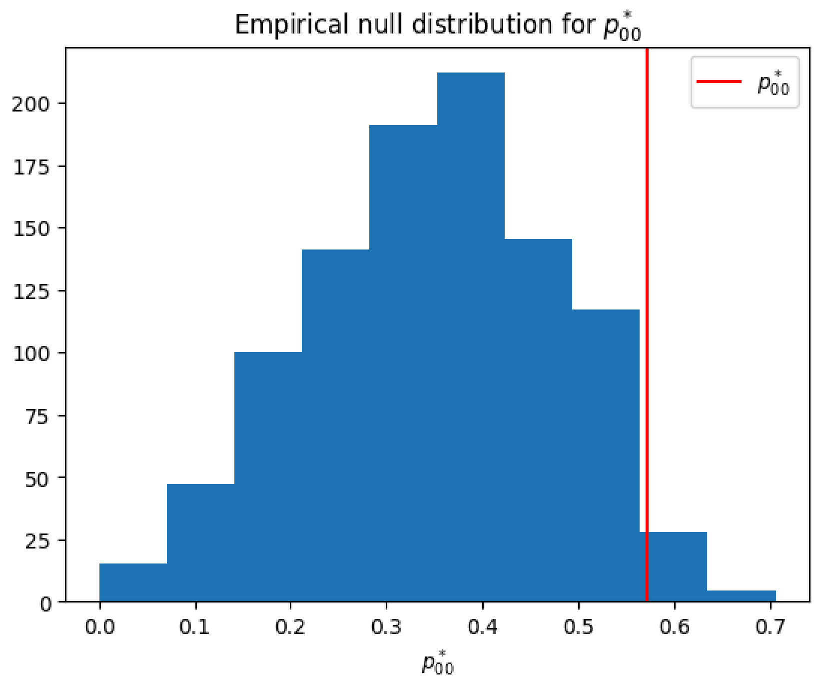
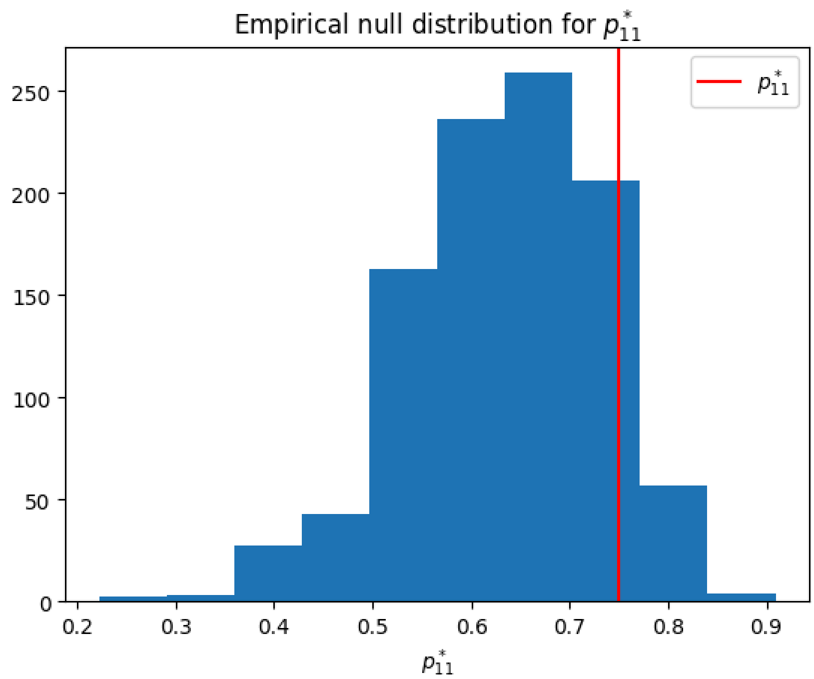
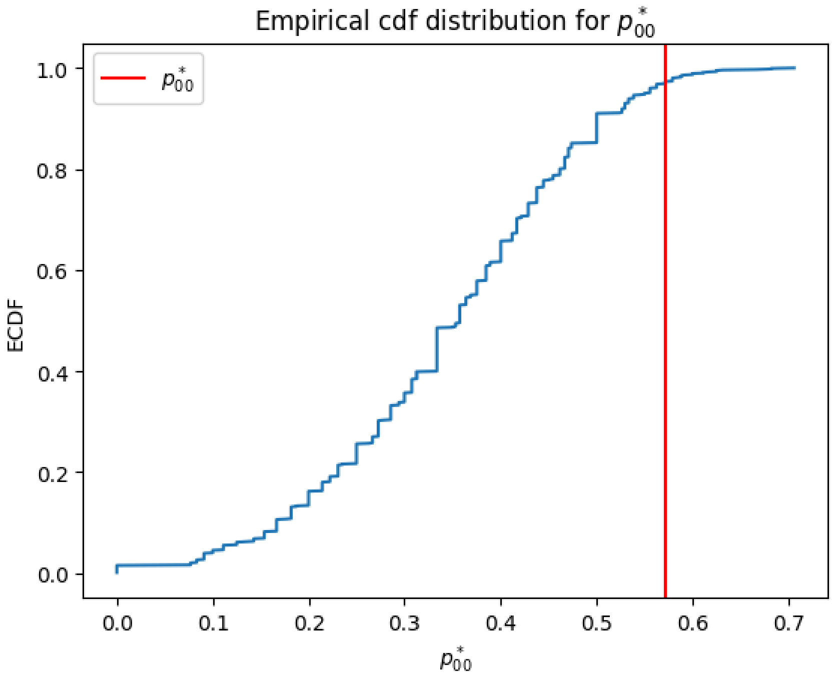
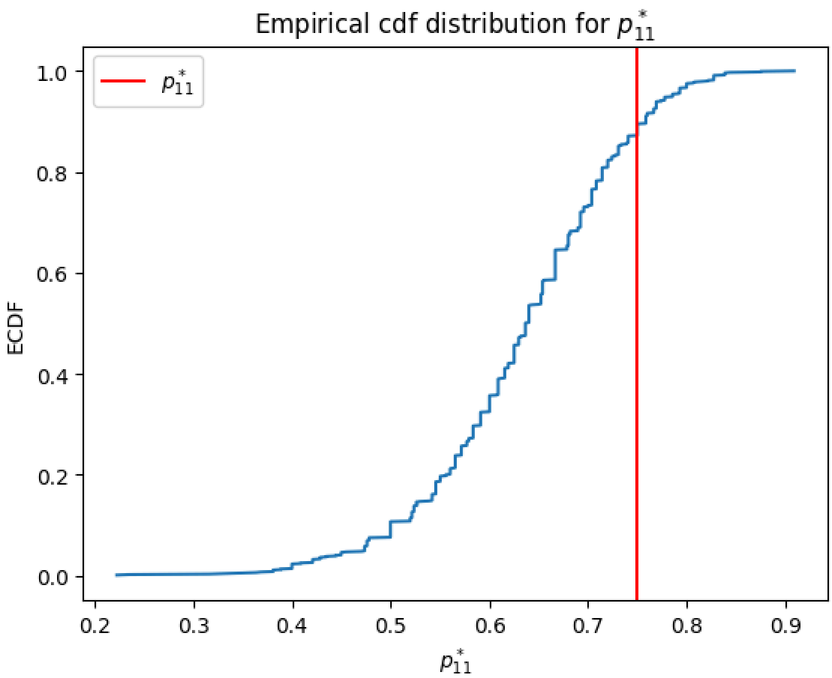
HW1

b)


Empirical null distribution for p^*_{00}
p^*_{00}
0
25
50
75
100
125
150
175
200
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
p^*_{00}


Empirical null distribution for p^*_{11}
p^*_{11}
0
50
100
150
200
250
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
p^*_{11}





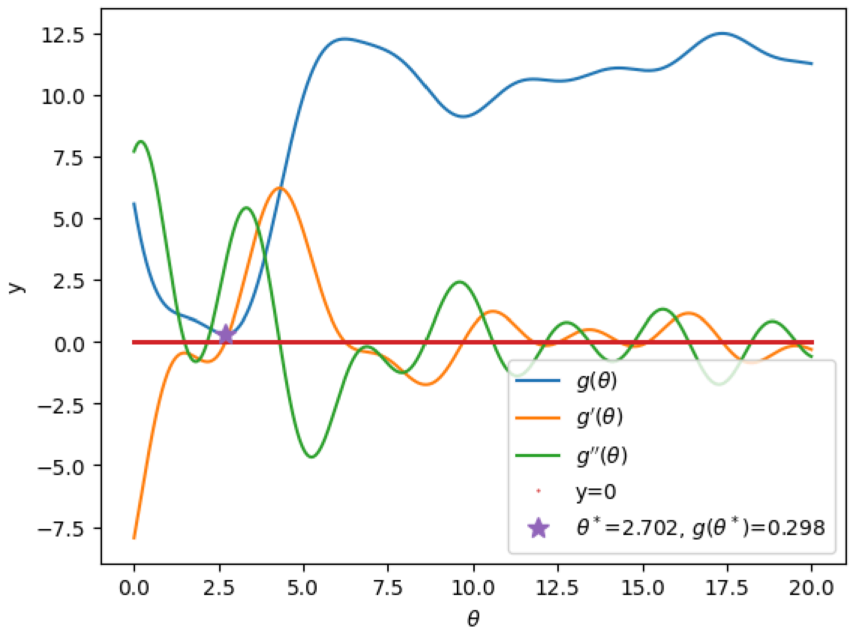

```
For p_00, the empirical p-value of 0.027 indicates that the observed transition probability is unlikely to arise from an i.i.d.
sequence, providing evidence for a first-order Markov model. Assuming an alpha of 0.05 we reject our null of a simple coin flip
model.

For p_11, the empirical p-value of 0.108 indicates that we do not have statistical significance, assuming an alpha of 0.05, to r
eject the null hypotesis that the data was generated via an simple coin flip model.
```

HW2

b)

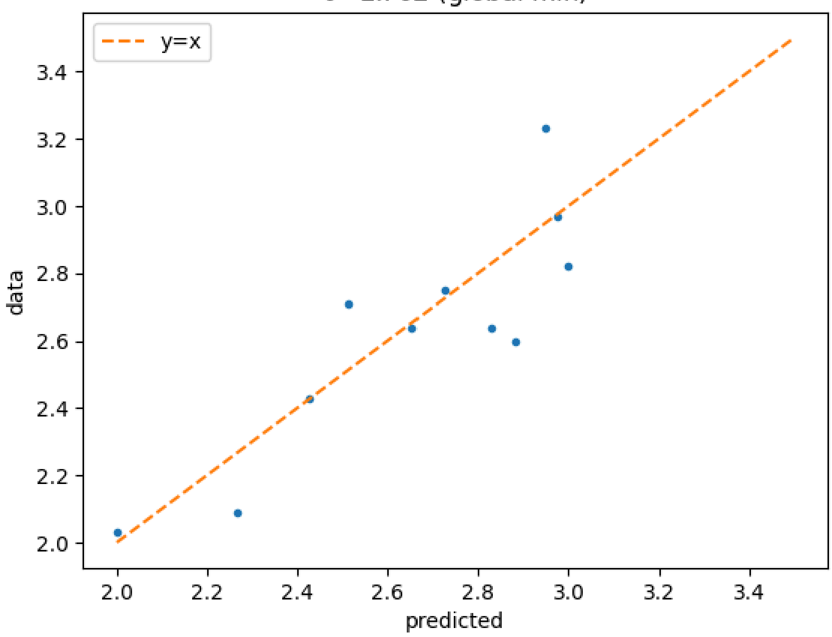
```
All optima are: [ 0.298,  9.104, 10.547, 10.971]
Respectively, the theta values corresponding to all optima are: [ 2.702,  9.712, 12.553, 15.181]
```

```
<matplotlib.legend.Legend at 0x113d109a0>
```

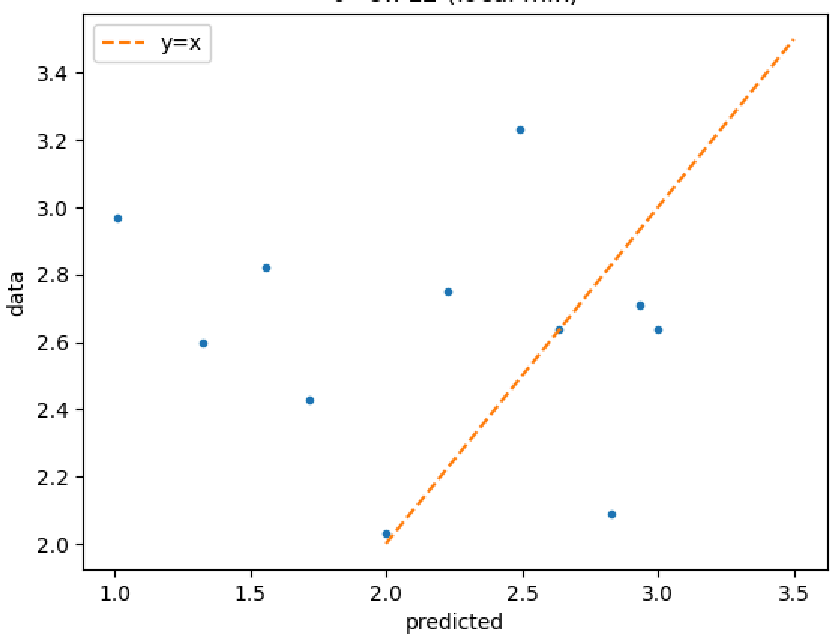


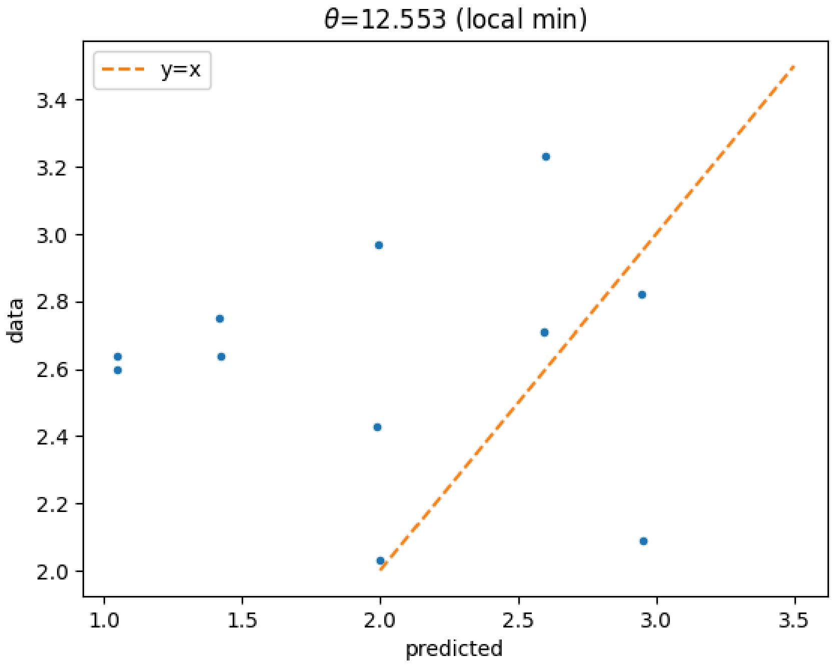
c)

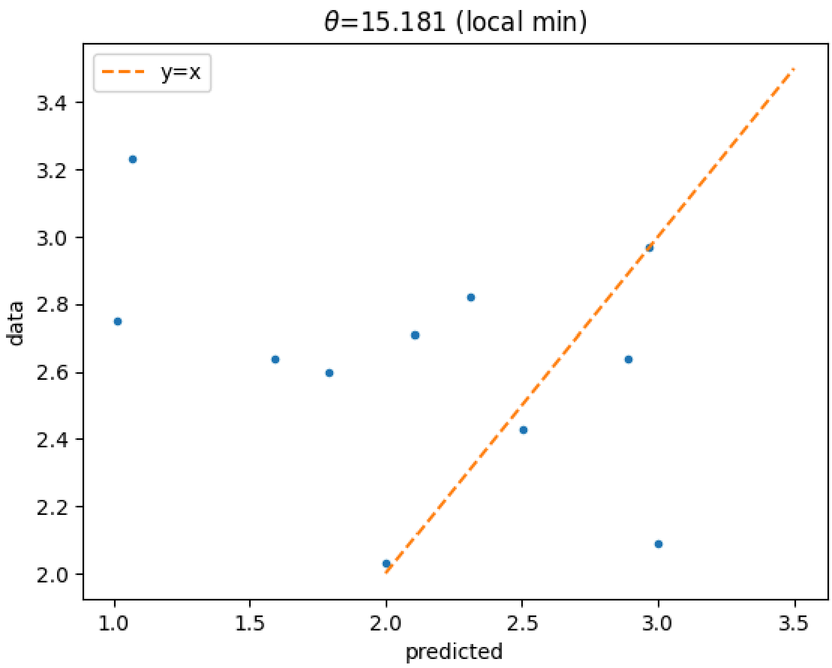
θ=2.702 (global min)



θ=9.712 (local min)








```
Newton method: theta=2.702120, g(theta)=0.298142
Global minimum value parameter for the objective function from the plot: theta=2.702103, g(theta)=0.298142
```

```
'The solution for the objective function theta* associated with the local minimum after running the Newton method with the start
ing point theta_0 equals to 10: theta=9.711984372034813, g(theta)=9.10391889626'
```

As we see, the Newton method executed for the starting point $\theta_0 = 3$ aligns with the values of the parameter θ associated with the global minimum, and finds the global optimum correctly.

Despite this fact, the Newton method does not guarantee to find the global optimum as it depends on the starting point. It might find the local minimums which does not model our data correctly. For example, for starting $\theta_0 = 10$, it finds the local optimum: $g(\theta^*) \approx 9.1$

Additionally, even if the method finds the parameters for the global minimum, potentially the precision error might happen. For instance, when it converges in the vicinity of the global minimum, but does not give the exact solution.

HW3

$y_i = x_{i|\theta} + \eta_i$ and $\eta_i \sim Norm(0, \sigma_i^2)$

$L(y_i|\theta) = \prod_{i=1}^{n} p(y_i|\theta)$

where $p(y_i|\theta) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(\frac{-(y_i - x_{i|\theta})^2}{2\sigma_i^2})$

$L(y_i|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(\frac{-(y_i - x_{i|\theta})^2}{2\sigma_i^2}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1 \cdot \ldots \cdot \sigma_n} \exp(-\sum_{i=1}^{n} \frac{(y_i - x_{i|\theta})^2}{2\sigma_i^2})$

$l(y_i|\theta) = \log L(y_i|\theta) = \underbrace{\frac{n}{2} \log(\frac{1}{2\pi})}_{constant} - \underbrace{\sum_{i=1}^{n} \log(\sigma_i)}_{no\ \theta} - \underbrace{\sum_{i=1}^{n} \frac{(y_i - x_{i|\theta})^2}{2\sigma_i^2}}_{2\ is\ a\ constant}$

As $\theta^* = \arg\min_\theta(-l(y_i|\theta))$

Therefore

As $\theta^* = \arg\min_\theta(\sum_{i=1}^{n} \frac{(y_i - x_{i|\theta})^2}{\sigma_i^2})$

The objective function is $g(\theta) = \sum_{i=1}^{n} \frac{(y_i - x_{i|\theta})^2}{\sigma_i^2}$

HW4

$N = 2$

$t = (1, 3)$

$y = (6, 7)$

$(\theta_1, \theta_2)^\top = (1, 1)$

$x_i = \theta_1 e^{\theta_2 t_i}$

$g(\theta) = \sum_{i=1}^{N} (x_{i|\theta} - y_i)^2 = \sum_{i=1}^{N} (\theta_1 e^{\theta_2 t_i} - y_i)^2$

$\frac{dg(\theta)}{d\theta 1} = 2\sum_{i=1}^{N} (\theta_1 e^{\theta_2 t_i} - y_i) e^{\theta_2 t_i} = 2\sum_{i=1}^{N} (1e^{1t_i} - y_i) e^{t_i} = 2(e - 6)e + 2(e^3 - 7)e^3$

$\frac{dg(\theta)}{d\theta 2} = 2\sum_{i=1}^{N} (\theta_1 e^{\theta_2 t_i} - y_i) e^{\theta_2 t_i} \theta_1 t_i = 2\sum_{i=1}^{N} (e^{t_i} - y_i) e^{t_i} t_i = 2e(e - 6) + 6e^3(e^3 - 7)$

$\frac{d^2 g(\theta)}{d\theta_1^2} = 2\sum_{i=1}^{N} e^{\theta_2 t_i} = 2e^2 + 2e^{2 \cdot 3} = 2(e^2 + e^6)$

$\frac{d^2 g(\theta)}{d\theta_1 d\theta_2} = 2\sum_{i=1}^{N} (\theta_1 e^{2\theta_2 t_i} 2t_i - y_i e^{\theta_2 t_i} t_i) = 2\sum_{i=1}^{N} (1e^{2t_i} 2t_i - y_i e^{1t_i} t_i) = 2(2e^2 - 6e) + 2(6e^6 - 7 \cdot 3e^3) = 2(2e^2 - 6e) + 2(6e^6 - 21e^3)$

$\frac{d^2 g(\theta)}{d\theta_2^2} = 2\sum_{i=1}^{N} (\theta_1^2 e^{2\theta_2 t_i} 2t_i^2 - y_i \theta_1 e^{\theta_2 t_i} t_i^2) = 2(1e^2 2 - 6e) + 2(e^{2 \cdot 3} 2 \cdot 9 - 7e^3 9) = 2(2e^2 - 6e + 18e^6 - 63e^3)$

```
d g(theta)/ d theta1: 507.8188003171955
d g(theta)/ d theta2: 1559.1389404388808

d^2 g(theta)/ d theta1^2: 821.6356991833313
d^2 g(theta)/ d theta1 dtheta2: 3994.4898135931526
d^2 g(theta)/ d theta1^2: 11989.595755871029
```

$$H = \begin{bmatrix} \frac{d^2 g(\theta)}{d\theta_1^2} & \frac{d^2 g(\theta)}{d\theta_1 \theta_2} \\ \frac{d^2 g(\theta)}{d\theta_1 \theta_2} & \frac{d^2 g(\theta)}{d\theta_2^2} \end{bmatrix} = \begin{bmatrix} 821.6 & 3994.5 \\ 3994.5 & 11989.6 \end{bmatrix}$$

If we consider H as: $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$

$$H^{-1} = \frac{1}{a \cdot c - b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix} = \frac{1}{821.6 \cdot 11989.6 - 3994.5^2} \begin{bmatrix} 11989.6 & -3994.5 \\ -3994.5 & 821.6 \end{bmatrix} = \frac{1}{-6105374.8} \begin{bmatrix} 11989.6 & -3994.5 \\ -3994.5 & 821.6 \end{bmatrix} = \begin{bmatrix} -0.001964 & 0.00065426 \\ 0.00065426 & -0.0001346 \end{bmatrix}$$

$$\theta^{s+1} = \theta^s + \Delta\theta = \theta^s - H^{-1} \cdot \nabla g(\theta^s) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -0.001964 & 0.00065426 \\ 0.00065426 & -0.0001346 \end{bmatrix} \begin{bmatrix} 507.82 \\ 1559.14 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.0228 \\ 0.1224 \end{bmatrix} = \begin{bmatrix} 0.9772 \\ 0.8776 \end{bmatrix}$$