simulated dataset 1