

UNIVERSITÀ DEGLI STUDI DI FIRENZE

CURRICULUM DATA SCIENCE

INFERENZA STATISTICA BAYESIANA

---

## Quaderno degli esercizi

---

*Studente*

FILIPPO MAMELI

`filippo.mameli@stud.unifi.it`

Anno accademico 2017-2018

# 1 Prima Parte

## 2 Seconda Parte

### Esercizio 11.2 pag. 246 Hoff

Randomized block design: researchers interested in identifying the optimal planting density for a type of perennial grass performed the following randomized experiment: ten different plots of land were each divided into eight subplots, and planting densities of 2, 4, 6 and 8 plants per square meter were randomly assigned to the subplots, so that there are two subplots at each density in each plot. At the end of the growing season the amount of plant matter yield was recorded in metric tons per hectare. These data appear in the file `pdensity.dat`. The researchers want to fit a model like  $y = \beta_1 + \beta_2 x + \beta_3 x^2 + \epsilon$ , where  $y$  is yield and  $x$  is planting density, but worry that since soil conditions vary across plots they should allow for some across-plot heterogeneity in this relationship. To accommodate this possibility we will analyze these data using the hierarchical linear model described in Section 11.1.

1. Before we do a Bayesian analysis we will get some ad hoc estimates of these parameters via least squares regression. Fit the model  $y = \beta_1 + \beta_2 x + \beta_3 x^2 + \epsilon$  using OLS for each group, and make a plot showing the heterogeneity of the least squares regression lines. From the least squares coefficients find ad hoc estimates of  $\theta$  and  $\Sigma$ . Also obtain an estimate of  $\sigma^2$  by combining the information from the residuals across the groups.
2. Now we will perform an analysis of the data using the following distributions as prior distributions:

$$\Sigma^{-1} \sim \text{Wishart}(4, \hat{\Sigma}^{-1})$$

$$\theta \sim \text{multivariate normal}(\hat{\theta}, \hat{\Sigma})$$

$$\sigma^2 \sim \text{inverse-gamma}(1, \hat{\sigma}^2)$$

where  $\hat{\theta}, \hat{\Sigma}, \hat{\sigma}^2$  are the estimates you obtained in a). Note that this analysis is not combining prior information with information from the data, as the “prior” distribution is based on the observed data. However,

such an analysis can be roughly interpreted as the Bayesian analysis of an individual who has weak but unbiased prior information.

3. Use a Gibbs sampler to approximate posterior expectations of  $\beta$  for each group  $j$ , and plot the resulting regression lines. Compare to the regression lines in a) above and describe why you see any differences between the two sets of regression lines.
4. From your posterior samples, plot marginal posterior and prior densities of  $\theta$  and the elements of  $\Sigma$ . Discuss the evidence that the slopes or intercepts vary across groups.
5. Suppose we want to identify the planting density that maximizes average yield over a random sample of plots. Find the value  $x_{max}$  of  $x$  that maximizes expected yield, and provide a 95% posterior predictive interval for the yield of a randomly sampled plot having planting density  $x_{max}$ .

### **Premessa e indicazioni generali**

Si vuole valutare la relazione tra raccolto e densità di piante relativamente a 10 lotti di terra osservati nel seguente modo:

- 10 lotti di terra.
- Un lotto comprende altri 8 sottolotti.
- Densità di piantagione pari a 2, 4, 6 e 8 sono assegnate in maniera casuale tra gli 8 sottolotti in modo tale che ogni lotto ha due sottolotti di ognuna delle densità.

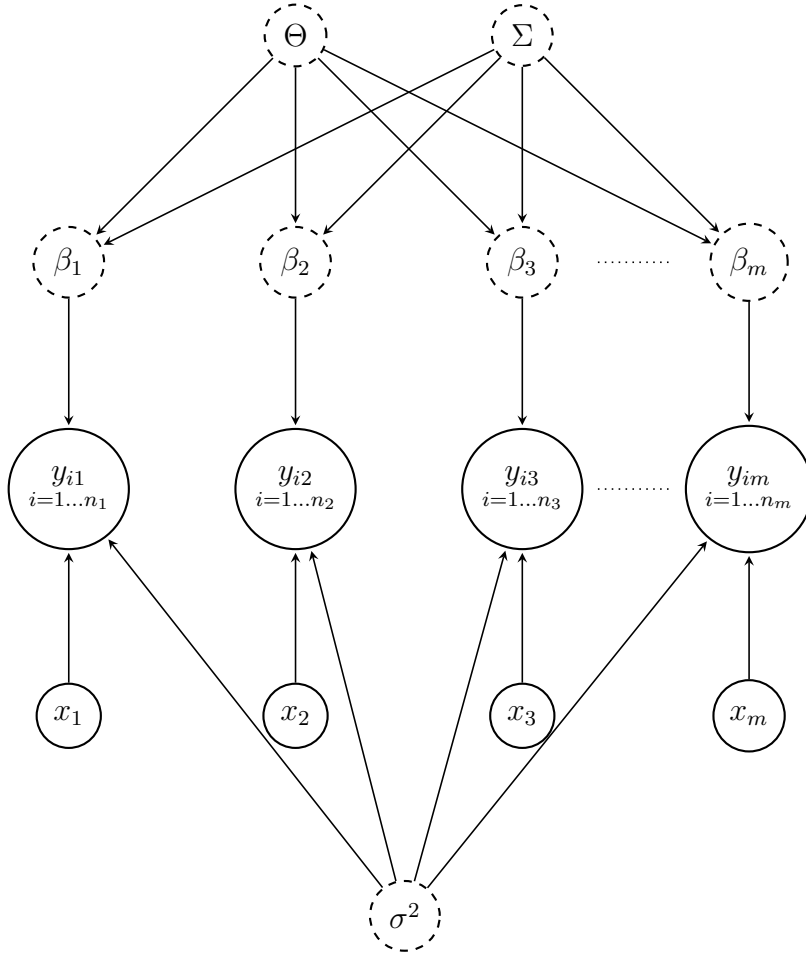
È richiesta l'analisi della relazione tra raccolto e densità mediante un modello di regressione lineare quadratica e tenendo conto della variabilità tra gruppi in termini di condizioni di suolo. Per come è costruito il disegno e per come è formulato il modello procediamo nell'analisi attraverso un modello di regressione lineare gerarchico. Il setting del modello è:

- $Y_{ij} = \beta_j^T x_{ij} + \varepsilon_{ij};$
- $\varepsilon_{ij} | \sigma^2 \sim i.i.d.N(0, \sigma^2).$

cioè:

- $Y_j|\beta_j, X_j, \sigma^2 \sim N_{n_j}(X_j\beta_j, \sigma^2 I) \quad i = 1, \dots, n; \quad j = 1, \dots, m;$
- $\beta_j|\theta, \Sigma \stackrel{iid}{\sim} N_p(\theta, \Sigma);$
- $\theta|\mu_0, \Lambda_0 \sim N_p(\mu_0, \Lambda_0);$
- $\Sigma \sim \text{Inverse - Wishart}(\eta_0, S_0 - 1);$
- $\sigma^2|v_0, \sigma_0^2 \sim \text{Inverse - Gamma}(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2});$
- $Y_i \perp Y_j|\beta_1, \dots, \beta_m, \sigma^2, i \neq j.$

Il DAG relativo è mostrato in figura seguente.



In seguito si riporta il codice R.

```
1 rmvnorm<-function(n,mu, Sigma ) {
  p<-length(mu)
3   res<-matrix( 0, nrow=n, ncol=p)
  if ( n>0 & p>0 ) {
5     E<-matrix( rnorm(n*p),n, p)
     res<-t ( t (E%%chol ( Sigma ) ) +c (mu) )
7     #R <- chol(A)
  }
9   res
11 }

rinvwish <- function(n, nu0, iS0)
13 {
  sL0 <- chol(iS0)
15   S <- array(dim = c (dim(iS0), n))
  for (i in 1:n)
17   {
    Z <- matrix(rnorm(nu0 * dim(iS0)[1]), nu0, dim(iS0)[1]) %% sL0
19    S[, , i] <- solve(t(Z) %% Z)
  }
21   S[, , 1:n]
  }

23 rwish <- function(n, nu0, S0)
  {
25   sS0 <- chol(S0)
   S <- array(dim = c(dim(S0), n))
27   for (i in 1:n)
   {
29     Z <- matrix(rnorm(nu0 * dim(S0)[1]), nu0, dim(S0)[1]) %% sS0
     S[, , i] <- t(Z) %% Z
31   }
   S[, , 1:n]
33 }

35 #Lettura dei dati:

37 dati<-read.table("~/git/notebookInferenzaBayesiana/code/pdensity.dat
  ", header=TRUE)
```

```
> head (dati)
```

```
plot density yield
1          2      8.25
1          2      5.81
1          4      8.69
1          4      8.03
1          6      7.96
1          6      8.89
```

```
1 #Calcolo di quantita' utili per ogni gruppo ( numerosita', vettore
  delle osservazioni, matrice del modello e numero di parametri ):
ids<-unique (dati$plot )
3 m<-length ( ids )
  Y<-list() ; X<-list() ; N<-NULL
5 for ( j in 1:m){
  Y[[ j ]]<-dati [ dati [,1]== ids [ j ], 3]
7   N[j]<- sum( dati$plot==ids [ j ])
  xj<-dati [ dati [,1]== ids [ j ], 2]
9   X[[ j ]]<-cbind ( rep (1,N[ j ]), xj, xj^2 )
  }
11 p<-dim(X[[1]])[2]
```

```
> N
```

```
[1] 8 8 8 8 8 8 8 8 8 8
```

```
> Y[ [ 1 ] ]
```

```
[1] 8.25 5.81 8.69 8.03 7.96 8.89 6.13 9.40
```

```
> X[ [ 1 ] ]
```

```
      xj
[1,] 1 2 4
[2,] 1 2 4
[3,] 1 4 16
[4,] 1 4 16
[5,] 1 6 36
[6,] 1 6 36
```

```
[7,] 1 8 64
[8,] 1 8 64
```

### Punto a)

Stimiamo i coefficienti di regressione con il metodo dei minimi quadrati in modo indipendente per ciascuno dei 10 lotti.

```
S2.OLS<-BETA.OLS<-NULL
2 for( j in 1:m){
  fit<-lm(Y[[j]]~-1+X[[j]])
4  BETA.OLS<-rbind (BETA.OLS, c(fit$coef ) )
  S2.OLS<-c ( S2.OLS, summary( fit )$sigma ^2)
6 }

8 colnames (BETA.OLS)<-c ("1"," x "," x^2")
```

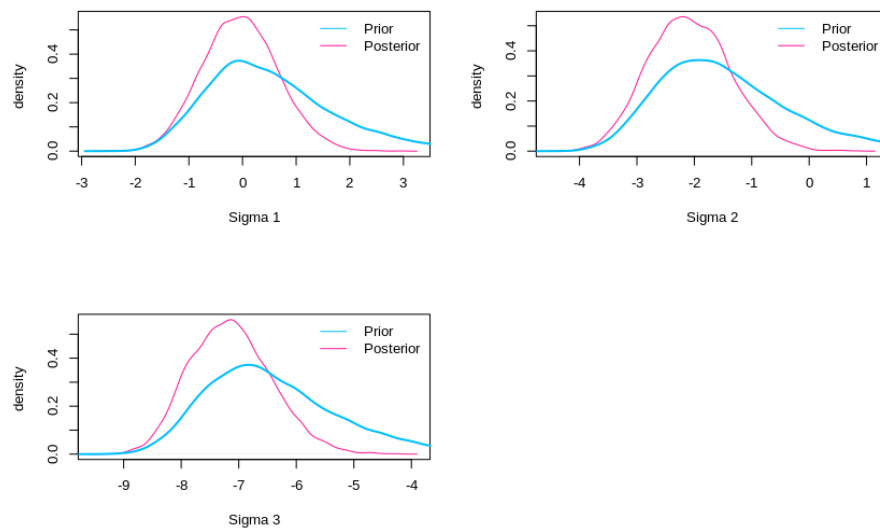
```
> BETA.OLS
```

	1	x	x^2
[1,]	4.84000	1.357250	-0.1243750
[2,]	4.53375	1.193375	-0.1290625
[3,]	2.07750	2.128250	-0.1643750
[4,]	2.60375	2.114875	-0.1928125
[5,]	3.57000	1.540500	-0.1500000
[6,]	1.47375	1.930875	-0.1215625
[7,]	3.96375	1.424875	-0.1278125
[8,]	0.52375	2.941875	-0.2653125
[9,]	3.36250	1.675500	-0.1400000
[10,]	1.73875	2.241125	-0.1771875

```
> S2.OLS
```

```
[1] 1.8005320 1.0760545 0.8134580 0.5019505 0.5886680 0.8074545 0.9575905
[8] 0.3965025 0.1328380 0.8030505
```

Rappresentiamo graficamente le 10 linee di regressione per valutare la variabilità tra gruppi. Riportiamo sul grafico anche la loro media.



**Figura 1:** Curve di regressione con stime OLS.

```
par (mfrow=c (1,2) )
2 plot ( range ( dati [,2]), range ( dati [,3]), type="n", xlab="
    planting density ", ylab="Expected yield ",main="Regression lines
    OLS")
for (j in 1:m) {
4   curve (BETA.OLS[j, 1] + BETA.OLS[j, 2] * x + BETA.OLS[j, 3] * x ^ 2
      , col = "deepskyblue", lty = 2, add = T)
}
6 BETA.OLS.MEAN <- apply (BETA.OLS, 2, mean)
curve (BETA.OLS.MEAN[1] + BETA.OLS.MEAN[2] * x + BETA.OLS.MEAN[3] *
      x ^ 2,lwd = 2, col = "deeppink", add = T)

> BETA.OLS.MEAN
```

```
      1      x      x^2
2.86875 1.85485 -0.15925
```



```
> cov(BETA.OLS)

      1      x      x^2
1 2.00120764 -0.69321313 0.044309549
x -0.69321313 0.27555421 -0.020742679
x 2 0.04430955 -0.02074268 0.001968451

> mean(S2.OLS)

[1] 0.7878099
```

Si osserva che:

- Tutte le curve hanno lo stesso andamento quadratico.
- Alcuni gruppi si discostano molto dalla media generale.

### Punto b)

Nel codice seguente impostiamo i parametri delle a priori semiconiugate per theta, Sigma e sigma2 come richiesto.

```
1 theta<-BETA.OLS.MEAN; Sigma<-cov(BETA.OLS) ; sigma2<-mean(S2.OLS)
  eta0 <-4; S0<-Sigma
3 mu0<-theta ; L0<-Sigma
  v0<-2; sigma20<-sigma2
```

### Punto c)

Impostiamo adesso i parametri delle a priori ed un Gibbs sampler, iniziando, in ordine, il numero di simulazioni, i valori iniziali dei parametri, e le strutture dati che serviranno per salvare i valori campionati dalle full conditional.

```
nsimul=10000
2 beta<-BETA.OLS
  Sigma<-cov(BETA.OLS)
4 sigma2<-mean(S2.OLS)

6 Sigma.post<-matrix (0,p,p)
```

```

BETA.pp<-THETA.POST<-S2.POST<-NULL
8 BETA.POST<-BETA.OLS*0
SIGMA.POST<-array (0, c(p,p, nsimul ) )
10 set.seed (1)

12 for ( s in 1: nsimul ) {
  for ( j in 1:m){
14     Vj<-solve ( solve (Sigma) + t(X[[j]])%*%X[[j]]/sigma2 )
     Ej<-Vj%*%( solve (Sigma)%*%theta + t(X[[j]])%*%Y[[j]]/sigma2 )
16     beta [ j,]<-rmvnorm(1,Ej, Vj)
  }
18
  #theta
20  Lm<- solve ( solve (L0) + m* solve (Sigma) )
  mum<- Lm%*%( solve (L0)%*%mu0 + solve (Sigma)%*%apply ( beta,2,sum)
    )
22  theta<-t (rmvnorm(1,mum,Lm) )

24  #Sigma
  mtheta<-matrix ( theta,m,p, byrow=TRUE)
26  Sigma<-solve ( rwish (1, eta0+m, solve ( S0+t ( beta-mtheta)%*%(
    beta-mtheta) ) ) )
  #Campioniamo la varianza residua:
28
  RSS<-0
30  for ( j in 1:m) {
    RSS<-RSS+sum((Y[[j]]-X[[j]]%*%beta[j,] )^2 )
32  }
  sigma2<-1/rgamma(1,( v0+sum(N) ) /2, (v0*sigma20+RSS)/2 )
34  #Immagazziniamo i valori appena campionati:
  S2.POST<-c(S2.POST, sigma2 ) ;THETA.POST<-rbind (THETA.POST, t (
    theta ) )
36  Sigma.post<-Sigma.post+Sigma ; BETA.POST<-BETA.POST+beta
  SIGMA.POST[, , s]<-Sigma
38  #Campioniamo dalla posterior pdeppinkictive dei coefficienti di
    regressione che ci servira' per il punto e).
  BETA.pp<-rbind (BETA.pp,rmvnorm(1, theta, Sigma) )
40 }

42 colnames (THETA.POST)<-c(" theta1 ", " theta2 ", " theta3 ")

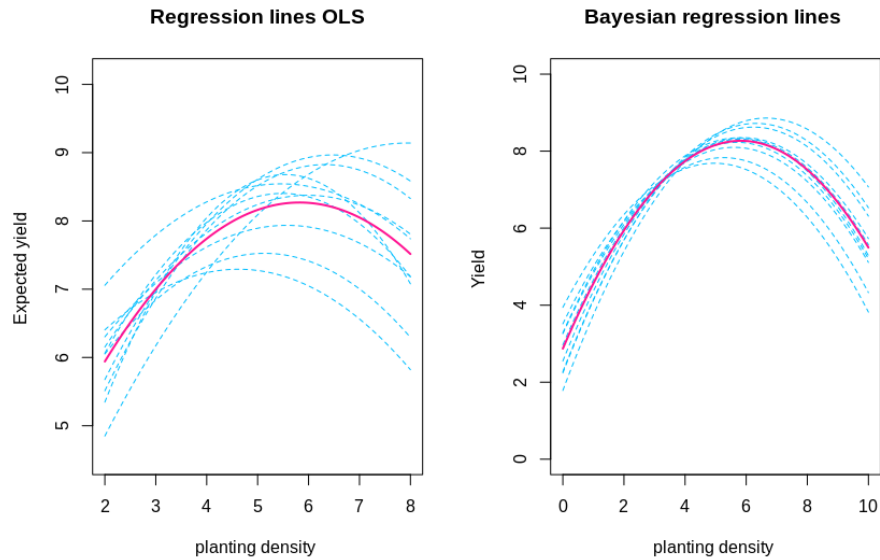
```

```

colnames (BETA.POST)<-colnames (BETA.pp)<-c(" beta1 ", " beta2 ", "
  beta3 ")
44 BETA.PM<-BETA.POST/nsimul
plot (range (c (0, 10)), range (c (0, 10)), type = "n", xlab = "
  planting density ", ylab = "Yield ", main = "Bayesian regression
  lines ")
46 for (j in 1:m) {
  curve (BETA.PM[j, 1] + BETA.PM[j, 2] * x + BETA.PM[j, 3] * x ^ 2,
    lty = 2, col = "deepskyblue ", add = T)
48 }
curve (mean(THETA.POST[, 1]) + mean(THETA.POST[, 2]) * x + mean(
  THETA.POST[, 3]) * x ^ 2, lwd = 2, add = T, col = "deeppink")

```

Osservando i due grafici a confronto si nota che il modello gerarchico permette di trarre informazioni dai gruppi, riportando le curve di regressione lungo la media. Dal momento che lavoriamo con gruppi tutti di piccola numerosità c'è una grande variabilità nelle stime OLS mentre nel caso del modello gerarchico i gruppi si influenzano in termini di informazioni e per l'effetto di shrinkage la stima OLS viene portata verso la stima media in modo uguale per tutti i gruppi perché hanno la stessa numerosità campionaria.



**Figura 2:** GLMM: stime dei minimi quadrati e stime bayesiane a confronto.

Controlliamo adesso la convergenza dell'algoritmo.

```
> effectiveSize (THETA.POST)
```

```
theta1  theta2  theta3
6938.395 7144.829 7501.045
```

#### Punto d)

Approssimazione delle distribuzioni a priori tramite simulazione Monte Carlo

```
1 n<-10000
3 THETA.PRIOR<-rmvnorm(n,mu0,L0) ; colnames (THETA.PRIOR)<-colnames (
  THETA.POST)
SIGMA.PRIOR<-rinvwish(n, eta0, solve( S0 ) )
```

```
> head(THETA.PRIOR)
```

	theta1	theta2	theta3
[1,]	-1.77780314	3.529232	-0.2613000
[2,]	3.47065885	1.752018	-0.1395134
[3,]	3.60717709	1.583506	-0.1350985
[4,]	0.07323521	2.928965	-0.2604753
[5,]	3.05388848	1.832459	-0.1571405
[6,]	2.27382048	1.786574	-0.1183835

```
> head(SIGMA.PRIOR)
```

```
[1] 3.63026418 -0.93710237 0.04650011 -0.93710237 0.28263581 -0.01833983
```

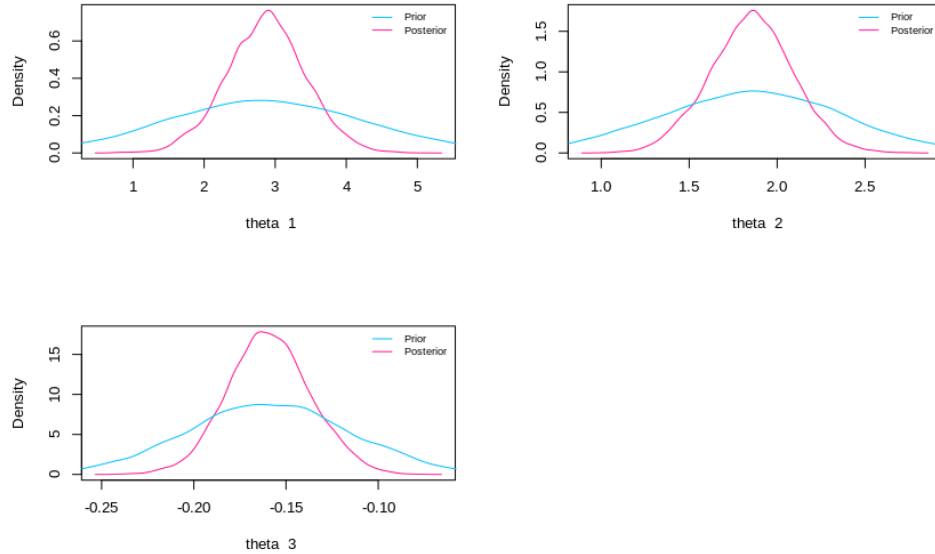
```
#A priori e a posteriori di theta a confronto ( plot ):
```

```
2 par (mfrow = c(2, 2))
  for (i in 1:3) {
4   plot (density (THETA.POST[, i]), xlab = paste (" theta ", i), main
        = "", col = "deeppink")
        lines (density (THETA.PRIOR[, i]), col = "deepskyblue ")
6   legend ("topright", legend = c(" Prior ", " Posterior "), col = c("
        deepskyblue ", " deeppink "), lty = 1, bty = "n", cex = 0.7)
  }
```

Si osserva che le distribuzioni a posteriori, pur essendo centrate sulla stessa media delle a priori (come ci aspettavamo dal momento che abbiamo centrato le prior nelle stime di massima verosimiglianza ), sono meno diffuse: in questo modo l'informazione a posteriori cambia nel senso che si dà maggiore probabilità ai valori che cadono attorno ad essa.

```
1 #Prior e posterior di Sigma a confronto ( plot ):
```

```
3 par (mfrow = c(2, 2))
  for (i in 1:3) {
5   plot (density (log(SIGMA.POST[i, i])), type = "l", xlab = paste ("
        Sigma", i), col = "deeppink", ylab = "density ", main = "")
        lines (density (log(SIGMA.PRIOR[i, i])), col = "deepskyblue ", lwd
        = 2)
7   legend("topright", legend = c(" Prior ", " Posterior "), col = c("
        deepskyblue ", " deeppink "), bty = "n", lty = 1)
  }
```



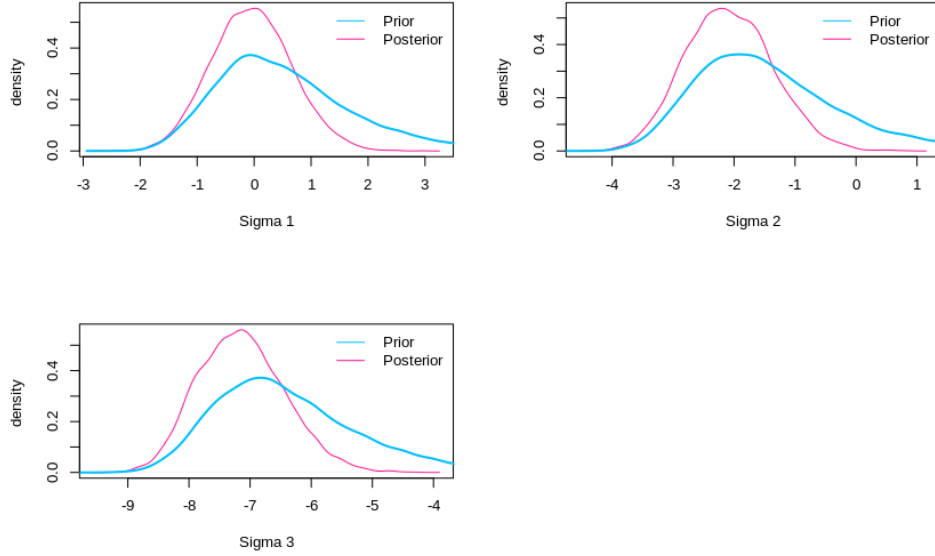
**Figura 3:** Densità a priori e a posteriori a confronto (media).

Abbiamo preso il logaritmo degli elementi di Sigma dal momento che sono valori molto bassi. Dal momento che è richiesto di commentare la variabilità relativa all'intercetta e alle pendenze nei gruppi, riportiamo i plot delle distribuzioni dei soli elementi della diagonale principale di Sigma.

La variabilità tra i gruppi delle intercette non è molto elevata, ma diminuisce ancora per il primo e per il secondo coefficiente.

### Punto e)

Si vuole trovare la densità di piante che massimizza il raccolto atteso per un campione di lotti. Ricordiamo che stiamo lavorando con un modello gerarchico e quindi vogliamo tenere in considerazione anche la variabilità tra gruppi: ci serviamo a questo scopo della distribuzione predittiva a posteriori dei coefficienti di regressione (una normale multivariata con vettore delle medie e matrice di varianza e covarianza pari quelli estratti ad ogni passo dell'algoritmo) per cui ogni vettore di coefficienti estratto rappresenta il vettore dei coefficienti di regressione per un gruppo futuro. Disponiamo già



**Figura 4:** Densità a priori e a posteriori a confronto (variabilità e covariabilità).

di tale distribuzione dal momento che è stata calcolata durante l'algoritmo. Necessario adesso confrontare 4 distribuzioni: quelle dei valori attesi del raccolto di un generico lotto, una per ogni valore osservato della covariata  $x$ . Anche in questo caso vogliamo tenere in considerazione la variabilità tra lotti e per questo approssimiamo le distribuzioni usando la predittiva a posteriori dei coefficienti di regressione appena discussa. Cerchiamo il valore atteso di  $y$  date le  $y$  passate e le  $x$ . Un modo per farlo è campionare i beta dalla loro posterior predictive e campionare le  $y$  dati i valori della  $x$  e quindi per ogni valore della  $x$  potevamo avere un valore atteso. Facciamo la media delle distribuzioni dei valori attesi. Campioniamo i beta tilde: i beta per un gruppo futuro. Per ogni valore di questo beta campinato abbiamo un valore del valore atteso per un  $y$  futuro. Alla fine per ogni  $x$  si genera una distribuzione del valore atteso della  $y$  futuro (4 vettori). Confrontiamo le 4 distribuzioni ( per es con la media o vedendo la probabilità che una sia maggiore dell'altra ). Prendendo il valore medio per ogni valore di beta tilde e per  $x$  e la varianza a posteriori campiono un valore  $y$ . Così si incorpora anche l'incertezza derivante dai gruppi.

```

x0<-c (2,4,6,8)
2 raccolto<-NULL
  for ( i in x0){
4   x<-c (1, i, i ^2)
    raccolto<-cbind ( raccolto,BETA.pp%*%x)
6 }
colnames ( raccolto )<-x0

```

```
> head( raccolto )
```

```

           2           4           6           8
[1,] 5.989406 7.758386 8.317531 7.666840
[2,] 6.033526 7.921516 8.605911 8.086712
[3,] 6.350655 7.597301 7.712147 6.695192
[4,] 6.888414 7.776763 7.907158 7.279599
[5,] 5.973396 7.669427 8.482208 8.411740
[6,] 6.247427 7.685203 7.955526 7.058398

```

Confrontiamo adesso le quattro distribuzioni dei valori attesi, graficamente e con un indice sintetico, la media:

```

1 par (mfrow = c (1, 1))
  hist (raccolto [, 1], prob = T, main = "Raccolto atteso a posteriori
    ", xlab = "Raccolto ", xlim = c (2, 13), ylim = c (0, 1.5), col
    = "deepskyblue")
3 lines (density (raccolto [, 1]), col = "deepskyblue")
  hist (raccolto [, 2], prob = T, col = "deeppink ", add = T)
5 lines (density (raccolto [, 2]), col = "deeppink")
  hist (raccolto [, 3], prob = T, col = " chartreuse ", add = T)
7 lines (density (raccolto [, 3]), col = " chartreuse ")
  hist (raccolto [, 4], prob = T, col = " darkorange ", add = T)
9 lines (density (raccolto [, 4]), col = "darkorange ")
  legend ("topright",legend = c("x=2", "x=4", "x=6", "x=8"),col = c("
    deepskyblue", " deeppink ", "chartreuse", " lightsalmon "), lty =
    1)

```

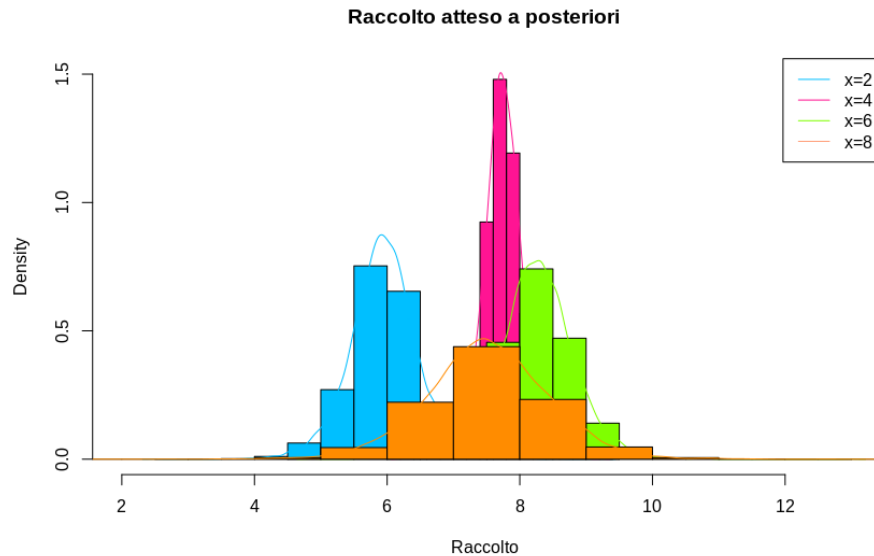
```
> apply ( raccolto,2,mean)
```

```

           2           4           6           8
5.942984 7.739188 8.264142 7.517844

```





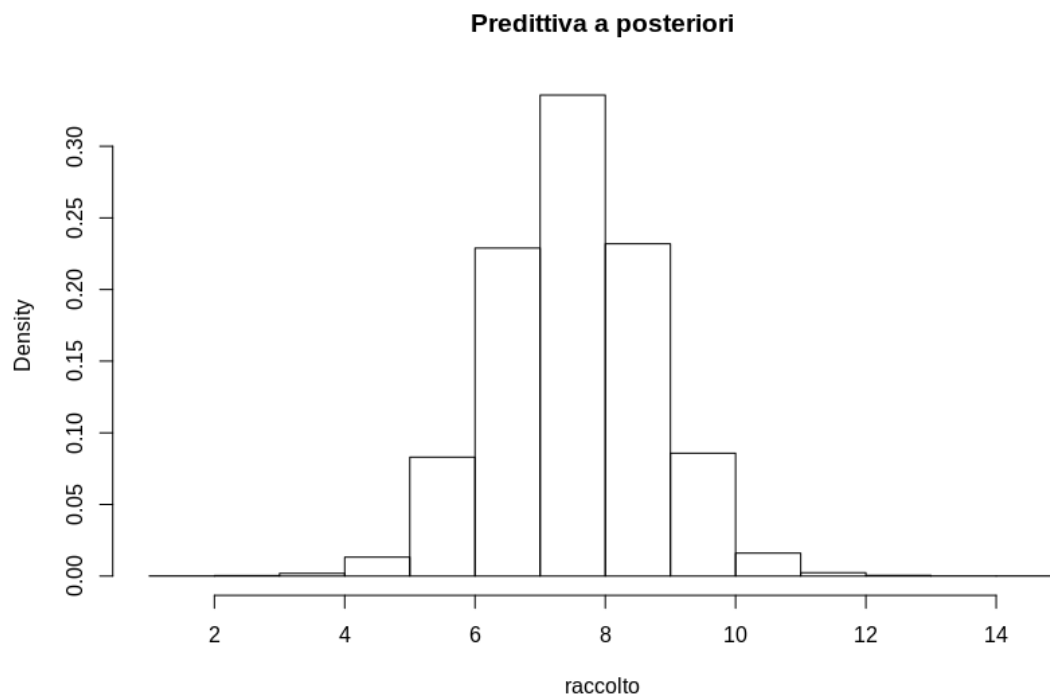
**Figura 5:** Valori attesi a posteriori per più valori di x a confronto.

Si osserva che il valore della covariata che massimizza il raccolto atteso per un generico lotto è  $x=6$ . Procediamo quindi considerando tale valore per il predittore lineare. Si vuole infine approssimare la distribuzione predittiva a posteriori per un generico lotto avendo una densità di piantagioni pari a 6. La logica è la stessa di quella usata per la predittiva a posteriori dei coefficienti di regressione, con la differenza che adesso siamo al livello più basso della gerarchia e quindi si aggiunge una parte di variabilità dovuta alle osservazioni campionarie. Per ogni vettore dei coefficienti estratto dalla predittiva a posteriori e per ogni elemento estratto dalla a posteriori della varianza residua (anche questo fatto già fatto nell'algoritmo) campioniamo un valore del raccolto da una normale con media pari al predittore lineare e varianza pari alla varianza residua. Riportiamo infine un plot della densità e l'intervallo di confidenza richiesto per tale distribuzione:

```
y.pred<-rnorm( nsimul,BETA.pp*%x, S2.POST)
```

```
> quantile (y.pred, c (0.025,0.975))
```

```
      2.5%      97.5%
5.213170 9.855576
```



**Figura 6:** Predittiva a posteriori.

```
1 hist(y.pred, prob=T, main="Predittiva a posteriori ", xlab="raccolto")  
   lines (y.pred )
```