Btry 694 Class Homework 5

11/2/98

1 Problem 5.4

Krishanu Maulik

Consider a finite space $\mathcal{X}$ and a function H defined on $\mathcal{X}$.

(a) Show that

$$\lim_{\lambda \to \infty} \frac{\exp\{\lambda H(x)\}}{\sum_{\mathcal{X}} \exp\{\lambda H(x)\}} = \begin{cases} \frac{1}{M} & \text{if } x \in \mathcal{O}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathcal{O}$ is the set of global minima of H and M is the cardinal of $\mathcal{O}$.

(*Hint*: Introduce $\omega = \min_{\mathcal{X}} H(x)$.)

(b) Show that, for $x \in \mathcal{O}$, $\pi_\lambda(x) \propto \exp\{\lambda H(x)\}$ increases and, for $x \notin \mathcal{O}$, $\pi_\lambda(x)$ decreases eventually.

Note :

The problem as given is wrong and there can be two modifications to the problem.

Version 1:

Consider a finite space $\mathcal{X}$ and a function H defined on $\mathcal{X}$.

(a) Show that

$$\lim_{\lambda \to -\infty} \frac{\exp\{\lambda H(x)\}}{\sum\limits_{\mathcal{X}} \exp\{\lambda H(y)\}} = \begin{cases} \frac{1}{M} & \text{if } x \in \mathcal{O}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathcal{O}$ is the set of global minima of H and M is the cardinal of $\mathcal{O}$.

(b) Show that, for $x \in \mathcal{O}$, $\pi_\lambda(x) \propto \exp\{\lambda H(x)\}$ increases and, for $x \notin \mathcal{O}$, $\pi_\lambda(x)$ decreases eventually.

Solution:

Let $\omega = \min\limits_{\mathcal{X}} H(x)$.
Let $h(x) = H(x) - \omega$. Then $h(x) \geq 0$ and equality holds iff $x \in \mathcal{O}$.

$$\therefore \lim_{\lambda \to -\infty} \exp\{\lambda h(x)\} \quad = \quad \begin{cases} 1 & \text{if } x \in \mathcal{O}, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

Let

$$\pi_\lambda(x) = \frac{\exp\{\lambda H(x)\}}{\sum\limits_{\mathcal{X}} \exp\{\lambda H(y)\}}.$$

$$\begin{aligned} \therefore \pi_\lambda(x) &= \frac{\exp\{\lambda h(x)\}}{\sum\limits_{\mathcal{X}} \exp\{\lambda h(y)\}}, \text{ [dividing numerator and denominator by } \exp(\omega)] \\ &= \frac{\exp\{\lambda h(x)\}}{\sum\limits_{\mathcal{O}} \exp\{\lambda h(y)\} + \sum\limits_{\mathcal{X} \setminus O} \exp\{\lambda h(y)\}} \end{aligned}$$

$$= \frac{\exp\{\lambda h(x)\}}{M + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}}, \text{ [as } h(y) = 0, y \in \mathcal{O}].$$

$$\begin{aligned}
\therefore \lim_{\lambda \to -\infty} \frac{\exp\{\lambda H(x)\}}{\sum\limits_{\mathcal{X}} \exp\{\lambda H(y)\}} &= \lim_{\lambda \to -\infty} \frac{\exp\{\lambda h(x)\}}{M + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}} \\
&= \frac{\lim\limits_{\lambda \to -\infty} \exp\{\lambda H(x)\}}{M + \lim\limits_{\lambda \to -\infty} \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}} \\
&= \frac{\lim\limits_{\lambda \to -\infty} \exp\{\lambda H(x)\}}{M + \sum\limits_{\mathcal{X}\setminus O} \lim\limits_{\lambda \to -\infty} \exp\{\lambda h(y)\}} \text{ [as } \mathcal{X} \setminus O \text{ is finite]}
\end{aligned}$$

$$= \frac{\lim\limits_{\lambda \to -\infty} \exp\{\lambda H(x)\}}{M} \text{ [using (1)]}$$

$$= \begin{cases} \frac{1}{M} & \text{if } x \in \mathcal{O} \\ 0 & \text{otherwise} \end{cases} \quad \text{[using (1)]}$$

If $x \in \mathcal{O}$, $\pi_\lambda(x) = \frac{1}{M + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}}$. Thus $\pi_\lambda(x)$ increases as λ decreases, since $h(y) > 0$, for $y \in \mathcal{X} \setminus O$.

If $x \in \mathcal{X} \setminus O$,

$$\pi_\lambda(x) = \frac{\exp\{\lambda h(x)\}}{M + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}}.$$

$$\therefore \pi_\lambda^{'}(x) = \frac{\exp\{\lambda h(x)\}}{[M + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}]^2}[Mh(x) + \sum_{\mathcal{X}\setminus O}\{(h(x) - h(y)) \exp(\lambda h(y))\}],$$

which has same sign as

$$Mh(x) + \sum_{\mathcal{X}\setminus O} [(h(x) - h(y)) \exp\{\lambda h(y)\}].$$

But this expression goes to $Mh(x) > 0$ as $\lambda \to -\infty$, since $h(y) > 0$, for $y \in \mathcal{X} \setminus O$. So eventually, for all small λ, the derivative will be positive and hence as $\lambda \downarrow -\infty$, $\pi_\lambda(x)$ will decrease.

Version 2:

Consider a finite space $\mathcal{X}$ and a function H defined on $\mathcal{X}$.

(a) Show that

$$\lim_{\lambda\to\infty} \frac{\exp\{\lambda H(x)\}}{\sum_{\mathcal{X}} \exp\{\lambda H(y)\}} = \begin{cases} \frac{1}{M} & \text{if } x \in \mathcal{O}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathcal{O}$ is the set of global maxima of H and M is the cardinal of $\mathcal{O}$.

(b) Show that, for $x \in \mathcal{O}$, $\pi_\lambda(x) \propto \exp\{\lambda H(x)\}$ increases and, for $x \notin \mathcal{O}$, $\pi_\lambda(x)$ decreases eventually.

Solution:

Let $\omega = \max_{\mathcal{X}} H(x)$.
Let $h(x) = H(x) - \omega$. Then $h(x) \leq 0$ and equality holds iff $x \in \mathcal{O}$.

$$\therefore \lim_{\lambda\to\infty} \exp\{\lambda h(x)\} \;=\; \begin{cases} 1 & \text{if } x \in \mathcal{O}, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

Let

$$\pi_\lambda(x) = \frac{\exp\{\lambda H(x)\}}{\sum\limits_{\mathcal{X}} \exp\{\lambda H(y)\}}.$$

$$\begin{aligned}
\therefore \pi_\lambda(x) &= \frac{\exp\{\lambda h(x)\}}{\sum\limits_{\mathcal{X}} \exp\{\lambda h(y)\}}, \text{ [dividing numerator and denominator by } \exp(\omega)] \\
&= \frac{\exp\{\lambda h(x)\}}{\sum\limits_{\mathcal{O}} \exp\{\lambda h(y)\} + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}} \\
&= \frac{\exp\{\lambda h(x)\}}{M + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}}, \text{ [as } h(y) = 0, y \in \mathcal{O}].
\end{aligned}$$

$$\begin{aligned}
\therefore \lim_{\lambda\to\infty} \frac{\exp\{\lambda H(x)\}}{\sum\limits_{\mathcal{X}} \exp\{\lambda H(y)\}} &= \lim_{\lambda\to\infty} \frac{\exp\{\lambda h(x)\}}{M + \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}} \\
&= \frac{\lim\limits_{\lambda\to\infty} \exp\{\lambda H(x)\}}{M + \lim\limits_{\lambda\to\infty} \sum\limits_{\mathcal{X}\setminus O} \exp\{\lambda h(y)\}} \\
&= \frac{\lim\limits_{\lambda\to\infty} \exp\{\lambda H(x)\}}{M + \sum\limits_{\mathcal{X}\setminus O} \lim\limits_{\lambda\to\infty} \exp\{\lambda h(y)\}} \text{ [as } \mathcal{X} \setminus O \text{ is finite]} \\
&= \frac{\lim\limits_{\lambda\to\infty} \exp\{\lambda H(x)\}}{M} \text{ [using (2)]} \\
&= \begin{cases} \frac{1}{M} & \text{if } x \in \mathcal{O} \\ 0 & \text{otherwise} \end{cases} \text{ [using (2)]}
\end{aligned}$$

If $x \in \mathcal{O}$, $\pi_\lambda(x) = \frac{1}{M + \sum\limits_{\mathcal{X} \setminus O} \exp\{\lambda h(y)\}}$. Thus $\pi_\lambda(x)$ increases as λ increases, since $h(y) < 0$, for $y \in \mathcal{X} \setminus O$.

If $x \in \mathcal{X} \setminus O$,

$$\pi_\lambda(x) = \frac{\exp\{\lambda h(x)\}}{M + \sum\limits_{\mathcal{X} \setminus O} \exp\{\lambda h(y)\}}.$$

$$\therefore \pi_\lambda^{'}(x) = \frac{\exp\{\lambda h(x)\}}{[M + \sum\limits_{\mathcal{X} \setminus O} \exp\{\lambda h(y)\}]^2}[Mh(x) + \sum_{\mathcal{X} \setminus O} \{(h(x) - h(y)) \exp(\lambda h(y))\}],$$

which has same sign as

$$Mh(x) + \sum_{\mathcal{X} \setminus O} [(h(x) - h(y)) \exp\{\lambda h(y)\}].$$

But this expression goes to $Mh(x) < 0$ as $\lambda \to \infty$, since $h(y) < 0$, for $y \in \mathcal{X} \setminus O$. So eventually, for all large λ, the derivative will be negative and hence as $\lambda \uparrow \infty$, $\pi_\lambda(x)$ will decrease.

2 Problem 5.5

Matt Tom

Suppose that we are trying to solve the system $Ax = b$. If we let $H = I - A$, we get that

$$x - Hx = b$$

and

$$x = Hx + b.$$

Suppose we know that the magnitude of the largest eigenvalue of H is less than 1, and suppose that the direct product of the eigenspaces is the whole space. Let $\{x_n\}$ be a sequence of vectors such that $x_{n+1} = Hx_n + b$, and let x^* be the solution to the equation $Ax = b$.

Then we know that

$$\begin{aligned} |x^* - x_{n+1}|^2 &= |x^* - Hx_n - b|^2 \\ &= |x^* - (I - A)x_n - Ax^*|^2 \cdot \\ &= |(I - A)(x^* - x_n)|^2 \end{aligned}$$

Because the eigenvalues of H have magnitude less than 1, we can conclude that

$$|x^* - x_{n+1}|^2 = |(I - A)(x^* - x_n)|^2 < |x^* - x_n|^2$$

The sequence $\{x_n\}$ converges faster than ${\lambda_1^{-}}^{n}$, so it goes to 0. So, $\{x_n\}$ converges to x^*.

Let A be a Markov chain transition matrix P. We want to calculate the expectation of

$$V_k = \left(\prod_{i=1}^{k} \frac{h_{x_i x_{i+1}}}{p_{x_i x_{i+1}}} \right) b_{k+1}.$$

Its expectation is

$$E(V_k) = \sum_{x_1} \cdots \sum_{x_k} \left(\prod_{i=1}^{k} \frac{h_{x_i x_{i+1}}}{p_{x_i x_{i+1}} p_{x_i x_{i+1}}} \right) b_{x_{k+1}} = b_{x_{k+1}} \sum_{x_1} \cdots \sum_{x_k} \prod_{i=1}^{k} h_{x_i x_{i+1}}$$

This is just the x_{k+1}st element of $H^k b$.

Suppose that we have a positive definite matrix B with a largest eigenvalue λ_1, and that we want to solve the system $By = b$. Let $A = B \cdot \frac{I}{\lambda_1}$ and $x = \lambda_1 y$. Then we can solve the system $Ax = b$ and then solve for y.

Let $H = I - A$. To solve $Ax = b$, start with a vector x_0 and calculate the sequence $x_{n+1} = Hx_n + b$. When the sequence finally converges to an x^*, let $y = x^*/\lambda_1$ be your output.

As far as flops go, this algorithm is terrible. Multiplying H by x_n takes $2n^2$ flops. At m iterations that's $2mn^2$. And to think that all you really should have to do to solve $Ax = b$ is a Gaussian elimination.

3 Problem 5.8

George Casella

Problem 5.8: Proof of Theorem 5.2.7, Error Terms Not Ignored We want to prove that

$$\lim_{\lambda\to\infty} \frac{\int_\Theta \theta\, e^{\lambda h(\theta)}\, d\theta}{\int_\Theta e^{\lambda h(\theta)}\, d\theta} = \theta^* .$$

The following proof is about as rigorous as I can handle. If you really want all of the details on the error terms, see Tierney and Kadane (1986 *JASA* page 82) or Schervisch 1995 *Theory of Statistics* Springer-Verlag, Section 7.4.

Let $\hat\theta$ satisfy $h'(\hat\theta) = 0$ and $h''(\hat\theta) < 0$, so $\hat\theta$ is an interior maximum. A Taylor expansion gives

$$h(\theta) = h(\hat\theta) + h''(\hat\theta)(\theta-\hat\theta)^2/2 + h'''(\hat\theta)(\theta-\hat\theta)^3/6 + O(|\theta-\hat\theta|^4)$$

where the error term satisfies

$$\lim_{\theta\to\hat\theta} \frac{O(|\theta-\hat\theta|^4)}{|\theta-\hat\theta|^4} = c = \text{ constant.}$$

We now can write

$$\begin{aligned}
\int \theta e^{\lambda h(\theta)} d\theta &= \int \theta \left[e^{\lambda\left(h(\hat\theta)+h''(\hat\theta)(\theta-\hat\theta)^2/2+h'''(\hat\theta)(\theta-\hat\theta)^3/6+O(|\theta-\hat\theta|^4)\right)}\right] d\theta \\
&= e^{\lambda h(\hat\theta)} \int \theta e^{\lambda h''(\hat\theta)(\theta-\hat\theta)^2/2} e^{\lambda h'''(\hat\theta)(\theta-\hat\theta)^3/6} e^{\lambda O(|\theta-\hat\theta|^4)} d\theta.
\end{aligned}$$

We can think of $O(|\theta-\hat\theta|^4)$ as looking like $c(|\theta-\hat\theta|^4)$ for some constant c. (There are higher order polynomial terms inside the O term, but the development below will show that these terms will go to zero faster than the fourth order term.) Now use the fact that a Taylor expansion around zero gives $e^{ax^n} = 1 + ax^n + O(x^{n+1})$ to write

$$e^{\lambda O(|\theta-\hat{\theta}|^4)} = 1 + \lambda O(|\theta - \hat{\theta}|^5)$$

and the integral becomes

$$\int \theta e^{\lambda h(\theta)} d\theta \;=\; e^{\lambda h(\hat{\theta})} \int \theta e^{\lambda h''(\hat{\theta})(\theta-\hat{\theta})^2/2} e^{\lambda h'''(\hat{\theta})(\theta-\hat{\theta})^3/6} \left[1 + \lambda O(|\theta - \hat{\theta}|^5)\right] d\theta.$$

Use this same Taylor expansion again to write

$$e^{\lambda h'''(\hat{\theta})(\theta-\hat{\theta})^3/6} = 1 + \lambda h'''(\hat{\theta})(\theta - \hat{\theta})^3/6 + \lambda O(|\theta - \hat{\theta}|^6)$$

and substitute into the integral to get

$$\begin{aligned}\int \theta e^{\lambda h(\theta)} d\theta \;&=\; e^{\lambda h(\hat{\theta})} \int \theta e^{\lambda h''(\hat{\theta})(\theta-\hat{\theta})^2/2} \left[1 + \lambda h'''(\hat{\theta})(\theta - \hat{\theta})^3/6 + \lambda O(|\theta - \hat{\theta}|^6)\right] \\ &\quad \times \left[1 + \lambda O(|\theta - \hat{\theta}|^5)\right] d\theta \\ &=\; e^{\lambda h(\hat{\theta})} \int \theta e^{\lambda h''(\hat{\theta})(\theta-\hat{\theta})^2/2} \left[1 + \lambda h'''(\hat{\theta})(\theta - \hat{\theta})^3/6 + \lambda O(|\theta - \hat{\theta}|^5)\right] d\theta.\end{aligned}$$

Next, write $\theta = (\theta - \hat{\theta}) + \hat{\theta}$ and, letting $t = \theta - \hat{\theta}$ we have

$$\begin{aligned}\int \theta e^{\lambda h(\theta)} d\theta \;&=\; e^{\lambda h(\hat{\theta})}\hat{\theta} \int e^{\lambda h''(\hat{\theta})t^2/2} \left[1 + \lambda h'''(\hat{\theta})t^3/6 + \lambda O(|t|^5)\right] dt \\ &\quad + e^{\lambda h(\hat{\theta})} \int t e^{\lambda h''(\hat{\theta})t^2/2} \left[1 + \lambda h'''(\hat{\theta})t^3/6 + \lambda O(|t|^5)\right] dt.\end{aligned}$$

We can now evaluate all of the integrals. The terms with t and t^3 in the integrand are zero by symmetry, and the error terms are bounded by the fourth order integral

$$\int \lambda M|t|^4 e^{\lambda h''(\hat{\theta})t^2/2} dt = \lambda M \frac{3}{\lambda^{5/2}[-h''(\hat{\theta})]^3} = \frac{3M}{\lambda^{3/2}[-h''(\hat{\theta})]^3} = \frac{M}{\lambda^{3/2}},$$

where M is a generic constant that may depend on $\hat{\theta}$. This is an application of the integral result that

$$\int_0^\infty x^a e^{-bx^2/2} dx = \frac{2^{\frac{a-1}{2}}\Gamma(\frac{a+1}{2})}{b^{\frac{a+1}{2}}},$$

which also shows why the higher order terms can be ignored. Substituting this in the above gives

$$\int \theta e^{\lambda h(\theta)} d\theta = e^{\lambda h(\hat{\theta})} \hat{\theta} \left[\sqrt{-2\pi h''(\hat{\theta})} + \frac{M}{\lambda^{3/2}} \right].$$

A similar development shows that the denominator is

$$\int e^{\lambda h(\theta)} d\theta = e^{\lambda h(\hat{\theta})} \left[\sqrt{-2\pi h''(\hat{\theta})} + \frac{M}{\lambda^{3/2}} \right],$$

and combining these last two approximations shows that

$$\frac{\int \theta e^{\lambda h(\theta)} d\theta}{\int e^{\lambda h(\theta)} d\theta} = \hat{\theta} \left[1 + \frac{M}{\lambda^{3/2}} \right],$$

establishing the limit. Now you know why I didn't want to do this in class.

Lastly, note that if we replace θ in the numerator integrand by some function $b(\theta)$, the limit would be the same. We would just have to do another Taylor series on $b(\theta)$. This is how Corollary 5.2.8 is established.

4 Problem 5.16

Bárbara Patricia González

In the exponential family, EM computations are somewhat simplified. Show that if the complete data density f is of the form

$$f(y, z|\theta) = h(y, z) \exp\{\eta(\theta) T(y, z) - B(\theta)\},$$

then we can write

$$Q(\theta|\theta*, y) = \mathbb{E}_{\theta*}[\log h(y, Z)] + \sum \eta_i(\theta) \mathbb{E}_{\theta*}[T_i|y] - B(\theta).$$

Deduce that calculating the complete data MLE only involves the simpler expectation $\mathbb{E}_{\theta*}[T_i|y]$.

4.1 Solution

We know that

$$Q(\theta|\theta*, y) = \mathbb{E}_{\theta*}[\log f(y, Z|\theta, y)].$$

But we have that

$$\log f(y, Z|\theta) = \log h(y, Z) + \sum \eta_i(\theta) T_i(y, Z) - B(\theta).$$

So, taking expectations conditioning on y at $\theta = \theta*$, we finally get

$$Q(\theta|\theta*, y) = \mathbb{E}_{\theta*}[\log h(y, Z)] + \sum \eta_i(\theta)\mathbb{E}_{\theta*}[T_i|y] - B(\theta).$$

Now, the MLE for the complete data is $\mathrm{a}rg\max_{\theta} Q(\theta|\theta*, y)$, and since

$$\mathbb{E}_{\theta*}[\log h(y, Z)]$$

does not depend on θ we only have to maximize

$$\sum \eta_i(\theta)\mathbb{E}_{\theta*}[T_i|y] - B(\theta),$$

and that only involves the simpler expectation $\mathbb{E}_{\theta*}[T_i|y]$.

5 Problem 5.17

Mitch Warachka

Kullback - Leibler information or distance

$$E_f[log(f(x)/g(x))] = \int log\left[\frac{f(x)}{g(x)}\right] f(x)dx \tag{3}$$

a) show information nonnegative using Jensen's Inequality

f convex (exponential) E[f(x)] $\geq$ f(E[x])

f concave (log), just reverse inequality E[f(x)] $\leq$ f(E[x])

$$0 \leq exp^{E_f(log[log(f(x)/g(x))])} \leq E_f\left[exp^{log(log(f(x)/g(x)))}\right] = E_f\left[log(f(x)/g(x)\right] \tag{4}$$

if f(x)=g(x), then

$$E_f[log(f(x)/f(x))] = E_f[log(1)] = E_f[0] = 0 \tag{5}$$

b) using part a,

$$E_f[log(f(x)/g(x)] \geq 0 \Rightarrow E_f[log(f(x)) - log(g(x))] \geq 0 \Rightarrow E_f[log(f(x))] \geq E_f[log(g(x))] \tag{6}$$

to show 5.3.18, let

$$h(\mathbf{z}, \hat{\theta}_{(\mathbf{j})}, \hat{\theta}_{(\mathbf{j+1})}, \mathbf{x}) = \frac{\mathbf{k}(\mathbf{z}|\hat{\theta}_{\mathbf{j+1}}, \mathbf{x})}{\mathbf{k}(\mathbf{z}|\hat{\theta}_{(\mathbf{j})}, \mathbf{x})} \tag{7}$$

now using the property that log is concave with the concave version of Jensen's inequality

$$E_{\hat{\theta}_{(j)}}\left[log(h(\mathbf{z}, \hat{\theta}_{(\mathbf{j})}, \hat{\theta}_{(\mathbf{j+1})}, \mathbf{x}))\middle|\hat{\theta}_{(j)}, \mathbf{x}\right] \leq log\left[E_{\hat{\theta}_{(j)}}\left[h(\mathbf{z}, \hat{\theta}_{(\mathbf{j})}, \hat{\theta}_{(\mathbf{j+1})}, \mathbf{x})\middle|\hat{\theta}_{(j)}, \mathbf{x}\right]\right] \tag{8}$$

6 Problem 5.20

Søren Hvidkjær

Consider an AR(p) model

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \epsilon_k, \tag{9}$$

with $\epsilon_k \sim \mathcal{N}(0, \sigma)$, observed for $t = p+1, \ldots, m$. The future values $X_{m+1}, \ldots, X_n$ are considered as missing data. The initial values $x_1, \ldots, x_p$ are taken to be zero.

(a) Give the expression of the observed and complete likelihoods.

Solving this by one-step ahead predictions, we note that conditional on the first $t-1$ observations, the tth observation is Gaussian with mean

$$E(X_t | x_{t-1}, \ldots, x_1) = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p},$$

and variance σ^2. Hence, the distribution is only dependent on the last p observations and we can write the density of the tth observation as

$$f(x_t | x_{t-1}, \ldots, x_1, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2}{2\sigma^2}\right],$$

where $\theta = [\phi_1, \ldots, \phi_p, \sigma]'$. The observed data likelihood is then

$$L(\theta|\mathbf{x}) = \left(2\pi\sigma^2\right)^{-(m-p)/2} \prod_{t=p+1}^{m} \exp\left[\frac{-(x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2}{2\sigma^2}\right].$$

Likewise, the complete data likelihood, corresponding to having observed the full sample, is

$$L^c(\theta|\mathbf{x}, \mathbf{z}) = \left(2\pi\sigma^2\right)^{-(n-p)/2} \prod_{t=p+1}^{n} \exp\left[\frac{-(x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2}{2\sigma^2}\right]. \tag{10}$$

(b) Give the conditional maximum likelihood estimators of ϕ, σ and $\mathbf{z} = (X_{m+1}, \ldots, X_n)$, that is the maximum likelihood estimator when the two other parameters are fixed.

Taking logs in (10), we get

$$\begin{aligned} \log L^c(\theta|\mathbf{x}, \mathbf{z}) &= -\frac{n-p}{2}\log 2\pi - \frac{n-p}{2}\log\sigma^2 \\ &\quad - \sum_{t=p+1}^{n} \left[\frac{(x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2}{2\sigma^2} \right]. \end{aligned} \tag{11}$$

To get the MLE of ϕ, maximizing (11) corresponds to minimizing

$$\sum_{t=p+1}^{n} (x_t - \sum_{j=1}^{p} \phi_j x_{t-j})^2,$$

which is the OLS estimator when regressing x_t on p of its own lagged values.
To get MLE of σ:

$$\begin{aligned} \frac{\partial \log L^c}{\partial \sigma^2} &= -\frac{n-p}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{t=p+1}^{n} (x_t - \sum_{j=1}^{p} \phi_j x_{t-j})^2 = 0 \\ &\Updownarrow \\ \hat{\sigma}^2 &= \frac{1}{n-p} \sum_{t=p+1}^{n} (x_t - \sum_{j=1}^{p} \phi_j x_{t-j})^2 \end{aligned}$$

To get MLE of $\mathbf{z}$, minimize

$$\sum_{t=m+1}^{n} (x_t - \sum_{j=1}^{p} \phi_j x_{t-j})^2,$$

which we can set to zero by recursively setting x_t equal to its expected value, given $\mathbf{x}$.

(c) Detail the E and M steps of the EM algorithm in this setup, when applied to the future values of $\mathbf{z}$ and when σ is fixed.

It is useful to write the pth order difference equation in (9) as the first order difference equation

$$\xi_t \quad = \quad F\xi_{t-1} + v_t,$$

defined by (with $p = 2$ for ease of notation)

$$\begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$$

Then we can write any $\xi_t,\ t = m+1, \ldots$ in terms of the innovations and ξ_m, and the first element in ξ_t can be written as

$$x_t = \sum_{j=1}^{p} f_{1j}^{(t-m+1)} x_{m-j} + \sum_{j=0}^{t-m} f_{11}^{(j)} \epsilon_{t-j}, \tag{12}$$

where $f_{1,a}^{(b)}$ denotes the $(1,a)$th element in F^b.

Taking expected values in (11) with respect to $k(\mathbf{z}|\theta_0, \mathbf{x})$ and disregarding constants:

$$\begin{aligned} Q(\theta|\theta_0, \mathbf{x}) \quad = \quad & -\frac{1}{2\sigma^2} \sum_{t=p+1}^{m} (x_t - \sum_{j=1}^{p} \phi_j x_{t-j})^2 \\ & -\frac{1}{2\sigma^2} \sum_{t=m+1}^{n} E[(x_t - \sum_{j=1}^{p} \phi_j x_{t-j})^2 | \theta_0, \mathbf{x}], \end{aligned}$$

Then maximize Q in θ to get

$$\begin{aligned} & \sum_{t=p+1}^{m} x_{t-i}(x_t - \sum_{j=1}^{p} \phi_j x_{t-j}) \\ & + \sum_{t=m+1}^{n} (E[x_{t-i}(x_t - \sum_{j=1}^{p} \phi_j x_{t-j}))|\theta_0, \mathbf{x}] = 0, \end{aligned}$$

for $i = 1, \ldots, p$. Using (12) to solve for the expected value to get

$$\sum_{t=p+1}^{m} x_{t-i}(x_t - \sum_{j=1}^{p} \phi_j x_{t-j})$$
$$+ \sum_{t=m+1}^{n} (K_1 + \sum_{j=0}^{t-i-m} f_{11}^{2(j)} \sigma^2 - \sum_{l=1}^{p} \phi_l (K_{2l} + \sum_{j=0}^{t-m-(i \vee l)} \sigma^2)) = 0, \qquad (13)$$

where K_c are constants. The EM algoritm is then, at ith step:

1. Obtain numerical values in (13) using $\theta_{(i)}$.

2. Solve for $\theta_{(i+1)}$.

(d) Compare with the Kalman filter prediction.

Equation (9) can be written in state-space form as

$$\begin{aligned} \xi_{t+1} &= F\xi_t + v_{t+1} \\ x_t &= H'\xi_t. \end{aligned}$$

For ease of notation, let $p = 2$. Then the state equation is defined as before and the observation equation is the identity

$$x_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}.$$

Let $\hat{\xi}_{t+1|t}$ and $P_{t+1|t}$ be the conditional mean and variance of ξ_{t+1}, given by

$$\begin{aligned} P_{t+1|t} &= \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \\ \hat{\xi}_{t+1|t} &= F\hat{\xi}_{t|t-1} + FP_{t+1|t}H(H'P_{t|t-1}H)^{-1}(x_t - H'\hat{\xi}_{t|t-1}) \end{aligned}$$

$$
\begin{aligned}
&= F\hat{\xi}_{t|t-1} + FH(x_t - H'\hat{\xi}_{t|t-1}) \\
\hat{\xi}_{p+1|p} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\end{aligned}
$$

These equations define, with the corresponding identity for $\hat{x}_{t+1|t}$, the Kalman filter.

The distribution of x_t is then

$$x_t|x_{t-1},\ldots,x_1 \sim \mathcal{N}(H'\hat{\xi}_{t|t-1}, \sigma^2),$$

and the corresponding sample log likelihood is (without constants)

$$-\frac{1}{\sigma^2}\sum_{t=p+1}^{m}(x_t - H'\hat{\xi}_{t|t-1})^2.$$

In general, we would maximize the log likelihood by numerical methods to get an estimate of F (and H). However in the AR(p) case, $H'\hat{\xi}_{t|t-1}$ is simply $\sum_{j=1}^{p}\phi_j x_{t-j}$, and we can do OLS. It is then straightforward to use the estimated parameter values for forecasting x_t for $t = m+1,\ldots,n$.

In both the EM algoritm and the Kalman filter we generally use iterative methods, but we're solving a simple problem at each step in the EM algoritm, while the iterations in the Kalman filter typically requires calculation the inverse of matrix of the second order partials. Further, the likelihood always increases in the EM algoritm, while this is not nescessarily the case in the numerical optimization. On the other hand, the likelihood may only increase very slowly in the EM algoritm.

7 Problem 5.22

Jianghong Wang

We observe independent Bernoulli variables $X_1,\ldots,X_n$, which depend on unobservable variables Z_i distributed independently as $N(\zeta,\sigma^2)$ where

$$X_i = \begin{cases} 0, & \text{if } Z_i \le u; \\ 1, & \text{if } Z_i \ge u. \end{cases}$$

Assuming that u is known, we are interested in obtaining ML estimates of ζ and σ^2.

(a) : Show that the likelihood function is

$$p^S(1-p)^{n-S}$$

where $S = \sum x_i$ and

$$p = P(z_i > u) = \Phi(\frac{\zeta - u}{\sigma})$$

Proof:

$$\begin{aligned} P(X_i = 1) &= P(Z_i > u) = P(\frac{Z_i - \zeta}{\sigma} > \frac{u - \zeta}{\sigma}) \\ &= \Phi(-\frac{u - \zeta}{\sigma}) = \Phi(\frac{\zeta - u}{\sigma}). \end{aligned}$$

Let

$$p = \Phi(\frac{\zeta - u}{\sigma})$$

then

$$P(X_i = 0) = 1 - p$$

So, the likelihood function is

$$\begin{aligned} P(X_1, \ldots, X_n) &= \Pi_{i=1}^{S} P(X_i = 1) \Pi_{i=1}^{n-s} P(X_i = 0) \\ &= p^S(1-p)^{n-S} \end{aligned}$$

(b): If we consider $z_1, \ldots, z_n$ to be the complete data, show that the complete data likelihood is

$$\Pi_{i=1}^{n} \frac{1}{\sqrt{(2\pi)}\sigma} exp(-\frac{1}{2\sigma^2}(z_i - \zeta)^2)$$

and the expected complete data log-likelihood is

$$-\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (E(Z_i^2|x_i) - 2\zeta E(Z_i|x_i) + \zeta^2)$$

Proof: $Z_i \sim N(\zeta, \sigma^2)$, so

$$f(z_i|\zeta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp - \frac{1}{2\sigma^2}(z_i - \zeta)^2$$

then the likelihood function of $Z_1, \ldots, Z_n$ is

$$\begin{aligned} f(Z_1, \ldots, Z_n|\zeta, \sigma^2) &= \Pi_{i=1}^{n} f(Z_i|\zeta, \sigma^2) \\ &= \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2\sigma^2}(z_i - \zeta)^2) \end{aligned}$$

The log-likelihood function is

$$\begin{aligned} log f(z_1, \ldots, z_n) &= log \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2\sigma^2}(z_i - \zeta)^2) \\ &= \sum_{i=1}^{n} [log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}(z_i - \zeta)^2] \\ &= -\frac{1}{2} \sum_{i=1}^{n} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \zeta)^2 \\ &= -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \zeta)^2 \end{aligned}$$

So, the expected complete data log-likelihood is

$$E log f(z_1, \ldots, z_i) = E[-\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \zeta)^2]$$

where the expectation is with respect to $k(Z|\zeta, \sigma^2, X)$. Then,

$$
\begin{aligned}
Elogf(z_1,\ldots,z_n) &= E[-\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i-\zeta)^2] \\
&= -\frac{n}{2}log(2\pi\sigma^2) - E[\frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i-\zeta)^2] \\
&= -\frac{n}{2}log(2\pi\sigma^2) - \int\frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i-\zeta)^2k(z_i|\zeta,\sigma^2,x_i)dz_i \\
&= -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\int(z_i-\zeta)^2k(z_i|x_i)dz_i \\
&= -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\int(z_i^2+\zeta^2-2\zeta z_i)k(z_i|x_i)dz_i \\
&= -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[\int z_i^2k(z_i|x_i)dz_i - \int 2\zeta z_ik(z_i|x_i)dz_i + \int\zeta^2k(z_i|x_i)dz_i] \\
&= -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[E(z_i^2|x_i) - 2\zeta E(z_i|x_i) + \zeta^2]
\end{aligned}
$$

(c): Show that the EM sequence is given by

$$\zeta_{(j+1)} = \frac{1}{n}\sum_{i=1}^{n}t_i(\zeta_j, \sigma_j^2)$$

$$\sigma^2_{(j+1)} = \frac{1}{n}[\sum_{i=1}^{n}v_i(\zeta_{(j)},\sigma^2_{(j)}) - \frac{1}{n}(\sum_{i=1}^{n}t_i(\zeta_{(j)},\sigma^2_{(j)}))^2]$$

where,

$$t_i(\zeta,\sigma^2) = E[Z_i|x_i,\zeta,\sigma^2]$$

$$v_i(\zeta,\sigma^2) = E[Z_i|x_i,\zeta,\sigma^2]$$

Proof: From (b), the expected log-likelihood function is

$$Q((\zeta,\sigma^2)|(\zeta_0,\sigma_0^2),X) = -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(E(z_i|x_i) - 2\zeta E(z_i|x_i) + \zeta^2)$$

By EM algorithm, the EM sequence $\theta_{(j+1)}$ is defined as the value of θ maximazing $Q(\theta|\theta_{(j)}, X)$, that is

$$Q(\theta_{(j+1)}|\theta_{(j)}, X) = maxQ(\theta|\theta_{(j)}, X)$$

Here, $\theta = (\zeta, \sigma^2)$

Let $\frac{\partial Q}{\partial \zeta} = 0$ and get

$$\begin{aligned}
\frac{\partial Q}{\partial \zeta} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2E(z_i|x_i) + 2\zeta) = 0 \\
&\Leftrightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(E(z_i|x_i) - \zeta) = 0 \\
&\Leftrightarrow \sum_{i=1}^{n} E(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)}) = \sum_{i=1}^{n}\zeta = n\zeta_{(j+1)}
\end{aligned}$$

$\Rightarrow$

$$\zeta_{(j+1)} = \frac{1}{n}\sum_{i=1}^{n} E(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)})$$

Let $\frac{\partial Q}{\partial \sigma^2} = 0$ and get

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma^2} &= -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}[E(z_i^2|x_i, \zeta_{(j)}, \sigma^2_{(j)}) - 2\zeta E(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)}) + \zeta^2_{(j+1)}] = 0 \\
&\Leftrightarrow -\frac{n}{\sigma^2} + \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}[E(z_i^2|x_i, \zeta_{(j)}, \sigma^2_{(j)}) - 2\zeta_{(j+1)}E(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)}) + \zeta^2_{(j+1)}] = 0
\end{aligned}$$

$\Rightarrow$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}[E(z_i^2|x_i, \zeta_{(j)}, \sigma^2_{(j)}) - 2\zeta_{(j+1)}E(\zeta_i|x_i, \zeta_{(j)}, \sigma^2_{(j)}) + \zeta^2_{(j+1)}]$$

Relace $\zeta_{(j+1)}$ with $\frac{1}{n}\sum_{i=1}^{n} E(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)})$, $\Rightarrow$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} E(z_i^2|x_i, \zeta_{(j)}, \sigma^2_{(j)}) - \frac{2}{n}\sum_{i=1}^{n} E(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)})\frac{1}{n}\sum_{i=1} nE(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)}) + \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{n}\sum_{i=1}^{n} E(z_i|x_i, \zeta_{(j)}, \sigma^2_{(j)}))^2$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} E(z_i^2|x_i,\zeta_{(j)},\sigma^2_{(j)}) - \frac{2}{n}(\sum_{i=1}^{n} E(z_i|x_i,\zeta_{(j)},\sigma^2_{(j)}))^2 + \frac{1}{n^2}(\sum_{i=1}^{n} E(z_i|x_i,\zeta_{(j)},\sigma^2_{(j)}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} E(z_i^2|x_i,\zeta_{(j)},\sigma^2_{(j)}) - \frac{1}{n^2}(\sum_{i=1}^{n} E(z_i|x_i,\zeta_{(j)},\sigma^2_{(j)}))^2
\end{aligned}
$$

Thus,

$$\sigma^2_{(j+1)} = \frac{1}{n}[\sum_{i=1}^{n} E(z_i^2|x_i,\zeta_{(j)},\sigma^2_{(j)}) - \frac{1}{n}(\sum_{i=1}^{n} E(z_i|x_i,\zeta_{(j)},\sigma^2_{(j)})]$$

Let

$$t_i(\zeta,\sigma^2) = E(z_i^2|x_i,\zeta,\sigma^2)$$

$$v_i(\zeta,\sigma^2) = E(z_i^2|x_i,\zeta,\sigma^2)$$

then, the EM sequence is given by

$$\zeta_{(j+1)} = \frac{1}{n}\sum_{i=1}^{n} t_i(\zeta_{(j)},\sigma^2_{(j)})$$

$$\sigma^2_{(j)} = \frac{1}{n}[\sum_{i=1}^{n} v_i(\zeta_{(j)},\sigma^2_{(j)}) - \frac{1}{n}(\sum_{i=1}^{n} t_i(\zeta_{(j)},\sigma^2_{(j)}))^2]$$

8 Problem 5.23

Baojun Song

(b).

$$\hat{Q}(\theta|\hat{\theta}_{(j)},x) = \frac{1}{m}\sum_{i=1}^{m} logL^c(\theta|x,z_i)$$

where z_i are iid from $k(z|x,\hat{\theta}_{(j)})$. By the *law of large number*,

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} logL^c(\theta|x,z_i) = E_{\hat{\theta}_{(j)}}[logL^c(\theta|x,z)|\hat{\theta}_{(j)},x] = Q(\theta|\hat{\theta}_{(j)},x)$$

where the expectation is with respect to $k(z|x, \theta_{(j)})$.Thus

$$\lim_{m\to\infty} \hat{Q}(\theta|\hat{\theta}_{(j)}, x) = Q(\theta|\hat{\theta}_{(j)}, x).$$

(c).

$$|\hat{Q}(\theta|\hat{\theta}_{(j)}^{(m)}, x) - Q(\theta|\hat{\theta}_{(j)}, x)| = |\hat{Q}(\theta|\hat{\theta}_{(j)}^{(m)}, x) - \hat{Q}(\theta|\hat{\theta}_{(j)}, x) + \hat{Q}(\theta|\hat{\theta}_{(j)}, x) - Q(\theta|\hat{\theta}_{(j)}, x)|$$

$$\leq |\hat{Q}(\theta|\hat{\theta}_{(j)}^{(m)}, x) - \hat{Q}(\theta|\hat{\theta}_{(j)}, x)| + |\hat{Q}(\theta|\hat{\theta}_{(j)}, x) - Q(\theta|\hat{\theta}_{(j)}, x)|.$$

Based on part (a) and part b, we have

$$\lim_{m\to\infty} (\hat{Q}(\theta|\hat{\theta}_{(j)}^{(m)}, x) - \hat{Q}(\theta|\hat{\theta}_{(j)}, x) = 0,$$

and

$$\lim_{m\to\infty} \hat{Q}(\theta|\hat{\theta}_{(j)}, x) - Q(\theta|\hat{\theta}_{(j)}, x) = 0,$$

which yields

$$\lim_{m\to\infty} \hat{Q}(\theta|\hat{\theta}_{(j)}^{(m)}, x) - Q(\theta|\hat{\theta}_{(j)}, x) = 0$$

as desired.

9 Problem 5.25

Dennis Carlson

For the situation of Example 5.3.7, data $(x_1, x_2, x_3, x_4) = (125, 18, 20, 34)$ are collected.

(a) Use the EM algorithm to find the MLE of θ.

(b) Use the Monte Carlo EM algoritm to find the MLE of θ. Compare your results to those of part (a).

Solutions:

We have that

$$\begin{aligned} f(x,z|\theta) &= \frac{n!}{z_1!z_2!x_2!x_3!x_4!}\left(\frac{1}{2}\right)^{z_1}\left(\frac{\theta}{4}\right)^{z_2}\left(\frac{(1-\theta)}{4}\right)^{x_2}\left(\frac{(1-\theta)}{4}\right)^{x_3}\left(\frac{\theta}{4}\right)^{x_4} \\ &= \frac{n!}{z_1!z_2!x_2!x_3!x_4!}\left(\frac{1}{2}\right)^{z_1}\left(\frac{\theta}{4}\right)^{z_2+x_4}\left(\frac{(1-\theta)}{4}\right)^{x_2+x_3} \\ &\propto \theta^{z_2+x_4}(1-\theta)^{x_2+x_3} \end{aligned}$$

and,

$$\begin{aligned} g(x|\theta) &= \frac{n!}{x_1!x_2!x_3!x_4!}\left(\frac{1}{2}+\frac{\theta}{4}\right)^{x_1}\left(\frac{(1-\theta)}{4}\right)^{x_2}\left(\frac{(1-\theta)}{4}\right)^{x_3}\left(\frac{\theta}{4}\right)^{x_4} \\ &= \frac{n!}{x_1!x_2!x_3!x_4!}\left(\frac{2+\theta}{4}\right)^{x_1}\left(\frac{\theta}{4}\right)^{x_4}\left(\frac{(1-\theta)}{4}\right)^{x_2+x_3} \\ &\propto (2+\theta)^{x_1}\theta^{z_2+x_4}(1-\theta)^{x_2+x_3} \end{aligned}$$

Moreover, with some trivial calculations, it can be shown that:

$$\begin{aligned} k(z_2|\theta,x) &= \frac{x_1!}{z_2!(x_1-z_2)!}\left(\frac{\theta}{(2+\theta)}\right)^{z_2}\left(\frac{2}{(2+\theta)}\right)^{x_1-z_2} \\ &\sim \text{Bin}(x_1, \frac{\theta}{(2+\theta)}) \end{aligned}$$

Th expected complete log-likelihood function is

$$\begin{aligned} E_{\theta_0}[(Z_2+x_4)\log\theta+(x_2+x_3)\log(1-\theta)] &= [E_{\theta_0}(Z_2)+E_{\theta_0}(x_4)]\log\theta+[E_{\theta_0}(x_2)+E_{\theta_0}(x_3)]\log(1-\theta) \\ &= \left(\left(\frac{\theta_0}{(2+\theta_0)}\right)x_1+x_4\right)\log\theta+(x_2+x_3)\log(1-\theta) \end{aligned}$$

We want to maximize this: differentiating, and setting equal to zero yields that θ is maximized at

$$\theta_1 = \frac{\left(\frac{\theta}{(2+\theta)}\right) x_1 + x_4}{\left(\frac{\theta}{(2+\theta)}\right) x_1 + x_2 + x_3 + x_4}$$

Using the EM algorithm yields the MLE of $\theta^* = 0.62682149787098$, after just over 15 iterations.

(b) We need to estimate $E_{\theta_0}(Z_2)$, which, by Monte Carlo methods is estimated by $z_m = \sum_{i=1}^{m} z_i$, where $z_i \sim \text{Bin}(x_1, \frac{\theta}{(2+\theta)})$. Attempting to find the MLE is more difficult in this case. If we choose a constant m, we get a certain degree of precision, and then jump around that precision due to the variability in the Monte Carlo sample. Taking our m to be exponential (as recommended by the text), yields a better result, although to get the precision desired would require a computer more powerful than the one currently at my disposal.

10 Problem 5.27

Daniel Fink Mark Release Recapture(MRR) Problem based on Ex 5.3.8 Assume that animal i can be captured at time j in one of m locations/states. The location/state of animals is a multinomial random variable,

$$H \sim M_m(\theta_1, ..., \theta_m)$$

and we model the number of animals caught in location/state k as a binomial random variable

$$X \sim B(p_k).$$

For this problem, we have 18 individuals, 5 time periods, and 5 locations. The data for the first individual is

$$H = (2, 2, 0, 0, 0)$$

$$X = (1, 1, 0, 0, 0)$$

The Data for this problem are presented below where each row represents one individuals capture history over the six time periods:

$$\left(\begin{pmatrix} 2 & 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 \\ 4 & 1 & 1 & 2 & 0 \\ 4 & 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 5 & 0 \\ 4 & 2 & 0 & 0 & 0 \\ 5 & 5 & 5 & 5 & 0 \\ 2 & 2 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 1 & 0 & 0 \\ 4 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 4 & 0 \\ 2 & 2 & 0 & 0 & 0 \end{pmatrix} \right)$$

The likelihood function is

$$\begin{aligned} L(\theta, p|x, y) &= \\ &= \sum_z L(\theta, p|x, y, z) \\ &= \sum_z \prod_{k=1}^{5} p_k^{(\text{no ind cap k})} (1 - p_k)^{nt - (\text{no ind cap k})} \theta_k^{(\text{no ind in k})} \end{aligned}$$

where the sum over z represents the expectation over all the states that could be visited.

10.1 A

Using the MCEM algorithm calculate the MLE's for θ and p.

This calculation is done by

(0) Choose a initial value of θ^0

(1) imputing the expected missing data values based on θ^0 into the data matrix

(2) Calculate θ^1 based on the 'full' data matrix

(3) Goto (1) until convergence is achieved

The results of this simulation after 10000 iterations are

$\theta = (0.44514, 0.32783, 0.00133, 0.10234, 0.12346)$
Proportion of individuals captured in each state is
$(0.55293, 0.54824, 0.00000, 0.56812, 0.56079)$.

10.2 B

Assume that location 5 represents the death state. Use the MCEM algorithm to calculate the MLE's and compare these results to those in part (a).
This calculation is performed as above with the following constraint: state 5 is an absorbing state for each individual. Thus, it is expected that the probability of being in state five should be greater than in part A. As a consequence of this, the probability of being in all the other states will decrease relative to part a. The results of this simulation after 10000 iterations are
$\theta = (0.34640, 0.25281, 0.00043, 0.07089, 0.32957)$
Proportion of individuals captured in each state is
$(0.71228, 0.71223, 0.00000, 0.82806, 0.20697)$.