this is because the indicator function does not depend contentiously on the values of μ_k and θ. Hence $\mathbb{1}_{\mu_k > \theta}$ is roughly constant over the posterior distribution.

To investigate how this variance decays with N, the method above was repeated 500 times with $N = 20, 30, 50, 100, 150, 250$ using Code 5. The sample variances were then normalised (sample variance at $N = 20$ is 1) such that they could be compared when plotted in Figures 2 and 3 below.

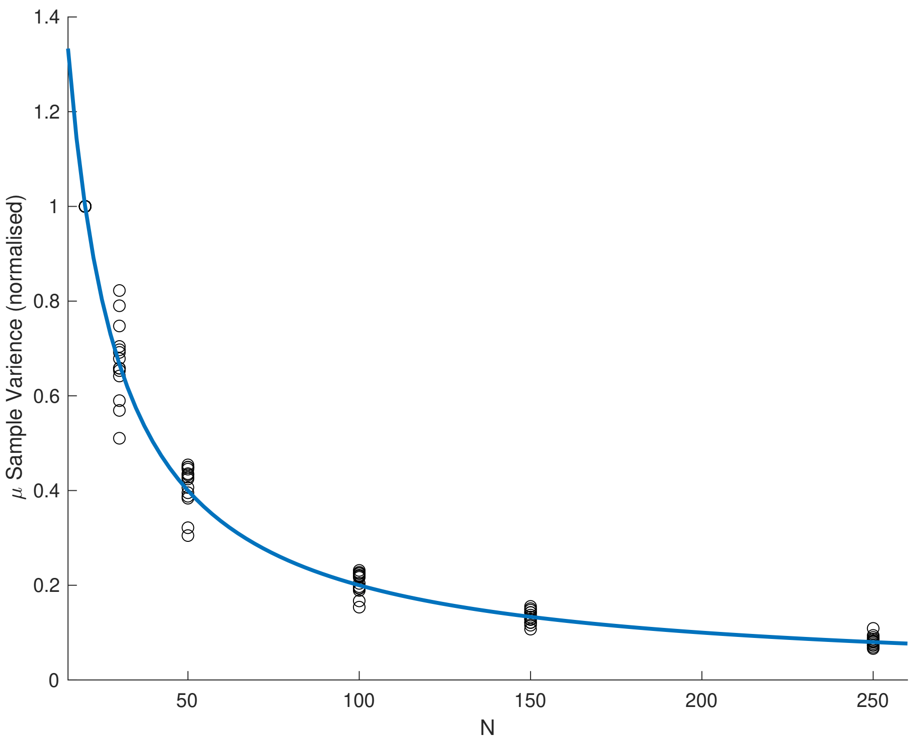

Figure 2: Sample variance of μ_k for different N

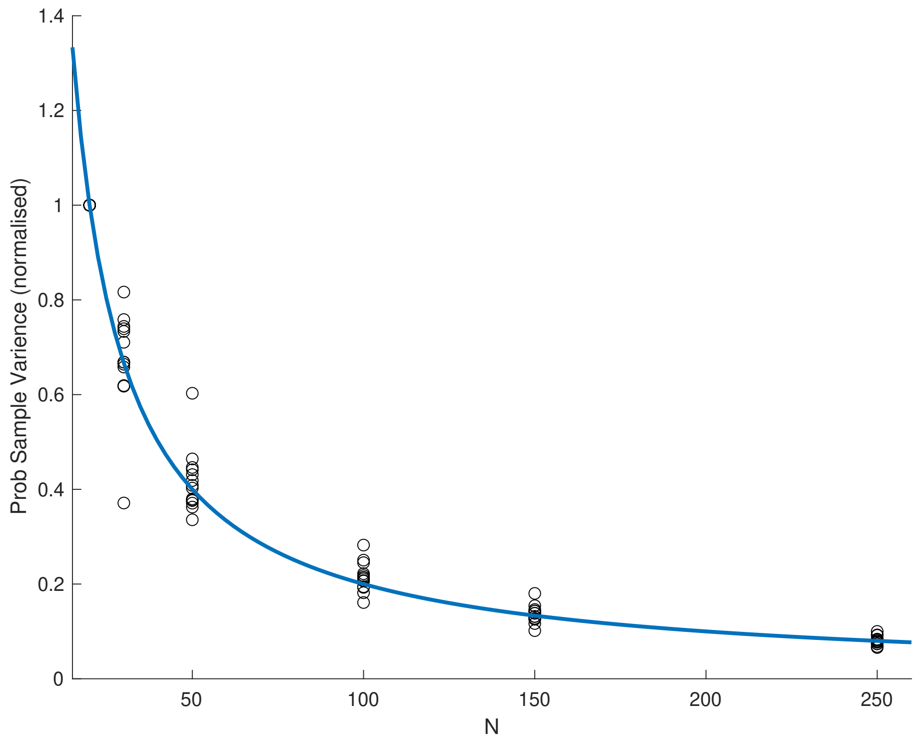

Figure 3: Sample variance of $\mathbb{P}(\mu_k > \theta|\mathbf{y})$ for different N

Notice how both sample variances decay like $1/N$. This is a direct consequence of the Central Limit Theorem. It states that if $X_i \sim X$ are i.i.d and $\mathrm{Var}(X) < \infty$, then

$$\frac{1}{N}\sum_{i=1}^{N} X_i - \mathbb{E}(X) \xrightarrow{d} \mathrm{N}\left(0, \frac{\mathrm{Var}(X)}{N}\right) \tag{13}$$

in the limit as $N \to \infty$. We know that the sample variance of any i.i.d random variables converges almost surely to the true variance. Since the X_i^2 are uniformly integrable, then

$$\mathrm{Var}\left(\frac{\sum_i^N \left(X_i - \bar{X}\right)^2}{N-1}\right) \to \frac{\mathrm{Var}(X)}{N} \tag{14}$$

asymptotically. This is exactly what we see in the two above figures; that as N increases, the variation from the normalised hyperbola shrinks, with all data points practically lying on it at $N = 250$.

Question 7

Sample variances were calculated by Code 7, referenced on page 14, with $N = 30$, while $M = 0, 10, 100$. The value for N was chosen deliberately small to highlight any changes in the distributions for small M. The simulations were repeated 500 times as before, and the results are tabulated in Table 5 below.

k	μ_k			$\mathbb{P}(\mu_k > \theta\|\mathbf{y})$		
	$M = 0$	$M = 10$	$M = 100$	$M = 0$	$M = 10$	$M = 100$
1	1.843564	2.237448	1.939471	1.081051e-4	1.182632e-4	1.569183e-4
2	0.4648628	0.4351083	0.4309788	3.686212e-3	3.447980e-3	3.992411e-3
3	0.6294533	0.7317089	0.6772295	2.928572e-3	2.954687e-3	3.086511e-3
4	0.6383631	0.7017769	0.6644687	4.627673e-3	4.874081e-3	4.215626e-3
5	0.2314946	0.1679683	0.2017004	3.287063e-4	2.038922e-4	2.483367e-4
6	2.331541	2.60194	2.796847	1.832304e-3	1.972389e-3	1.696811e-3
7	0.7538931	0.7933827	0.8928876	3.805348e-3	3.677091e-3	4.001117e-3
8	0.2473297	0.2971523	0.2557318	5.900512e-4	5.645557e-4	5.807659e-4
9	0.8379703	0.9233271	0.9659404	5.199091e-3	5.755350e-3	5.377025e-3
10	0.6053968	0.6235268	0.7439646	5.815631e-5	1.014072e-4	1.624360e-4
11	0.2069356	0.1995997	0.1818190	8.819906e-4	7.993943e-4	8.766466e-4
12	0.9143364	0.9594655	0.9342799	6.825206e-3	6.413231e-3	6.973787e-3
13	0.9178443	0.8146838	0.8955065	1.961630e-3	1.876802e-3	1.898352e-3
14	0.2940941	0.3144930	0.2827326	2.813431e-3	2.648746e-3	2.713315e-3

Table 5: Sample variance in μ_k and $\mathbb{P}(\mu_k > \theta|\mathbf{y})$ for different M

There is very little change in the variance for μ_k, suggesting that the distribution settles very quickly. This was intended, since the starting value $\mathbf{x}^0$ was chosen to be as close to the posterior distribution as possible (with no prior knowledge of $\mathbf{y}$).
The variance in $\mathbb{P}(\mu_k > \theta|\mathbf{y})$ is also largely unaffected by the value of M with a few sporadic large deviations. These can be ignored, since the method used to calculate $\mathbb{P}(\mu_k > \theta|\mathbf{y})$ does not depend continuously on the values of μ_k and θ. Thus, when N is small, the values calculated in different simulations can change by 'large' (discontinuous) amounts.

Question 8

The chain was run independently 500 times using a variety of starting states when $N = 30$ and $M = 70, 0$. The sample means and variances of μ_k were calculated, as well as the distance (2-norm) between these vectors and the ones in Tables 2 and 5 respectively (in particular the calculated mean and $M = 100$ variance of μ_k). These norms are designed to indicate if the two distributions are similar by checking if the sample means and variances are close to the ones previously in the project. For the sake of notation, let m_M denote the norm in the difference of means and v_M likewise for the variance when the value M is used. Also let $\mu_k^{(i)}$ denote the state x_k^i, $\sigma_k^{(i)} = x_{K+k}^i$ and $\theta^{(i)} = x_{2K+1}^i$. The following data tabulated in Table 6 was produced by Code 7.

Starting State	m_{70}	v_{70}	m_0	v_0
$\mu_k^{(0)} = 60, \sigma_k^{(0)} = 100, \theta^{(0)} = 60$	0.97288	0.53985	1.1936	0.64422
$\mu_k^{(0)} = 1000, \sigma_k^{(0)} = 0.001, \theta^{(0)} = 1000$	0.93635	0.71239	0.87688	0.44730
$\mu_k^{(0)} = -500, \sigma_k^{(0)} = 500, \theta^{(0)} = -500$	1.0400	0.67274	703.40	467.64
$\mu_k^{(0)} = -500, \sigma_k^{(0)} = 500, \theta^{(0)} = 60$	0.95777	0.62181	1.6842	0.78523

Table 6: Effect of different starting states on MCMC

We see that columns 2 and 3 match regardless of the starting state, confirming that the chain tends to the posterior distribution regardless of the starting distribution. Secondly notice that for our original starting state, $m_0 \approx m_{70}$ and $v_0 \approx v_{70}$, matching with the conclusion of Question 7 that in this case the posterior settles very quickly. However the values of m_0 and v_0 vary largely for different starting states, and this will be discussed in greater detail below by considering the conditional distributions in Question 2.

Affects of varying $\mu_k^{(0)}$: the way the chain is set up means that varying $\mu_k^{(0)}$ has no effect, since the conditional distributions of $\mu_k|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \theta$ do not depend on $\mu_k^{(0)}$. Therefore these parameters can be ignored

Affects of varying $\sigma_k^{(0)}$: if $\sigma_k^{(0)}$ is chosen to be very large, then $\mu_k^{(1)}$ is centred around $\theta^{(0)}$. Therefore if $\theta^{(0)}$ is chosen to be far away from the true expectation of μ_k, then there will be a large error. This error will also cause the distribution of $\sigma_k^{(1)}$ to become shifted, and will cause a (partial) error in the value of $\theta^{(1)}$, but this is less important, since we average over the values of μ_k. Therefore we expect slow convergence to the posterior (consistent with the above). Conversely if $\sigma_k^{(0)}$ is chosen to be small, then $\mu_k^{(1)}$ is distributed as a sharp spike around the value $\sum_{t=1}^{T} y_{kt}/T$. This is very close to the mean of the posterior distribution, and so there is almost no error, and thus $\theta^{(1)}$ and $\sigma_k^{(1)}$ will be distributed correctly as well. This causes the distribution to settle very quickly. (consistent with the above).

Affects of varying $\theta^{(0)}$: We've seen that the value of $\theta^{(0)}$ only has an effect on the chain if $\sigma_k^{(0)}$ is chosen to be large. If this is the case then we require $\theta^{(0)}$ to be a reasonable estimate for the average of all μ_k. This way any individual μ_k will (hopefully) not be too far away from it's true mean, and hence the distribution settle relatively quickly. Compare the last two starting states in Table 6. The results confirm exactly this point; that when $\theta^{(0)}$ is near the average the posterior settles far quicker.

We conclude that the value of M is only important when the starting state has $\sigma^{(0)}$ chosen to be large and $\theta^{(0)}$ chosen far away from the average.

Code

Gibbs Sampler

```
digits(8)
Data={"Team","2024","2023","2022","2021","2020"
"Arsnal",83,90,78,87,81
"Asten Villa",47,56,45,50,60
"Blackborn Rovers",42,44,60,46,56
"Boltin Wandrers",58,53,44,40,60
"Charlston Athletic",46,53,49,44,45
"Chelsea Buns",95,79,67,64,71
"Evraton",61,39,59,43,46
"Fullem",44,52,48,44,53
"Livurpule",58,60,64,80,54
"Manchester Ununited",77,75,83,77,70
"Middlesbro",55,48,49,45,51
"Newcassel Divided",44,56,69,71,64
"Slothampton",32,47,52,45,55
"Tottenham Coldspur",52,45,50,50,59};
K=size(Data,1)-1;
T=size(Data,2)-1;

sigma_0=10;
alpha_0=10^-5;
beta_0=10^-3;
mu_0=60;
tau_0=20;

StartVal=zeros(2*K+1,1);
StartVal(1:2*K+1)=60;
StartVal(K+1:2*K)=100;

N=5;
tic
Sim=Gibbs(Data,sigma_0,alpha_0,beta_0,mu_0,tau_0,N,StartVal)
latex(sym(vpa(Sim)))
toc

function Simulations = Gibbs(Data,sigma_0,alpha_0,beta_0,mu_0,tau_0,N,StartVal)
K=size(Data,1)-1;
T=size(Data,2)-1;

%Standardizes parameters
sigma_0=sigma_0^2;
tau_0=tau_0^2;

Simulations=zeros(2*K+1,N+1);
Simulations(:,1)=StartVal;
Counter=2;

while Counter<=N+1
    lineLength = fprintf('%.1f%% complete',Counter*100/N);
    SubCounter=1;
    while SubCounter<=2*K+1
        if SubCounter<=K
```

```
            %   SIMULATING MU PARAMETERS
            sigma_k=Simulations(SubCounter+K,Counter-1);
            sumData=sum([Data{SubCounter+1,2:T+1}]);
            theta=Simulations(2*K+1,Counter-1);
            Mean=(sigma_k^(-1)*sumData+theta*sigma_0^(-1))/(T*sigma_k^(-1)+sigma_0^(-1));
            Var=1/(T*sigma_k^(-1)+sigma_0^(-1));
            Simulations(SubCounter,Counter)=normrnd(Mean,sqrt(Var));
        elseif SubCounter<=2*K
            %   SIMULATING SIGMA PARAMETERS
            sumData=sum(([Data{SubCounter-K+1,2:T+1}]-Simulations(SubCounter-K,Counter)).^2);
            Mean=alpha_0+T/2;
            Var=beta_0+sumData/2;
            Simulations(SubCounter,Counter)=(gamrnd(Mean,1/Var))^-1;
        else
            %   SIMULATING THETA PARAMETER
            sumData=sum(Simulations(1:K,Counter));
            Mean=(sigma_0^-1*sumData+mu_0*tau_0^-1)/(K*sigma_0^-1+tau_0^-1);
            Var=1/(K*sigma_0^-1+tau_0^-1);
            Simulations(SubCounter,Counter)=normrnd(Mean,sqrt(Var));
        end
        SubCounter=SubCounter+1;
    end
    fprintf(repmat('\b',1,lineLength))
    Counter=Counter+1;
end
end
```

Code 4

```
digits(8)

sigma_0=10;
alpha_0=10^-5;
beta_0=10^-3;
mu_0=60;
tau_0=20;

StartVal=zeros(2*K+1,1);
StartVal(1:2*K+1)=60;
StartVal(K+1:2*K)=100;

tic
N=250;
Sim=Gibbs(Data,sigma_0,alpha_0,beta_0,mu_0,tau_0,N,StartVal);
toc

PosteriorMean=sum(Sim(:,2:end),2)./N;
EstimatedMeans=zeros(K,4);
EstimatedMeans(:,1)=PosteriorMean(1:K);
EstimatedMeans(:,3)=PosteriorMean(K+1:2*K);
EstimatedMeans(:,2)=sum(cell2mat(Data(2:K+1,2:T+1)),2)/T;
EstimatedMeans(:,4)=sum((cell2mat(Data(2:K+1,2:T+1))-EstimatedMeans(:,2)).^2,2)/(T-1);
sum(EstimatedMeans(:,1))

disp(EstimatedMeans)
latex(sym(vpa(EstimatedMeans)))

thetaData=Sim(end,2:end);
mid=sum(thetaData)/N

XLower=min(thetaData)-2;
XHigher=max(thetaData)+2;

h=figure;
histogram(thetaData,'Normalization','pdf')
xlabel('\theta')
ylabel('Frequency Density')
print(h,'Image_1','-depsc')
```

Code 5

```
digits(8)

sigma_0=10;
alpha_0=10^-5;
beta_0=10^-3;
mu_0=60;
tau_0=20;

List=[20,30,50,100,150,250];
NCounter=1;
MeanProb=zeros(K,size(List,2));
MeanMu=zeros(K,size(List,2));
VarProb=zeros(K,size(List,2));
VarMu=zeros(K,size(List,2));

while NCounter<=size(List,2)
N=List(1,NCounter);
Number_of_Runs=500;

Prob=zeros(K,Number_of_Runs);
Mu=zeros(K,Number_of_Runs);

SuperCounter=1;
while SuperCounter<=Number_of_Runs

LineLength=fprintf('List: %d, Run: %.1f%% complete',N,SuperCounter*100/Number_of_Runs);
Sim=Gibbs(Data,sigma_0,alpha_0,beta_0,mu_0,tau_0,N);

Counter=2;
Carry=zeros(K,1);
while Counter<=N+1
    SubCounter=1;
    while SubCounter<=K
        if Sim(SubCounter,Counter)>Sim(2*K+1,Counter)
            Carry(SubCounter,1)=Carry(SubCounter,1)+1;
        end
        SubCounter=SubCounter+1;
    end
    Counter=Counter+1;
end
Prob(:,SuperCounter)=Carry/N;
Carry=sum(Sim,2);
Mu(:,SuperCounter)=Carry(1:K,1)/N;
fprintf(repmat('\b',1,LineLength))
SuperCounter=SuperCounter+1;
end

MeanProb(:,NCounter)=sum(Prob,2)/Number_of_Runs;
VarProb(:,NCounter)=sum((Prob-MeanProb(:,NCounter)).^2,2)/(Number_of_Runs-1);
MeanMu(:,NCounter)=sum(Mu,2)/Number_of_Runs;
VarMu(:,NCounter)=sum((Mu-MeanMu(:,NCounter)).^2,2)/(Number_of_Runs-1);
NCounter=NCounter+1;
end
disp(MeanMu)
```

```
disp(VarMu)
disp(MeanProb)
disp(VarProb)

figure
Counter=1;
while Counter<=K
    scatter(List,VarMu(Counter,:)/VarMu(Counter,1),'k')
    hold on
    Counter=Counter+1;
end
x=linspace(List(1,1)-5,List(1,end)+10);
y=List(1,1)./x;
plot(x,y,'LineWidth',2)
xlim([List(1,1)-5,List(1,end)+10])
xlabel('N')
ylabel('\mu Sample Varience (normalised)')
print('Image_2','-depsc')

figure
Counter=1;
while Counter<=K
    scatter(List,VarProb(Counter,:)/VarProb(Counter,1),'k')
    hold on
    Counter=Counter+1;
end
x=linspace(List(1,1)-5,List(1,end)+10);
y=List(1,1)./x;
plot(x,y,'LineWidth',2)
xlim([List(1,1)-5,List(1,end)+10])
xlabel('N')
ylabel('Prob Sample Varience (normalised)')
print('Image_3','-depsc')
```

Code 7

```
digits(8)
K=size(Data,1)-1;
T=size(Data,2)-1;

sigma_0=10;
alpha_0=10^-5;
beta_0=10^-3;
mu_0=60;
tau_0=20;

CalculatedMean=[81.4968825139968;52.1355802069355;50.672260968429;52.5080817745987;...
    48.1389564329937;69.784992226184;51.6374396731938;48.7565228578938;...
    62.2933104774652;74.386133868267;49.8382175215858;59.9520363668573;...
    48.2862648696441;51.5355709368392];
CalculatedVar=[1.9394709;0.43097878;0.67722952;0.66446875;0.20170042;2.7968475;0.89288769;...
    0.25573182;0.96594047;0.74396463;0.18181906;0.93427989;0.89550653;0.2827326];

StartingValue=zeros(2*K+1,1);
StartingValue(1:2*K+1)=60;
StartingValue(K+1:2*K)=100;
StartingValue(2*K+1)=60;

List=[0,70];
N=30;

NCounter=1;
MeanProb=zeros(K,size(List,2));
MeanMu=zeros(K,size(List,2));
VarProb=zeros(K,size(List,2));
VarMu=zeros(K,size(List,2));

while NCounter<=size(List,2)
M=List(1,NCounter);
Number_of_Runs=500;

Prob=zeros(K,Number_of_Runs);
Mu=zeros(K,Number_of_Runs);

SuperCounter=1;
while SuperCounter<=Number_of_Runs

LineLength=fprintf('List: %d, Run: %.1f%% complete',M+N,SuperCounter*100/Number_of_Runs);
Sim=Gibbs(Data,sigma_0,alpha_0,beta_0,mu_0,tau_0,M+N,StartingValue);

Counter=2+M;
Carry=zeros(K,1);
while Counter<=M+N+1
    SubCounter=1;
    while SubCounter<=K
        if Sim(SubCounter,Counter)>Sim(2*K+1,Counter)
            Carry(SubCounter,1)=Carry(SubCounter,1)+1;
        end
        SubCounter=SubCounter+1;
    end
```

```
    Counter=Counter+1;
end
Prob(:,SuperCounter)=Carry/N;
Carry=sum(Sim(:,M+2:M+N+1),2);
Mu(:,SuperCounter)=Carry(1:K,1)/N;
fprintf(repmat('\b',1,LineLength))
SuperCounter=SuperCounter+1;
end

MeanProb(:,NCounter)=sum(Prob,2)/Number_of_Runs;
VarProb(:,NCounter)=sum((Prob-MeanProb(:,NCounter)).^2,2)/(Number_of_Runs-1);
MeanMu(:,NCounter)=sum(Mu,2)/Number_of_Runs;
VarMu(:,NCounter)=sum((Mu-MeanMu(:,NCounter)).^2,2)/(Number_of_Runs-1);
NCounter=NCounter+1;
end
disp(MeanMu)
disp(VarMu)
latex(sym(vpa(VarMu)))

Norms=[List;vecnorm(MeanMu-CalculatedMean);vecnorm(VarMu-CalculatedVar)];
disp(Norms)
latex(sym(vpa(Norms)))

disp(MeanProb)
disp(VarProb)
latex(sym(vpa(VarProb)))
```