1. (a)

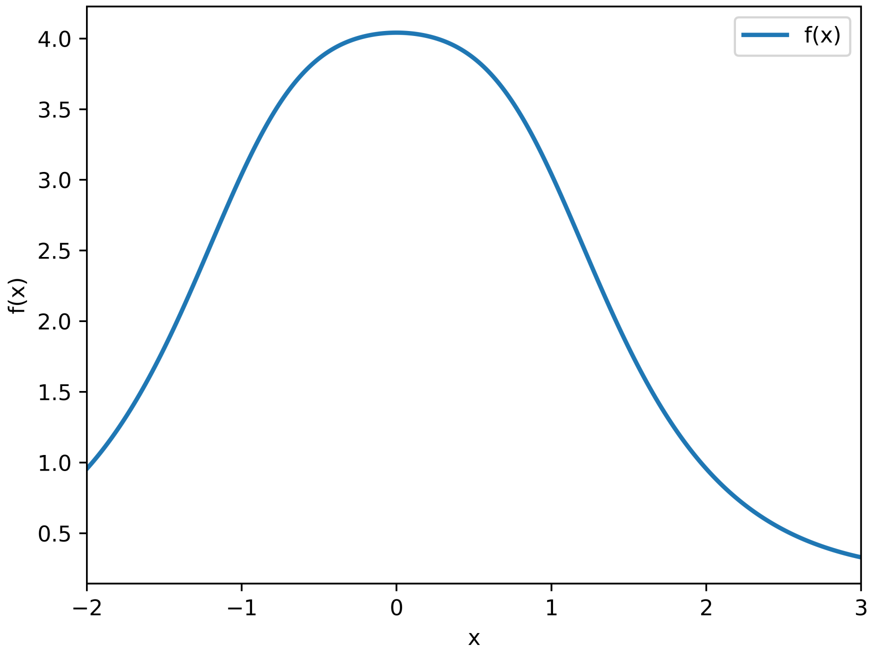

Figure 1: Q1(a)

(b)

Table 1: Gaussian quadrature for $n = 20$

		Values																				Integral
Legendre	Nodes	-0.993	-0.964	-0.912	-0.839	-0.746	-0.636	-0.511	-0.374	-0.228	-0.077	0.077	0.228	0.374	0.511	0.636	0.746	0.839	0.912	0.964	0.993	11.82
	Weights	0.018	0.041	0.063	0.083	0.102	0.118	0.132	0.142	0.149	0.153	0.153	0.149	0.142	0.132	0.118	0.102	0.083	0.063	0.041	0.018	
Chebyshev 1	Nodes	-0.997	-0.972	-0.924	-0.853	-0.760	-0.649	-0.522	-0.383	-0.233	-0.078	0.078	0.233	0.383	0.522	0.649	0.760	0.853	0.924	0.972	0.997	11.83
	Weights	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	0.157	
Chebyshev 2	Nodes	-0.989	-0.956	-0.901	-0.826	-0.733	-0.623	-0.500	-0.365	-0.223	-0.075	0.075	0.223	0.365	0.500	0.623	0.733	0.826	0.901	0.956	0.989	11.82
	Weights	0.003	0.013	0.028	0.047	0.069	0.091	0.112	0.130	0.142	0.149	0.149	0.142	0.130	0.112	0.091	0.069	0.047	0.028	0.013	0.003	
Jacobi	Nodes	-0.984	-0.947	-0.890	-0.814	-0.720	-0.612	-0.490	-0.358	-0.218	-0.073	0.073	0.218	0.358	0.490	0.612	0.720	0.814	0.890	0.947	0.984	11.81
	Weights	0.001	0.005	0.014	0.029	0.049	0.072	0.097	0.119	0.136	0.145	0.145	0.136	0.119	0.097	0.072	0.049	0.029	0.014	0.005	0.001	

2. (a) The parameters of interest are $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_u^2, \sigma_\epsilon^2)^\intercal$. Consider the complete data $(\boldsymbol{Y}, \boldsymbol{u}) = (\{y_{ij}\}_{i=1,j=1}^{I,J}, \{u_i\}_{i=1}^{I})$, the complete-data likelihood has the form,

$$
\begin{aligned}
f(\boldsymbol{Y}, \boldsymbol{u}|\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2) &= \prod_{i=1}^{I}\left[f(u_i|\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2)\prod_{j=1}^{J} f(y_{ij}|u_i, \beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2)\right] \\
f(y_{ij}|u_i, \beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2) &= \frac{1}{\sqrt{2\pi}\sigma_\epsilon} e^{-\frac{(y_{ij}-\beta_0-\beta_1 x_{ij}-u_i)^2}{2\sigma_\epsilon^2}} \\
f(u_i|\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2) &= \frac{1}{\sqrt{2\pi}\sigma_u} e^{-\frac{u_i^2}{2\sigma_u^2}} \\
\log f(\boldsymbol{Y}, \boldsymbol{u}|\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2) &= -\frac{I}{2}\log(\sigma_u^2) - \frac{IJ}{2}\log(\sigma_\epsilon^2) - \sum_{i=1}^{I}\frac{u_i^2}{2\sigma_u^2} \\
&\quad - \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(y_{ij}-\beta_0-\beta_1 x_{ij}-u_i)^2}{2\sigma_\epsilon^2} + \text{unrelated terms}
\end{aligned}
$$

We can then derive the conditional distribution of $\boldsymbol{u}$,

$$
u_i|\boldsymbol{Y}, \beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2 \sim N(\frac{J\sigma_u^2(\bar{y}_{i+}-\beta_0-\beta_1\bar{x}_{i+})}{J\sigma_u^2+\sigma_\epsilon^2}, \frac{\sigma_u^2\sigma_\epsilon^2}{J\sigma_u^2+\sigma_\epsilon^2})
$$

where $\bar{y}_{i+} = \sum_{j=1}^{J} y_{ij}/J$ and $\bar{x}_{i+} = \sum_{j=1}^{J} x_{ij}/J$. Then we have

$$
\begin{aligned}
E[u_i|\boldsymbol{\theta}^{(t)}] &= \frac{J\sigma_u^2(\bar{y}_{i+}-\beta_0-\beta_1\bar{x}_{i+})}{J\sigma_u^2+\sigma_\epsilon^2} \\
E[u_i^2|\boldsymbol{\theta}^{(t)}] &= \frac{J^2\sigma_u^4(\bar{y}_{i+}-\beta_0-\beta_1\bar{x}_{i+})^2}{(J\sigma_u^2+\sigma_\epsilon^2)^2} + \frac{\sigma_u^2\sigma_\epsilon^2}{J\sigma_u^2+\sigma_\epsilon^2}
\end{aligned}
$$

So we can get the Q-function

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= -\frac{I}{2}\log(\sigma_u^2) - \frac{IJ}{2}\log(\sigma_\epsilon^2) - \sum_{i=1}^{I}\frac{E[u_i^2|\boldsymbol{\theta}^{(t)}]}{2\sigma_u^2} + \text{unrelated terms} \\
&\quad - \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(y_{ij}-\beta_0-\beta_1 x_{ij})^2 - 2(y_{ij}-\beta_0-\beta_1 x_{ij})E[u_i|\boldsymbol{\theta}^{(t)}] + E[u_i^2|\boldsymbol{\theta}^{(t)}]}{2\sigma_\epsilon^2} \\
&= -\frac{I}{2}\log(\sigma_u^2) - \frac{IJ}{2}\log(\sigma_\epsilon^2) - \sum_{i=1}^{I}\frac{E[u_i^2|\boldsymbol{\theta}^{(t)}]}{2\sigma_u^2} \\
&\quad - \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(y_{ij}-\beta_0-\beta_1 x_{ij}-E[u_i|\boldsymbol{\theta}^{(t)}])^2 + Var[u_i|\boldsymbol{\theta}^{(t)}]}{2\sigma_\epsilon^2} + \text{unrelated terms}
\end{aligned}
$$

So the M-step of EM algorithm is performed as following. Firstly, the $\beta_0^{(t+1)}$ and $\hat{\beta}_1^{(t+1)}$ is obtained as the OLSE of the regression model $y_{ij} - E[u_i|\boldsymbol{\theta}^{(t)}] = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$. And then obtain

$$\sigma_\epsilon^{2\,(t+1)} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(y_{ij} - E[u_i|\boldsymbol{\theta}^{(t)}] - \beta_0^{(t+1)} - \beta_1^{(t+1)} x_{ij})^2 + Var[u_i|\boldsymbol{\theta}^{(t)}]}{IJ}$$

$$\sigma_u^{2\,(t+1)} = \sum_{i=1}^{I} \frac{E[u_i^2|\boldsymbol{\theta}^{(t)}]}{I}$$

(b) After 100 iterations, we derive the results shown in Table 2.

Table 2: Simulation results of ($\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}_\epsilon^2$, $\hat{\sigma}_u^2$)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_u^2$
Mean	1.0091	0.9931	0.2558	0.2456
Std.	0.1003	0.1512	0.0588	0.0377

3. (a) For the complete data (y_i, u_i), where u_i indicates which distribution y_i is from, the complete-data likelihood has the form,

$$g(\{y_i, u_i\}_{i=1}^n|\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{k} [\omega_j f_j(y_i)]^{I(u_i=j)}, \tag{0.1}$$

where $\boldsymbol{\theta}$ is the unknown parameter and

$$f_j(y_i) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_i-\mu_j)^2}{2\sigma_j^2}\right\}.$$

(b) The conditional distribution of y_i given $u_i = j$ is f_j, i.e.,

$$y_i|u_i = j \sim N(\mu_j, \sigma_j^2).$$

Thus, the marginal distribution of y_i has the form,

$$p(y_i) = \sum_{j=1}^{k} p(y_i|u_i = j)p(u_i = j) = \sum_{j=1}^{k} \omega_j f_j(y_i) = \sum_{j=1}^{k} \frac{\omega_j}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_i-\mu_j)^2}{2\sigma_j^2}\right\}$$

(c)

It seems that the observations have three clusters centering at -2.5, -0.7 and 1.8 respectively.

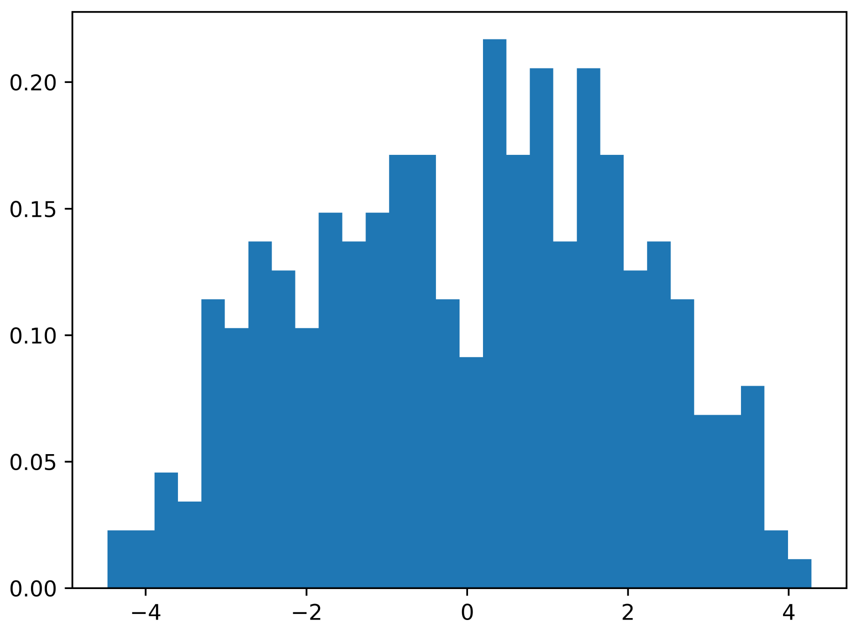

Figure 2: Q3(c)

(d) Given $k = 3, \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1, \omega_1 = \omega_2 = \omega_3 = 1/3$, the complete-data likelihood is,

$$g(\{y_i, u_i\}_{i=1}^n | \mu_1, \mu_2, \mu_3) = \prod_{i=1}^n \prod_{j=1}^3 \left[\frac{1}{3} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(y_i - \mu_j)^2}{2} \right\} \right]^{I(u_i = j)}.$$

Let $\boldsymbol{\mu}^{(0)}$ be the current estimate of $\boldsymbol{\mu}$, then

$$\begin{aligned} Q(\boldsymbol{\mu}|\boldsymbol{\mu}^{(0)}) &= E_{\boldsymbol{\mu}^{(0)}} \left[\log \left\{ g(\{y_i, u_i\}_{i=1}^n | \boldsymbol{\mu}) \right\} | \boldsymbol{y} \right] \\ &= \sum_{i=1}^n \sum_{j=1}^3 E_{\boldsymbol{\mu}^{(0)}}[I(u_i = j)|y_i] \log f_j(y_i) + \text{unrelated terms} \\ &= -\sum_{i=1}^n \sum_{j=1}^3 P_{\boldsymbol{\mu}^{(0)}}[u_i = j|y_i] \frac{(y_i - \mu_j)^2}{2} + \text{unrelated terms}, \end{aligned}$$

where

$$P_{\boldsymbol{\mu}^{(0)}}[u_i = j|y_i] = \frac{\exp\left\{ -\frac{(y_i - \mu_j^{(0)})^2}{2} \right\}}{\sum_{k=1}^3 \exp\left\{ -\frac{(y_i - \mu_k^{(0)})^2}{2} \right\}}$$

In the M-step, we can derive the updated parameter $\boldsymbol{\mu}^{(1)}$ by

$$\mu_j^{(1)} = \frac{\sum_{i=1}^n P_{\boldsymbol{\mu}^{(0)}}[u_i = j|y_i] y_i}{\sum_{i=1}^n P_{\boldsymbol{\mu}^{(0)}}[u_i = j|y_i]}, \quad j = 1, 2, 3.$$

Using the EM algorithm, the updates reach convergence after 41 iterations (tolerance $= 10^{-5}$) with

$$\hat{\boldsymbol{\mu}} = (-2.1680, 0.1710, 2.0124)^\intercal.$$

(e) Given $k = 3, \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, the complete-data likelihood is,

$$g(\{y_i, u_i\}_{i=1}^n | \mu_1, \mu_2, \mu_3, \omega_1, \omega_2, \omega_3) = \prod_{i=1}^n \prod_{j=1}^3 \left[\omega_j \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(y_i - \mu_j)^2}{2} \right\} \right]^{I(u_i = j)}.$$

Let $\boldsymbol{\mu}^{(0)}, \boldsymbol{\omega}^{(0)}$ be the current estimate of $\boldsymbol{\mu}, \boldsymbol{\omega}$, respectively, then

$$\begin{aligned} Q(\boldsymbol{\mu}, \boldsymbol{\omega}|\boldsymbol{\mu}^{(0)}, \boldsymbol{\omega}^{(0)}) &= E_{\boldsymbol{\mu}^{(0)}, \boldsymbol{\omega}^{(0)}} \left[\log \left\{ g(\{y_i, u_i\}_{i=1}^n | \boldsymbol{\mu}, \boldsymbol{\omega}) \right\} | \boldsymbol{y} \right] \\ &= \sum_{i=1}^n \sum_{j=1}^3 E_{\boldsymbol{\mu}^{(0)}, \boldsymbol{\omega}^{(0)}}[I(u_i = j)|y_i] \{\log \omega_j + \log f_j(y_i)\} + \text{unrelated terms} \\ &= \sum_{i=1}^n \sum_{j=1}^3 P_{\boldsymbol{\mu}^{(0)}, \boldsymbol{\omega}^{(0)}}[u_i = j|y_i] \left\{ -\frac{(y_i - \mu_j)^2}{2} + \log \omega_j \right\} + \text{unrelated terms}, \end{aligned}$$

where

$$P_{\boldsymbol{\mu}^{(0)},\boldsymbol{\omega}^{(0)}}[u_i = j|y_i] = \frac{\omega_j^{(0)} \exp\left\{-\frac{(y_i-\mu_j^{(0)})^2}{2}\right\}}{\sum_{k=1}^{3} \omega_k^{(0)} \exp\left\{-\frac{(y_i-\mu_k^{(0)})^2}{2}\right\}}$$

In the M-step, we can derive the updated parameter $\boldsymbol{\mu}^{(1)}, \boldsymbol{\omega}^{(1)}$ by

$$\begin{aligned}
\omega_j^{(1)} &= \frac{1}{n}\sum_{i=1}^{n} P_{\boldsymbol{\mu}^{(0)},\boldsymbol{\omega}^{(0)}}[u_i = j|y_i] \\
\mu_j^{(1)} &= \frac{\sum_{i=1}^{n} P_{\boldsymbol{\mu}^{(0)},\boldsymbol{\omega}^{(0)}}[u_i = j|y_i]y_i}{\sum_{i=1}^{n} P_{\boldsymbol{\mu}^{(0)},\boldsymbol{\omega}^{(0)}}[u_i = j|y_i]}, \quad j = 1, 2, 3.
\end{aligned}$$

Using the EM algorithm, the updates reach convergence after 363 iterations (tolerance $= 10^{-5}$) with

$$\hat{\boldsymbol{\mu}} = (-2.1985, 0.0901, 1.9810)^{\intercal}, \hat{\boldsymbol{\omega}} = (0.3219, 0.3301, 0.3480)^{\intercal}.$$