Archit Murkunde – 2845576
email: a.u.murkunde@student.vu.nl

## Question 1:

**Question 1**: Using the training data, make a table with the descriptive statistics (number of observations, means, medians, standard deviations, skewness, *excess* kurtosis, minimum, maximum) with a column for each of the four patients. Provide this table in the report.

Answer:

Table 1: Descriptive Statistics

| | No of Obs | Mean | Median | Std Dev | Skewness | Excess Kurtosis | Min | Max |
|---|---|---|---|---|---|---|---|---|
| **Patient 1** | 2000 | 69.211311 | 69.026500 | 6.157390 | 0.073437 | -0.252180 | 51.041275 | 87.324332 |
| **Patient 2** | 2000 | 74.506205 | 74.592051 | 5.782674 | 0.018712 | -0.069098 | 56.678226 | 93.344119 |
| **Patient 3** | 2000 | 86.092599 | 85.948986 | 5.225615 | 0.049079 | -0.000813 | 68.504259 | 102.644232 |
| **Patient 4** | 2000 | 78.620294 | 78.552662 | 5.156608 | -0.042983 | -0.109199 | 61.628978 | 94.387767 |

Note: The data shown is based on heart rate measurements for four different patients.

## Question 2:

Let $x_{i,t}$ denote the average heart rate of patient $i$ observed at the 10-minute interval at time $t$. Consider the following statistical model for this heart rate:

$$x_{i,t} = \mu_{i,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim NID(0, \sigma_i^2),$$
$$\mu_{i,t+1} = \omega_i + \alpha_i x_{i,t} + \beta_i \mu_{i,t}$$

Here, $\mu_{i,t}$ is to be interpreted as the *slowly varying* true average heart and $\epsilon_{i,t}$ as a collection of external stimuli that pushed the heart rate away from this average during the specific interval at time $t$.

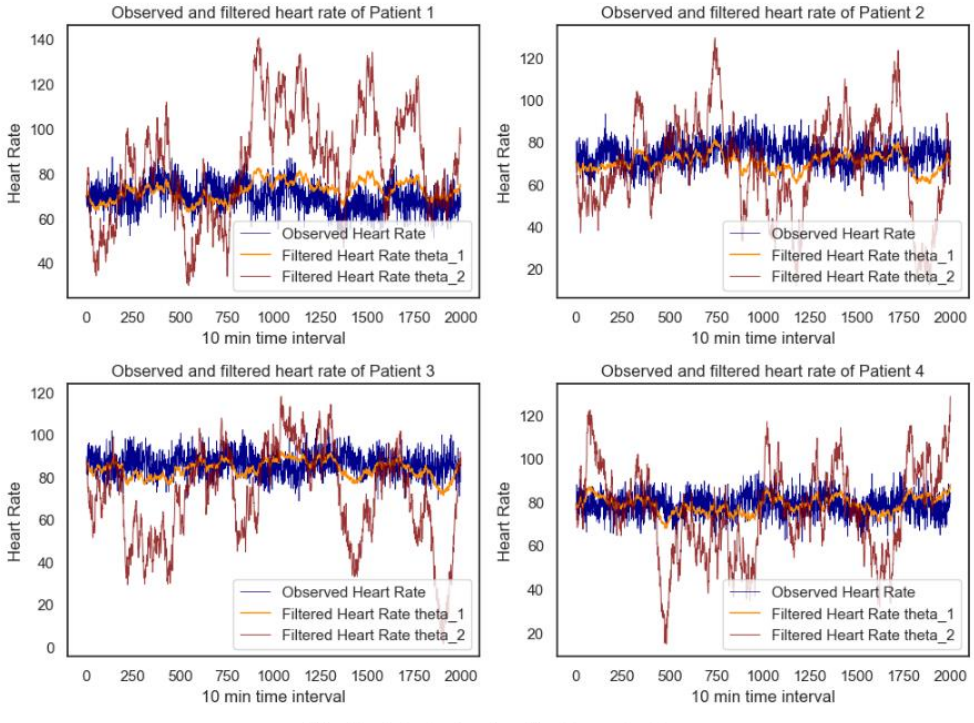
**Question 2**: The parameters $\theta_i = (\omega_i, \alpha_i, \beta_i, \sigma_i)$ are unknown and will need to be estimated from the data. However, to get an idea on how these parameters influence the filtered heart rates, consider the following hypothetical values:

(i) $\theta_i^1 = (0.01 \cdot \bar{x}_i, 0.15, 0.84, 4)$,

(ii) $\theta_i^2 = (0.01 \cdot \bar{x}_i, 0.84, 0.15, 4)$,

for $i = 1, \ldots, 4$, where $\bar{x}_i = \frac{1}{100}\sum_{t=1}^{100} x_{i,t}$. Setting $\mu_{i,1} = \bar{x}_i$, compute using the training data for each hypothetical $\theta_i$, the filtered heart rates $\mu_{i,t}$. Create a $2 \times 2$ grid, and plot for each patient the original heart rate data and the filtered heart rates corresponding to $\theta_i^1$ and $\theta_i^2$ in a single subplot. Comment on the differences between these two filters and how these are related to the choice of $\alpha_i$ and $\beta_i$. Provide your opinion on which filter seems more useful for the detection of persistent elevated heart rates.

Answer:




Filtered heart rate signals use two different parameter sets.

Archit Murkunde – 2845576
email: a.u.murkunde@student.vu.nl

Comments & Opinion:

From the four plots above, it is evident that the second parameter set (with higher alpha than beta) is highly sensitive to changes in the heart rates of the patients, responding quickly to even minor fluctuations. This results in the filter closely following the observed heart rate. In contrast, the first parameter set (with higher beta than alpha) shows a smoother response, with less sensitivity to rapid changes, as it retains previous heart rate values more strongly, leading to a more gradual and stable filtering effect.

For the detection of persistent elevated heart rates, the first parameter set is preferable. Its higher beta value ensures that the filter smooths out short-term fluctuations and places more emphasis on the long-term trend, making it less reactive to temporary spikes. The lower alpha value further contributes to the filter's stability, helping to avoid false positives due to transient changes. This combination makes the first parameter set more reliable for detecting sustained increases in heart rate, which is crucial for identifying potential health concerns.

**Question 3:**

**Question 3**: To estimate the parameters $\theta_i$, we are going to implement the method of maximum likelihood. Derive and argue (e.g. based on the lecture slides of week 1) that the log-likelihood is given by

$$L(\theta_i \,|\, x_{i,1},\ldots,x_{i,T}) = \sum_{t=1}^{T} \left\{ -\log\sqrt{2\pi\sigma_i^2} - \frac{(x_t-\mu_t)^2}{2\sigma_i^2} \right\}.$$

Answer:

Given a set of observations $x_1, \ldots, x_T$ from the time series, we can write the likelihood function $L(\theta|x_1, \ldots, x_T)$ as the product of conditional and marginal densities:

$$f(x_1, x_2, \ldots, x_T; \theta) = f(x_1; \theta) \prod_{t=2}^{T} f(x_t|x_{t-1}; \theta)$$

For the AR(1) model, the conditional density of $x_t$ given $x_{t-1}$ is a Gaussian distribution:

$$x_t|x_{t-1} \sim \mathcal{N}(\phi x_{t-1}, \sigma_\epsilon^2)$$

The density function is given by:

$$f(x_t|x_{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{(x_t - \phi x_{t-1})^2}{2\sigma_\epsilon^2}\right)$$

Taking the logarithm of the likelihood function, we get the log-likelihood function:

$$\log L(\theta|x_1, \ldots, x_T) = \sum_{t=2}^{T} \log f(x_t|x_{t-1}; \theta)$$

Substituting the expression for the conditional density:

$$\log L(\theta|x_1, \ldots, x_T) = \sum_{t=2}^{T} \left(-\frac{1}{2}\log(2\pi\sigma_\epsilon^2) - \frac{(x_t - \phi x_{t-1})^2}{2\sigma_\epsilon^2}\right)$$

Now, for the second form of the log-likelihood function, where $x_{t-1}$ is replaced by $\mu_t$, we have:
The log-likelihood function is:

$$L(\theta_i|x_{i,1}, \ldots, x_{i,T}) = \sum_{t=1}^{T} \left(-\log\sqrt{2\pi\sigma_i^2} - \frac{(x_t - \mu_t)^2}{2\sigma_i^2}\right)$$

Archit Murkunde – 2845576
email: a.u.murkunde@student.vu.nl

## Question 4:

**Question 4**: Again setting $\mu_{i,1} = \bar{x}_i$, code up the log-likelihood function. In the report, provide the value you obtain for the log-likelihood for patient 1 using the training data, evaluated at $\tilde{\theta} = (0.7, 0.15, 0.84, 4)$.

Answer:

The log-likelihood value for patient 1 using the given parameter set (0.7, 0.15, 0.84, 4) is **−6173.7754**

## Question 5:

**Question 5**: Compute the maximum likelihood estimator $\hat{\theta}_i$ of $\theta_i$ for $i = 1, \ldots, 4$ based on the training dataset. Report your estimates in a nice table, were each column corresponds to a patient and each row corresponds to a parameter. Add one extra row in which you report the optimal value of the log-likelihood.

Answer:

Table 2: Maximum Likelihood Estimators

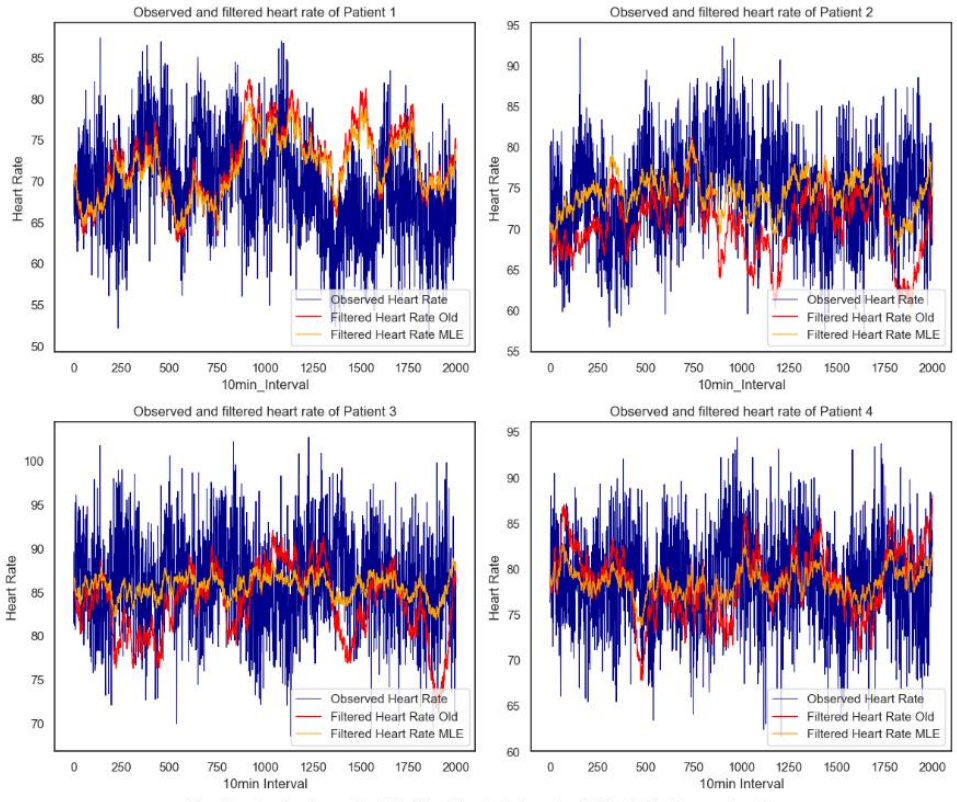
| | Patient 1 | Patient 2 | Patient 3 | Patient 4 |
|---|---|---|---|---|
| **Omega** | 0.5712 | 1.4378 | 2.9508 | 1.7407 |
| **Alpha** | 0.0923 | 0.0904 | 0.0672 | 0.0698 |
| **Beta** | 0.8995 | 0.8903 | 0.8986 | 0.9081 |
| **Sigma** | 4.9677 | 4.9827 | 4.9744 | 4.8922 |
| **Optimal Log-Likelihood** | -6040.5976 | -6056.4166 | -6055.7077 | -6010.1040 |

Note: The table presents the estimated parameters (Omega, Alpha, Beta, and Sigma) and the optimal log-likelihood values for four patients.

## Question 6:

**Question 6**: For each patient, use the training data to create a plot with the observed heart rates $x_{i,t}$, the heart rates filtered based on $\theta_i^1$ as in Question 2(i) and the heart rates filtered based on the MLE $\hat{\theta}_i$ from Question 5. Store these plots in a $(2 \times 2)$ grid and copy this to your report.

Answer:



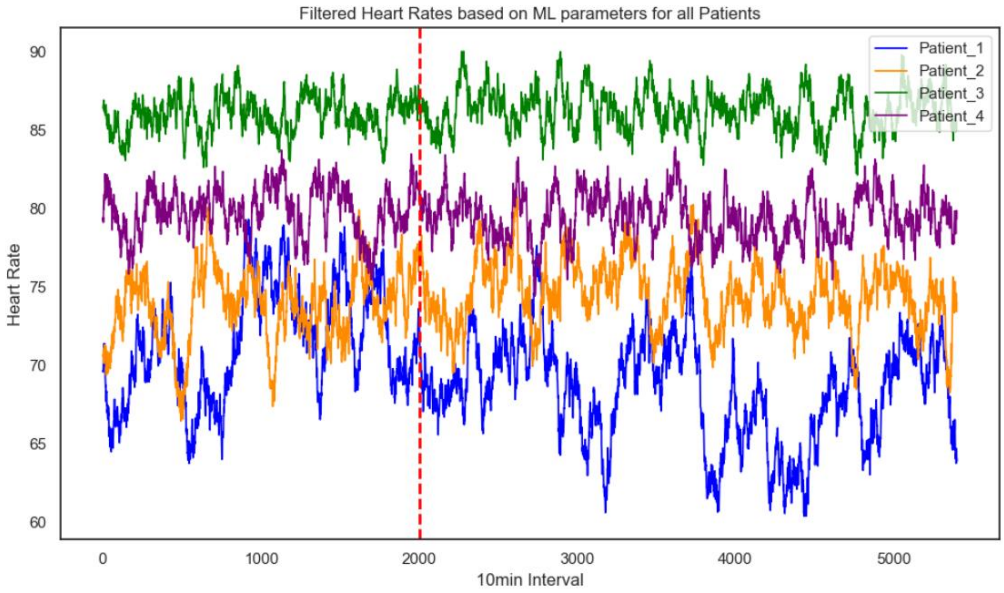

The plots show the observed and the filtered heart rate based on MLE and default parameter set.

Archit Murkunde – 2845576
email: a.u.murkunde@student.vu.nl

**Question 7:**

**Question 7**: Finally, let us consider now the full sample of data ranging from $t = 1, \ldots, 5,400$. For all four patients, compute the filtered heart rates based on the maximum likelihood estimators from Question 5. Plot the four filtered sequences of heart rates in a single plot. Put a vertical line in your plot at $t = 2,000$, the end of your estimation sample. Copy the plot to your report and comment on whether a certain patient may require a follow-up.

Answer:




Observation:

From the above plot, Patients 1 and 2 show noticeable fluctuations in their heart rates over time, with significant dips and spikes, particularly after t = 2000. These variations suggest some instability in their heart rates, which could raise concerns for cardiovascular health. As a result, these two patients should be given priority for follow-up to ensure that no serious health issues are developing.

On the other hand, Patients 3 and 4 have consistently elevated heart rates, but they remain stable without major fluctuations. While their heart rates are higher, the steady pattern makes them less of an immediate concern compared to Patients 1 and 2. Therefore, follow-up should primarily focus on Patients 1 and 2 to check for any potential health risks.