Esercizio Bayesian

Giulio Vannini

14 giugno 2017

Dati Forniti:

Y <- c(7, 8, 9, 4, 7, 1, 8, 5, 6, 7, 0, 1, 6, 7, 9)  
N <- sum(xtabs(~Y))  
  
alpha0 <- 10 #totale mail arrivate in   
beta0 <- 3 #settimane  
alpha1 <- alpha0 + sum(Y)  
beta1 <- beta0 + N

1. Implementazione del modello

library(rstan)  
data <- list(N, Y)  
  
modPG <- '  
data {  
int N;  
int Y[N];  
int alpha1;  
int beta1;  
}  
transformed data{  
  
}  
  
parameters {  
real<lower=0> Lambda;  
  
}  
transformed parameters {  
  
}  
model {  
Lambda ~ gamma(alpha1,beta1);  
Y ~ poisson(Lambda);  
}  
'  
  
require(rstan)  
iniTime <- date()  
mod <- stan\_model(model\_code = modPG)

## In file included from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/base.hpp:28:0,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array.hpp:21,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/numeric/odeint/util/multi\_array\_adaption.hpp:29,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/numeric/odeint.hpp:61,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/prim/arr/functor/integrate\_ode\_rk45.hpp:13,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/prim/arr.hpp:36,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/prim/mat.hpp:235,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/mat.hpp:9,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math.hpp:4,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/src/stan/model/model\_header.hpp:4,  
## from file3010451f73fd4.cpp:8:  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp: In static member function 'static void boost::multi\_array\_concepts::detail::idgen\_helper<N>::call(Array&, const IdxGen&, Call\_Type)':  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:42:43: warning: typedef 'index\_range' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index\_range index\_range;  
## ^  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:43:37: warning: typedef 'index' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index index;  
## ^  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp: In static member function 'static void boost::multi\_array\_concepts::detail::idgen\_helper<0ull>::call(Array&, const IdxGen&, Call\_Type)':  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:53:43: warning: typedef 'index\_range' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index\_range index\_range;  
## ^  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:54:37: warning: typedef 'index' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index index;  
## ^  
## In file included from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/core.hpp:42:0,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/mat.hpp:4,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math.hpp:4,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/src/stan/model/model\_header.hpp:4,  
## from file3010451f73fd4.cpp:8:  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/core/set\_zero\_all\_adjoints.hpp: At global scope:  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/core/set\_zero\_all\_adjoints.hpp:14:17: warning: 'void stan::math::set\_zero\_all\_adjoints()' defined but not used [-Wunused-function]  
## static void set\_zero\_all\_adjoints() {  
## ^

endTime <- date();c(iniTime = iniTime, endTime = endTime)

## iniTime endTime   
## "Thu Jun 15 12:58:27 2017" "Thu Jun 15 12:59:03 2017"

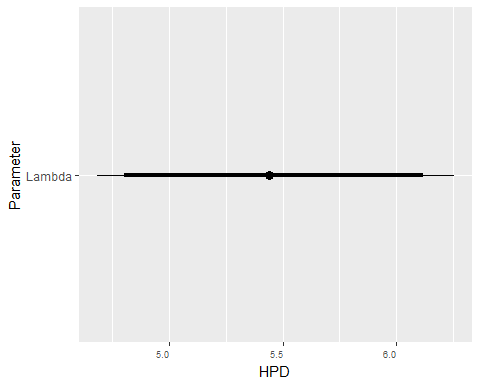
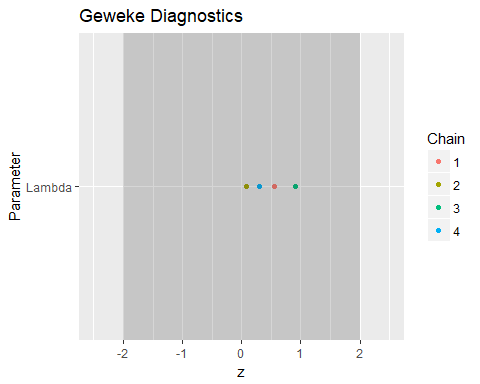
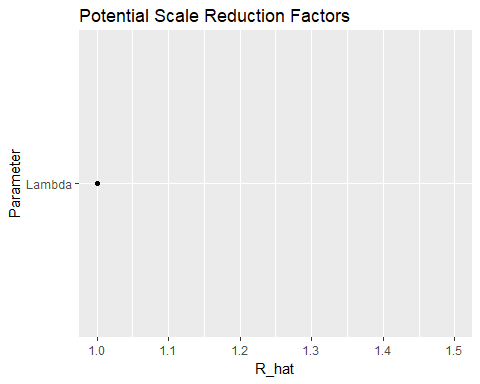
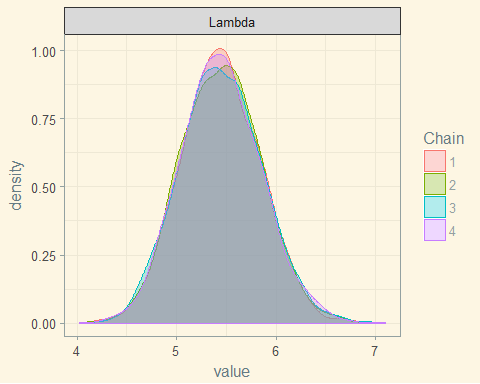
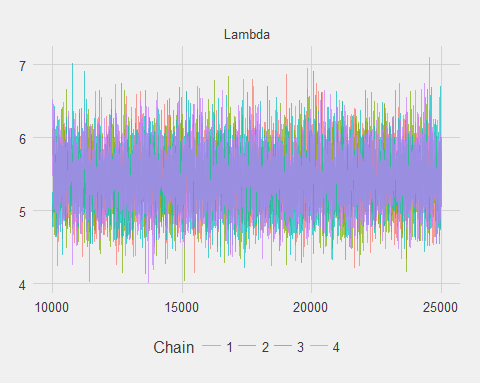
1. Ottieni un campione simulato dalla a-posteriori

library(rstudioapi)  
system.time(  
 fit <<- sampling(mod,  
 data = data,  
 pars = c("Lambda"), #a noi interessa capire la distribuzione di Lambda, non di Y.   
#Perchè dopo che si campiona lambda si capisce come distribuisce la Y  
 chains = 4,  
 iter = 25000,  
 warmup = 10000,  
 thin = 5,  
 cores = 4,  
 seed = 19861986,  
 control = list(max\_treedepth = 10,  
 adapt\_delta = 0.8)  
 )  
)

## user system elapsed   
## 0.93 0.52 7.67

1. Esplora le diagnostiche di output e discuti i risultati

## # A tibble: 12,000 ž 4  
## Iteration Chain Parameter value  
## <dbl> <int> <fctr> <dbl>  
## 1 1 1 Lambda 5.076613  
## 2 2 1 Lambda 5.019012  
## 3 3 1 Lambda 4.790655  
## 4 4 1 Lambda 5.224815  
## 5 5 1 Lambda 5.083689  
## 6 6 1 Lambda 5.354110  
## 7 7 1 Lambda 5.455608  
## 8 8 1 Lambda 6.453824  
## 9 9 1 Lambda 5.661801  
## 10 10 1 Lambda 6.200843  
## # ... with 11,990 more rows



## # A tibble: 1 ž 6  
## Parameter low Low median High high  
## \* <fctr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Lambda 4.680829 4.80044 5.44176 6.119154 6.255451

## Length Class Mode   
## 1 stanfit S4

1. Riassumi le caratteristiche principali della distribuzione a-posteriori

A vedere dal traceplot e dal density plot la sovrapposizione della 4 catene è buona. Magari dal density plot soprattutto si nota qualche discrepanza in più Gelman invece mostra che non ci sono fattori di potenziale riduzione di scala per Lambda. Il che avalla la tesi sulla discreta bontà del modello. Bontà avallata anche da Geweke, dato che tutte le catene sono comprese su valori Z fra -2 e 2 Il grafico Caterpillar mostra invece che la credibilità di lambda è alta fra circa 4,5 e 6,25 gg\_pairs non funziona perchè servono almeno due parametri (in questo caso abbiamo solo Lambda!) La funzione summary() aiuta a monitorare numericamente quanto detto fin'ora: vale a dire che le catene si sovrappongono e non ci sono problemi con il modello.

1. Ottieni la distribuzione prededittiva della futura osservazione con un campione di dimensione 10k.

modPGp <- '  
data {  
int N;  
int Y[N];  
int alpha1;  
int beta1;  
}  
parameters {  
real<lower=0> Lambda;  
  
}  
transformed parameters {  
  
}  
model {  
Lambda ~ gamma(alpha1,beta1);  
Y ~ poisson(Lambda);  
}  
generated quantities{  
real Y\_predict;  
Y\_predict = poisson\_rng(Lambda);  
}  
'  
iniTime <- date()  
mod <- stan\_model(model\_code = modPGp)

## In file included from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/base.hpp:28:0,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array.hpp:21,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/numeric/odeint/util/multi\_array\_adaption.hpp:29,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/numeric/odeint.hpp:61,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/prim/arr/functor/integrate\_ode\_rk45.hpp:13,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/prim/arr.hpp:36,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/prim/mat.hpp:235,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/mat.hpp:9,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math.hpp:4,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/src/stan/model/model\_header.hpp:4,  
## from file30104514d6683.cpp:8:  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp: In static member function 'static void boost::multi\_array\_concepts::detail::idgen\_helper<N>::call(Array&, const IdxGen&, Call\_Type)':  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:42:43: warning: typedef 'index\_range' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index\_range index\_range;  
## ^  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:43:37: warning: typedef 'index' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index index;  
## ^  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp: In static member function 'static void boost::multi\_array\_concepts::detail::idgen\_helper<0ull>::call(Array&, const IdxGen&, Call\_Type)':  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:53:43: warning: typedef 'index\_range' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index\_range index\_range;  
## ^  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/BH/include/boost/multi\_array/concept\_checks.hpp:54:37: warning: typedef 'index' locally defined but not used [-Wunused-local-typedefs]  
## typedef typename Array::index index;  
## ^  
## In file included from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/core.hpp:42:0,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/mat.hpp:4,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math.hpp:4,  
## from C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/src/stan/model/model\_header.hpp:4,  
## from file30104514d6683.cpp:8:  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/core/set\_zero\_all\_adjoints.hpp: At global scope:  
## C:/Users/GiulioVannini/Documents/R/win-library/3.3/StanHeaders/include/stan/math/rev/core/set\_zero\_all\_adjoints.hpp:14:17: warning: 'void stan::math::set\_zero\_all\_adjoints()' defined but not used [-Wunused-function]  
## static void set\_zero\_all\_adjoints() {  
## ^

endTime <- date();c(iniTime = iniTime, endTime = endTime)

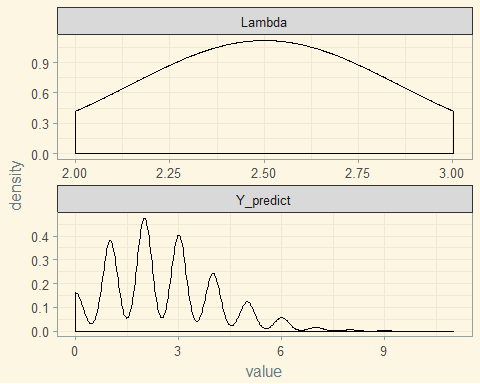
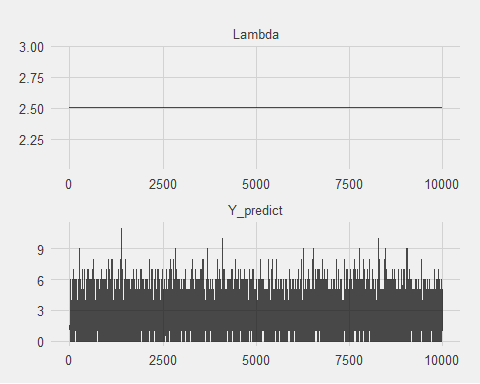
## iniTime endTime   
## "Thu Jun 15 12:59:15 2017" "Thu Jun 15 12:59:51 2017"

system.time(  
 fit <<- sampling(mod,  
 data = data,  
 pars = c("Lambda", "Y\_predict"),  
 chains = 1,  
 iter = 10000,  
 warmup = 0,  
 thin = 1,  
 cores = 4,  
 seed = 19861986,  
 control = list(max\_treedepth = 10,  
 adapt\_delta = 0.8)  
 )  
)

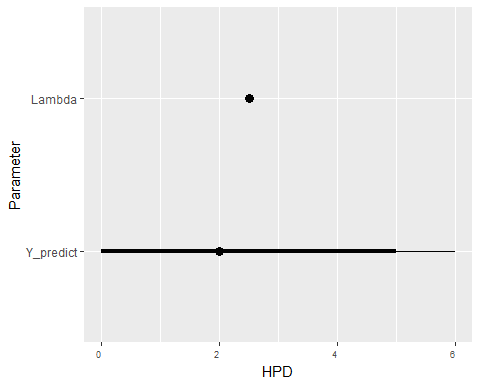
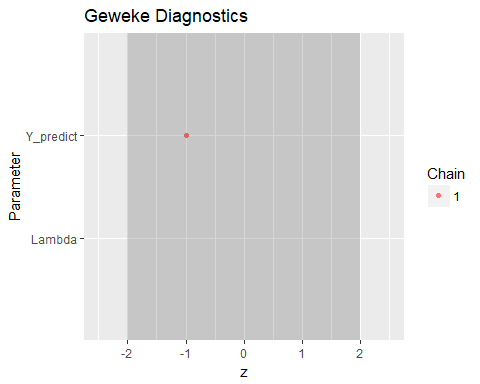
##   
## SAMPLING FOR MODEL '1e127d7a2b782c761884efcff0b2ffaa' NOW (CHAIN 1).  
## WARNING: No variance estimation is  
## performed for num\_warmup < 20  
##   
##   
## Chain 1, Iteration: 1 / 10000 [ 0%] (Sampling)  
## Chain 1, Iteration: 1000 / 10000 [ 10%] (Sampling)  
## Chain 1, Iteration: 2000 / 10000 [ 20%] (Sampling)  
## Chain 1, Iteration: 3000 / 10000 [ 30%] (Sampling)  
## Chain 1, Iteration: 4000 / 10000 [ 40%] (Sampling)  
## Chain 1, Iteration: 5000 / 10000 [ 50%] (Sampling)  
## Chain 1, Iteration: 6000 / 10000 [ 60%] (Sampling)  
## Chain 1, Iteration: 7000 / 10000 [ 70%] (Sampling)  
## Chain 1, Iteration: 8000 / 10000 [ 80%] (Sampling)  
## Chain 1, Iteration: 9000 / 10000 [ 90%] (Sampling)  
## Chain 1, Iteration: 10000 / 10000 [100%] (Sampling)  
## Elapsed Time: 0 seconds (Warm-up)  
## 0.076 seconds (Sampling)  
## 0.076 seconds (Total)

## user system elapsed   
## 0.2 0.0 0.2

## # A tibble: 20,000 ž 4  
## Iteration Chain Parameter value  
## <dbl> <int> <fctr> <dbl>  
## 1 1 1 Lambda 2.501049  
## 2 2 1 Lambda 2.501049  
## 3 3 1 Lambda 2.501049  
## 4 4 1 Lambda 2.501049  
## 5 5 1 Lambda 2.501049  
## 6 6 1 Lambda 2.501049  
## 7 7 1 Lambda 2.501049  
## 8 8 1 Lambda 2.501049  
## 9 9 1 Lambda 2.501049  
## 10 10 1 Lambda 2.501049  
## # ... with 19,990 more rows



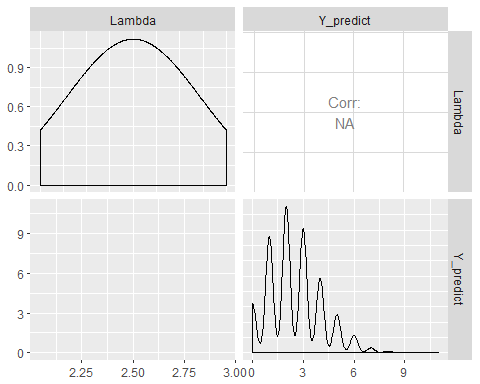
## Warning: Removed 1 rows containing missing values (geom\_point).



## # A tibble: 2 ž 6  
## Parameter low Low median High high  
## \* <fctr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Lambda 2.501049 2.501049 2.501049 2.501049 2.501049  
## 2 Y\_predict 0.000000 0.000000 2.000000 5.000000 6.000000

## Warning in cor(x, y, method = method, use = use): la deviazione standard Š  
## zero

## Warning: Computation failed in `stat\_density2d()`:  
## bandwidths must be strictly positive



Riassumo le caratteristiche della distribuzione predittiva:

# Riassumi le sue caratteristiche.  
median(outSim$value)

## [1] 2.501049

mean(outSim$value)

## [1] 2.496824

min(outSim$value)

## [1] 0

max(outSim$value)

## [1] 11

range(outSim$value)

## [1] 0 11

var(outSim$value)

## [1] 1.260453

sd(outSim$value)

## [1] 1.122699

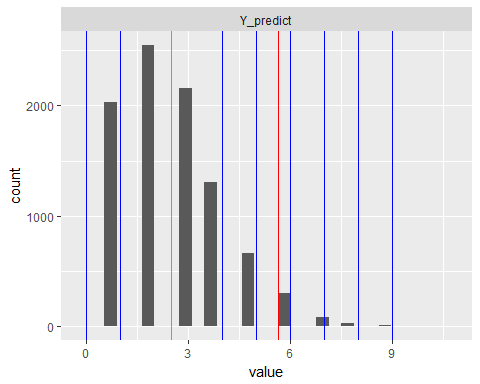
summary(fit)

## Length Class Mode   
## 1 stanfit S4

summary(outSim)

## Iteration Chain Parameter value   
## Min. : 1 Min. :1 Lambda :10000 Min. : 0.000   
## 1st Qu.: 2501 1st Qu.:1 Y\_predict:10000 1st Qu.: 2.000   
## Median : 5000 Median :1 Median : 2.501   
## Mean : 5000 Mean :1 Mean : 2.497   
## 3rd Qu.: 7500 3rd Qu.:1 3rd Qu.: 2.501   
## Max. :10000 Max. :1 Max. :11.000

1. Confronta i risultati ottenuti via simulazione MCMC con i risultati esatti sopra riportati



Riguardo al grafico finale: Blue - Valori di Y osservati Rosso - Media dei valori di Y osservati Verde - Media di Y\_predict Mediamente è stato osservato l'arrivo di 5,6 email al giorno. Il valore sta dentro alla distribuzione che risulta dal modello predittivo. Si può dire che ci si può aspettare fra le 5 e le 6 email al giorno. Guardando anche i valori osservati si può dire che in generale non ci sono "fughe" verso valori improbabili. Direi che il modello è accettabile.