ENGG 5202: Assignment #1

Due on Thursday, March 3, 2015

Kai Chen

Problem 1

The likelihood of θ_1

$$p(D_1|\theta_1) = p(x_1|\omega_1, \theta_1)p(x_2|\omega_1, \theta_1)$$

The log-likelihood

$$l(\theta_1) = \log p(x_1|\omega_1, \theta_1) = \log p(x_1|\omega_1, \theta_1) + \log p(x_2|\omega_1, \theta_1)$$

We determine θ_1 by maximizing $l(\theta_1)$

$$\hat{\theta_1} = \arg\max l(\theta_1)$$

Let

$$\nabla l(\theta_1) = 0$$

By substituting symbols with numeral values

$$\begin{aligned} l(\theta_1) &= 2\log\frac{2}{\theta_1} + \log(1 - \frac{2}{\theta_1}) + \log(1 - \frac{5}{\theta_1}) \\ \nabla l(\theta_1) &= -\frac{4}{\theta_1} + \frac{4}{\theta_1(\theta_1 - 2)} + \frac{10}{\theta_1(\theta_1 - 5)} \end{aligned}$$

We get

$$\theta_1 = 8 \quad or \quad \theta_1 = 2.5$$

However, $p(x|\omega_1) = 0$ when $x > \theta_1$ according to the densities form, if $\theta_1 = 2.5$, then $D_1 = \{2, 5\}$ will not occur. Thus $\theta_1 = 8$.
Similarly, we can calculate $\theta_2 \approx 14.2$

Problem 2

2.1

Figure 1 shows $p(x|\theta)$ versus x for $\theta = 1$.
Figure 2 shows $p(x|\theta)$ versus θ for $x = 2$.

2.2

The log-likelihood

$$\begin{aligned} l(\theta) &= \log\prod_{k=1}^{n} p(x_k|\theta) \\ &= \sum_{k=1}^{n}\log\theta \mathrm{e}^{-\theta x_k} \end{aligned}$$

The gradient of $l(\theta)$

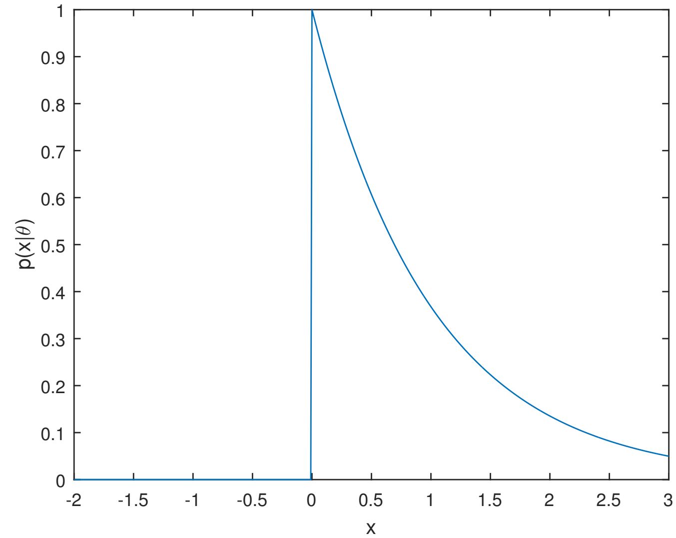

Figure 1: $p(x|\theta)$ versus x for $\theta = 1$

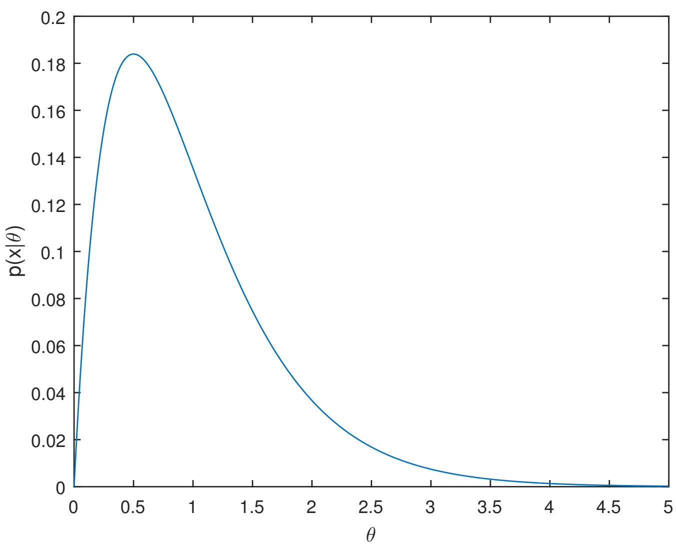

Figure 2: $p(x|\theta)$ versus θ for $x = 2$

$$\begin{aligned}\nabla l(\theta) &= \sum_{k=1}^{n} \frac{(1-\theta x_k)\mathrm{e}^{-\theta x_k}}{\theta \mathrm{e}^{-\theta x_k}} \\ &= \sum_{k=1}^{n} \frac{1-\theta x_k}{\theta} \\ &= \frac{n - \theta \sum_{k=1}^{n} x_k}{\theta}\end{aligned}$$

Let $\nabla l(\theta) = 0$, we can calculate

$$\hat{\theta} = \frac{n}{\sum_{k=1}^{n} x_k}$$

2.3

According to the law of large numbers, when n is very large, the sample average converges to the expected value.

$$\frac{\sum_{k=1}^{n} x_k}{n} \to \mathcal{E}(x) = \int_0^{\infty} \theta x \mathrm{e}^{-\theta x} \,\mathrm{d}x = \frac{1}{\theta}$$

So $\hat{\theta}$ approach to the true θ when n is very large. If the samples are generated from $p(x|\theta)$ with $\theta = 1$, the maximum-likelihood estimate $\hat{\theta}$ for large n is 1.

Problem 3

3.1

$$p(x_k|\theta^{(t)}) = \sum_{j=1}^{m} \sum_{i=1}^{l} p_j^{(t)} q_i^{(t)} N(x_k; \mu_j^{(t)}, {\sigma_i^{(t)}}^2)$$

$$p(z_k, y_k, x_k|\theta^{(t)}) = p_{z_k} q_{y_k} N(x_k; \mu_{z_k}^{(t)}, \sigma_{y_k}^{(t)})$$

$$\begin{aligned}p(z_k, y_k; x_k, \theta^{(t)}) &= \frac{p(z_k, y_k, x_k|\theta^{(t)})}{p(x_k|\theta^{(t)})} \\ &= \frac{p_{z_k} q_{y_k} N(x_k; \mu_{z_k}^{(t)}, \sigma_{y_k}^{(t)})}{\sum_{j=1}^{m} \sum_{i=1}^{l} p_j \, q_i N(x_k; \mu_j^{(t)}, {\sigma_i^{(t)}}^2)}\end{aligned}$$

3.2

$$l_c(x_1, \dots, x_n, z_1, \dots, z_n, y_1, \dots, y_n; \theta) = \sum_{k=1}^{n} \log(p_{z_k} q_{y_k} N(x_k; \mu_{z_k}^{(t)}, \sigma_{y_k}^{(t)}))$$

$$
\begin{aligned}
Q(\theta;\theta^{(t)}) &= \mathcal{E}\{l_c(x_1,\ldots,x_n,z_1,\ldots,z_n,y_1,\ldots,y_n;\theta)|x_1,\ldots,x_n,\theta^{(t)}\} \\
&= \sum_{k=1}^{n} \mathcal{E}\{\log(p_{z_k} q_{y_k} N(x_k;\mu_{z_k}^{(t)},\sigma_{y_k}^{(t)}))|x_k,\theta^{(t)}\} \\
&= \sum_{k=1}^{n}\sum_{j=1}^{m}\sum_{i=1}^{l} P(z_k=j,y_k=i|x_k,\theta^{(t)}) \log(p_j\, q_i N(x_k;\mu_j,\sigma_i^2))
\end{aligned}
$$

3.3

$$
\text{Let} \frac{\partial Q(\theta;\theta^{(t)})}{\partial \mu_j} = \sum_{k=1}^{n}\sum_{i=1}^{l} P(z_k=j,y_k=i|x_k,\theta^{(t)}) \frac{x_k-\mu_j}{\sigma_i^2} = 0
$$

$$
\mu_j = \frac{\sum_{k=1}^{n}\sum_{i=1}^{l} P(z_k=j,y_k=i|x_k,\theta^{(t)})\frac{x_k}{\sigma_i^2}}{\sum_{k=1}^{n}\sum_{i=1}^{l} P(z_k=j,y_k=i|x_k,\theta^{(t)})\frac{1}{\sigma_i^2}}
$$

3.4

The approximation quality is significantly worse when $l = 1$ than $l > 1$. This is because the data is sampled from a true mixture model with 3 mean parameters and 2 variance parameters, while a mixture model with single variance parameter and many mean parameters has weaker ability to fit the data well due to the degree of freedom. Instead, a mixture model which can adjust multiple variance parameters as well as mean parameters can do better.

Problem 4

4.1

One sequence generated by my program: 2 1 1 1 1 2 2 2 3 1

4.2

The most likely sequence of hidden states: 1 2 2 2 2 1 1 1 1 2
The true hidden states: 2 1 2 2 2 2 2 1 1 1
Codes have been submitted to Piazza.