University of Amsterdam

Theory of Complex Systems

On Statistical Inference and Modeling Data with the Ising Model

September 28, 2025

Student:
Kattelijn Bouma
12862444

Lecturer:
dr. Clelia de Mulatier

Course code:
5284THCS6Y

1 Modeling the activity of a single neuron

The dataset "Data_neuron.txt" contains the neuronal activity of one neuron, i.e. the value of the successive times at which the neuron spikes. The time unit is the millisecond.

Q1

In Figure 1, we present a histogram illustrating the distribution $P(\tau)$ of time intervals τ between consecutive spikes obtained from the dataset "Data_neuron.txt". Additionally, we determined the minimum time interval, denoted as the refractory period $\tau_0 \approx 1.90$ ms.

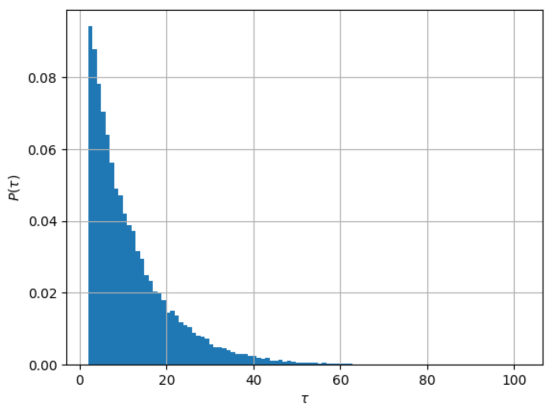


Figure 1: Histogram illustrating the distribution $P(\tau)$ of the time intervals τ between successive spikes.

Q2

In Figure 2 the natural logarithm of $P(\tau)$ is plotted illustrating that the decay of the distribution of inter-spike intervals is indeed exponential for small values of τ. We fitted an exponential function to the data and measured the decay rate $\lambda \approx 0.12$.

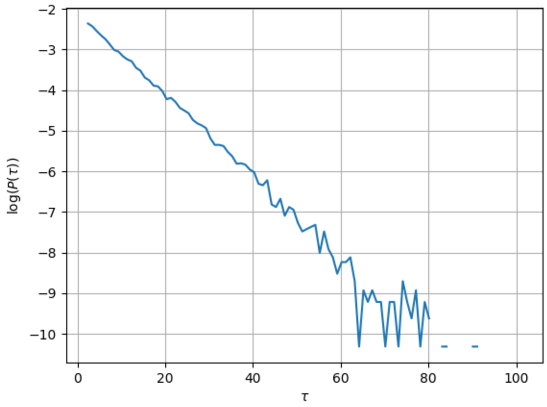


Figure 2: Natural logarithm of $P(\tau)$ as a function of τ.

Q3

The distribution of the waiting times τ between two consecutive events of a Poisson process with rate λ is the exponential distribution

$$P(\tau) = \lambda e^{-\lambda\tau}.$$

For a delayed Poisson process with a refractory period τ_0, the distribution of time intervals τ can now be expressed as

$$P(\tau) = \begin{cases} 0 & \text{if } \tau < \tau_0 \\ \lambda e^{-\lambda\tau} & \text{otherwise} \end{cases}. \tag{1}$$

In Figure 3 we again plotted the distribution $P(\tau)$ obtained from the dataset with the fitted model distribution overlaid.

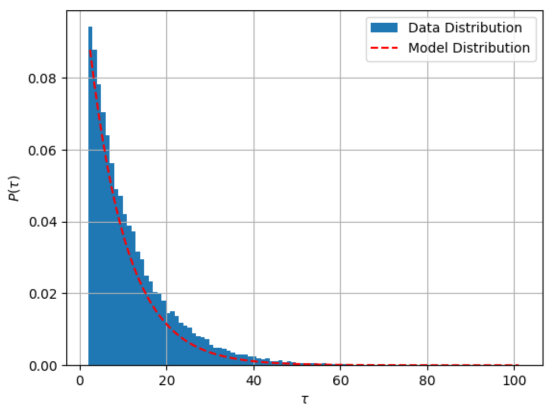


Figure 3: Histogram of distribution $P(\tau)$ with fitted model distribution overlaid.

Q4

We can generate another 1000 spike times using the exponential distribution with a mean of $\frac{1}{\lambda}$ and ensuring that the generated inter-spike interval is greater than the refractory period.

Q5

The average spike rate $f = \frac{1}{\langle\tau\rangle}$ of the neuron in the data is approximately 0.084 spikes/ms.

2 Modeling binary data with the Ising model

2.1 Pairwise spin model

We take the probability distribution to have the general form of an Ising model

$$p_{\boldsymbol{g}}(\boldsymbol{s}) = \frac{1}{Z(\boldsymbol{g})} \exp\left(\sum_{i=1}^{n} h_i s_i + \sum_{pair(i,j)} J_{ij} s_i s_j \right), \tag{2}$$

where n is the number of spin variables, where $pair(i,j)$ denotes a summation over all possible pairs of distinct spin variables, where $\boldsymbol{g} = (h_1, \cdots, h_n, J_{1,2}, \cdots, J_{n-1,n})$ is a vector of (real) parameters, and where $Z(\boldsymbol{g})$ is a normalization factor.

The general goal is to infer the set of parameters

$$\boldsymbol{g} = (h_1, \cdots, h_n, J_{1,2}, \cdots, J_{n-1,n}) \tag{3}$$

that is the most appropriate to model the data, i.e. to find the parameters $\boldsymbol{g}$ for which the probability distribution in Eq. (2) best fits the data.

Q1.1

Assuming n is the number of spin variables, there are $\frac{n(n-1)}{2}$ terms in the sum over the $pair(i,j)$, thus the vector $\boldsymbol{g}$ defined in (3) contains $n + \frac{n(n-1)}{2}$ parameters and we can rewrite the sum over the $pair(i,j)$ as a double sum over i and j as follows

$$\sum_{pair(i,j)} J_{ij} s_i s_j = \sum_{i}^{n-1} \sum_{j=i+1}^{n} J_{ij} s_i s_j.$$

Q1.2

For a system with $n = 3$ spins, the terms in the exponential of Eq. (2) are given by

$$\sum_{i=1}^{3} h_i s_i + \sum_{i=1}^{2} \sum_{j=i+1}^{3} J_{ij} s_i s_j = h_1 s_1 + h_2 s_2 + h_3 s_3 + J_{1,2} s_1 s_2 + J_{1,3} s_1 s_3 + J_{2,3} s_2 s_3.$$

Q1.3

In Eq. (2), we can recognize the Boltzmann distribution

$$P(\boldsymbol{s}) = \frac{1}{Z} \exp(-\beta E(\boldsymbol{s})), \tag{4}$$

in which the parameter β was taken equal to 1. The energy function is thus given by

$$E(\boldsymbol{s}) = -\left(\sum_{i=1}^{n} h_i s_i + \sum_{pair(i,j)} J_{ij} s_i s_j \right) \tag{5}$$

and the partition function is then given by

$$Z = \sum_{\boldsymbol{s}} \exp(-E(\boldsymbol{s})). \tag{6}$$

Q1.4

If h_i is positive, a spin s_i will tend to turn positive as this will minimize the associated energy $-h_i s_i$. Similarly, if J_{ij} is positive configurations of (s_i, s_j) where $s_i s_j$ is positive will minimize the coupling energy $-J_{ij} s_i s_j$.

The sign of a local field parameter h_i indicates a tendency for judge i to vote conservatively (positive) or in favor of liberal-oriented decisions (negative). The sign of a coupling parameter J_{ij} indicates if judges i and j tend to vote in the same or opposite direction.

2.2 Observables

The important observables of the system are the local average magnetisations $\langle s_i \rangle$ and the local correlations $c_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle$, where the angle brackets $\langle A(\boldsymbol{s}) \rangle$ denotes the ensemble average (or thermal average) of the microscopic quantity $A(\boldsymbol{s})$.

Q2.1

Given a stationary probability distribution of the state $p_{\boldsymbol{g}}(\boldsymbol{s})$, $\langle s_i \rangle$ and $\langle s_i s_j \rangle$ are defined as

$$\langle s_i \rangle = \sum_{\boldsymbol{s}} p_{\boldsymbol{g}}(\boldsymbol{s}) s_i \quad \text{and} \quad \langle s_i s_j \rangle = \sum_{\boldsymbol{s}} p_{\boldsymbol{g}}(\boldsymbol{s}) s_i s_j. \tag{7}$$

Q2.2

Considering a dataset $\hat{\boldsymbol{s}}$ composed of N independent observations of the spins: $\hat{\boldsymbol{s}} = (\boldsymbol{s}^{(1)}, \cdots, \boldsymbol{s}^{(N)})$, we compute the empirical averages of s_i and $s_i s_j$ as follows

$$\langle s_i \rangle_D = \frac{1}{N} \sum_{k=1}^{N} s_i^{(k)} \quad \text{and} \quad \langle s_i s_j \rangle_D = \frac{1}{N} \sum_{k=1}^{N} s_i^{(k)} s_j^{(k)}. \tag{8}$$

Q2.3

The empirical probability $p_D(\boldsymbol{s})$ to observe the state $\boldsymbol{s}$ in the data is given by the ratio of the number of times $K(\boldsymbol{s})$ that the state $\boldsymbol{s}$ appears in the data to the total number of datapoints:

$$p_D(\boldsymbol{s}) = \frac{K(\boldsymbol{s})}{N}. \tag{9}$$

We can re-write $\langle s_i \rangle_D$ as

$$\begin{aligned} \langle s_i \rangle_D &= \frac{1}{N} \sum_{k=1}^{N} s_i^{(k)} \\ &= \frac{1}{N} \sum_{\boldsymbol{s}} K(\boldsymbol{s}) s_i \\ &= \sum_{\boldsymbol{s}} \frac{K(\boldsymbol{s})}{N} s_i \\ &= \sum_{\boldsymbol{s}} P_D(\boldsymbol{s}) s_i. \end{aligned}$$

As the number of datapoints N increases, the empirical distribution $p_D(\boldsymbol{s})$ converges to the true distribution $p_{\boldsymbol{g}}(\boldsymbol{s})$, thus

$$\begin{aligned} \lim_{N\to\infty} \langle s_i \rangle_D &= \lim_{N\to\infty} \sum_{\boldsymbol{s}} P_D(\boldsymbol{s}) s_i \\ &= \sum_{\boldsymbol{s}} p_{\boldsymbol{g}}(\boldsymbol{s}) s_i \\ &= \langle s_i \rangle. \end{aligned}$$

2.3 Maximum Entropy models

We will show that the probability distribution defined in Eq. (2) is indeed the (most general) probability distribution that maximizes the Shannon entropy $S[p_{\boldsymbol{g}}(\boldsymbol{s})]$ given a set of constraints.

Q3.1

Consider a spin system with stationary probability distribution $p(\boldsymbol{s})$. The Shannon entropy is defined as

$$S[p(\boldsymbol{s})] = -k_B \sum_{\boldsymbol{s}} p(\boldsymbol{s}) \log p(\boldsymbol{s}), \tag{10}$$

where as mentioned before for the Boltzmann distribution we will take $k_B = 1$.

Q3.2

We are looking for the set of 2^n probabilities $p(\boldsymbol{s})$ that maximizes $S[p(\boldsymbol{s})]$ given the constraints:

$$\sum_{\boldsymbol{s}} p(\boldsymbol{s}) = 1 \quad \text{and} \quad \sum_{\boldsymbol{s}} p(\boldsymbol{s}) s_i(\boldsymbol{s}) = \langle s_i \rangle_D \quad \text{and} \quad \sum_{\boldsymbol{s}} p(\boldsymbol{s}) s_i(\boldsymbol{s}) s_j(\boldsymbol{s}) = \langle s_i s_j \rangle_D. \tag{11}$$

In total there is a number of $1 + n + \frac{n(n-1)}{2}$ constraints.

Q3.3

We introduce an auxiliary function

$$\begin{aligned} U[p(\boldsymbol{s})] = S[p(\boldsymbol{s})] &+ \lambda_0 \left(\sum_{\boldsymbol{s}} p(\boldsymbol{s}) - 1 \right) + \sum_{i=1}^{n} \alpha_i \left(\sum_{\boldsymbol{s}} p(\boldsymbol{s}) s_i(\boldsymbol{s}) - \langle s_i \rangle_D \right) \\ &+ \sum_{pair(i,j)}^{n} \eta_{ij} \left(\sum_{\boldsymbol{s}} p(\boldsymbol{s}) s_i(\boldsymbol{s}) s_j(\boldsymbol{s}) - \langle s_i s_j \rangle_D \right) \end{aligned} \tag{12}$$

To find $p(\boldsymbol{s})$ one must maximize this auxiliary function with respect to the 2^n probabilities $p(\boldsymbol{s})$. The derivative of $U[\boldsymbol{p}]$ with respect to $p(\boldsymbol{s})$ consists of the four terms

$$\begin{aligned} \frac{\partial S[\boldsymbol{p}]}{\partial p_{\boldsymbol{s}}} &= \frac{\partial \sum_{\boldsymbol{s}} \boldsymbol{p} \log \boldsymbol{p}}{\partial p_{\boldsymbol{s}}} = -\log p_{\boldsymbol{s}} - 1 \\ \frac{\partial \lambda_0 \left(\sum_{\boldsymbol{s}} \boldsymbol{p} - 1 \right)}{\partial p_{\boldsymbol{s}}} &= \frac{\partial \lambda_0 \sum_{\boldsymbol{s}} \boldsymbol{p}}{\partial p_{\boldsymbol{s}}} = \lambda_0 \\ \frac{\partial \sum_{i=1}^{n} \alpha_i \left(\sum_{\boldsymbol{s}} \boldsymbol{p} s_i(\boldsymbol{s}) - \langle s_i \rangle_D \right)}{\partial p_{\boldsymbol{s}}} &= \frac{\partial \sum_{i=1}^{n} \alpha_i \sum_{\boldsymbol{s}} \boldsymbol{p} s_i(\boldsymbol{s})}{\partial p_{\boldsymbol{s}}} = \sum_{i=1}^{n} \alpha_i s_i(\boldsymbol{s}) \\ \frac{\partial \sum_{pair(i,j)}^{n} \eta_{ij} \left(\sum_{\boldsymbol{s}} \boldsymbol{p} s_i(\boldsymbol{s}) s_j(\boldsymbol{s}) - \langle s_i s_j \rangle_D \right)}{\partial p_{\boldsymbol{s}}} &= \frac{\partial \sum_{pair(i,j)}^{n} \eta_{ij} \sum_{\boldsymbol{s}} \boldsymbol{p} s_i(\boldsymbol{s}) s_j(\boldsymbol{s})}{\partial p_{\boldsymbol{s}}} = \sum_{pair(i,j)} \eta_{ij} s_i(\boldsymbol{s}) s_j(\boldsymbol{s}), \end{aligned}$$

where we use that $\frac{\partial \sum_{\boldsymbol{s}} \boldsymbol{p}}{\partial p_{\boldsymbol{s}}} = 1$. We can conclude

$$\frac{\partial U[\boldsymbol{p}]}{\partial p_{\boldsymbol{s}}} = -\log p_{\boldsymbol{s}} - 1 + \lambda_0 + \sum_{i=1}^{n} \alpha_i s_i(\boldsymbol{s}) + \sum_{pair(i,j)} \eta_{ij} s_i(\boldsymbol{s}) s_j(\boldsymbol{s}).$$

Q3.4

We find the most general expression of $p_{\boldsymbol{s}}$ with maximal entropy that satisfies the constraints in Eq. (11) by setting the derivative to zero.

$$\begin{aligned}
0 &= \frac{\partial U[\boldsymbol{p}]}{\partial p_{\boldsymbol{s}}} \\
0 &= -\log p_{\boldsymbol{s}} - 1 + \lambda_0 + \sum_{i=1}^{n} \alpha_i s_i(\boldsymbol{s}) + \sum_{pair(i,j)} \eta_{ij} s_i(\boldsymbol{s}) s_j(\boldsymbol{s}) \\
\log p_{\boldsymbol{s}} &= -1 + \lambda_0 + \sum_{i=1}^{n} \alpha_i s_i(\boldsymbol{s}) + \sum_{pair(i,j)} \eta_{ij} s_i(\boldsymbol{s}) s_j(\boldsymbol{s}) \\
p_{\boldsymbol{s}} &= \exp\left(-1 + \lambda_0 + \sum_{i=1}^{n} \alpha_i s_i(\boldsymbol{s}) + \sum_{pair(i,j)} \eta_{ij} s_i(\boldsymbol{s}) s_j(\boldsymbol{s})\right) \\
p_{\boldsymbol{s}} &= \exp(-1 + \lambda_0) \exp\left(\sum_{i=1}^{n} \alpha_i s_i(\boldsymbol{s}) + \sum_{pair(i,j)} \eta_{ij} s_i(\boldsymbol{s}) s_j(\boldsymbol{s})\right)
\end{aligned}$$

This expression for $p_{\boldsymbol{s}}$ is Eq. (2), where

$$\lambda_0 = \log\left(\frac{1}{Z}\right) + 1 \quad \text{and} \quad \alpha_i = h_i \quad \text{and} \quad \eta_{ij} = J_{ij}.$$

2.4 Statistical inference: model with no couplings

We now consider the model with no couplings (all the $J_{ij} = 0$)

$$p_{\boldsymbol{g}}(\boldsymbol{s}) = \frac{1}{Z(\boldsymbol{g})} \exp\left(\sum_{i=1}^{n} h_i s_i\right). \tag{13}$$

The vector $\boldsymbol{g} = (h_1, \cdots, h_n)$ now only contains n local field parameters.

Q4.1

The probability distribution $p_{h_i}(s_i)$ for the spin variable s_i is given by

$$p_{h_i}(s_i) = \frac{1}{Z(h_i)} \exp(h_i s_i) = \frac{\exp(h_i s_i)}{\sum_{\boldsymbol{s}} \exp(-E(s_i))} = \frac{\exp(h_i s_i)}{\sum_{\boldsymbol{s}} \exp(h_i s_i)}.$$

We can now write the joint probability distribution as a product of a probability distribution over each spin variable

$$\begin{aligned}
p_{\boldsymbol{g}}(\boldsymbol{s}) &= \frac{1}{Z(\boldsymbol{g})} \exp\left(\sum_{i=1}^{n} h_i s_i\right) \\
&= \frac{1}{\sum_{\boldsymbol{s}} \exp(\sum_{i=1}^{n} h_i s_i)} \exp\left(\sum_{i=1}^{n} h_i s_i\right) \\
&= \frac{1}{\sum_{\boldsymbol{s}} \prod_{i=1}^{n} \exp(h_i s_i)} \prod_{i=1}^{n} \exp(h_i s_i) \\
&= \frac{1}{\prod_{i=1}^{n} \sum_{\boldsymbol{s}} \exp(h_i s_i)} \prod_{i=1}^{n} \exp(h_i s_i) \\
&= \prod_{i=1}^{n} \frac{\exp(h_i s_i)}{\sum_{\boldsymbol{s}} \exp(h_i s_i)} = \prod_{i=1}^{n} p_{h_i}(s_i),
\end{aligned}$$

showing that the variables are independent from each other.

Q4.2

Suppose h_i satisfies the constraint $\langle s_i \rangle_D = \langle s_i \rangle$, then

$$\begin{aligned}
\langle s_i \rangle_D &= \langle s_i \rangle \\
&= \sum_{\boldsymbol{s}} p(\boldsymbol{s}) s_i \\
&= \sum_{\boldsymbol{s}} \left(\prod_{i=1}^{n} p_{h_i}(s_i) \right) s_i \\
&= \sum_{\boldsymbol{s}} \left(\prod_{i=1}^{n} \frac{\exp(h_i s_i)}{\sum_{\boldsymbol{s}} \exp(h_i s_i)} \right) s_i \\
&= \frac{e^{h_i} - e^{-h_i}}{e^{h_i} + e^{-h_i}} \\
&= \frac{2 \sinh h_i}{2 \cosh h_i} \\
&= \tanh h_i
\end{aligned}$$

and thus

$$h_i = \tanh^{-1}(\langle s_i \rangle_D). \tag{14}$$

Q4.3

In Eq. (14), we observe that:

- if $\langle s_i \rangle_D > 0$, then the inferred h_i is also positive. If $\langle s_i \rangle_D > 0$, it means that the empirical average of s_i over the dataset is positive, indicating that the judge i tends to vote more conservatively on average. In this case, the inferred h_i is also positive, which aligns with the interpretation from Question Q1.4 that the judge's tendency is towards conservative voting.
- reciprocally, if $\langle s_i \rangle_D < 0$, then the inferred h_i is also negative. If $\langle s_i \rangle_D > 0$, it means that the empirical average of s_i over the dataset is negative, indicating that the judge i tends to vote more liberal on average. In this case, the inferred h_i is also negative, which aligns with the interpretation from Question Q1.4 that the judge's tendency is towards voting in favor of liberal-oriented decisions.

2.5 Statistical inference: maximizing the log-likelihood function

We introduce the *log-likelihood function*:

$$\mathcal{L}(\boldsymbol{g}) = \log P_{\boldsymbol{g}}(\hat{\boldsymbol{s}}), \tag{15}$$

where $P_{\boldsymbol{g}}(\hat{\boldsymbol{s}})$ is the probability that the model $p_{\boldsymbol{g}}(\boldsymbol{s})$ produces the dataset $\hat{\boldsymbol{s}} = (\boldsymbol{s}^1, \cdots, \boldsymbol{s}^N)$. Note that $\mathcal{L}(\boldsymbol{g})$ is a function of the parameters $\boldsymbol{g}$. The inference procedure therefore consists in finding the value $\boldsymbol{g}$ of the parameters that maximizes $\mathcal{L}(\boldsymbol{g})$.

Q5.1

We assume that, in the dataset $\hat{\boldsymbol{s}}$, all the datapoints are independently sampled from an underlying distribution $p_{\boldsymbol{g}}(\boldsymbol{s})$. Then

$$\begin{aligned}
\mathcal{L}(\boldsymbol{g}) &= \log P_{\boldsymbol{g}}(\hat{\boldsymbol{s}}) \\
&= \log\left(\prod_{\boldsymbol{s}} p_{\boldsymbol{g}}(\boldsymbol{s})^{K(\boldsymbol{s})}\right) \\
&= \sum_{\boldsymbol{s}} \log\left(p_{\boldsymbol{g}}(\boldsymbol{s})^{K(\boldsymbol{s})}\right) \\
&= \sum_{\boldsymbol{s}} K(\boldsymbol{s}) \log p_{\boldsymbol{g}}(\boldsymbol{s}) \\
&= N \sum_{\boldsymbol{s}} \frac{K(\boldsymbol{s})}{N} \log p_{\boldsymbol{g}}(\boldsymbol{s}) \\
&= N \sum_{\boldsymbol{s}} p_D(\boldsymbol{s}) \log p_{\boldsymbol{g}}(\boldsymbol{s}),
\end{aligned}$$

where $p_D(\boldsymbol{s}) = \frac{K(\boldsymbol{s})}{N}$ is the empirical distribution over the states with $K(\boldsymbol{s})$ the number of times the datapoint $\boldsymbol{s}$ occurs in the dataset.

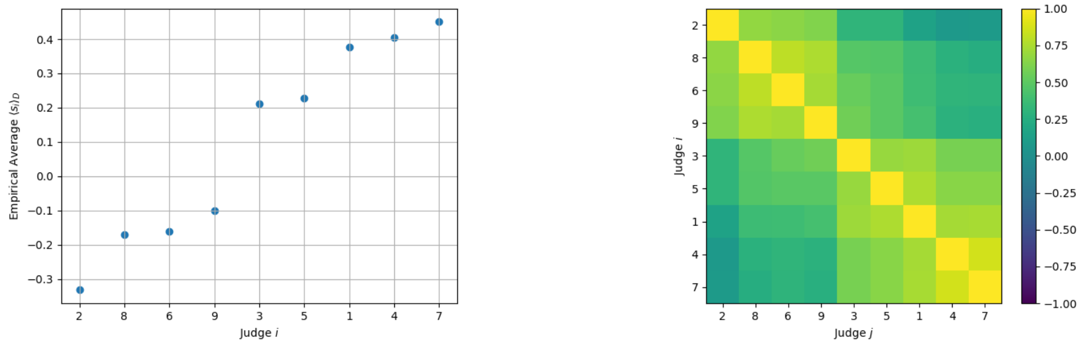
3 Application to the analysis of the US supreme Court

Q6.1

For the US Supreme court dataset there is a number $n = 9$ of spin variables, representing the nine judges. A system with $n = 9$ spins can be in $2^9 = 512$ different states. The dataset contains a total number $N = 895$ of datapoints, but only $N_{max} = 128$ different states are observed.

Q6.3

Figure 4 depicts the behaviour of the empirical averages $\langle s_i \rangle_D$ and $\langle s_i s_j \rangle_D$. In Figure 4a the empirical average $\langle s_i \rangle_D$ is plotted as a function of i with the label i ordered so that the values of $\langle s_i \rangle_D$ are ordered from the smallest (negative value) to the largest (positive value). Keeping the new ordering, Figure 4b shows a heatmap of the empirical average matrix $\langle s_i s_j \rangle_D$.

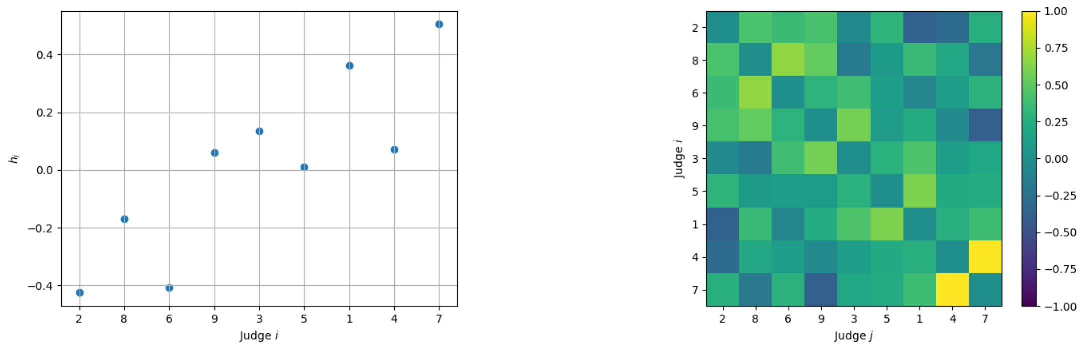


(a) (b)

Figure 4: Empirical average $\langle s_i \rangle_D$ plotted as a function of i (left) and heatmap of empirical average matrix $\langle s_i s_j \rangle_D$ (right).

These figures provide insights into the voting behavior of individual judges and the relationships between their voting tendencies. We have ordered the judges such that judges with the most liberal averages are on the left, while judges with the most conservative averages are on the right. The heatmap shows that judges with similar tendencies (either liberal or conservative) tend to vote similarly.

Q6.4



(a) (b)

Figure 5: Fitted vector of h_i's (left) and heatmap of the fitted matrix of J_{ij}'s (right).

Figure 5 provides us with more insights into the voting behaviour of the nine judges. Figure 5a provides insights into the individual bias or tendency of each judge, while the heatmap in Figure 5b shows the pairwise interactions between the voting tendencies of different judges. Comparing Figure 4 and Figure 5 we notice some interesting behaviour. For example, the coupling parameter $J_{4,7}$ is quite high. This indicates a strong influence between judge 4 and 7, which is further underscored by the disparity between the individual tendency h_4 and their observed empirical average $\langle s_4 \rangle_D$. It seems that judge 4 is markedly impacted by the voting behavior of judge 7.

Q6.5

Figure 6 shows how well the model with the fitted parameters h_i and J_{ij} captures the empirical probabilities of the observed states. Ideally, we would expect the points to lie close to the line $y = x$, indicating perfect agreement between the model predictions and the empirical observations. Deviations from this line suggest discrepancies between the model and the data.

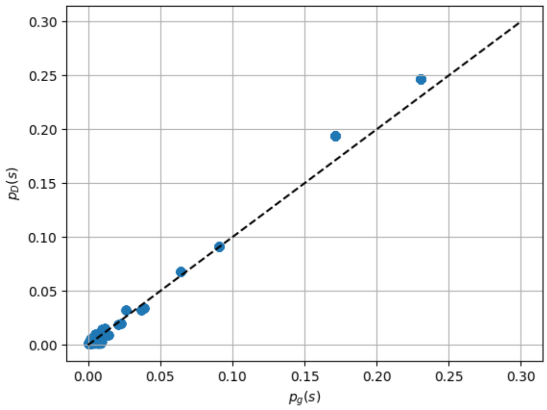
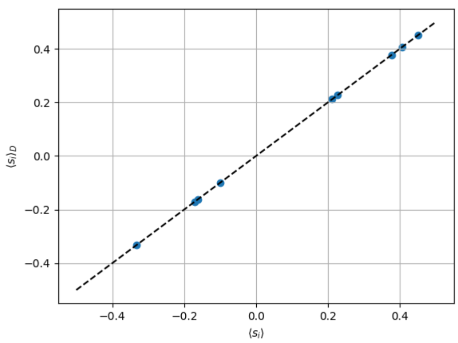
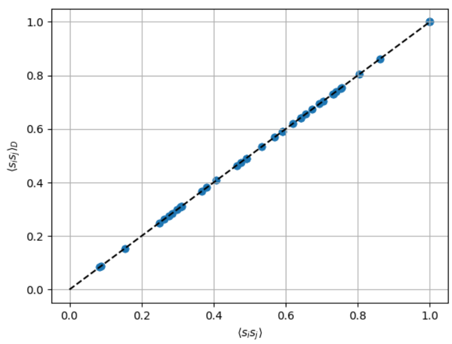


Figure 6: Empirical probability of the state, $p_D(\boldsymbol{s})$ plotted against the model probability $p_{\boldsymbol{g}}(\boldsymbol{s})$ with the fitted parameters (h_i and J_{ij})

Q6.6



(a)



(b)

Figure 7: Empirical averages $\langle s_i \rangle_D$ and $\langle s_i s_j \rangle_D$ from the data against averages $\langle s_i \rangle$ and $\langle s_i s_j \rangle$ from the fitted model.

From the scatter plots in Figure 7, we can assess how well the fitted model captures both the individual tendencies of each spin and the interactions between pairs of spins. Ideally, the points should cluster around the line $y = x$, indicating good agreement between the data and the fitted model.

Q6.7

The probability that k judges vote conservative is given by

$$P_I(k) = \sum_S \prod_{i \in S} p_i(+1) \prod_{j \notin S} (1 - p_j(+1)), \tag{16}$$

where we sum over all possible selections S of k judges and $p_i(+1)$ is the probability that judge i votes conservative. In the dataset, 5 judges have votes that are more conservative on average (see Figure 4a), therefore we expect the highest value of $P_I(k)$ to be at $k = 5$. We use $\langle s_i \rangle_D$ as an approximation for $p_i(+1)$ and plot the values of $P_I(k)$ as a function of k, see Figure 8.

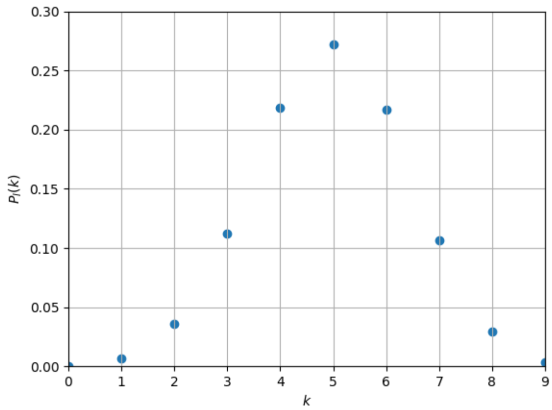


Figure 8: Probabilities $P_I(k)$ plotted as a function of k.

Q6.8

To compute the values of $P_D(k)$ from the data for different values of k, we count the number of occurrences of k conservative votes in the dataset and divide it by N.

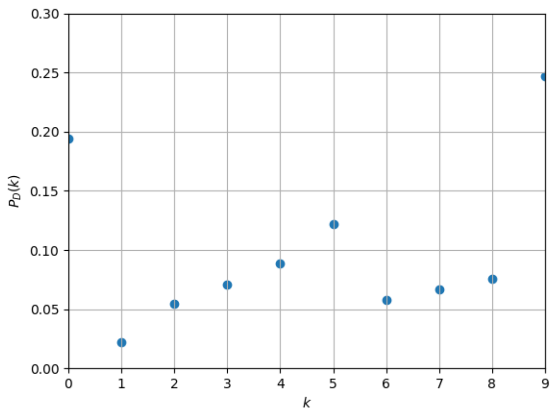


Figure 9: Probabilities $P_D(k)$ plotted as a function of k.

In Figure 9 the computed probabilities $P_D(k)$ are plotted as a function of k. Comparing these probabilities with the probabilities $P_I(k)$ from Figure 8, we note some important differences. Where the maximum for $P_I(k)$ is at $k = 5$, the highest values for $P_D(k)$ are actually at $k = 0$ and $k = 9$. This discrepancy suggests that the independent model, where judges vote independently of each other, does not accurately capture the voting behavior observed in the data. The high probabilities $P_D(0)$ and $P_D(9)$ imply that a substantial number of data points contain either only liberal or only conservative votes, indicating some degree of unanimity or polarization in the voting patterns.

Q6.9

To compute the values of $P_P(k)$ we sum the probabilities $p_{\boldsymbol{g}}(\boldsymbol{s})$ of all possible states $\boldsymbol{s}$ that have exactly k conservative votes

$$P_P(k) = \sum_{\substack{\boldsymbol{s} \\ \sum s_i = k}} p_{\boldsymbol{g}}(\boldsymbol{s}). \tag{17}$$

The resulting probabilities $P_P(k)$ are plotted in Figure 10.

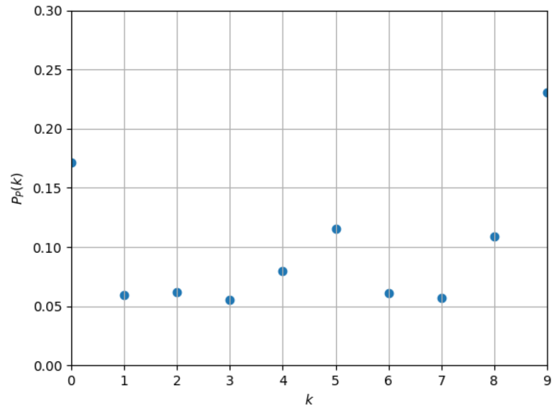


Figure 10: Probabilities $P_P(k)$ plotted as a function of k.

In Figure 11 we compare the three probability curves $P_I(k)$, $P_D(k)$ and $P_P(k)$. We can conclude that the probability distribution $P_P(k)$ obtained from fitting the Ising model with pairwise coupling best represents the probability curve $P_D(k)$ obtained directly from the data.

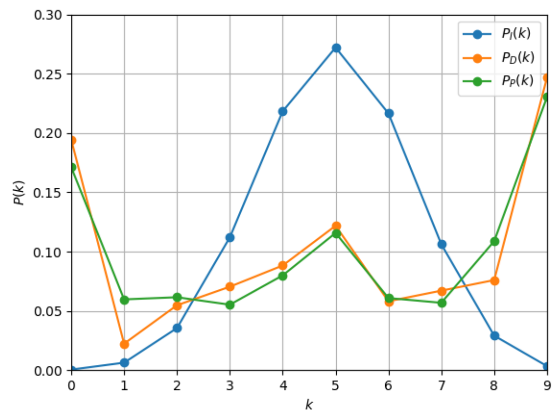


Figure 11: Probabilities $P_I(k)$, $P_D(k)$ and $P_P(k)$ plotted as functions of k.