统计中的计算方法第二次作业（修订）

傅长青　信息与计算科学　13300180003

July 4, 2016

1. 状态转移矩阵 $A = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}$ **状态显示矩阵** $B = \begin{pmatrix} 0.5 & 0.5 \\ 0.8 & 0.2 \end{pmatrix}$ **计算 HTTT 出现的概率。**

解：默认 $\mathbb{P}(s_1 = F) = \mathbb{P}(s_1 = B) = 0.5$, 归纳得到每一步的隐状态概率

$$\pi^{(k+1)} = \pi^{(k)} A \equiv (0.5, 0.5),\ \forall k \in \mathbb{Z}^+$$

已知 $x_1, \ldots, x_n$, 设

$$f_l(i) := \mathbb{P}(x_1, \ldots, x_i, s_i = l)$$

第一步为 H, 隐状态为 F,H 概率分别为

$$f_F(1) := \mathbb{P}(x_1, s_1 = F) = 0.5 \times 0.5 = 0.25 \quad f_B(1) := \mathbb{P}(x_1, s_1 = B) = 0.8 \times 0.5 = 0.4$$

$$i.e.\ f(1) = (0.25, 0.4)$$

由公式

$$f_l(i) = e_l(x_i) \sum_{k=F,B} f_k(i-1) a_{kl}$$

$$i.e.\ f(i) = B(:, x_i). * Af(i-1)$$

(此处符号使用 Matlab 习惯) 依次得到显状态为 HT,HTT,HTTT, 最后一步隐状态为 F,H 的概率为

$$f_F(2) = 0.155, f_B(2) = 0.068$$

$$f_F(3) = 0.0601, f_B(3) = 0.02056$$

$$f_F(4) = 0.022142, f_B(4) = 0.0072752$$

结果为

$$\mathbb{P}(X = HTTT) = f_F(4) + f_B(4) = 0.0294172$$

用 R 语言实现：

```
forward <- function(x, pi0, T, E, step = length(x)){
##given hmm parameters, compute the prob of showing emmision states x (
    up to "step" steps) with hidden status s(s is a vector) at this step
    , using the forward algorithm
##output: P(x(1),...x(step),s(step))
  if (step > length(x))
    rstop("step␣is␣greater␣than␣length␣of␣x")
  f = pi0 * E[,x[1]]
  for (i in 2:step)
    f = E[,x[i]]*(f%*%T)
  return(f)
```

```
}
evaluation <- function(x, pi0, T, E){
##P(x)
  sum(forward(x, pi0, T, E, length(x)))
}
x = c(1,2,2,2)
pi0 = c(0.5,0.5)
T = matrix(c(0.6,0.4,0.4,0.6),2)
E = matrix(c(0.5,0.8,0.5,0.2),2)
prob1 = evaluation(x,pi0,T,E)
```

结果：

```
> prob1
[1] 0.0294172
```

2. 每个隐状态为 B 的概率。

用 R 语言实现：

```
backward <- function(x,pi0,T,E,step = length(x)){
  ##output: P(x(i+1),...|s(i))
  n = length(x)
  if(step > n)
    return("errror:step is greater than length of x")
  if(step == n)
    return(c(1,1))
  f = t(T %*% E[,x[n]])
  i = n-1
  while(i >= step){
    f = f*t(T%*%E[,x[i]])
    i=i-1
  }
  return(f)
}
forward_backward <-function(x,pi0,T,E,step = length(x)){
  forward(x,pi0,T,E,step)*backward(x,pi0,T,E,step)
}
prob2 = lapply(1:4,forward_backward,x=x,pi0=pi0,T=T,E=E)
```

结果：

```
> prob2
[[1]]
           [,1]        [,2]
[1,] 0.00204464 0.001832491

[[2]]
           [,1]        [,2]
[1,] 0.00850516 0.002228224

[[3]]
           [,1]        [,2]
[1,] 0.00867844 0.002105344

[[4]]
         [,1]      [,2]
[1,] 0.022142 0.0072752
```

结果解释:list 中每一个元素代表每一步各状态出现的概率。

3. 隐状态为 FFBB 的概率

用 R 语言实现：

```
decoding <- function(x,pi0,T,E,s){
##P(s|x)
  f = function(i){forward_backward(x,pi0,T,E,i)[s[i]]}
  exp(sum(log(sapply(1:length(x),f)))) / evaluation(x,pi0,T,E)
}
s = c(1,1,2,2)
prob3 = decoding(x,pi0,T,E,s)
```

结果：

```
> prob3
[1] 9.054534e-09
```

结果解释:

$$\begin{aligned}\mathbb{P}(s|x) &= \frac{\prod_i \mathbb{P}(s_i,x)}{\mathbb{P}(x)} \\ &= 0.00204464 * 0.00850516 * 0.002105344 * 0.0072752/0.0294172 \\ &= 9.054534e-09\end{aligned}$$

4. 求最优的隐状态路径。

用 R 语言实现：

```
viterbi <- function(x,pi0,T,E){
##optimized state sequence given observed sequence
##argmax_s{P(s|x)}
  L = length(x) ##length of observed sequence
  S = dim(T)[1] ##S states in total
  s_ = rep(0,L) #initialize
  v_ = matrix(0,L+1,S)##v_k(i)=max_(s_1,...s_i-1)P(s_1,...s_i=k,x_1,...x
     _i)
  v_[1,1] = 1;
  for(i in 1:L){
    m = which.max(v_[i,]) ##max(v)=v[m]
    v_[i+1,] = v_[i,m] * T[m,] * E[,x[i]]
  }
  s_[L] = which.max(v_[L+1,])
  i=L
  while(i > 1){
    s_[i-1]=which.max(v_[i,] * T[,s_[i]])
    #print(v_[i,] * T[,s_[i]])
    i = i-1
  }
  s_
}
prob4 = viterbi(x,pi0,T,E)
```

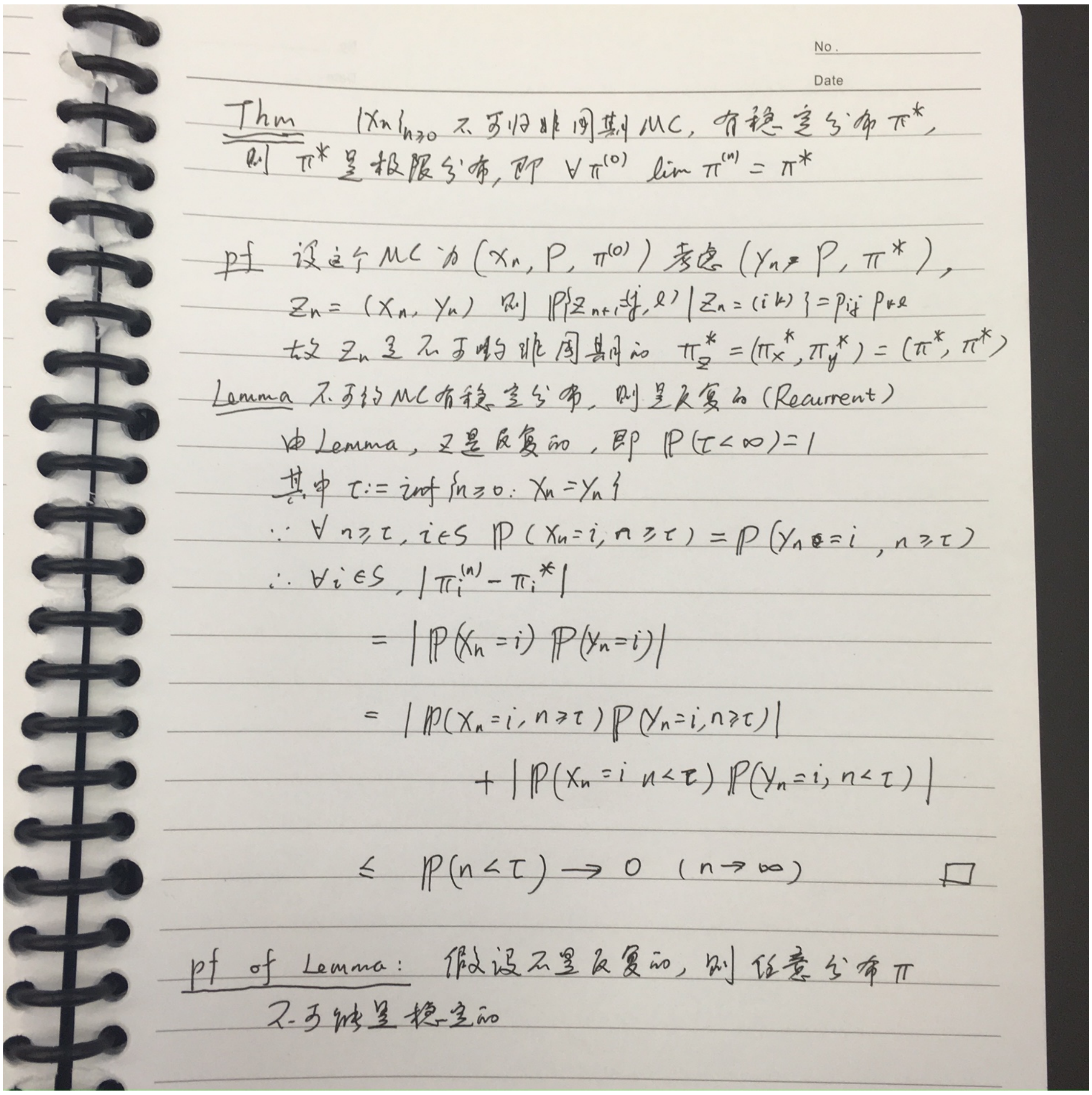
结果：

```
> prob4
[1] 1 1 1 1
```

结果解释:1 为 F，2 为 T，最优隐状态链为 FFFF。

5. 见第三次作业 revision.pdf

6. 如图



Thm $\{X_n\}_{n\geq 0}$ 不可约非周期 MC，有稳定分布 π^*，
则 π^* 是极限分布，即 $\forall \pi^{(0)}$ $\lim \pi^{(n)} = \pi^*$

pf 设这个 MC 为 $(X_n, P, \pi^{(0)})$ 考虑 (Y_n, P, π^*)，
$Z_n = (X_n, Y_n)$ 则 $\mathbb{P}\{Z_{n+1}=(j,l) \mid Z_n=(i,k)\} = p_{ij}p_{kl}$
故 Z_n 是不可约的非周期的 $\pi_Z^* = (\pi_X^*, \pi_Y^*) = (\pi^*, \pi^*)$
Lemma 不可约的 MC 有稳定分布，则是反复的 (Recurrent)
由 Lemma，Z 是反复的，即 $\mathbb{P}(\tau<\infty)=1$
其中 $\tau := \inf\{n\geq 0: X_n = Y_n\}$
$\because \forall n\geq\tau, i\in S$ $\mathbb{P}(X_n=i, n\geq\tau) = \mathbb{P}(Y_n=i, n\geq\tau)$
$\therefore \forall i\in S, |\pi_i^{(n)} - \pi_i^*|$

$= |\mathbb{P}(X_n=i)\ \mathbb{P}(Y_n=i)|$

$= |\mathbb{P}(X_n=i, n\geq\tau)\ \mathbb{P}(Y_n=i, n\geq\tau)|$
$+ |\mathbb{P}(X_n=i\ n<\tau)\ \mathbb{P}(Y_n=i, n<\tau)|$

$\leq \mathbb{P}(n<\tau) \to 0 \quad (n\to\infty)$ □

pf of Lemma: 假设不是反复的，则任意分布 π 不可能是稳定的