

# Komputime bayesian, përafruese për procese me memorie të gjatë

---

Ejona KODRA<sup>1</sup>

**Abstrakt:** Metoda “Approximate Bayesian Computation” ose metoda likelihood – free përdoret për të gjetur në mënyrë aprosimative distribucionin posterior, pa përdorur funksionin likelihood (në raste kur ky funksion nuk është i disponueshëm ose kur është e vështirë që të përlllogaritet). Metoda ABC për herë të parë është përdorur në gjenetikë (Tavare, 1997) dhe më pas është përshtatur për statistikën. Paraqesim këtu metodën ABC si një algoritëm accept–reject bazuar në simulimet e një grupi të madh të dhënash. Ideja mbi të cilën bazohet ky algoritëm është ajo e llogaritjes dhe krahasimit të distancës midis summary statistic e të dhënave të observuara dhe të atyre të gjeneruara nga simulimet. Nëse distanca është më e vogël se ajo e një niveli të caktuar, atëherë vektori me parametra pranohet, në rast të kundërt algoritmi fillon nga e para. Këtu do të cekem përciptazi dhe koncepte si densiteti spektral dhe perodogrami, të cilët do të përdoren për të përshkruar më mirë proceset me memorie të gjatë. Qëllimi ynë është përafrimi i koeficientëve të një modeli ARFIMA, duke simuluar fillimisht për një proces ARFIMA (0, d, 0) dhe më pas jemi përpjekur për të përafruar koeficientët e një modeli ARFIMA (1,1,1).

**Fjalë kyçe:** Distribucioni posterior, likelihood–free, approximate bayesian computation, memorie e gjatë

**Abstract:** The Approximate Bayesian Computation method or the likelihood – free method is used to find approximation of the posterior distribution without making explicit use of the likelihood function (in cases where this function is not available or difficult to calculate). The ABC method was developed for the first time in genetics (Tavare, 1997) and then used in statistics. We present here the ABC method as a accept–reject algorithm based on the simulation of a large group of data. The idea beneath the algorithm is that of the calculation

---

<sup>1</sup> Kompania e auditimit Deloitte: ekodra@deloitte.al

and comparison of the distance between the summary statistics of the observed data with those generated from the simulation. If such distance is less than a certain fixed level, then the vector of parameters is accepted, otherwise the algorithm start from the beginning. Here will be briefly introduced the concepts of spectral density and periodogram to better describe long memory processes. Our aim is to approximate the coefficients of an ARFIMA model, simulating in the first place for ARFIMA(0, d, 0) and then we tried to approximate the coefficient of an ARFIMA (1,1,1).

**Key words:** Posterior distribution, likelihood-free, approximate bayesian computation, long memory

## 1. Metoda ABC

Në inferencën *bayesiane* (Sisson, 2010), parametri me interes  $\theta$  nuk është më një vlerë fikse që ne duam të stimulojmë, por kthehet në një variabël rëndom me prioritet distribucionin e saj  $\pi(\theta)$  (i cili bazohet mbi informacionin që ne disponojmë në atë moment). Qasja *bayesiane* është që të gjejë distribucionin posterior të  $\theta$  duke marrë si të dhënë observimet e të dhëna  $y = (y_1, \dots, y_n)$ ,

$$\pi(\theta|y) \propto \pi(\theta) \pi(y|\theta)$$

Inferenca për parametrin  $\theta$  bazohet në distribucionin posterior dhe simulimet numerike nevojiten që të përcaktohen integracionet e nevojshme për karakteristikat e poteries. Në rastet kur funksioni *likelihood*  $\pi(y|\theta)$  nuk mund të përftohet në mënyrë analitike ose është i vështirë që të përlogaritët, atëherë implementohet metoda “Approximate Bayesian computation”.

### 1.1 Origjina e ABC-së

Kjo metodë u përmend për herë të parë nga Rubin (1994). Tavare (1997), e paraqiti metodën ABC si një teknikë refuzuese e cila linte mënjane llogaritjen e funksionit *likelihood*, duke simuluar nga distribucioni dhe algoritmi që vijon, implementohet (Marin, 2011). Algoritmi origjinal i ABC-së është një metodë *accept-reject*, ku parametri  $\theta$  gjenerohet nga distribucioni prior  $\pi(\theta)$  dhe niveli i pranimit varet nga fakti sesa simulimet reciproke do të jenë të “ngjashme” me të dhënat e observuara. Japim më poshtë algoritmin:

---

#### Algorithm 1 Likelihood-free rejection sampler 1

---

```

for i = 1 tek N do
  repeat
    Gjenero  $\theta'$  nga prior distribution  $\pi(\theta)$ 
    Gjenero të dhënat  $x$  nga likelihood  $\pi(x|\theta')$ 
    until  $x \approx y$ 

  vendos  $\theta_i = \theta'$ 
end for

```

---

Për çdo vektor parametrash kandidat  $\theta'$  i gjeneruar nga distribucioni prior, një set të dhënash gjenerohet nga funksioni *likelihood*  $x \sim \pi(x|\theta')$ . Në rast se të dhënat e gjeneruara dhe ato të observuara janë të ngjashme, në mënyrë të tillë që  $x \approx y$ , atëherë vlerat e  $\theta'$  ruhen. Në rast të kundërt, nëse  $x$  nuk është i ngjashëm me  $y$ , atëherë me shumë mundësi  $\theta'$  nuk ka gjeneruar të dhënat e observuara nga modeli dhe eliminohet si vlerë. Vektori i parametrave të pranuar këtu mund të konsiderohet sikur të jetë derivuar nga distribucioni posterior  $\pi(\theta|y)$ . Në këtë mënyrë, vlerësimi i funksioni *likelihood*  $\pi(y|\theta')$  zëvendësohet nga një stimë e afërsisë së setit të të dhënave të simuluar  $x \sim \pi(x|\theta')$  me të dhënat e observuara  $y$ .

## 1.2 Summary statistic për një ARFIMA

Në rastet kur seria e të dhënave paraqet një dimension të lartë, atëherë algoritmit të mësipërm i bëhen disa ndryshime. Hyn këtu koncepti i *summary statistic* i cili përdoret për të përmbledhur informacion mbi një set të dhënash, në mënyrë të tillë që të përmbledhë të dhënat e observuara dhe ato të përfutuara nga modeli dhe t'i projektojë në një dimension më të reduktuar. Ideja bazë mbi të cilën mbështetet versioni më i zgjeruar i algoritmit i përgatitur nga Pritchard (1999) është ajo që kombinimi i përdorimit të një *summary statistic* dhe një epsilon të vogël, duhet të japë një përafrim të mirë të distribucionit posterior.

---

### Algorithm 2 Likelihood-free rejection sampler 2

---

```

for  $i = 1$  tek  $N$  do
  repeat
    Gjenero  $\theta'$  nga prior distribution  $\pi(\theta)$ 
    Gjenero të dhënat  $x$  nga likelihood  $\pi(x|\theta')$ 
    until  $\rho(\eta(x), \eta(y)) \leq \epsilon$ 
    vendos  $\theta_i = \theta'$ 
  end for

```

---

ku  $\eta(x)$  tregon një *summary statistic*,  $\rho > 0$  tregon një matës distance dhe  $\epsilon > 0$  tregon një nivel tolerance që së bashku përcaktojnë saktësinë e algoritmit. Vlerat e pranuar të  $\theta$  nuk vijnë nga distribucioni posterior, por nga një përafrim i tij (Wilkinson R., 2008)

$$\pi(\theta | \rho(\eta(x), \eta(y)) \leq \epsilon)$$

Për  $\epsilon = 0$  algoritmi është i saktë dhe jep vlera nga distribucioni posterior  $\pi(\theta|y)$ , në anën tjetër kur  $\epsilon \rightarrow \infty$  algoritmi jep vlera nga distribucioni prior. Vlera të vogla të  $\epsilon$  çojnë në një përafrim më të mirë të distribucionit posterior, por gjithashtu sjellin edhe një nivel pranimi më të vogël se sa për  $\epsilon$  me vlera më të mëdha.

#### *Periodogrami*

Një seri të dhënash mund të përshkruhet edhe si një shumë e peshuar e funksioneve periodike, dhe njihet si analiza spektrale ose domain i frekuencave. Duke qenë se objekti i këtij artikulli nuk është ai i shpjegimit të funksionit të spektrit, po të sjellim këtu vetëm atë pjesë që është më e nevojshme (për një thellim të njohurive mbi këtë temë mund të konsultohet

Hamilton (1994). Por nëse kemi një funksion të spektrit ai do të ishte:

$$s_y(\omega) = \frac{1}{2\pi} \left\{ \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(\omega j) \right\}$$

Ku  $\gamma_j$  është funksioni i autokovariancës i shprehur ndryshe si  $\text{Cov}(X_t + X_{t-j})$ , e cila tregon varësinë e të dhënave  $X_t + X_{t-j}$  dhe është indikator i kohëzgjatjes së memories së procesit. Nga ky ekuacion mund të derivojmë disa karakteristika si për shembull ato që lidhen me funksionin  $\cos(\omega)$ , si fakti që është një funksion tek dhe periodik. Stima e një spektri mund të përftohet edhe nga periodogrami, i cili ka të njëjtat karakteristika me spektrin dhe me ekuacion si më poshtë:

$$\hat{s}(\omega) = \frac{1}{2\pi} \left\{ \hat{\gamma}_0 + 2 \sum_{j=1}^{\infty} \hat{\gamma}_j \cos(\omega j) \right\}$$

## 2. Memoria e gjatë

Le të jetë  $\gamma_t$  një proces me autokorrelacion  $\rho_j$  në lag  $j$ , atëherë mund të themi që një proces përmban memorie të gjatë (McLeod & Hipel, 1978)

$$\lim_{n \rightarrow \infty} \sum_{i=-n}^n |\rho_j| = \infty$$

pra autokorrelacioni në vlerë absolute nuk është i mblledhshëm. E shprehur në një mënyrë ndryshe, memoria e gjatë implikon faktin që një proces ka një dipendencë të madhe nga koha, e kundërta në rastin kur

$$\lim_{n \rightarrow \infty} \sum_{i=-n}^n |\rho_j| = k$$

këtu themi që procesi ka memorie të shkurtër. Brockwell & Davis (1998) e forcojnë akoma më shumë përkufizimin e një procesi me memorie të gjatë në

$$\rho_j \sim C|j|^{2d-1} \quad j \rightarrow \infty$$

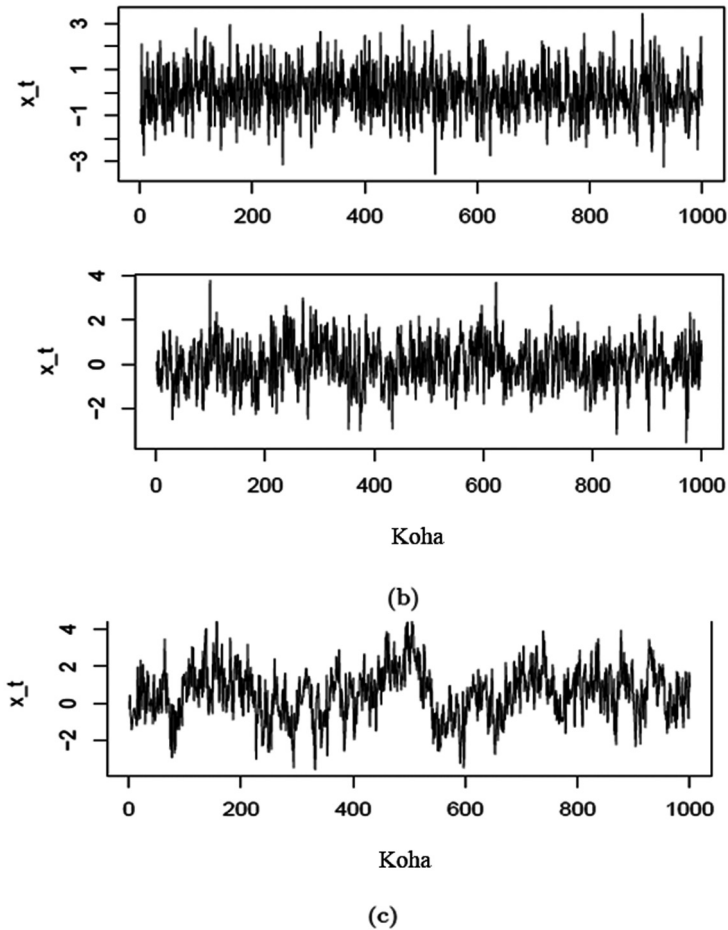
$$\text{për } C \neq 0 \text{ dhe } 0 < d < \frac{1}{2}.$$

**Self – similarity.** Një përshkrim i memories së gjatë nuk është i plotë pa përmendur konceptin e *self-similarity*. Përkufizimi i përdorur këtu është marrë nga Beran (1994).

**Përkufizimi 1.**  $\gamma_t$  është një proces stokastik i vazhdueshëm.  $\gamma_t$  është *self-similar* me parameter  $H$ , nëse për  $c > 0$ , procesi i ri  $c^{-H} \gamma_{ct}$  është i barabartë në distribucion me origjinalin  $\gamma_t$ .

Kjo do të thotë që për çdo sekuençë kohore dhe kostante  $c$  pozitive,  $c^{-H} (\gamma_{ct1}, \gamma_{ct2}, \dots, \gamma_{ctk})$  ka të njëjtin distribucion si  $(\gamma_{t1}, \gamma_{t2}, \dots, \gamma_{tk})$ . Në kontrast me procesin *self-similar* deterministik, nuk do të thotë që imazhi përsërit veten sa më afër që ne i afrohem, por është impresion i përgjithshëm ai që ngelet i njëjti. Në figurën 1 tregohet grafiku i një ARIMA (0, H -1/2, 0) me parametër  $H = 0.5; 0.7; 0.9$ . Me rritjen e parametrit, observohen periudha të gjata me vlera të

larta të ndjekura nga periudha të gjata me vlera të ulëta. Duke parë në periudha më të shkurtra duket sikur seria e të dhënave karakterizohet nga një trend ose cikël, por duke e parë në tërësinë e saj nuk vërehen as trende e as cikle (Beran, 1994) (veçanti kjo e proceseve me memorie të gjatë).



**Figura 1** Seri të simuluar nga një ARIMA(0,d,0) fraksional me  $H=0.5$  në figurën 1a,  $H=0.7$  në figurën 1b dhe  $H=0.9$  në figurën 1c

### 3. Modeli ARFIMA(0,1,0)

Modeli fraksional ARIMA, përdoret për modelimin e proceseve me memorie të gjatë. Parametri  $d$  lejon që të kapë sjelljen e modeleve me memorie të gjatë, pa hasur problemet që mund të krijohen nga përdorimi i një ARMA, si vlera të larta parametrash dhe numër i madh i tyre (Granger Joyeux, 1980 dhe Hosking, 1981). Japim këtu shembullin e një *white noise* fraksional. Nëse një seri të dhënash ka nevojë që të diferencohet, që të rrisë qëndrueshmërinë (një seri të dhënash thuhet që është e qëndrueshme nëse vlera e pritur e serisë është konstante),

do të thotë që seria origjinale ka një variancë infinite. Procesi i integruar më elementar, është shëtitja aleatore *random walk* ose ARIMA (0, 1, 0)

$$(1 - B)x_t = \varepsilon_t$$

i cili është një proces i integruar i nivelit të parë. Forma e përgjithshme e integruar e një ARFIMA (0, d, 0) është

$$\nabla^d x_t = (1 - B)^d x_t = \varepsilon_t$$

Nëse supozojmë që d mund të marrë edhe vlera fraksionale atëherë procesi quhet i integruar në mënyrë fraksionale. Fillojmë implementimin e metodës ABC me procesin me memorie të gjatë më të thjeshtë, *fractional white noise* ARFIMA (0,1,0). Aplikimi që do të tregohet testohet duke ndryshuar vlerat e gjatësisë së kampionit, numrit të simulimeve, numrit të *summary statistics* dhe nivelit të tolerancës. Të dhënat janë të gjeneruara nëpërmjet programit R me komandë:

```
>d = 0.3
>y = arima.sim(n=samplesize, list(d))
```

Vlerat e d-së zgjidhen midis intervalit (0,0.5), në mënyrë të tillë që të respektojnë kushtin e qëndrueshmërisë, gjatësia e kampionit vendoset 1024. Moledelet ARFIMA mund të studiohen në domainin e frekuencave dhe karakteristika e tyre kryesore është që kanë një pol rreth frekuencës zero. Periodogrami mund jetë një stimator i mirë i densitetit spektral dhe në këtë perspektivë është menduar që të merret si një sufficient summary statistic vektori me 5 vlerat e para të priodogramit të llogaritur sipas frekuencave të Fourier-it, duke përdorur komandën në R (program statistikor):

```
>ff=abs(fft(y)/sqrt(samplesize))^2
>stat.oss=ff[1:5]
```

Në këtë mënyrë kemi ndërtuar objektin e parë, *stat.oss*. Parametri d simulohet nga një variabël uniformë me prior distribution (0,0.5). Numri i simulimeve është n.sim = 10<sup>k</sup>

```
>n.sim=10^k
>d.sim=c()
>d.sim=c(d.sim,runif(n.sim, min=0, max=0.5))
>prior.sim=data.frame(d.sim)
```

Tani procedojmë me llogaritjen e vlerave të simuluar të summary statistics. Këto vlera simulohen si ato të vërtetat dhe përdoret i njëjti funksion:

```
>stat.simulated=matrix(NA,nrow=length(d.sim),ncol=5)
>for(i in 1:length(d.sim))
+y.new=arima.sim(n=samplesize,list(d=d.sim[i]))
+y.new=y.new[-c(1)] +ff=abs(fft(y.new)/sqrt(samplesize))^2
+stat.simulated[i,]=(y.new)[1:5].
```

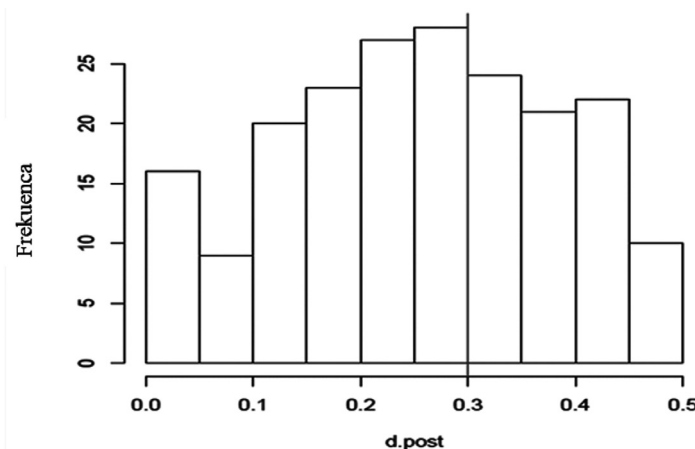
Duke përdorur funksionin *abc*

```
b>post.value=abc(target=stat.oss, param=prior.
sim,sumstat=
+data.frame(stat.simulated),tol=0.01, method="rejection")
```

simulimet janë përsëritur për një kampion me  $n=1024$  dhe me  $n=8194$ , me vlera tolerance 0.01, 0.02, 0.03 dhe me një numër total simulimesh që fillojnë nga  $10^4$ ,  $10^5$  dhe  $10^6$ . Përafrimi më i mirë i probabilitetit posterior është arritur me një vlerë  $n.sim =$  dhe  $tol = 0.02$ . Funkzioni *summary* jep vlerat e mëposhtme:

```
Call: abc(target = stat.oss, param = prior.sim, sumstat
= data.frame(stat.simulated), tol = 0.02, method =
"rejection")
Data:
abc.outunadj.values (200 posterior samples)
              d.sim
Min:          0.0066
2.5% Perc    0.0168
Median:      0.2636
Mean:        0.2566
Mode:        0.2678
97.5% Perc.: 0.4780
Max.:        0.4990
```

**Figura 2** Distribucioni posterior i d-së



Siç e shikojmë edhe nga grafiku, arrijmë të marrim vlera të përafërta me vlerën e vërtetë të d-së. Duke rritur numrin e simulimeve, periodogrami bëhet homogjen dhe nuk tregon vlera të mira të përafërta.

#### 4. Modeli ARFIMA (1, 1, 1)

Këtu përdorim metodën përafruese Bayesiane, në stimën e parametrave të modelit ARFIMA (1,1,1). Japim një përkufizim të këtij modeli

$$\phi_1(1 - B)^d X_t = \psi_1 \varepsilon_t$$

```

Të dhënat gjenerohen sipas:
>sample=1024
>t=c(0.18,0.3,0.4)
>y=arima.sim(n=sample,list(ar=t[1],d=t[2],ma=t[3]))
>y=y(-c[1])
>ff=abs(fft(y)/sqrt(sample))^2
>stat.oss=ff[1:5]

```

Ku t-ja është vektori i parametrave të modelit. Duke qenë se kushti i qëndrueshmërisë së një modeli ARMA përfitohet për  $|\phi| < 1$  dhe kushti i invertibilitetit përfitohet për  $|\theta| < 1$ , prior distribucion nga e cila  $\phi$ ,  $d$  dhe  $\theta$  do të gjenerohen janë përkatësisht

- $\phi \sim U(-1,1)$
- $d \sim U(0,0.5)$
- $\theta \sim U(-1,1)$

```

>n.sim=10k
>d.sim=c()
>phi.sim=c()
>theta.sim=c()
>d.sim=c(d.sim,runif(n.sim, min=0, max=0.5))
>phi.sim=c(phi.sim,runif(n.sim, min=-1, max=1))
>theta.sim=c(theta.sim,runif(n.sim, min=-1, max=1))
>prior.sim=data.frame(phi.sim,d.sim,theta.sim)

```

Edhe në këtë rast procedura testohet për numra të ndryshëm simulimesh, si në rastin e *fractional white noise*. Për një përafrim më të mirë të vlerave reale do të përdorim si *summary statistic* për koeficientët e modelit ARMA, pesë vlerat e para të funksionit të autokorrelacionit dhe për parametrin  $d$  do të përdoren pesë vlerat e para të periodogramit. Në këtë rast vektori i *summary statistic* është

```

>stat.oss=c(acf(y,plot=F)acf[1:5],ff[1:5])

```

ku *summary statistic* e simuluar janë

```

>comp.acf=5
>comp.për=5
>comp=comp.acf + comp.për
>stat.simulated=matrix(NA,nrow=length(d.sim),ncol=comp)
>for(i in 1:length(d.sim)){
+y.new=arima.sim(n=sample,list(ar=phi.sim[i],d=d.
sim[i],ma=theta.sim[i]))
+y.new=y.new[-c(1)] ff=abs(fft(y.new)/sqrt(sample))^2
stat.simulated[i,]=c(acf(y.new,plot=F)acf[1:5],ff[1:comp.
për]) }

```

Për një numër total prej 10.000 simulimesh dhe me një tolerancë prej 0.01, komanda *summary* e përdorur për *post.values* jep rezultatet e mëposhtme:

```

Call:
abc(target = stat.oss, param = prior.sim, sumstat = data.
frame(stat.simulated), tol = 0.01, method = «rejection»)

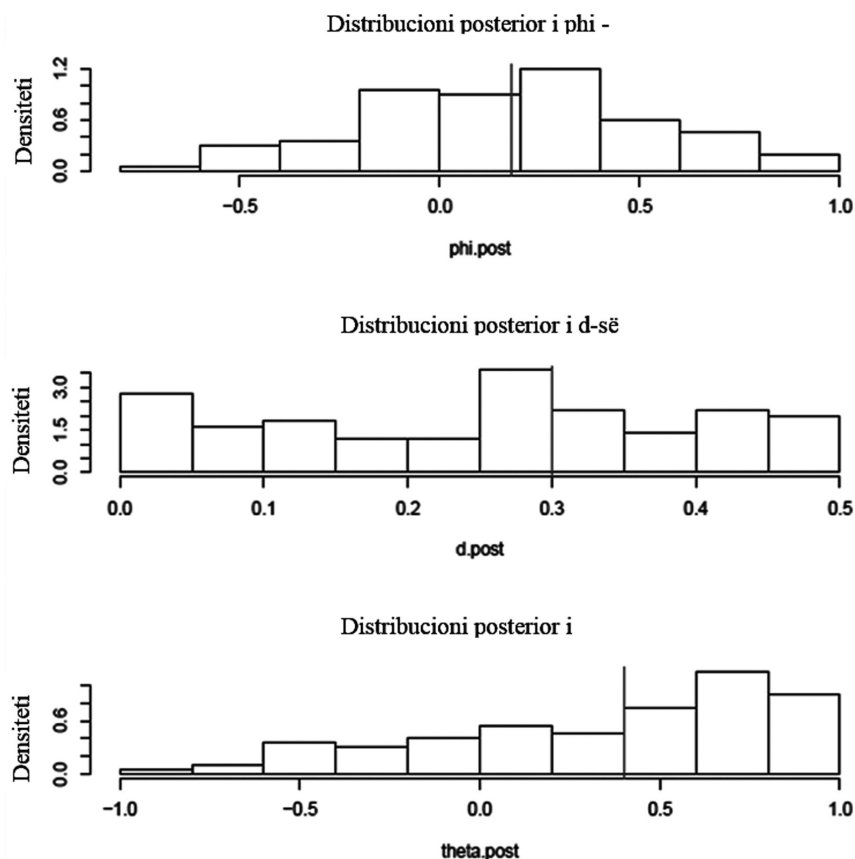
```



Data:  
 abc.outunadj.values (100 posterior samples)

	phi.sim	d.sim	theta. sim
Min	-0.7823	0.0028	-0.8795
2.5%	-0.4996	0.0145	-0.6092
Perc.:			
Median:	0.1973	0.2765	0.5295
Mean:	0.1647	0.2524	0.3519
Mode:	0.2278	0.2949	0.6728
97.5%	0.8171	0.4940	0.9819
Perc.			
Max:	0.9412	0.4991	0.9976

**Figura 3** Distribuzioni posterior i phi, d,  $\theta$



## 5. Përmbyllje

Avantazhi i përdorimit të metodës ABC është ai i të qenurit një metodë shumë intuitive dhe me një implementim të thjeshtë. Nga ana tjetër, në mënyrë që të përftohet një përafrim sa më i mirë i të dhënave duhet të zgjidhen sufficient summary statistic me dimension të vogël, gjë që në problemet reale do të përbënte një vështirësi. Ne ilustruam rastin e metodës rejection, që është ajo më e thjeshta, gjë që mund të ndikojë në mirësinë e përafrimit.

## Bibliografia

- Beran, J. (1994) "Statistics for long memory processes." Chapman and Hall.
- Hamilton, J.D, (1994) "Time series analysis". Princeton Univ. Press,
- Hosking, J.R.M, "Fractional Integration". *Biometrika*, 68 (1), 165 - 176.
- Granger, C.W. and Joyeux, R. "An introduction to long-memory time series models and fractional differencing". *Journal of Time Series Analysis* 1, 15 - 29.
- Marin, J.M., Pudlo, P., Robert, C.P, and Ryder, R., "Approximate Bayesian Computational methods." *Statistics and Computing* 21, 289 - 291.
- McLeod, A. I. & Hipel, K. W (1987) Preservation of the rescaled adjusted range, part 1. A reassessment of the Hurst phenomenon *Water resources research* 14
- Pritchard J, (1999) "Population growth of human Y chromosomes: a study of Y chromosomes microsatellites." *Molecular biology and evolution* 16,1791-1798
- Sisson, S.A., and Fan, Y. (2011), "Likelihood-free Markov chain Monte-Carlo. Handbook of Markov Chain Monte Carlo", Eds. S.P. Brooks, A. Gelman, G. Jones and X.L. Meng, Chapman and Hall.
- Tavare S, Balding DJ, Griffiths RC (1997) "Inferring Coalescence Times From DNA Sequence Data" *Genetics* 145:505-518
- Wilkinson, R. D, (2008) "ABC gives exact results under the assumption of model error." Technical report, Dept. of Probability and Statistics, University of Sheffield