UQ 6310 Homework4, Yu Wang

Assuming that the sampling distributions of $E1, E2, E3, E4, N(Noise)$ follow the normal distributions, where the mean values of the five sampling distributions here are 30000, 25000, 32000, 27000, 0.05, and the values of the standard deviations are 3000, 2700, 3200, 2500, 0.02 respectively. Also the prior distributions for the above 5 variables are assumed to be the normal distributions, where the mean values here are 25000, 20000, 28000, 22000, 0.07, and the values of the standard deviations are 1000, 2000, 3000, 2000, 0.02. We sampling 200 turns. For each turn, 1000 data points are generated from the five sampling distributions where the probability of each data point in the sampling distribution is evaluated at the same time. For each data point, likelihood is calculated from the normal distribution where the mean value is the default observed value and the standard deviation is the value of the N variable. The probability of the posterior distribution is calculated as the multiplication between the probability of the prior distribution and the likelihood. Dividing the probability of the posterior distribution by the probability of the sampling distribution, we get the accepted ratio.

Based on the accepted ratio, a random number is generated from the uniform distribution between $[0, 1]$ and is used to determine whether or not the current data point is accepted. If so, it is added to the current data set. Every time after sampling 1000 data points and filtering the data points with the lower accepted ratio, numerical integration is used to calculate the KL divergence value between the prior distribution and the posterior distribution for each variable, from which we can quantify whether the posterior distribution is stable. The results are as follows:

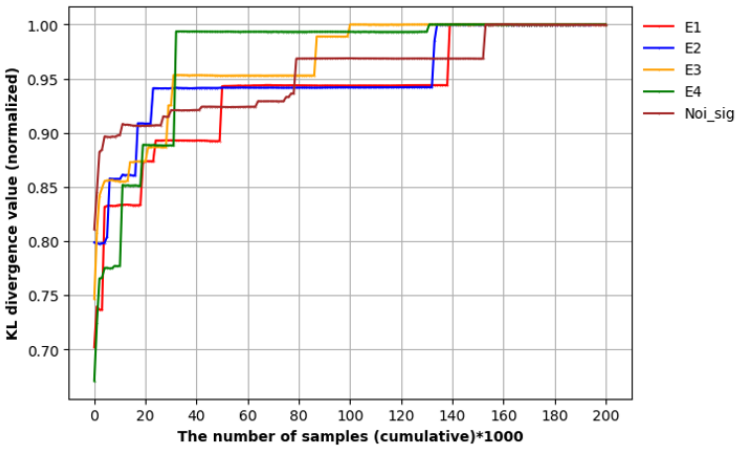
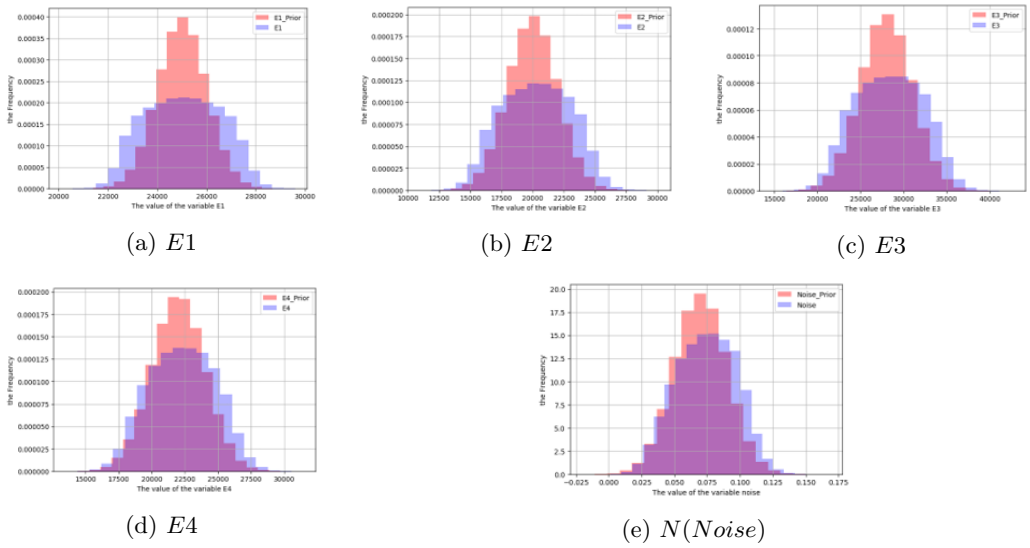


Figure 1: K-L divergence score between the prior and the posterior pdf for each variable

From Fig.(1), it can be clearly seen that the posterior distributions of all 5 variables are stable after the number of samples exceeds 160000. Based on the above result, the prior

distribution and the posterior distribution for each variable are shown as follows:



(a) $E1$ (b) $E2$ (c) $E3$

(d) $E4$ (e) $N(Noise)$

Figure 2: Prior distribution and Posterior distribution after 160000 samples

In summary, there exists a difference between the prior distribution and the posterior distribution, which means our initial guess about the distribution for each variable is not accuracy. After combining information from the data points generated by physics model, the posterior distribution can be more representative.