

esercizi

Esercizio 1

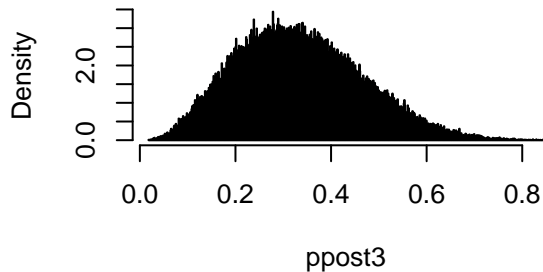
Si svolgano le simulazioni dell'esperimento Bayes'billard come riportato nel file delle dispense della parte delle applicazioni considerando i seguenti punti: ##1.1 Si disegnino in un'unica finestra grafica gli istogrammi delle densità condizionate di $(p|X = 3)$ e $(p|x = 4)$, $(p|x = 5)$ e $(p|x = 6)$ (si inseriscano i titoli ai grafici e si salvi la figura);

```
nsim <- 10^6
n <- 10
p <- runif(nsim)

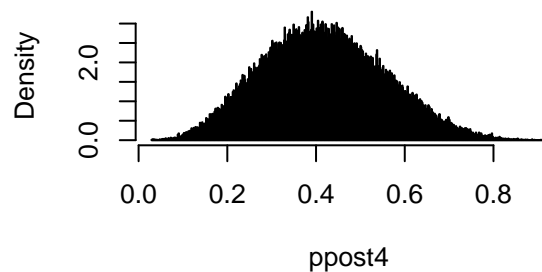
xcond <- rbinom(nsim, size = n, prob = p)
ppost3 <- p[xcond==3]
ppost4 <- p[xcond==4]
ppost5 <- p[xcond==5]
ppost6 <- p[xcond==6]

par(mfrow=c(2,2))
hist(ppost3, col="pink", breaks = 1000, freq = F)
hist(ppost4, col="pink", breaks = 1000, freq = F)
hist(ppost5, col="pink", breaks = 1000, freq = F)
hist(ppost6, col="pink", breaks = 1000, freq = F)
```

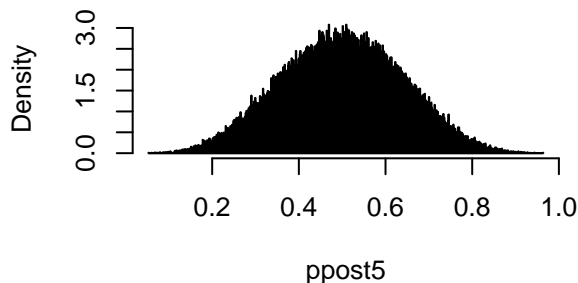
Histogram of ppost3



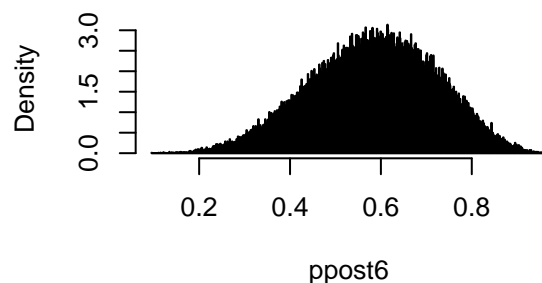
Histogram of ppost4



Histogram of ppost5



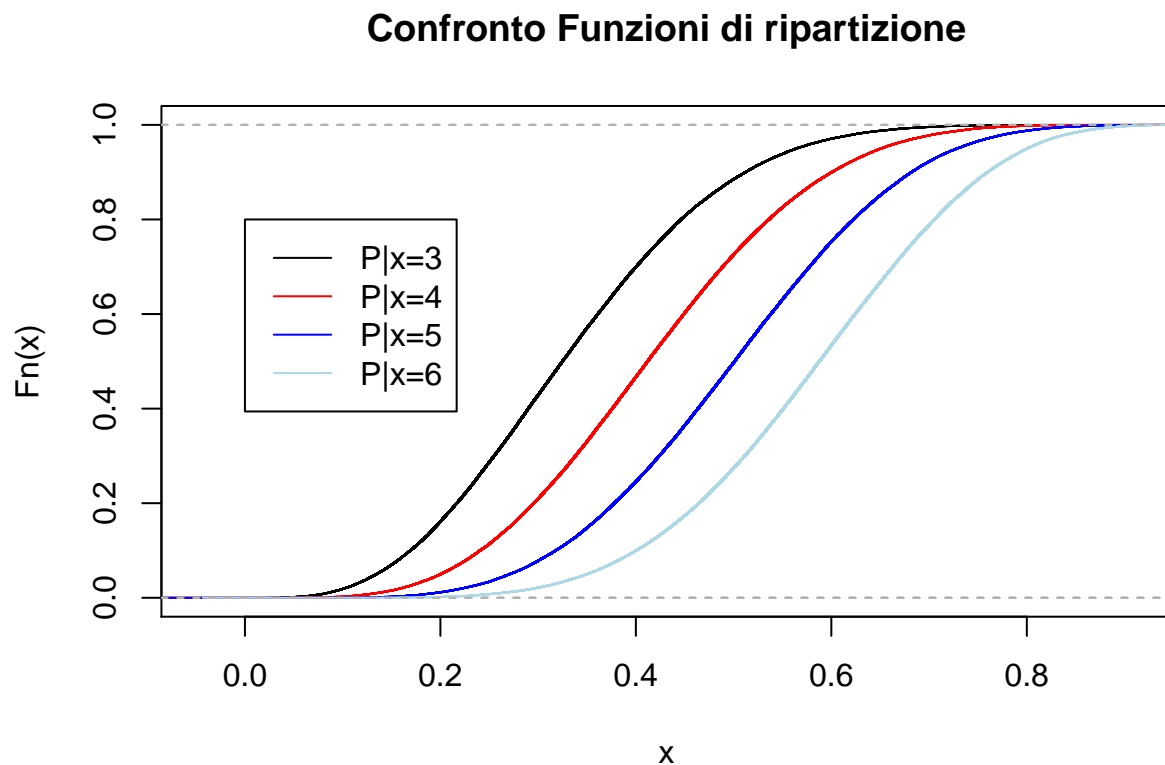
Histogram of ppost6



1.2

Si disegnino le funzioni di ripartizione empiriche delle osservazioni del punto precedente in un'unica finestra grafica e si riportino insieme alla legenda;

```
par(mfrow=c(1,1))
plot(ecdf(ppost3), main="Confronto Funzioni di ripartizione")
plot(ecdf(ppost4),add=T, col="red")
plot(ecdf(ppost5),add=T, col="blue")
plot(ecdf(ppost6),add=T, col="lightblue")
legend(0,0.8,c("P|x=3","P|x=4", "P|x=5", "P|x=6"),
      col=c("black","red","blue","lightblue"), lty=1)
```



1.3

Si calcolino in base alle conoscenze teoriche i parametri delle distribuzioni disegnate al punto 1

```
#beta 4,8  beta 5,7  beta 6,6  beta 7,5
```

1.4

Si riporti figura di cui al punto 1 in cui viene anche sovrapposta all'istogramma la densità calcolata al punto precedente (curve(d..), add=TRUE)

```

par(mfrow=c(2,2))
hist(ppost3, col="pink", breaks = 1000, freq = F)
curve(dbeta(x,4,8), lwd=3, col=2, main="Dist f(p|x=3)", add=T)

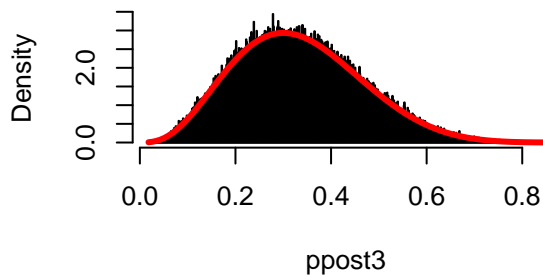
hist(ppost4, col="pink", breaks = 1000, freq = F)
curve(dbeta(x,5,7), lwd=3, col=2, main="Dist f(p|x=4)", add=T)

hist(ppost5, col="pink", breaks = 1000, freq = F)
curve(dbeta(x,6,6), lwd=3, col=2, main="Dist f(p|x=5)", add=T)

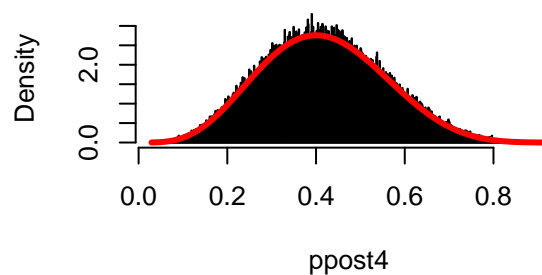
hist(ppost6, col="pink", breaks = 1000, freq = F)
curve(dbeta(x,7,5), lwd=3, col=2, main="Dist f(p|x=6)", add=T)

```

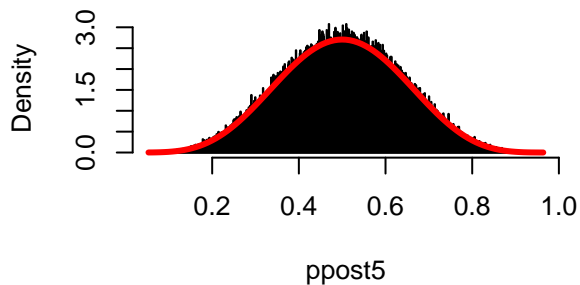
Histogram of ppost3



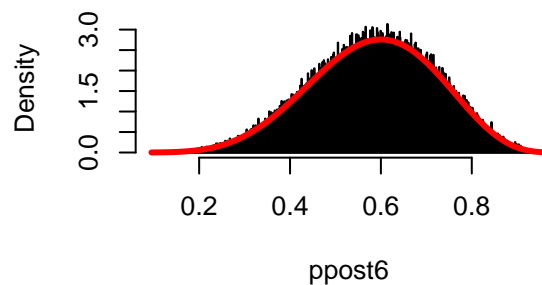
Histogram of ppost4



Histogram of ppost5



Histogram of ppost6



1.5

Si calcoli il valore atteso e la moda per ogni distribuzione disegnata al punto 1

```

#atteso: k+a / n+a+b
#moda: k+a-1 / n+a+b
#E3=4/12 E4=5/12 E5=6/12 E6=7/12
#M3=3/12 4/12 5/12 6/12

```

Esercizio 2

Si supponga di disporre di 2 monete: una truccata con probabilità di testa pari a $3/4$ e l'altra non truccata. Si sceglie una moneta tra le due senza riconoscerle. Supponendo di lanciare 3 volte la moneta scelta ed osservando 3 teste, si determini la probabilità di aver scelto la moneta non truccata.

A evento 3 volte testa Mnt moneta non truccata Mt moneta truccata

$p(\text{Mnt}=\text{testa})=0.5$ $p(\text{Mt}=\text{testa})=0.75$ Si cerca $P(\text{Mt}|\text{A})$

$p(\text{A}|\text{Mnt})$ e $P(\text{A}|\text{Mt})=(3/4)^3$

Applicando Bayes $P(\text{Mnt}|\text{A})=P(\text{A}|\text{Mnt}) \cdot p(\text{Mnt}) / p(\text{A})$

Applico regola prob complessiva $P(\text{A})=P(\text{A}|\text{Mnt})P(\text{Mnt})+p(\text{A}|\text{Mt})P(\text{Mt})$

$P(\text{Mnt}|\text{A})=P(\text{A}|\text{Mnt})P(\text{Mnt}) / P(\text{A})=0.5^3 \cdot 0.5 / 0.5^3 \cdot 0.5 + 3/4^3 \cdot 0.5=0.23$

Esercizio 3

Si disegnino in un'unica finestra grafica costituita da 6 righe e sei colonne le seguenti forme assunte dalla densità della variabile casuale Beta al variare dei parametri: $(0.5,0.5)$, $(0.5,1)$, $(1,0.5)$, $(1,1)$, $(1,2)$, $(2,1)$, $(2,2)$, $(2,6)$, $(6,2)$, $(3,3)$, $(3,8)$, $(8,3)$.

Esercizio 4

4.1

Si svolga il punto 1 dell'esercizio 1 considerando $n = 50$ e $k = 35, 37, 40, 43$. Si riportino le densità corrispondenti e si calcolino le quantità di cui al punto 5 dell'esercizio 1.

4.2

Si calcoli per ogni situazione disegnata $E(p)$, $E(p|X = k)$ e la moda della distribuzione a posteriori. Si nota che le distribuzioni sono più o meno influenzate dalla distribuzione iniziale (a priori) uniforme?

Esercizio 5

Supponendo una distribuzione discreta per $p(x, ??)$ e tale che il modello sia il seguente quando $?? \in \{1, 2, 3\}$:

5.1

Si determini la stima di massima verosimiglianza di $??$ per ogni valore di x .

5.2

Considerando le seguenti probabilità a priori per $??$ $p(??) = 0.2$, $p(??) = 0.5$ e $p(??) = 0.3$ e supponendo di osservare $x = 3$ si determini la probabilità a posteriori di $??$.

Esercizio 15

Si considerino i files dei dati denominati path.Rdata e freq.Rdata che riguardano dei punteggi assegnati a 1000 pazienti in 6 visite consecutive e le relative frequenze (dati simulati). La valutazione del medico circa lo stato di salute psicologica dell'individuo è stata espressa rispetto alla seguente scala ordinale: 0 = lieve, 1 = moderata, 2 = grave.

15.1

Si descrivano i dati disponibili e si identifichino i tre pattern di risposta aventi le frequenze più elevate. Carico i dati.

```
load("path.RData")
load("freq.RData")

m <- cbind(S,fr)
df <- as.data.frame(m)

#Identifico le 3 frequenze più elevate
head(sort(fr,decreasing = T))

## [1] 63 26 25 22 22 22

#Identifico i pattern relativi alle frequenze più elevate
subset(df, fr==63 | fr==26 | fr==25)

##      V1 V2 V3 V4 V5 V6 fr
## 5     0  0  0  0  0  0 63
## 22    0  0  1  0  0  0 25
## 77    1  0  0  0  0  0 26
```

I tre pattern con frequenze più alte sono: 63 soggetti con valutazione ritenuta lieve in tutte le visite mediche effettuate; 25 soggetti con valutazione sempre lieve tranne nella terza visita in cui è stata registrata una situazione di moderata entità; 26 soggetti con valutazione moderata alla prima osservazione e nel resto del periodo valutazione lieve.

15.2

Si riporti la frequenza di coloro che risultano avere la malattia in forma moderata durante tutte le visite.

```
subset(df,df$V1==1 & df$V2==1 & df$V3==1 & df$V4==1 & df$V5==1 & df$V6==1)

##      V1 V2 V3 V4 V5 V6 fr
## 146    1  1  1  1  1  1  2
```

Solo 2 soggetti hanno la malattia in forma moderata durante tutte le visite.

15.3

Si utilizzi la libreria LMest per stimare un modello latente di Markov omogeneo nel tempo selezionando un numero di stati latenti da 1 a 3 in base ad uno dei criteri d'informazione introdotti nella parte teorica.

```
require(LMest)

## Loading required package: LMest
```

```
## Loading required package: MASS
## Loading required package: MultiLCIRT
## Loading required package: limSolve
#Fisso i 3 differenti livelli di k
#k identifica il numero di stati latenti da testare
k1 <- 1
k2 <- 2
k3 <- 3
mod1 <- est_lm_basic(S = S, yv = fr, k=k1,
                     mod=1,
                     start = 0)
mod2 <- est_lm_basic(S = S, yv = fr, k=k2,
                     mod=1,
                     start = 0)
```

| ## | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
|----|-------|-------|-------|-------|----------|------------|-------------|
| ## | mod | k | start | step | lk | lk-lko | discrepancy |
| ## | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| ## | 1 | 2 | 0 | 0 | -6180.79 | | |
| ## | 1 | 2 | 0 | 10 | -5403.7 | 0.968487 | 0.00455607 |
| ## | 1 | 2 | 0 | 20 | -5398.64 | 0.314753 | 0.00260694 |
| ## | 1 | 2 | 0 | 30 | -5396.78 | 0.105721 | 0.00142988 |
| ## | 1 | 2 | 0 | 40 | -5396.2 | 0.0338325 | 0.000689589 |
| ## | 1 | 2 | 0 | 50 | -5395.98 | 0.0155933 | 0.000553488 |
| ## | 1 | 2 | 0 | 60 | -5395.85 | 0.0108121 | 0.000502042 |
| ## | 1 | 2 | 0 | 70 | -5395.75 | 0.00925196 | 0.000465103 |
| ## | 1 | 2 | 0 | 80 | -5395.67 | 0.008485 | 0.000452311 |
| ## | 1 | 2 | 0 | 90 | -5395.58 | 0.00792127 | 0.000440995 |
| ## | 1 | 2 | 0 | 100 | -5395.51 | 0.00741767 | 0.000427197 |
| ## | 1 | 2 | 0 | 110 | -5395.44 | 0.00694112 | 0.000412145 |
| ## | 1 | 2 | 0 | 120 | -5395.37 | 0.00648525 | 0.000396582 |
| ## | 1 | 2 | 0 | 130 | -5395.31 | 0.00604974 | 0.000380943 |
| ## | 1 | 2 | 0 | 140 | -5395.25 | 0.00563533 | 0.000365486 |
| ## | 1 | 2 | 0 | 150 | -5395.2 | 0.00524265 | 0.000350365 |
| ## | 1 | 2 | 0 | 160 | -5395.14 | 0.00487202 | 0.00033567 |
| ## | 1 | 2 | 0 | 170 | -5395.1 | 0.00452339 | 0.000321458 |
| ## | 1 | 2 | 0 | 180 | -5395.05 | 0.00419647 | 0.000307758 |
| ## | 1 | 2 | 0 | 190 | -5395.01 | 0.00389069 | 0.000294589 |
| ## | 1 | 2 | 0 | 200 | -5394.98 | 0.00360535 | 0.000281955 |
| ## | 1 | 2 | 0 | 210 | -5394.94 | 0.00333959 | 0.000269856 |
| ## | 1 | 2 | 0 | 220 | -5394.91 | 0.0030925 | 0.000258286 |
| ## | 1 | 2 | 0 | 230 | -5394.88 | 0.00286307 | 0.000247235 |
| ## | 1 | 2 | 0 | 240 | -5394.85 | 0.00265031 | 0.000236688 |
| ## | 1 | 2 | 0 | 250 | -5394.83 | 0.0024532 | 0.000226632 |
| ## | 1 | 2 | 0 | 260 | -5394.8 | 0.00227073 | 0.000217048 |
| ## | 1 | 2 | 0 | 270 | -5394.78 | 0.00210193 | 0.000207918 |
| ## | 1 | 2 | 0 | 280 | -5394.76 | 0.00194585 | 0.000199225 |
| ## | 1 | 2 | 0 | 290 | -5394.74 | 0.0018016 | 0.00019095 |
| ## | 1 | 2 | 0 | 300 | -5394.73 | 0.00166832 | 0.000183072 |
| ## | 1 | 2 | 0 | 310 | -5394.71 | 0.00154519 | 0.000175575 |
| ## | 1 | 2 | 0 | 320 | -5394.7 | 0.00143147 | 0.000168439 |
| ## | 1 | 2 | 0 | 330 | -5394.68 | 0.00132642 | 0.000161646 |
| ## | 1 | 2 | 0 | 340 | -5394.67 | 0.0012294 | 0.000155179 |

| | | | | | | | |
|----|-------|-------|-------|-------|----------|-------------|-------------|
| ## | 1 | 2 | 0 | 350 | -5394.66 | 0.00113979 | 0.000149022 |
| ## | 1 | 2 | 0 | 360 | -5394.65 | 0.001057 | 0.000143741 |
| ## | 1 | 2 | 0 | 370 | -5394.64 | 0.000980508 | 0.000139344 |
| ## | 1 | 2 | 0 | 380 | -5394.63 | 0.000909825 | 0.000135028 |
| ## | 1 | 2 | 0 | 390 | -5394.62 | 0.000844494 | 0.000130798 |
| ## | 1 | 2 | 0 | 400 | -5394.61 | 0.000784097 | 0.000126655 |
| ## | 1 | 2 | 0 | 410 | -5394.6 | 0.000728246 | 0.000122601 |
| ## | 1 | 2 | 0 | 420 | -5394.6 | 0.000676587 | 0.000118639 |
| ## | 1 | 2 | 0 | 430 | -5394.59 | 0.00062879 | 0.00011477 |
| ## | 1 | 2 | 0 | 440 | -5394.58 | 0.000584555 | 0.000110993 |
| ## | 1 | 2 | 0 | 450 | -5394.58 | 0.000543604 | 0.000107311 |
| ## | 1 | 2 | 0 | 460 | -5394.57 | 0.000505682 | 0.000103721 |
| ## | 1 | 2 | 0 | 470 | -5394.57 | 0.000470553 | 0.000100226 |
| ## | 1 | 2 | 0 | 480 | -5394.56 | 0.000438002 | 9.68235e-05 |
| ## | 1 | 2 | 0 | 490 | -5394.56 | 0.000407831 | 9.35137e-05 |
| ## | 1 | 2 | 0 | 500 | -5394.55 | 0.000379855 | 9.02955e-05 |
| ## | 1 | 2 | 0 | 510 | -5394.55 | 0.000353908 | 8.71681e-05 |
| ## | 1 | 2 | 0 | 520 | -5394.55 | 0.000329834 | 8.41303e-05 |
| ## | 1 | 2 | 0 | 530 | -5394.54 | 0.000307491 | 8.11808e-05 |
| ## | 1 | 2 | 0 | 540 | -5394.54 | 0.000286748 | 7.83183e-05 |
| ## | 1 | 2 | 0 | 550 | -5394.54 | 0.000267485 | 7.55413e-05 |
| ## | 1 | 2 | 0 | 560 | -5394.54 | 0.00024959 | 7.28483e-05 |
| ## | 1 | 2 | 0 | 570 | -5394.53 | 0.00023296 | 7.02376e-05 |
| ## | 1 | 2 | 0 | 580 | -5394.53 | 0.000217502 | 6.77077e-05 |
| ## | 1 | 2 | 0 | 590 | -5394.53 | 0.000203129 | 6.52568e-05 |
| ## | 1 | 2 | 0 | 600 | -5394.53 | 0.00018976 | 6.28832e-05 |
| ## | 1 | 2 | 0 | 610 | -5394.53 | 0.000177321 | 6.05851e-05 |
| ## | 1 | 2 | 0 | 620 | -5394.52 | 0.000165745 | 5.83607e-05 |
| ## | 1 | 2 | 0 | 630 | -5394.52 | 0.000154968 | 5.62083e-05 |
| ## | 1 | 2 | 0 | 640 | -5394.52 | 0.000144932 | 5.41261e-05 |
| ## | 1 | 2 | 0 | 650 | -5394.52 | 0.000135583 | 5.21123e-05 |
| ## | 1 | 2 | 0 | 660 | -5394.52 | 0.000126872 | 5.0165e-05 |
| ## | 1 | 2 | 0 | 670 | -5394.52 | 0.000118753 | 4.82826e-05 |
| ## | 1 | 2 | 0 | 680 | -5394.52 | 0.000111183 | 4.64632e-05 |
| ## | 1 | 2 | 0 | 690 | -5394.51 | 0.000104124 | 4.50191e-05 |
| ## | 1 | 2 | 0 | 700 | -5394.51 | 9.75388e-05 | 4.36438e-05 |
| ## | 1 | 2 | 0 | 710 | -5394.51 | 9.13941e-05 | 4.2318e-05 |
| ## | 1 | 2 | 0 | 720 | -5394.51 | 8.5659e-05 | 4.10396e-05 |
| ## | 1 | 2 | 0 | 730 | -5394.51 | 8.03046e-05 | 3.98066e-05 |
| ## | 1 | 2 | 0 | 740 | -5394.51 | 7.53043e-05 | 3.86171e-05 |
| ## | 1 | 2 | 0 | 750 | -5394.51 | 7.06334e-05 | 3.74692e-05 |
| ## | 1 | 2 | 0 | 760 | -5394.51 | 6.62691e-05 | 3.63612e-05 |
| ## | 1 | 2 | 0 | 770 | -5394.51 | 6.21901e-05 | 3.52914e-05 |
| ## | 1 | 2 | 0 | 780 | -5394.51 | 5.83768e-05 | 3.42582e-05 |
| ## | 1 | 2 | 0 | 790 | -5394.51 | 5.48109e-05 | 3.32603e-05 |
| ## | 1 | 2 | 0 | 793 | -5394.51 | 5.37868e-05 | 3.29675e-05 |
| ## | ----- | ----- | ----- | ----- | ----- | ----- | ----- |

```
mod3 <- est_lm_basic(S = S, yv = fr, k=k3,
  mod=1,
  start = 0)
```

| ## | mod | k | start | step | lk | lk-lko | discrepancy |
|----|-------|-------|-------|-------|-------|--------|-------------|
| ## | ----- | ----- | ----- | ----- | ----- | ----- | ----- |

| | | | | | | | |
|----|---|---|---|-----|----------|-------------|-------------|
| ## | 1 | 3 | 0 | 0 | -5991.12 | | |
| ## | 1 | 3 | 0 | 10 | -5400.28 | 0.729972 | 0.0110875 |
| ## | 1 | 3 | 0 | 20 | -5397.1 | 0.144131 | 0.00425537 |
| ## | 1 | 3 | 0 | 30 | -5396.34 | 0.0445448 | 0.00202783 |
| ## | 1 | 3 | 0 | 40 | -5396.03 | 0.0240953 | 0.00110125 |
| ## | 1 | 3 | 0 | 50 | -5395.83 | 0.0174333 | 0.000705541 |
| ## | 1 | 3 | 0 | 60 | -5395.68 | 0.014196 | 0.000694674 |
| ## | 1 | 3 | 0 | 70 | -5395.55 | 0.0122228 | 0.00069355 |
| ## | 1 | 3 | 0 | 80 | -5395.43 | 0.01085 | 0.000689151 |
| ## | 1 | 3 | 0 | 90 | -5395.33 | 0.00981276 | 0.000683496 |
| ## | 1 | 3 | 0 | 100 | -5395.24 | 0.00898562 | 0.000677307 |
| ## | 1 | 3 | 0 | 110 | -5395.15 | 0.0083005 | 0.000670749 |
| ## | 1 | 3 | 0 | 120 | -5395.07 | 0.00771656 | 0.000663798 |
| ## | 1 | 3 | 0 | 130 | -5395 | 0.00720743 | 0.000656388 |
| ## | 1 | 3 | 0 | 140 | -5394.93 | 0.00675529 | 0.000648473 |
| ## | 1 | 3 | 0 | 150 | -5394.86 | 0.00634768 | 0.000640032 |
| ## | 1 | 3 | 0 | 160 | -5394.8 | 0.00597575 | 0.000631074 |
| ## | 1 | 3 | 0 | 170 | -5394.74 | 0.00563311 | 0.000621627 |
| ## | 1 | 3 | 0 | 180 | -5394.69 | 0.00531511 | 0.000611733 |
| ## | 1 | 3 | 0 | 190 | -5394.64 | 0.00501831 | 0.00060144 |
| ## | 1 | 3 | 0 | 200 | -5394.59 | 0.00474012 | 0.000590801 |
| ## | 1 | 3 | 0 | 210 | -5394.54 | 0.00447857 | 0.000579869 |
| ## | 1 | 3 | 0 | 220 | -5394.5 | 0.00423213 | 0.000568694 |
| ## | 1 | 3 | 0 | 230 | -5394.46 | 0.00399956 | 0.000557326 |
| ## | 1 | 3 | 0 | 240 | -5394.42 | 0.00377986 | 0.000545808 |
| ## | 1 | 3 | 0 | 250 | -5394.38 | 0.00357218 | 0.000534183 |
| ## | 1 | 3 | 0 | 260 | -5394.35 | 0.00337577 | 0.000522488 |
| ## | 1 | 3 | 0 | 270 | -5394.31 | 0.00319 | 0.000510758 |
| ## | 1 | 3 | 0 | 280 | -5394.28 | 0.00301426 | 0.000499025 |
| ## | 1 | 3 | 0 | 290 | -5394.25 | 0.00284803 | 0.000487315 |
| ## | 1 | 3 | 0 | 300 | -5394.23 | 0.00269081 | 0.000475654 |
| ## | 1 | 3 | 0 | 310 | -5394.2 | 0.00254212 | 0.000464064 |
| ## | 1 | 3 | 0 | 320 | -5394.18 | 0.00240153 | 0.000452564 |
| ## | 1 | 3 | 0 | 330 | -5394.15 | 0.00226861 | 0.000441173 |
| ## | 1 | 3 | 0 | 340 | -5394.13 | 0.00214298 | 0.000429906 |
| ## | 1 | 3 | 0 | 350 | -5394.11 | 0.00202424 | 0.000418776 |
| ## | 1 | 3 | 0 | 360 | -5394.09 | 0.00191206 | 0.000407795 |
| ## | 1 | 3 | 0 | 370 | -5394.07 | 0.00180607 | 0.000396974 |
| ## | 1 | 3 | 0 | 380 | -5394.05 | 0.00170596 | 0.00038632 |
| ## | 1 | 3 | 0 | 390 | -5394.04 | 0.00161141 | 0.000375842 |
| ## | 1 | 3 | 0 | 400 | -5394.02 | 0.00152212 | 0.000365545 |
| ## | 1 | 3 | 0 | 410 | -5394.01 | 0.00143783 | 0.000355436 |
| ## | 1 | 3 | 0 | 420 | -5393.99 | 0.00135825 | 0.000345518 |
| ## | 1 | 3 | 0 | 430 | -5393.98 | 0.00128313 | 0.000335795 |
| ## | 1 | 3 | 0 | 440 | -5393.97 | 0.00121224 | 0.000326269 |
| ## | 1 | 3 | 0 | 450 | -5393.96 | 0.00114533 | 0.000316943 |
| ## | 1 | 3 | 0 | 460 | -5393.94 | 0.0010822 | 0.000307818 |
| ## | 1 | 3 | 0 | 470 | -5393.93 | 0.00102264 | 0.000298894 |
| ## | 1 | 3 | 0 | 480 | -5393.92 | 0.00096644 | 0.000290171 |
| ## | 1 | 3 | 0 | 490 | -5393.91 | 0.000913429 | 0.000281651 |
| ## | 1 | 3 | 0 | 500 | -5393.91 | 0.000863425 | 0.00027333 |
| ## | 1 | 3 | 0 | 510 | -5393.9 | 0.000816263 | 0.00026521 |
| ## | 1 | 3 | 0 | 520 | -5393.89 | 0.000771785 | 0.000257287 |
| ## | 1 | 3 | 0 | 530 | -5393.88 | 0.00072984 | 0.000249561 |

| | | | | | | | |
|----|---|---|---|------|----------|-------------|-------------|
| ## | 1 | 3 | 0 | 540 | -5393.87 | 0.000690287 | 0.000242029 |
| ## | 1 | 3 | 0 | 550 | -5393.87 | 0.000652992 | 0.00023469 |
| ## | 1 | 3 | 0 | 560 | -5393.86 | 0.000617829 | 0.000227541 |
| ## | 1 | 3 | 0 | 570 | -5393.86 | 0.000584678 | 0.000220578 |
| ## | 1 | 3 | 0 | 580 | -5393.85 | 0.000553424 | 0.000213801 |
| ## | 1 | 3 | 0 | 590 | -5393.84 | 0.000523962 | 0.000207205 |
| ## | 1 | 3 | 0 | 600 | -5393.84 | 0.000496189 | 0.000200787 |
| ## | 1 | 3 | 0 | 610 | -5393.83 | 0.000470011 | 0.000194545 |
| ## | 1 | 3 | 0 | 620 | -5393.83 | 0.000445336 | 0.000188474 |
| ## | 1 | 3 | 0 | 630 | -5393.83 | 0.000422078 | 0.000182573 |
| ## | 1 | 3 | 0 | 640 | -5393.82 | 0.000400158 | 0.000176838 |
| ## | 1 | 3 | 0 | 650 | -5393.82 | 0.000379499 | 0.000171264 |
| ## | 1 | 3 | 0 | 660 | -5393.81 | 0.000360028 | 0.000165849 |
| ## | 1 | 3 | 0 | 670 | -5393.81 | 0.000341677 | 0.00016059 |
| ## | 1 | 3 | 0 | 680 | -5393.81 | 0.000324383 | 0.000155483 |
| ## | 1 | 3 | 0 | 690 | -5393.8 | 0.000308084 | 0.000150523 |
| ## | 1 | 3 | 0 | 700 | -5393.8 | 0.000292723 | 0.000145709 |
| ## | 1 | 3 | 0 | 710 | -5393.8 | 0.000278246 | 0.000141037 |
| ## | 1 | 3 | 0 | 720 | -5393.8 | 0.000264602 | 0.000136502 |
| ## | 1 | 3 | 0 | 730 | -5393.79 | 0.000251743 | 0.000132102 |
| ## | 1 | 3 | 0 | 740 | -5393.79 | 0.000239624 | 0.000127834 |
| ## | 1 | 3 | 0 | 750 | -5393.79 | 0.0002282 | 0.000123694 |
| ## | 1 | 3 | 0 | 760 | -5393.79 | 0.000217433 | 0.000119678 |
| ## | 1 | 3 | 0 | 770 | -5393.78 | 0.000207284 | 0.000115784 |
| ## | 1 | 3 | 0 | 780 | -5393.78 | 0.000197717 | 0.000112008 |
| ## | 1 | 3 | 0 | 790 | -5393.78 | 0.000188697 | 0.000108347 |
| ## | 1 | 3 | 0 | 800 | -5393.78 | 0.000180194 | 0.000104799 |
| ## | 1 | 3 | 0 | 810 | -5393.78 | 0.000172177 | 0.000101359 |
| ## | 1 | 3 | 0 | 820 | -5393.77 | 0.000164617 | 9.80251e-05 |
| ## | 1 | 3 | 0 | 830 | -5393.77 | 0.000157488 | 9.47942e-05 |
| ## | 1 | 3 | 0 | 840 | -5393.77 | 0.000150765 | 9.16635e-05 |
| ## | 1 | 3 | 0 | 850 | -5393.77 | 0.000144423 | 8.86301e-05 |
| ## | 1 | 3 | 0 | 860 | -5393.77 | 0.000138441 | 8.56911e-05 |
| ## | 1 | 3 | 0 | 870 | -5393.77 | 0.000132798 | 8.2844e-05 |
| ## | 1 | 3 | 0 | 880 | -5393.77 | 0.000127473 | 8.00859e-05 |
| ## | 1 | 3 | 0 | 890 | -5393.76 | 0.000122448 | 7.74144e-05 |
| ## | 1 | 3 | 0 | 900 | -5393.76 | 0.000117705 | 7.48269e-05 |
| ## | 1 | 3 | 0 | 910 | -5393.76 | 0.000113228 | 7.23209e-05 |
| ## | 1 | 3 | 0 | 920 | -5393.76 | 0.000109001 | 6.9894e-05 |
| ## | 1 | 3 | 0 | 930 | -5393.76 | 0.00010501 | 6.75439e-05 |
| ## | 1 | 3 | 0 | 940 | -5393.76 | 0.00010124 | 6.52681e-05 |
| ## | 1 | 3 | 0 | 950 | -5393.76 | 9.76787e-05 | 6.30646e-05 |
| ## | 1 | 3 | 0 | 960 | -5393.76 | 9.4314e-05 | 6.09311e-05 |
| ## | 1 | 3 | 0 | 970 | -5393.76 | 9.11342e-05 | 5.88655e-05 |
| ## | 1 | 3 | 0 | 980 | -5393.76 | 8.81284e-05 | 5.68657e-05 |
| ## | 1 | 3 | 0 | 990 | -5393.75 | 8.52864e-05 | 5.49296e-05 |
| ## | 1 | 3 | 0 | 1000 | -5393.75 | 8.25987e-05 | 5.30555e-05 |
| ## | 1 | 3 | 0 | 1000 | -5393.75 | 8.25987e-05 | 5.30555e-05 |

```
print(mod1)
```

```
## Call:
## est_lm_basic(S = S, yv = fr, k = k1, start = 0, mod = 1)
##
```

```
## Convergence info:
##      LogLik np      AIC      BIC
## [1,] -5409.461  2 10822.92 10832.74
```

```
print(mod2)
```

```
## Call:
## est_lm_basic(S = S, yv = fr, k = k2, start = 0, mod = 1)
##
## Convergence info:
##      LogLik np      AIC      BIC
## [1,] -5394.507  7 10803.01 10837.37
```

```
print(mod3)
```

```
## Call:
## est_lm_basic(S = S, yv = fr, k = k3, start = 0, mod = 1)
##
## Convergence info:
##      LogLik np      AIC      BIC
## [1,] -5393.754 14 10815.51 10884.22
```

Il miglior modello latente di Markov secondo il criterio di informazione BIC risulta quello con con 1 stato latente ($k=1$). Nonostante esso rilevi il BIC minore, conviene stimare il modello con $k=2$ stati latenti, il quale risulta preferibile in termini di AIC e loglik. Quindi il modello che meglio sembra adattare i dati è quello a 2 stati latenti, con tempi omogenei. Esso stima 7 parametri e impiega circa 709 passi per raggiungere la convergenza.

15.4

Si descrivano le stime dei parametri del modello scelto al punto precedente in relazione al contesto applicativo. I parametri da stimare per il modello a 2 stati latenti sono 7.

Inizialmente vediamo le stime per il modello di misura al fine di caratterizzare i gruppi.

```
mod2$Psi #modello di misura --> caratterizzazione dei gruppi
```

```
## , , item = 1
##
##      state
## category    1      2
##      0 0.6666222 0.47570362
##      1 0.2092284 0.45205642
##      2 0.1241494 0.07223996
```

Il primo gruppo raccoglie per la maggior parte soggetti con malattia lieve, una minor parte di soggetti con malattia moderata e infine una minima parte di soggetti con malattia grave. Il secondo gruppo raccoglie invece per oltre il 90% soggetti egualmente distribuiti tra malattia di grado lieve e malattia di grado moderato, mentre trattiene solo il 7% di soggetti con malattia grave.

Vediamo ora le stime del modello latente relative alle probabilità iniziali.

```
mod2$piv #prob iniziali del modello latente
```

```
## [1] 0.7087739 0.2912261
```

La probabilità di appartenere in partenza al primo gruppo è del 71% contro la probabilità di appartenere in partenza al secondo gruppo del 29%.

Vediamo ora le stime del modello latente relative alle probabilità di transizione.

```
mod2$Pi #prob di transizione del modello latente
```

```
## , , time = 1
##
##      state
## state 1 2
##      1 0 0
##      2 0 0
##
## , , time = 2
##
##      state
## state      1      2
##      1 0.95153220 0.0484678
##      2 0.07477411 0.9252259
##
## , , time = 3
##
##      state
## state      1      2
##      1 0.95153220 0.0484678
##      2 0.07477411 0.9252259
##
## , , time = 4
##
##      state
## state      1      2
##      1 0.95153220 0.0484678
##      2 0.07477411 0.9252259
##
## , , time = 5
##
##      state
## state      1      2
##      1 0.95153220 0.0484678
##      2 0.07477411 0.9252259
##
## , , time = 6
##
##      state
## state      1      2
##      1 0.95153220 0.0484678
##      2 0.07477411 0.9252259
```

Poichè il tempo è ritenuto omogeneo, si ottiene la stessa tabella delle probabilità di transizione per tutti i tempi considerati. Dai risultati, leggendo le probabilità per riga, vediamo che chi parte in gruppo 1 ha un'altissima probabilità (95%) di restare in quel gruppo; allo stesso modo chi parte in gruppo 2 ha una altrettanto alta probabilità di rimanere nello stesso gruppo (93%). Questo ci porta a concludere che è molto difficile per un soggetto transitare da un gruppo all'altro, intesa come modificare la propria situazione relativa alla valutazione di malattia.

Esercizio 19

19.1

Si considerino i dati presenti nei files denominati pat1.Rdata e yv1.Rdata. I dati simulati riguardano le configurazioni di risposta e le rispettive frequenze di 500 giovani donne nella fascia di età compresa tra i 9 e 31 anni le quali sono state sottoposte per 3 occasioni temporali consecutive alla misurazione del livello di minerali presenti nelle ossa. Supponendo che i livelli (0,1,2,3) della variabile siano codificati nel seguente modo: $\leq 0:00362$, $(0:0036; 0:0214]$, $(0:0214; 0:0650]$, $(0:0650; 0:2199]$ i cui livelli sono densità stimate del minerale con DXA. Si illustrino i dati e si commentino. Carico i dati e creo una matrice m.

```
load("pat1.Rdata")
load("yv1.Rdata")
m <- cbind(Ssim,yvmsim)
df <- as.data.frame(m)
sort(df$yvmsim, decreasing=T)
```

```
## [1] 31 29 24 22 22 21 20 18 16 15 15 14 14 14 12 12 10 9 9 9 9 8 8
## [24] 7 7 7 7 7 7 6 6 6 6 6 6 5 5 4 4 4 4 3 3 3 3 3
## [47] 3 3 2 2 2 2 2 1 1 1 1
```

```
subset(df,yvmsim==31 | yvmsim==29)
```

```
##      V1 V2 V3 yvmsim
## 18    2  1  1      31
## 33    3  0  0      29
```

#può essere utile mostrare head e tail per descrivere la matrice

I pattern con frequenza più numerosa sono quello che conta 31 soggetti che partono da un livello 2 di presenza di minerale alla prima visita per poi assestarsi alle visite successive al livello 1; parimenti quello che conta 29 soggetti che partono da un livello 3 per poi assestarsi al livello più basso. In generale i pattern più frequenti presentano livelli di minerale in ordine decrescente nel tempo.

19.2

Si consideri un modello latente di Markov per questi dati. Supponendo un modello con 4 stati latenti si illustrino le ipotesi del modello. Le ipotesi del modello sono 3 da commentare: l'uso di lm, k=4 e mod=1.

```
require(LMest)
k <- 4 #fisso gli stati latenti
mod <- est_lm_basic(S = Ssim, yv = yvmsim, k=k,
                    mod=1,
                    start = 0)
```

| ## | ----- | | ----- | | ----- | | ----- | | ----- | | ----- | | ----- | |
|----|-------|--|-------|--|-------|--|-------|--|----------|--|-----------|--|-------------|--|
| ## | mod | | k | | start | | step | | lk | | lk-lko | | discrepancy | |
| ## | ----- | | ----- | | ----- | | ----- | | ----- | | ----- | | ----- | |
| ## | 1 | | 4 | | 0 | | 0 | | -2390.18 | | | | | |
| ## | 1 | | 4 | | 0 | | 10 | | -1910.58 | | 1.10686 | | 0.0156966 | |
| ## | 1 | | 4 | | 0 | | 20 | | -1907.23 | | 0.171346 | | 0.00576775 | |
| ## | 1 | | 4 | | 0 | | 30 | | -1905.61 | | 0.158851 | | 0.00556042 | |
| ## | 1 | | 4 | | 0 | | 40 | | -1904.29 | | 0.0999905 | | 0.0053645 | |
| ## | 1 | | 4 | | 0 | | 50 | | -1903.67 | | 0.0387239 | | 0.0041855 | |
| ## | 1 | | 4 | | 0 | | 60 | | -1903.4 | | 0.0209126 | | 0.00291033 | |
| ## | 1 | | 4 | | 0 | | 70 | | -1903.22 | | 0.0168143 | | 0.00205395 | |

| | | | | | | | |
|----|---|---|---|-----|----------|------------|-------------|
| ## | 1 | 4 | 0 | 80 | -1903.06 | 0.0151237 | 0.00151709 |
| ## | 1 | 4 | 0 | 90 | -1902.91 | 0.0138404 | 0.00132805 |
| ## | 1 | 4 | 0 | 100 | -1902.78 | 0.0125142 | 0.0011344 |
| ## | 1 | 4 | 0 | 110 | -1902.67 | 0.0110741 | 0.000945659 |
| ## | 1 | 4 | 0 | 120 | -1902.56 | 0.00964616 | 0.00091345 |
| ## | 1 | 4 | 0 | 130 | -1902.47 | 0.0084141 | 0.000857693 |
| ## | 1 | 4 | 0 | 140 | -1902.39 | 0.00749745 | 0.000785707 |
| ## | 1 | 4 | 0 | 150 | -1902.32 | 0.00691305 | 0.000705202 |
| ## | 1 | 4 | 0 | 160 | -1902.26 | 0.00660412 | 0.000622479 |
| ## | 1 | 4 | 0 | 170 | -1902.19 | 0.00648635 | 0.000541739 |
| ## | 1 | 4 | 0 | 180 | -1902.13 | 0.00648119 | 0.000497276 |
| ## | 1 | 4 | 0 | 190 | -1902.06 | 0.00653045 | 0.000506851 |
| ## | 1 | 4 | 0 | 200 | -1902 | 0.0065976 | 0.000524829 |
| ## | 1 | 4 | 0 | 210 | -1901.93 | 0.00666306 | 0.000550641 |
| ## | 1 | 4 | 0 | 220 | -1901.86 | 0.00671854 | 0.000583364 |
| ## | 1 | 4 | 0 | 230 | -1901.79 | 0.00676277 | 0.000621957 |
| ## | 1 | 4 | 0 | 240 | -1901.73 | 0.00679871 | 0.000665445 |
| ## | 1 | 4 | 0 | 250 | -1901.66 | 0.00683194 | 0.000713034 |
| ## | 1 | 4 | 0 | 260 | -1901.59 | 0.00686979 | 0.000764156 |
| ## | 1 | 4 | 0 | 270 | -1901.52 | 0.00692083 | 0.000818468 |
| ## | 1 | 4 | 0 | 280 | -1901.45 | 0.00699453 | 0.000875812 |
| ## | 1 | 4 | 0 | 290 | -1901.38 | 0.00710104 | 0.000936171 |
| ## | 1 | 4 | 0 | 300 | -1901.31 | 0.00725097 | 0.000999596 |
| ## | 1 | 4 | 0 | 310 | -1901.24 | 0.00745523 | 0.00106614 |
| ## | 1 | 4 | 0 | 320 | -1901.16 | 0.00772478 | 0.00113579 |
| ## | 1 | 4 | 0 | 330 | -1901.08 | 0.00807036 | 0.00120832 |
| ## | 1 | 4 | 0 | 340 | -1901 | 0.00850191 | 0.00128327 |
| ## | 1 | 4 | 0 | 350 | -1900.91 | 0.00902764 | 0.00135971 |
| ## | 1 | 4 | 0 | 360 | -1900.82 | 0.00965227 | 0.00143614 |
| ## | 1 | 4 | 0 | 370 | -1900.72 | 0.010374 | 0.00151029 |
| ## | 1 | 4 | 0 | 380 | -1900.61 | 0.0111798 | 0.00157888 |
| ## | 1 | 4 | 0 | 390 | -1900.49 | 0.0120387 | 0.00163753 |
| ## | 1 | 4 | 0 | 400 | -1900.37 | 0.012894 | 0.00168074 |
| ## | 1 | 4 | 0 | 410 | -1900.23 | 0.0136574 | 0.00170222 |
| ## | 1 | 4 | 0 | 420 | -1900.09 | 0.0142099 | 0.00169578 |
| ## | 1 | 4 | 0 | 430 | -1899.95 | 0.0144197 | 0.00165679 |
| ## | 1 | 4 | 0 | 440 | -1899.81 | 0.014178 | 0.00158397 |
| ## | 1 | 4 | 0 | 450 | -1899.67 | 0.0134447 | 0.00148086 |
| ## | 1 | 4 | 0 | 460 | -1899.54 | 0.0122788 | 0.00135587 |
| ## | 1 | 4 | 0 | 470 | -1899.42 | 0.0108283 | 0.00122057 |
| ## | 1 | 4 | 0 | 480 | -1899.32 | 0.00928155 | 0.00108681 |
| ## | 1 | 4 | 0 | 490 | -1899.24 | 0.00780831 | 0.00096402 |
| ## | 1 | 4 | 0 | 500 | -1899.17 | 0.00652135 | 0.000857856 |
| ## | 1 | 4 | 0 | 510 | -1899.11 | 0.00546955 | 0.000770386 |
| ## | 1 | 4 | 0 | 520 | -1899.06 | 0.00465272 | 0.000701093 |
| ## | 1 | 4 | 0 | 530 | -1899.02 | 0.00404318 | 0.000648061 |
| ## | 1 | 4 | 0 | 540 | -1898.98 | 0.0036038 | 0.000689083 |
| ## | 1 | 4 | 0 | 550 | -1898.94 | 0.00329902 | 0.000731389 |
| ## | 1 | 4 | 0 | 560 | -1898.91 | 0.00309978 | 0.000773361 |
| ## | 1 | 4 | 0 | 570 | -1898.88 | 0.00298467 | 0.00081607 |
| ## | 1 | 4 | 0 | 580 | -1898.85 | 0.00293918 | 0.000860711 |
| ## | 1 | 4 | 0 | 590 | -1898.82 | 0.00295443 | 0.000908516 |
| ## | 1 | 4 | 0 | 600 | -1898.79 | 0.00302597 | 0.000960676 |
| ## | 1 | 4 | 0 | 610 | -1898.76 | 0.00315265 | 0.00101827 |

| | | | | | | | |
|----|-------|-------|-------|-------|----------|------------|-------------|
| ## | 1 | 4 | 0 | 620 | -1898.73 | 0.0033358 | 0.00108216 |
| ## | 1 | 4 | 0 | 630 | -1898.7 | 0.00357835 | 0.00115287 |
| ## | 1 | 4 | 0 | 640 | -1898.66 | 0.00388375 | 0.00123034 |
| ## | 1 | 4 | 0 | 650 | -1898.62 | 0.0042544 | 0.00131373 |
| ## | 1 | 4 | 0 | 660 | -1898.57 | 0.00468932 | 0.00140099 |
| ## | 1 | 4 | 0 | 670 | -1898.52 | 0.00518088 | 0.0014885 |
| ## | 1 | 4 | 0 | 680 | -1898.47 | 0.00571079 | 0.00157073 |
| ## | 1 | 4 | 0 | 690 | -1898.41 | 0.0062465 | 0.00164018 |
| ## | 1 | 4 | 0 | 700 | -1898.34 | 0.00674014 | 0.00168799 |
| ## | 1 | 4 | 0 | 710 | -1898.27 | 0.007133 | 0.00170523 |
| ## | 1 | 4 | 0 | 720 | -1898.2 | 0.00736746 | 0.00168511 |
| ## | 1 | 4 | 0 | 730 | -1898.13 | 0.00740419 | 0.00162535 |
| ## | 1 | 4 | 0 | 740 | -1898.05 | 0.00723736 | 0.00152963 |
| ## | 1 | 4 | 0 | 750 | -1897.98 | 0.00689869 | 0.00140716 |
| ## | 1 | 4 | 0 | 760 | -1897.92 | 0.00644687 | 0.00127029 |
| ## | 1 | 4 | 0 | 770 | -1897.85 | 0.00594812 | 0.00113128 |
| ## | 1 | 4 | 0 | 780 | -1897.8 | 0.00545822 | 0.000999701 |
| ## | 1 | 4 | 0 | 790 | -1897.75 | 0.00501324 | 0.000881268 |
| ## | 1 | 4 | 0 | 800 | -1897.7 | 0.00462928 | 0.000778178 |
| ## | 1 | 4 | 0 | 810 | -1897.65 | 0.00430756 | 0.000690512 |
| ## | 1 | 4 | 0 | 820 | -1897.61 | 0.00404068 | 0.000676093 |
| ## | 1 | 4 | 0 | 830 | -1897.57 | 0.00381779 | 0.000657088 |
| ## | 1 | 4 | 0 | 840 | -1897.54 | 0.00362791 | 0.000634351 |
| ## | 1 | 4 | 0 | 850 | -1897.5 | 0.00346165 | 0.000608778 |
| ## | 1 | 4 | 0 | 860 | -1897.47 | 0.00331178 | 0.000581242 |
| ## | 1 | 4 | 0 | 870 | -1897.43 | 0.00317319 | 0.000552548 |
| ## | 1 | 4 | 0 | 880 | -1897.4 | 0.00304254 | 0.000523407 |
| ## | 1 | 4 | 0 | 890 | -1897.37 | 0.00291781 | 0.000494414 |
| ## | 1 | 4 | 0 | 900 | -1897.34 | 0.00279791 | 0.000483228 |
| ## | 1 | 4 | 0 | 910 | -1897.32 | 0.00268233 | 0.000483287 |
| ## | 1 | 4 | 0 | 920 | -1897.29 | 0.00257087 | 0.000482132 |
| ## | 1 | 4 | 0 | 930 | -1897.27 | 0.00246352 | 0.000479857 |
| ## | 1 | 4 | 0 | 940 | -1897.24 | 0.00236029 | 0.00047656 |
| ## | 1 | 4 | 0 | 950 | -1897.22 | 0.00226123 | 0.000472343 |
| ## | 1 | 4 | 0 | 960 | -1897.2 | 0.00216631 | 0.000467305 |
| ## | 1 | 4 | 0 | 970 | -1897.18 | 0.00207548 | 0.000461542 |
| ## | 1 | 4 | 0 | 980 | -1897.16 | 0.00198865 | 0.000455144 |
| ## | 1 | 4 | 0 | 990 | -1897.14 | 0.0019057 | 0.000448194 |
| ## | 1 | 4 | 0 | 1000 | -1897.12 | 0.00182648 | 0.000440769 |
| ## | 1 | 4 | 0 | 1000 | -1897.12 | 0.00182648 | 0.000440769 |
| ## | ----- | ----- | ----- | ----- | ----- | ----- | ----- |

Alla luce delle osservazioni, i valori riscontrati del minerale nelle ossa sono misurati con errore da uno strumento di misurazione, perciò assumo che il processo stocastico di interesse sia quello latente. Ecco perchè uso lm invece che mc. L'ipotesi del modello riguardo al tempo è che è stato assunto con matrice di transizione omogenea nel tempo per parsimonia, perciò si sceglie mod=1 nella formula del modello. Infine sono stati ipotizzati k=4 stati latenti, pari al numero di categorie della variabile risposta. Il modello raggiunge la convergenza con un numero di iterazioni dell'algoritmo EM pari a 1000.

19.3

Si determinino i parametri stimati del modello e si riportino le stime con gli opportuni commenti anche rispetto alla distribuzione stimata della variabile risposta.

Per prima cosa guardo le informazioni di convergenza

```
print(mod)
```

```
## Call:
## est_lm_basic(S = Ssim, yv = yvmsim, k = k, start = 0, mod = 1)
##
## Convergence info:
##           LogLik np      AIC      BIC
## [1,] -1897.117 27 3848.234 3962.029
```

Il modello stima 27 parametri liberi; ha un BIC di 3962 e una loglik di -1897

Ora guardo le stime dei parametri per commentare il modello di misura.

```
summary(mod)
```

```
## Call:
## est_lm_basic(S = Ssim, yv = yvmsim, k = k, start = 0, mod = 1)
##
## Coefficients:
##
## Initial probabilities:
##      est_piv
## [1,] 0.1941
## [2,] 0.0000
## [3,] 0.4677
## [4,] 0.3382
##
## Transition probabilities:
##      state
## state    1      2      3      4
##   1 0.5307 0.1041 0.3445 0.0207
##   2 0.6895 0.3105 0.0000 0.0000
##   3 0.5244 0.2210 0.2546 0.0000
##   4 0.7177 0.0779 0.2044 0.0000
##
## Conditional response probabilities:
## , , item = 1
##
##      state
## category    1      2      3      4
##   0 0.6363 0.0665 0.0857 0.0000
##   1 0.3637 0.4577 0.2522 0.0016
##   2 0.0000 0.0000 0.6621 0.1471
##   3 0.0000 0.4759 0.0000 0.8514
```

La sezione interessante è quella relativa alla parte di misura, cioè la tabella delle conditional response probabilities. Questa tabella mi permette di caratterizzare gli stati latenti della catena. Lo stato latente 1 si caratterizza per chi ha livelli bassi del mineale, lo stato 2 per chi ha prevalentemente livello 1 e 3 perciò rappresenta un mix, lo stato 3 per chi ha livelli moderati, lo stato 4 per chi ha livelli medio-alti. Osservando le probabilità iniziali, osservo che la probabilità maggiore è per il gruppo con livello 3. In secondo luogo, per il gruppo con livelli alti. Osservando invece le probabilità di transizione: guardando la diagonale riesco a leggere le probabilità di rimanere nello stesso stato latente, e vedo che in generale le probabilità di rimanere sono moderate; osservando le transizioni noto che per chi parte da livelli medio-alti di minerale (e quindi è nello stato latente 3 e 4) è più probabile transitare verso livelli più bassi, mentre sono quasi nulle le probabilità di transitare verso livelli maggiori. Quindi la transizione verso destra è molto rara rispetto alla transizione

verso sinistra

Cerco ora i parametri del modello di misura, cioè la distribuzione a posteriori degli stati latenti per ogni configurazione e per ogni time occasion. PS. Round arrotonda i valori in questo caso alla terza cifra decimale.

```
m[1,]  
  
##          yvmsim  
##      1      0      0      14  
round(mod$V[1,,],digits = 3)
```

```
##      [,1] [,2] [,3]  
## [1,] 0.374 0.896 0.904  
## [2,] 0.000 0.039 0.021  
## [3,] 0.622 0.065 0.075  
## [4,] 0.004 0.000 0.000
```

Ad esempio, per la prima configurazione osservata che ha frequenza 14 e livelli di minerale ordinati 1-0-0, il valore osservato al primo tempo (1) ha una probabilità del 62% di appartenere allo stato latente 3, i valori osservati al secondo e terzo tempo (0) hanno prob del 90% di appartenere allo stato latente 1.

```
m[12,]  
  
##          yvmsim  
##      2      3      0      7  
round(mod$V[12,,],3)
```

```
##      [,1] [,2] [,3]  
## [1,] 0.000  0 0.955  
## [2,] 0.000  1 0.045  
## [3,] 0.946  0 0.000  
## [4,] 0.054  0 0.000
```

Per coloro che hanno pattern di risposta 2-3-0, la massima probabilità a posteriori su ogni tempo è associata agli stati latenti 3,2,1.

```
m[44,]  
  
##          yvmsim  
##      0      0      2      6  
round(mod$V[44,,],3)
```

```
##      [,1] [,2] [,3]  
## [1,] 0.759 0.943 0.000  
## [2,] 0.000 0.000 0.000  
## [3,] 0.241 0.057 0.988  
## [4,] 0.000 0.000 0.012
```

In base alle probabilità a posteriori osservo che il pattern latente è pari a 1-1-3.

19.4 esempio con aggiunta di covariate

Si consideri un altro campione composto da altre 145 giovani donne in cui è presente anche la covariata età (si considerino i pattern di risposta individuali del file Scov.Rdata e l'array per la covariata in base ai valori che variano nel tempo nel file Xc.Rdata). Si forniscano le statistiche descrittive dei valori della covariata per le 3 occasioni temporali di interesse. Si stimi il modello latente di Markov valutando la covariata rispetto al

modello di misura e fissando a 4 il numero di stati latenti. Si riportino le stime dei parametri e si commentino i risultati in particolare rispetto alla covariata disponibile.

Carico i dati

```
load("Scov.Rdata")
load("Xc.Rdata")
#aggr_data(S) #matrice delle frequenze dei pattern
```

Cerco età media delle donne in ogni tempo

```
round(apply(Xc,2,mean),0)
```

```
## [1] 16 19 22
```

L'età media al tempo 1 è di 16 anni, al tempo 2 è di 19 e al tempo 3 è di 22 anni

Inserimento della covariata nel modello per capire come agisce l'età.

```
require(LMest)
k <- 4
res4 <- est_lm_cov_manifest(S = S, X=Xc,
                           k = k,
                           mod = "LM",
                           tol = 10^-2)
```

```
## n. distinct covariate conf. = 23
```

```
## EM step:
```

```
## -----|-----|-----|-----|-----|-----|
##      k   |   start   |   step   |   lk      |   lk-lko   | discrepancy |
## -----|-----|-----|-----|-----|-----|
##          4 |           0 |           0 | -641.412 |             |
```

```
## Warning in mu + al %*% la: Recycling array of length 1 in vector-array arithmetic is deprecated.
## Use c() or as.vector() instead.
```

```
## Warning in al - al %*% la: Recycling array of length 1 in vector-array arithmetic is deprecated.
## Use c() or as.vector() instead.
```

```
## -----|-----|-----|-----|-----|-----|
```

```
print(res4)
```

```
## Call:
```

```
## est_lm_cov_manifest(S = S, X = Xc, k = k, mod = "LM", tol = 10^-2)
```

```
##
```

```
## Convergence info:
```

```
##      LogLik np      AIC      BIC
```

```
## [1,] -584.0621 19 1206.124 1262.682
```

```
summary(res4)
```

```
## Call:
```

```
## est_lm_cov_manifest(S = S, X = Xc, k = k, mod = "LM", tol = 10^-2)
```

```
##
```

```
## Coefficients:
```

```
##
```

```
## Vector of cut-points:
```

```
## [1] 4.3911 2.7756 0.9169
```

```
##
```

```
## Support points for the latent states:
```

```
## [1] -2.8593 -0.9389  0.8271  2.9852
##
## Estimate of the vector of regression parameters:
## [1] -0.1342
##
## Vector of initial probabilities:
## [1] 0.0978 0.4097 0.3732 0.1192
##
## Transition matrix:
##      [,1] [,2] [,3] [,4]
## [1,] 0.3847 0.1782 0.3039 0.1332
## [2,] 0.0378 0.7623 0.1105 0.0895
## [3,] 0.0798 0.1221 0.7507 0.0474
## [4,] 0.1255 0.2885 0.1515 0.4345
```

19 parametri liberi di variare La parte del modello latente (prob iniziali e matrice di transizione) non cambia rispetto a prima perchè non è influenzata dalle covariate Cut points sono ordinati, quindi gli stati latenti sono ordinati rispetto alle categorie della var risposta La covariata è continua, quindi ho un parametro di regressione che rappresenta il cambiamento della var risposta rispetto al cambiamento delle unità covariate: esso ha segno negativo, cioè l'effetto dell'età è negativo sul livello del minerale nelle ossa. Per valutare la magnitudine dell'effetto è meglio esponenziare il coefficiente: $\exp(-0.1342)=0.87 \rightarrow$ per ogni 3 anni in più d'età (step di passaggio tra un livello e l'altro di età rilevato), il livello di minerali decresce di 0.87. Quindi il tasso di decrescita è di 0.87 unità in meno di minerali per ogni 3 anni in più di anzianità.

Esercizio 20

Si considerino i dati presenti nei files denominati Sn.Rdata e Xn.Rdata che riguardano una valutazione soggettiva dello stato di salute di 134 anziani cittadini statunitensi effettuata in 8 interviste consecutive (sottoinsieme di dati riferiti a quelli della libreria LMest denominati data SRHS long riguardanti lo studio Health and Retirement Study). Carico i dati

```
load("Sn.Rdata")
load("Xn.Rdata")
```

20.1

Si descrivano le risposte riferite allo stato di salute (che secondo l'Organizzazione Mondiale della Sanità è definito come stato di completo benessere fisico, psichico e sociale e non semplice assenza di malattia) dell'unità con id=16, considerando che le risposte sono state fornite secondo la seguente scala ordinale: 0 (cagionevole), 1 (discreto), 2 (buono), 3 (molto buono), 4 (eccellente).

```
Sn[16,]
```

```
## [1] 2 2 2 1 1 2 3 1
```

Il soggetto 16esimo rivela nelle prime tre interviste di possedere uno stato di benessere buono, che muta in discreto nelle successive due interviste, per poi ritornare buono e molto buono; nell'ultima intervista il paziente dichiara nuovamente di sentirsi in uno stato di salute discreto.

20.2

Si descrivano le covariate disponibili sapendo che la prima covariata è riferita al genere (0 maschio e 1 femmina) e la seconda al titolo di studio (0 senza laurea, 1 con laurea). Per esplorare le frequenze relative alle

due covariate, seleziono per i 134 soggetti solo la risposta relativa alla prima colonna (alla prima intervista) poichè le variabili sono costanti nel tempo.

```
table(Xn[,1,1])
```

```
##
##  0  1
## 115 19
```

```
table(Xn[,1,2])
```

```
##
##  0  1
## 101 33
```

Il dataset contiene 115 uomini e 19 donne; 101 non laureati e 33 laureati

20.3

Si consideri un modello latente di Markov con le covariate che influenzano la parte del modello di misura. Supponendo un modello con 3 stati latenti si illustrino le ipotesi del modello e si stimi il modello.

```
require(LMest)
k <- 3
res3 <- est_lm_cov_manifest(S = Sn, X=Xn,
                           k = k,
                           mod = "LM")
```

```
## n. distinct covariate conf. = 4
## EM step:
```

| k | start | step | lk | lk-lko | discrepancy |
|---|-------|------|----------|-------------|-------------|
| 3 | 0 | 0 | -1533.26 | | |
| 3 | 0 | 10 | -1375.91 | 0.0949665 | 0.0349597 |
| 3 | 0 | 20 | -1375.62 | 0.00912203 | 0.0108022 |
| 3 | 0 | 30 | -1375.58 | 0.0018933 | 0.00312638 |
| 3 | 0 | 40 | -1375.57 | 0.00038364 | 0.00167968 |
| 3 | 0 | 50 | -1375.57 | 6.75168e-05 | 0.000738149 |
| 3 | 0 | 60 | -1375.57 | 1.37029e-05 | 0.000306671 |

```
## Warning in mu + al %*% la: Recycling array of length 1 in vector-array arithmetic is deprecated.
## Use c() or as.vector() instead.
```

```
## Warning in al - al %*% la: Recycling array of length 1 in vector-array arithmetic is deprecated.
## Use c() or as.vector() instead.
```

```
## -----|-----|-----|-----|-----|-----|
```

Abbiamo 60 iterazioni dell'algoritmo EM.

```
print(res3)
```

```
## Call:
## est_lm_cov_manifest(S = Sn, X = Xn, k = k, mod = "LM")
##
## Convergence info:
##      LogLik np      AIC      BIC
## [1,] -1375.566 14 2779.132 2819.702
```

```
summary(res3)
```

```
## Call:
## est_lm_cov_manifest(S = Sn, X = Xn, k = k, mod = "LM")
##
## Coefficients:
##
##   Vector of cut-points:
## [1]  4.0303  1.2194 -1.3332 -3.9646
##
##   Support points for the latent states:
## [1] -3.8350 -0.4261  3.2445
##
##   Estimate of the vector of regression parameters:
## [1] -0.6930  1.9844
##
##   Vector of initial probabilities:
## [1] 0.1939 0.5099 0.2962
##
##   Transition matrix:
##           [,1]  [,2]  [,3]
## [1,] 0.9037 0.0338 0.0624
## [2,] 0.0344 0.9101 0.0555
## [3,] 0.0038 0.1326 0.8635
```

Il modello stima 14 parametri liberi del modello. Come output del summary abbiamo: il vettore dei cut-points cioè le intercette $MU(y)$ cioè i punti in cui le categorie della var risposta vengono interrotte; i support points cioè le intercette riferite ai 3 stati latenti: osservandoli vediamo che gli stati sono ordinati in senso crescente, cioè il primo stato latente raggruppa soggetti con livelli di benessere minore; il vettore dei parametri di regressione che rappresentano gli odds, il primo riferito al sesso, il secondo alla laurea; le prob iniziali: vediamo che il 50% appartiene inizialmente allo stato latente intermedio, cioè di coloro che hanno benessere medio, mentre il 30% inizialmente si trova allo stato latente di coloro che stanno meglio; la matrice di transizione: guardando la diagonale osserviamo che le prob di rimanere nello stesso stato latente è molto alta per ogni stato.

Spiegazione dei parametri di regressione per le due covariate

```
exp(-0.6930)
```

```
## [1] 0.5000736
```

```
exp(1.9844)
```

```
## [1] 7.274681
```

L'odds delle donne rispetto agli uomini: $\exp(-0.69)=0.5 \rightarrow$ le donne percepiscono uno stato di salute meno buono nel tempo rispetto ai maschi, quasi la metà. L'odds dei laureati vs non laureati: $\exp(1.9844) = 7.27 \rightarrow$ i laureati percepiscono uno stato di salute molto migliore nel tempo rispetto ai non laureati, circa 7 volte maggiore