Theoretical Neuroscience II
Teacher: Prof. Jochen Braun
Assistant Teacher: Ehsan Kakaei
Name: Viviana Villafañe Barraza, Tasnuba Hossain Tazeen, Lee Po Shing
Date: 14th June, 2018

Exercise 7: Expectation Maximization

Introduction

When we look around the world, our eyes become filled with observations. These observations result from various causes, with each cause having different distributions of probability. In order to explain these observations, a *causal model* is useful. Causal models take into account true causes from the external world, and model causes generated by the internal causes, and forms an observation such as an image based on these causes. For developing a causal model, sufficient observations are required to reveal their statistical structure. The quality of a causal model is judged by their ability to reproduce this structure. Therefore a step for optimization is always necessary.

Causal models can be formulated using both *generative formulation* to describe a sensory scene in terms of causes) and *recognition formulation* (obtained by inverting generative formulation using Bayes' rule). Using recognition formulation, a recognition model can be generated in order to compute the most likely cause given a particular observation. From an ensemble of observations, and a generative model (obtained from generative formulation), we can assign expected probabilities of alternative causes and obtain an ensemble of assigned causes. This step is known as the *expectation step*. From the ensemble of assigned causes, we re-compute the statistics of causes in order to obtain an improved generative model. This step is known as the *maximization step.* The expectation and maximization steps help to generate a model with maximized expectations.

In this assignment, a sensory world with different observations was considered and a causal model was constructed and improved using expectation maximization in explaining these observations.

Method

In this exercise, a model was built aimed at explaining the distribution of a set of observations. A Matlab file called *observations.mat* was provided containing the values of 10000 observations (u_{obs}, sample in the code). The model was first constructed using manual work of trial-and-error. It was assumed that the observations were caused by different causes (X) with a normal distribution pattern. Then the model was optimized using the strategy of expectation maximization (details described below). A histogram of the observations was plotted in the range of -400 and 400 with bin width of 20, together with the calculations by the model before and after the optimisation steps for a visual validation of the model.

Based on the shape of the histogram, the number of causes X, and the corresponding probabilities or the relative area (γ_x, gamma in the code), means (μ_x, mu in the code) and variances (σ^2_x, var in the code) were constructed 'by hand' so that they generated a rough model briefly intimating the histogram. It was firstly hypothesized to have three different causes, which $X = \{A, B, C\}$. Next, μ_x were firstly estimated dependent on the peaks of the histogram; then σ^2_x on the width of the curves centred from μ_x; and lastly γ_x on the relative height on the curves. The γ_x were set in a way that the sum was one.

The next step was to optimize the model with the strategy of expectation maximisation, which involved alternative turns of the expectation step and the maximization step. For finding the expectation, i.e. the recognition probability $P(X|u_{obs})$ (P_X_u in the code), a Matlab function file called *Expect.m* was created after taking into account of the sample along with γ_x, μ_x and σ^2_x. The probabilities of the observations given by different causes $P(u_{obs}|X)$ (P_u_X in the code) were first calculated. Then $P(X|u_{obs})$ determined the probability that a given data point u was caused either by cause A, B or C and it was calculated based on the equation modified from Bayes' rule:

$$P(X|u_{obs}) = \frac{\gamma_X P(u_{obs}|X)}{\gamma_A P(u_{obs}|A) + \gamma_B P(u_{obs}|B) + \gamma_C P(u_{obs}|C)}$$

Once the expectation step was determined, the maximization step was performed to infer with γ_x, μ_x and σ^2_x using $P(X|u_{obs})$ and u_{obs}.The equations used were shown below:

$$\gamma_X = \sum_i^n \frac{P(X|u_{obs})}{n} \qquad \mu_X = \frac{1}{\gamma_X}\sum_i^n \frac{u_{obs}P(X|u_{obs})}{n} \qquad \sigma_X^2 = \frac{1}{\gamma_X}\sum_i^n \frac{(u_{obs}-\mu_X)^2 P(X|u_{obs})}{n}$$

For the maximization step, another Matlab file called *Maximize.m* was created. In the main script, a for-loop was used to perform the alternative processes of expectation step and maximization step. Since a model can never perfectly explain the observations, a break was introduced in the for-loop once the sum of the marginal changes of σ^2_x reached below 0.003. As mentioned, the graphic representation was plotted before and after the optimization, and the function plottingobservation inside the main script severed this purpose.

Results & Discussion

A model consisted of three causes could briefly account for the observations

Given a set of observations, a model which could explain the pattern of the observations was constructed. The first step was done "by hand" to create a rough model briefly mimicking the pattern of observations. The set of observations given was thought to be due to three different causes, namely *A*, *B* and *C*. Assuming these causes have normal distributions around, the next step was determining the parameters γ_x, μ_x and σ^2_x manually so that the of these distributions would adjust briefly to the pattern of the histogram obtained from the said observations. The following numbers for the parameters were determined to describe each normal distribution of causes:

$$\gamma_A = 0.2\ ;\ \mu_A = -110\ ;\ \sigma_A = 2500\ \ (red\ in\ Fig. 1.)$$

$$\gamma_B = 0.6\ ;\ \mu_B = 100\ ;\ \sigma_B = 900\ (yellow\ in\ Fig. 1.)$$

$$\gamma_C = 0.2\ ;\ \mu_C = -20\ ;\ \sigma_C = 2500\ (purple\ in\ Fig. 1.)$$

As seen in Fig.1, these parameters provided a roughly suitable model to explain the observed points. Since this was done by trial-and-error, there was little discrepancy like the distribution of predicted observation at the values around 120. As one may see, there was a slightly higher prediction then the actual number of observations with such values. These left room of improvement of the model and it was then done using expectation maximization as shown in the next session.

Theoretical Neuroscience II
Teacher: Prof. Jochen Braun
Assistant Teacher: Ehsan Kakaei
Name: Viviana Villafañe Barraza, Tasnuba Hossain Tazeen, Lee Po Shing
Date: 14th June, 2018

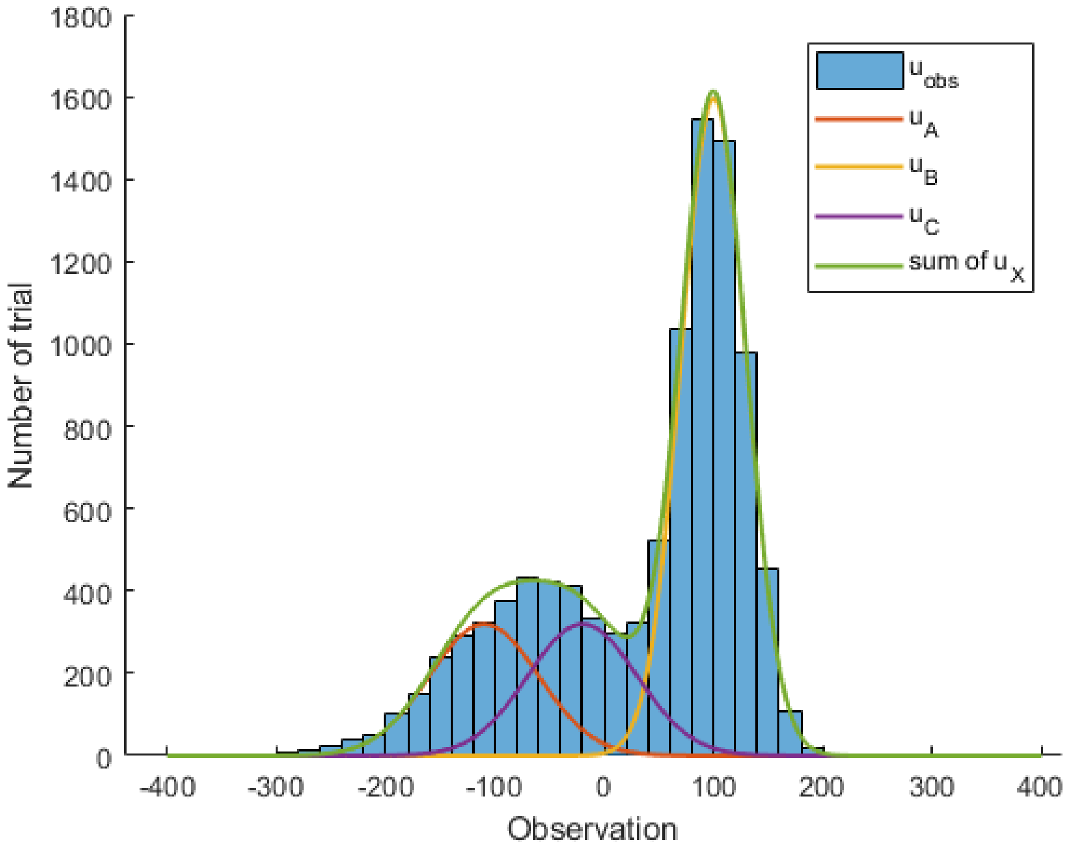

Fig. 1 Histogram of the set of observations and the predicted observations given by the model built by trial-and error. The three possible causes was hypothesized to be the origins of the observations (purple, red and yellow lines) with normal distributions. With trial-and-error, a model was constructed so that the sum of all three causes (green line) explained the observations preliminarily well.

Expectation maximisation improved the model in explaining the observations

With iteratively use of an expectation step and a maximization one, the parameters were adjusted each time to fit the observations better and better. The resulting distributions obtained after this iteration were plotted against the histogram of the observed data and can be seen in Fig. 2. After iteratively doing both expectation and maximization steps, we obtained a better description of the causes which could account for the observations obtained.

Theoretical Neuroscience II
Teacher: Prof. Jochen Braun
Assistant Teacher: Ehsan Kakaei
Name: Viviana Villafañe Barraza, Tasnuba Hossain Tazeen, Lee Po Shing
Date: 14th June, 2018

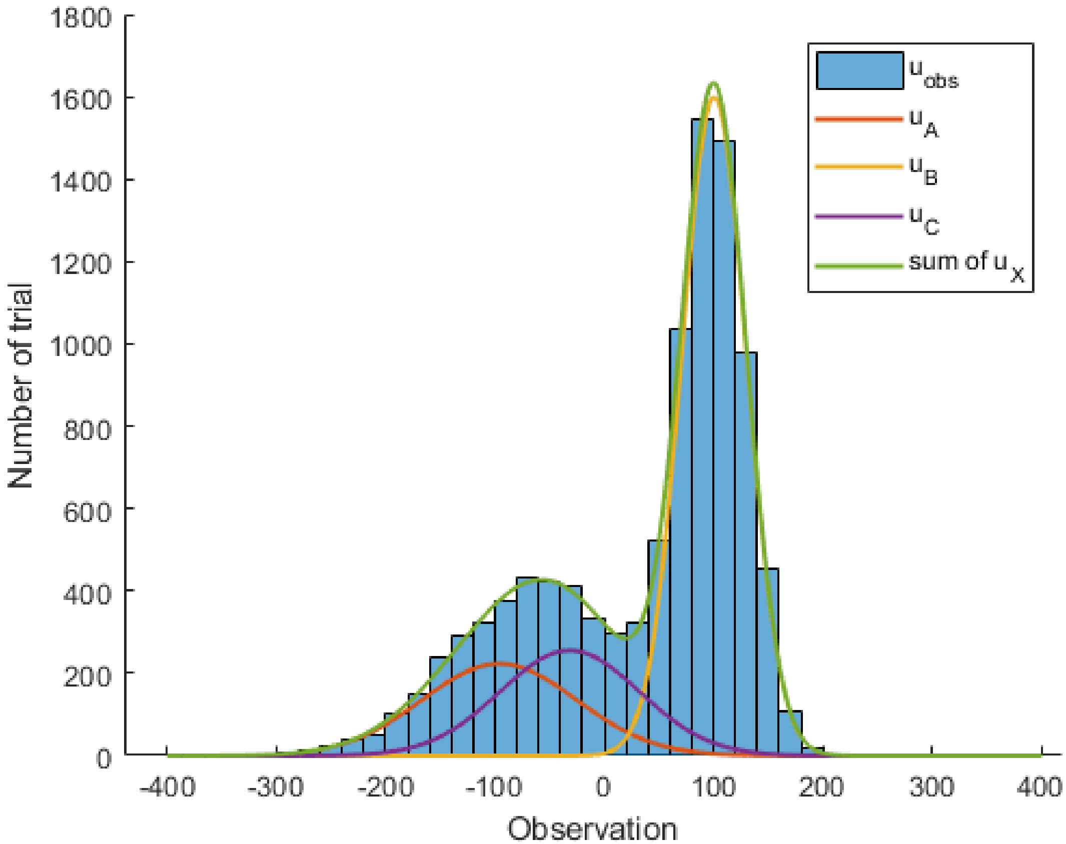

Fig. 2 Histogram of the set of observations and the predicted observations given by the model improved by expectation maximization. The three possible causes (purple, red and yellow lines) with parameters improved by expectation maximization could better explain the observations with their sum fitting the observations much closely.

The parameters obtained after expectation maximization were as followed. The differences with the parameters calculated by hand and after optimization were not very high for γ_x and μ_x. The parameter with the highest changes after applying expectation maximization was σ^2_x.

$$\gamma_A = 0.1967\, ;\ \mu_A = -101.01\, ;\ \sigma_A = 4640.5\ (red\ in\ Fig. 2.)$$

$$\gamma_B = 0.6037\, ;\ \mu_B = 100.24\, ;\ \sigma_B = 900.5\ (yellow\ in\ Fig. 2.)$$

$$\gamma_C = 0.1996\, ;\ \mu_C = -29.85\, ;\ \sigma_C = 3516.1\ (purple\ in\ Fig. 2.)$$