THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT6011/7611/8305 COMPUTATIONAL STATISTICS
(2018 Fall)

Assignment 3, due on November 30

All numerical computation MUST be conducted in Python, and attach the Python code.

1. Consider an integral
$$\int_{-1}^{5} \frac{\{\sin(x)\}^3 + 5}{\sqrt{x^4+1}} dx.$$
   (a) Plot the above integrand function in the range of $(-1, 5)$.
   (b) Use the Gaussian Legendre, Chebyshev 1, Chebyshev 2, and Jacobi quadratures respectively to approximate the integral with 5, 10, 20 nodes and weights, respectively. Present the nodes, weights, and the approximation results.

2. Use the dataset q2.csv, the observed data $\mathbf{y} = (y_1, \ldots, y_{1000})$ are from a mixture of normal distributions, i.e., $y_i \sim N(\mu_j, \sigma_j^2)$, $j = 1, \ldots, K$. Apply the EM algorithm to estimate the number of components K, and (μ_j, σ_j), $j = 1, \ldots, K$.

3. Use the EM algorithm to estimate the parameters in the random intercept model, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$,
$$\begin{aligned} &y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \epsilon_{ij}, \\ &\epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \\ &u_i \sim N(0, \sigma_u^2). \end{aligned}$$
   The unknown parameter vector $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_u^2, \sigma_\epsilon^2)^{\mathrm{T}}$.
   (a) Derive the Q-function and the M-step of the EM algorithm.
   (b) Conduct simulations as follows. Set the parameters $\beta_0 = 0.5$, $\beta_1 = 1$, $\sigma_u = 1$, $\sigma_\epsilon = 1$, $I = 100$, and $J = 5$. For each dataset, simulate x_{ij} from Uniform$(-1, 1)$, simulate ϵ_{ij} and u_i from the corresponding normal distributions, and then obtain y_{ij}. Use the EM algorithm to obtain the parameter estimates based on each simulated dataset. Repeat the simulation process 1000 times and present the bias (averaged over 1000 simulations) and standard deviation for $\boldsymbol{\theta}$. Comment on your findings.

4. In the hidden Markov model, consider tossing a fair die (with probabilities 1/6 for all the six side numbers $\{1, \ldots, 6\}$) and a loaded die (with probabilities

$\{1/12, 1/12, 1/12, 1/6, 1/4, 1/3\}$ for all the six numbers $\{1, \ldots, 6\}$). Suppose that we observe a sequence of tosses with numbers

$$\{251326344212463366565535614566523665561326345621443235213263461435421\}$$

determine the underlining coin status: fair or loaded, under each of the four transition matrices.

(a) If the transition matrix between the fair and loaded dice is

	Fair	Loaded
Fair	0.8	0.3
Loaded	0.2	0.7

determine the underlining coin status: fair or loaded.

(b) If the transition matrix between the fair and loaded dice is

	Fair	Loaded
Fair	0.9	0.15
Loaded	0.1	0.85

determine the underlining coin status: fair or loaded.

(c) If the transition matrix between the fair and loaded dice is

	Fair	Loaded
Fair	0.95	0.6
Loaded	0.05	0.4

determine the underlining coin status: fair or loaded.

(d) If the transition matrix between the fair and loaded dice is

	Fair	Loaded
Fair	0.5	0.5
Loaded	0.5	0.5

determine the underlining coin status: fair or loaded.