UQ 6310 Homework4

Assuming that $E1, E2, E3, E4, P, Noise$ follows the normal distribution, where the mean values are $30000, 25000, 32000, 27000, 2000, 200$, and the values of the standard deviation are $3000, 2700, 3200, 2500, 400, 10$ respectively. For each sampling, 100 data points are extracted from the above distribution. For each point, accepted ratio is calculated by dividing the posterior pdf by the pdf of the sampling distribution. Assuming $E1, E2, E3, E4, P, Noise$ are independent variables, the pdf of the sampling distribution is calculated by the product of the 6 pdfs.

After sampling 100 data points and add them to the current data set, numerical integration is used to calculate the KL divergence value between the sampling distribution and the posterior distribution for each variable, from which we can quantify whether the posterior distribution is stable. The results are as follows:

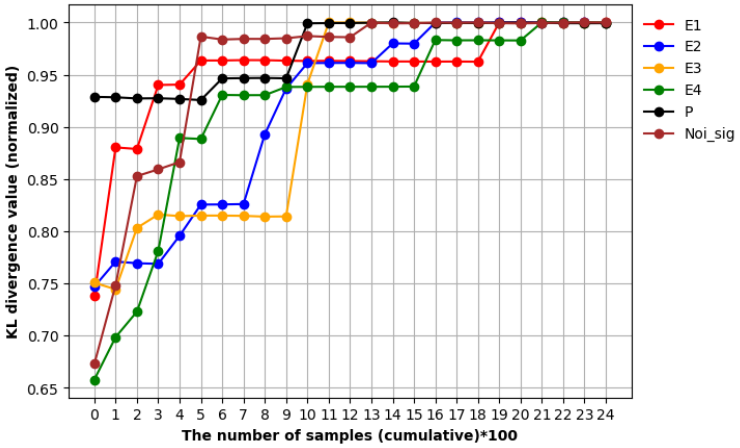
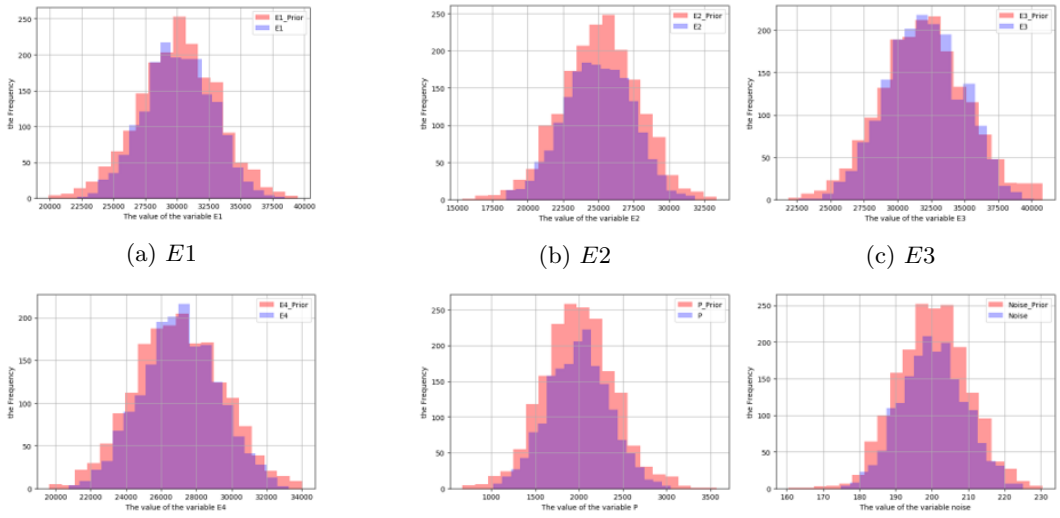


Figure 1: Normalized KL divergence value for each variable

From Fig.(1), it can be clearly seen that the posterior distributions of all 6 variables are stable after the number of samples exceeds 2100. Based on the results, the sampling distribution and the posterior distribution after 2100 data points are sampled are shown as follows:



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

(d) $E4$ (e) P (f) $Noise - Sigma$

Figure 2: Sampling distribution and Posterior distribution after 2500 samples

In summary, there exists a difference between the sampling (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.