

Time Series - Introduzione

Vers. 1.1.3

Gianluca Mastrantonio

gianluca.mastrantonio@polito.it

In questa parte del corso ci occuperemo di **serie temporali**. Quindi una serie di osservazioni prese nel tempo, in cui si suppone che ci sia delle dipendenza temporale, i.e., l'osservazione futura dipende dal passato. L'argomento è molto ampio e noi ci occuperemo solamente di una sua porzione, assumendo

- Osservazioni prese a intervalli regolari;
- Osservazioni continue con valori (almeno approssimativamente) in \mathcal{R} ;
- Osservazioni distribuite “marginalmente” come una normale.
- Processo Markoviani, i.e., il processo non dipende da tutto il passato ma solo dal passato recente;
- effetti stagionali
- processi stazionari, almeno dopo essere stati de-trendizzati e/o trasformati.

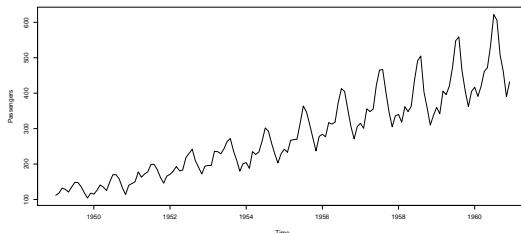


Figure: Serie temporale del numero di passeggeri

Nella figura si vedono chiaramente molte delle caratteristiche delle serie storiche

- Osservazioni che hanno un andamento generale ben definito temporalmente - aumendo quasi lineare (trend)
- Picchi e valli che si ripetono ciclicamente ogni 12 osservazioni (effetti stagionali)
- dipendenza temporale tra osservazioni vicini - fate attenzione che potrebbe essere un effetto del trend.

Possiamo decomporre la serie temporale nelle sue varie componenti, per poterla studiare meglio. Lo vediamo solo a livello descrittivo senza soffermarci sulle assunzioni o come funziona.

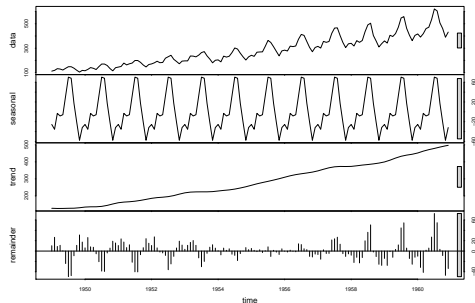


Figure: Decomposizione della serie temporale del numero di passeggeri

La decomposizione non è ottimale perchè l'effetto stagionale non è costante nel tempo (come viene assunto dalla funzione), e quindi rimane dell'effetto stagionale nel "reminder"

Delle volte si possono fare delle trasformazioni per aiutare il modello e migliorare i risultati. Per esempio modellizzare il log del numero dei passeggeri

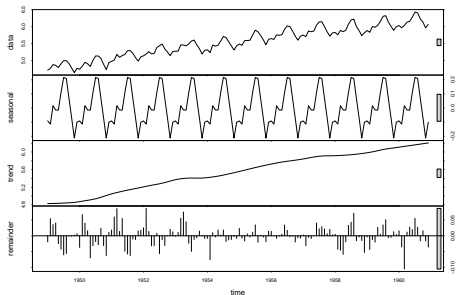


Figure: Decomposizione della serie temporale del (log-) numero di passeggeri

Adesso daremo una definizione di un particolare processo stocastico. Non andremo troppo nei dettagli ma definiremo solo le cose che saranno necessarie a noi.

Un processo aleatorio è una collezione di variabili aleatorie

$$\{X_t, t \in \mathcal{T}\}$$

indicizzate da un parametro $t \in \mathcal{T}$. Se \mathcal{T} è continuo allora il processo stocastico è a tempo continuo, altrimenti discreto.

Un processo stocastico può avere dimensione infinita ma si può dimostrare che è completamente caratterizzato dalla conoscenza della distribuzione di probabilità di ogni possibile insieme $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$, per ogni $t_1, t_2, \dots, t_n \in \mathcal{T}$ e ogni n (distribuzioni finito-dimensionali).

Nel corso vedremo quasi esclusivamente processi dove

- \mathcal{T} è il tempo;
- gli elementi di \mathcal{T} sono equispaziati;
- $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ sono distribuite normalmente (il processo si dice anche Gaussiano).

Attenzione: per semplificare la notazione, assumerò sempre che $\mathcal{T} \equiv \mathbb{Z}$ (gli interi). Si passa a insiemi diversi da \mathbb{Z} in maniera molto facile, ma complica la notazione senza nessun vero vantaggio. Quando è possibile lascio la notazione nel modo più generale possibile.

Oltretutto, X_t , X_{t_i} , X_i hanno lo stesso significato, e uso uno o l'altro per semplificare la notazione o per utilizzare la notazione classica di un particolare modello.

Vediamo alcune proprietà dei processi stocastici.

Definizione - Stazionarietà forte

Un processo si dice stazionario in senso forte se per ogni n naturale, e ogni n -upla $(t_1, t_2, \dots, t_n) \in \mathcal{T}^n$, e per ogni scelta di h tale che $t_i + h \in \mathcal{T}$, per ogni $i = 1, 2, \dots, n$, abbiamo

$$f(X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n) = f(X_{t_1+h} = x_1, X_{t_2+h} = x_2, \dots, X_{t_n+h} = x_n)$$

Abbiamo quindi che la distribuzione non cambia se trasliamo la serie di un tempo h . Nell'esempio precedente sembrerebbe che la serie non sia stazionaria.

Definizione - Stazionarietà debole

Un processo si dice stazionario in senso debole di ordine $s \in \mathbb{N}$ se per ogni $n \leq s$ naturale, e ogni n -upla $(t_1, t_2, \dots, t_n) \in \mathcal{T}^n$, e per ogni scelta di h tale che $t_i + h \in \mathcal{T}$, per ogni $i = 1, 2, \dots, n$, abbiamo

$$E(X_{t_1} X_{t_2} \dots X_{t_n}) = E(X_{t_1+h} X_{t_2+h} \dots X_{t_n+h}) < \infty$$

Nel caso in cui il processo sia Gaussiano, le distribuzioni finito-dimensionali sono completamente descritte dalla (funzione) media e (funzione) covarianza o correlazione. Nel caso in cui il processo sia stazionario debole di ordine 2 abbiamo che

$$\begin{aligned}E(X_{t_i}) &= E(X_{t_i+l}) \\ E(X_{t_i}X_{t_i+l}) &= E(X_{t_i+h}X_{t_i+l+h})\end{aligned}$$

assumendo naturalmente che $t_i, t_i + l, t_i + h + l$ siano tutti in \mathcal{T} .

Quindi la media è costante e il momento secondo dipende solo dal **lag** temporale.

Abbiamo allora che

$$Cov(X_{t_i}X_{t_i+l}) = E(X_{t_i}X_{t_i+l}) - E(X_{t_i})E(X_{t_i+l}) = \gamma(l)$$

e quindi la covarianza dipende solo dal lag-temporale, con $Var(X_{t_i}) = \gamma(0)$ e $\gamma(l) = \gamma(-l)$, dove $\gamma(\cdot)$ è la funzione di autocovarianza.

Possiamo definire la funzione di autocorrelazione come

$$\rho(l) = \frac{Cov(X_{t_i}, X_{t_i+l})}{Var(X_{t_i})} = \frac{\gamma(l)}{\gamma(0)}$$

Un plot in cui si mostra l'autocorrelazione in funzione del tempo è chiamato correlogramma. Si può calcolare anche il correlogramma e l'autocovarianza per serie non stazionarie, ma in questo caso dipenderà dal tempo t_i

$$\rho_{t_i}(l) = \frac{Cov(X_{t_i}, X_{t_i+l})}{\sqrt{Var(X_{t_i})Var(X_{t_i+l})}}$$

e

$$\gamma_{t_i}(l) = Cov(X_{t_i}, X_{t_i+l}).$$

Data una realizzazione di un processo stocastico $\mathbf{x} = (x_1, x_2, \dots, x_n)$ di $\mathbf{X} = (X_{t_1}, X_{t_2}, \dots, X_{t_n})$ con $t_1 < t_2 < \dots < t_n$ con $t_{i+1} - t_i = c$ per ogni $i = 1, \dots, n-1$, allora possiamo stimare $\gamma(l)$ con

$$\hat{\gamma}(l) = \frac{1}{n} \sum_{i=l+1}^n (x_i - \bar{x})(x_{i-l} - \bar{x})$$

che si dimostra essere uno stimatore “buono” da punto di vista dell'efficienza, e abbiamo

$$\hat{\rho}(l) = \frac{\sum_{i=l+1}^n (x_i - \bar{x})(x_{i-l} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Per l elevati lo stimatore non è “buono” visto che il numeratore viene calcolato con molti meno elementi del denominatore.

Il correlogramma campionario può essere calcolato per qualsiasi serie storica, anche una non stazionaria (in questo caso potrebbe non avere troppo senso). L'analisi del correlogramma può dare molte informazioni. Per esempio, riprendendo l'esempio precedente, abbiamo che il correlogramma è

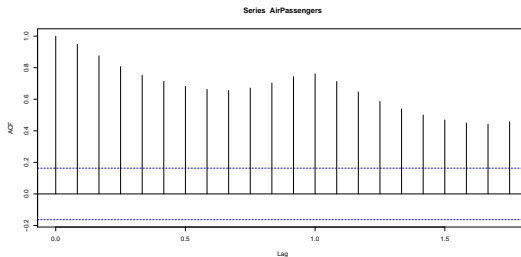


Figure: Correlogramma (campionario) della serie temporale del numero di passeggeri calcolate come se la serie fosse stazionaria

Vediamo l'effetto temporale (covarianza che ha massimi locali), ma molta della covarianza potrebbe dipendere dal trend (effetto temporale di lungo periodo).

Le linee orizzontali indica il limite entro il cui le correlazioni sono da considerarsi nulle

Possiamo estrarre il “reminder” dalla decomposizione precedente (sulla serie originale), e vedere la sua ACF

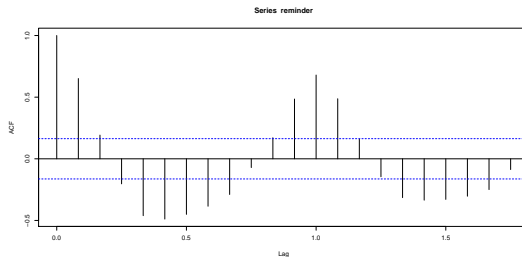


Figure: Correlogramma (campionario) del reminder della serie temporale del numero di passeggeri calcolate come se la serie fosse stazionaria

Vediamo un'altro esempio in cui, ogni 2 secondi, e per un totale di 128 punti, si misurava l'attività del cervello stimolando per 32 secondi la mano (con un pennello) e poi lo stimolo non veniva applicato per altri 32 secondi:

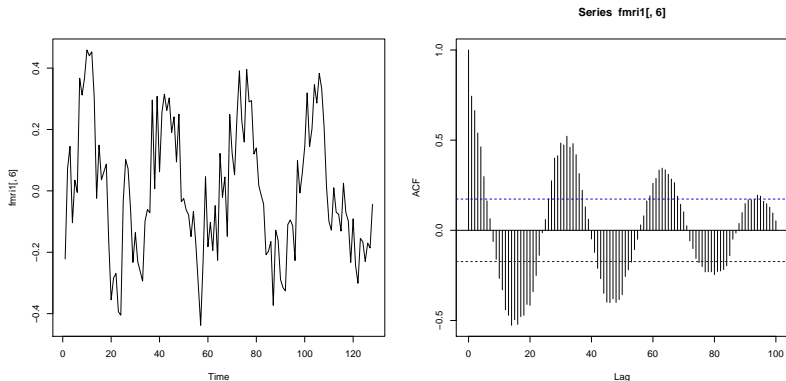


Figure: Osservazioni e correlogramma dell'esperimento sullo stimolo cerebrale

Dalla figura è chiaro come ci sia un effetto stagionale ogni 32 punti (covarianza che raggiunge massimo locale), ma la dipendenza comunque decresce con il tempo (come tendenza generale).

Possiamo anche introdurre una differente autocorrelazione, chiamata **autocorrelazione parziale**, che si definisce come

$$\psi(l) = \frac{Cov(X_{t_i}, X_{t_i+l} | X_{t_i+1}, X_{t_i+2}, \dots, X_{t_i+l-1})}{Var(X_{t_i} | X_{t_i+1}, X_{t_i+2}, \dots, X_{t_i+l-1})}$$

per le serie stazionarie, oppure

$$\psi_{t_i}(l) = \frac{Cov(X_{t_i}, X_{t_i+l} | X_{t_i+1}, X_{t_i+2}, \dots, X_{t_i+l-1})}{\sqrt{Var(X_{t_i+l} | X_{t_i+1}, X_{t_i+2}, \dots, X_{t_i+l-1}) Var(X_{t_i} | X_{t_i+1}, X_{t_i+2}, \dots, X_{t_i+l-1})}}$$

per quelle non stazionarie, che è una funzione di autocorrelazione depurata dall'effetto delle variabili che si trovano nel mezzo. Anche qui esiste un corrispettivo campionario. Per capirne l'utilità dobbiamo introdurre la proprietà di Markov, ma prima vediamo cosa viene se la calcoliamo per l'esempio precedente.

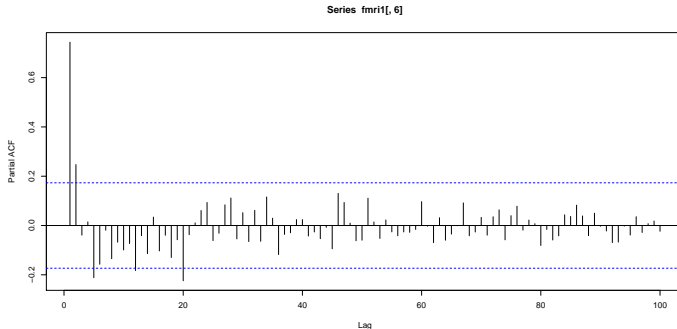


Figure: Correlogramma parziale dell'esperimento sullo stimolo cerebrale.

e come si vede la struttura di correlazione è molto diversa, e a ogni lag, ad eccezione del primo e forse del secondo, è molto piccola e probabilmente zero.

Definizione - Proprietà di Markov

Supponiamo che $t_1 < t_2 < \dots, < t_i$ un set di indici temporali, allora un processo ha la proprietà di Markov se

$$f(X_{t_i} \in A | X_{t_{i-1}} = x_{i-1}, X_{t_{i-2}} = x_{i-2}, \dots, X_{t_1} = x_1) = f(X_{t_i} \in A | X_{t_{i-1}} = x_{i-1})$$

A livello intuitivo ci dice che la legge di un processo a un dato tempo t_i non dipende dal suo intero passato, ma solo dal suo passato più recente. Un processo che ha la proprietà di Markov è una Markov chain.

Se un processo ha la proprietà di Markov permette di scrivere la densità congiunta in un maniera molto comoda.

Definiamo come $\mathbf{x}_j^i = (x_j, \dots, x_i)$

$$f(\mathbf{x}) = f(x_n | \mathbf{x}_1^{n-1}) f(x_{n-1} | \mathbf{x}_1^{n-2}) \dots f(x_3 | \mathbf{x}_1^2) f(x_2 | \mathbf{x}_1^1) f(x_1) =$$

$$f(x_n | x_{n-1}) f(x_{n-1} | x_{n-2}) \dots f(x_3 | x_2) f(x_2 | x_1) f(x_1) = f(x_1) \prod_{i=2}^n f(x_i | x_{i-1})$$

Fate attenzione che anche se il processo ha la proprietà di Markov, potremmo avere

$$0 \neq \gamma(h)$$

quindi la funzione di autocorrelazione non si annulla anche per lag maggiori di 1 (potenzialmente), dove

$$\gamma(h) = \int_{\mathbf{X}} \int_{\mathbf{X}} (x_{t_i} - E(X_{t_i}))(x_{t_i+h} - E(X_{t_i+h}))f(x_i, x_{i+h})dx_i dx_{i+h}$$

L'autocorrelazione parziale si annulla per lag maggiori di 1, ma la dimostrazione la facciamo quando introduciamo la proprietà di Markov più generale

La proprietà di Markov è troppo restrittiva e ci sono modi per estenderla. Noi useremo la seguente

Definizione - Proprietà di Markov di ordine h

Supponiamo che $t_1 < t_2 < \dots, < t_i$ sia un set di indici temporali, allora un processo ha la proprietà di Markov di ordine h se

$$f(X_{t_i} \in A | X_{t_i-1} = x_{i-1}, X_{t_i-2} = x_{i-2}, \dots, X_{t_2} = x_2, X_{t_1} = x_1) = \\ f(X_{t_i} \in A | X_{t_i-1} = x_{i-1}, X_{t_i-2} = x_{i-2}, \dots, X_{t_i-h} = x_{i-h})$$

Anche in questo caso è facile vedere come la proprietà ci dice che la distribuzione da un dato tempo, dipende solo dagli h tempi precedenti. Un processo con questa proprietà è una catena di Markov di ordine h .

Abbiamo che l'autocorrelazione parziale si annulla per lag maggiori di h . Questo si può vedere facilmente ricordando che se un processo è Markoviano di ordine h , abbiamo che

$$f(x_{t_i} | x_{t_i-1}, x_{t_i-2}, \dots, x_{t_1}) = f(x_{t_i} | x_{t_i-1}, x_{t_i-2}, \dots, x_{t_i-h})$$

Abbiamo che per $c > 0$

$$f(x_{i+h+c}, x_i | \mathbf{x}_{i+1}^{i+h+c-1}) = f(x_{i+h+c} | \mathbf{x}_{i+1}^{i+h+c-1}, x_i) f(x_i | \mathbf{x}_{i+1}^{i+h+c-1})$$

e per la proprietà di Markov abbiamo

$$f(x_{i+h+c} | \mathbf{x}_{i+1}^{i+h+c-1}, x_i) = f(x_{i+h+c} | \mathbf{x}_{i+c}^{i+h+c-1})$$

allora

$$f(x_{i+h+c}, x_i | \mathbf{x}_{i+1}^{i+h+c-1}) = f(x_{i+h+c} | \mathbf{x}_{i+c}^{i+h+c-1}) f(x_i | \mathbf{x}_{i+1}^{i+h+c-1})$$

che mostra come x_{i+h+c} e x_i sono indipendenti, condizionatamente a $\mathbf{x}_{i+1}^{i+h+c-1}$ (oppure, condizionatamente a $\mathbf{x}_{i+c}^{i+h+c-1}$), visto che la densità si fattorizza in due termini, e $f(x_{i+h+c} | \mathbf{x}_{i+c}^{i+h+c-1})$ non dipende da x_i .

I modelli per serie temporali

Attenzione:

- Da ora in poi, se non specificato diversamente, ogni volta che avremo una serie di elementi di un processo $\mathbf{X} = (X_{t_1}, X_{t_2}, \dots, X_{t_n})$, o una sua realizzazione $\mathbf{x} = (x_1, x_2, \dots, x_n)$, questi sono temporalmente equidistanti;
- Non definiremo mai direttamente la distribuzione delle X , ma sarà una conseguenza delle relazioni tra le varie componenti della serie temporale

Rumore Bianco

Il modello white noise è il modello più semplice e assume

$$x_t = w_t$$

$$W_t \stackrel{iid}{\sim} N(0, \sigma^2)$$

è immediato vedere come il white noise è semplicemente un modello in cui le componenti sono indipendenti. Abbiamo anche che

$$\mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

e l'autocorrelazione (parziale e no) è zero per qualsiasi lag, tranne 0. Il processo è stazionario del secondo ordine (è anche stazionario in senso forte).

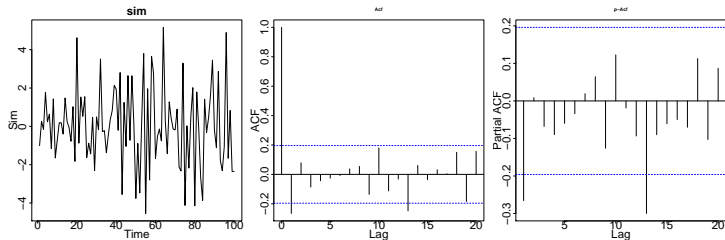


Figure: Simulazione di un white noise. ACF è calcolate come se la serie fosse stazionaria

Come potete vedere non è facilissimo dalla figura determinare se il processo ha componenti indipendenti o no.

Random Walk

Indichiamo con (w_1, \dots, w_n) un rumore bianco con $W_t \sim N(0, \sigma^2)$, allora una passeggiata aleatoria è definita come

$$x_t = x_{t-1} + w_t$$

$$x_1 = w_1$$

quindi al tempo t ci spostiamo di un passo random gaussiano dall'ultimo valore osservato.

Attenzione: Qui stiamo assumendo che la serie abbia inizio al tempo $t = 1$, perchè non ne possiamo fare a meno. Ma quando useremo i modelli AR(), MA() e ARMA(), faremo assunzioni diverse.

Possiamo facilmente calcolare la distribuzione condizionata di un tempo dato tutti gli altri, i.e., la distribuzione di $x_t | x_{t-1}, x_{t-2}, \dots, x_1$ che è Markoviana (del primo ordine) con

$$X_t | x_{t-1} \sim N(x_{t-1}, \sigma^2)$$

Per trovare la distribuzione dell'intero vettore dei parametri possiamo sostituire x_{t-1} con la sua espressione, ottenendo

$$x_t = (x_{t-2} + w_{t-1}) + w_t$$

poi andando a sostituire x_{t-2} con il suo valore, e così via, fino a ottenere

$$x_t = \sum_{i=1}^t w_i$$

quindi **marginalmente** abbiamo

$$X_t \sim N(0, t\sigma^2)$$

Abbiamo quindi che il processo non è stazionario neanche in senso debole (hanno varianze diverse). Se la serie non partisse al tempo 0, la varianza sarebbe infinita. Per trovare la distribuzione congiunta dobbiamo solo trovare la covarianza tra due tempi generici t e $t+k$, con $k > 0$, visto che la media della congiunta è composta dagli

elementi della marginale se i dati sono combinazioni lineari di normali. Dobbiamo calcolare

$$\gamma_t(k) = \text{Cov}(X_t, X_{t+k}) = \text{Cov}\left(\sum_{i=1}^t w_i, \sum_{j=1}^{t+k} w_j\right) = \sum_{i=1}^t \sum_{j=1}^{t+k} \text{Cov}(w_i, w_j) = t\sigma^2$$

e quindi

$$\mathbf{X} \sim N(\mathbf{0}, \Sigma)$$

con $[\Sigma]_{i,j} = \min\{i, j\}\sigma^2$.

La funzione di autocorrelazione è

$$\begin{aligned}\rho_t(k) &= \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+k})}} = \frac{t\sigma^2}{\sqrt{t\sigma^2(t+k)\sigma^2}} = \\ &= \frac{t\sigma^2}{\sigma^2\sqrt{t}\sqrt{t+k}} = \frac{1}{\sqrt{1+k/t}}\end{aligned}$$

mentre per l'autocorrelazione parziale abbiamo

$$\psi_t(1) = \rho_t(1) = \frac{1}{\sqrt{1+1/t}}$$

e

$$\psi_t(k) = \frac{Cov(X_t, X_{t+k} | X_{t+1}, \dots, X_{t+k-1})}{\sqrt{Var(X_t | X_{t+1}, \dots, X_{t+k-1}) Var(X_{t+k} | X_{t+1}, \dots, X_{t+k-1})}} = 0$$

con $k > 1$ visto che il processo è Markoviano

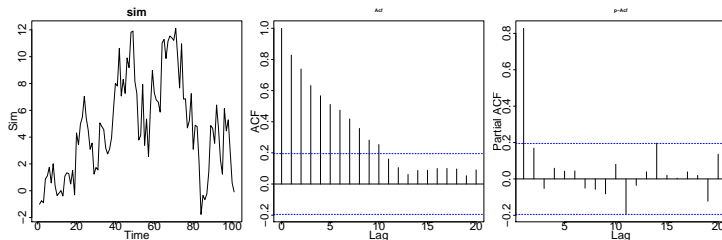


Figure: Simulazione di un random walk.

Fate attenzione che il correlogramma campionario da alcune indicazioni sul tipo di dipendenza, in questo caso molto forte, ma l'autocorrelazione dipende da t , attraverso t/k , e quindi quello della figura non si può considerare una stima di $\rho_t(k)$. Nella prossima figura vediamo delle stime MC dello stimatore campionario di $\rho_t(k)$ per vari t .

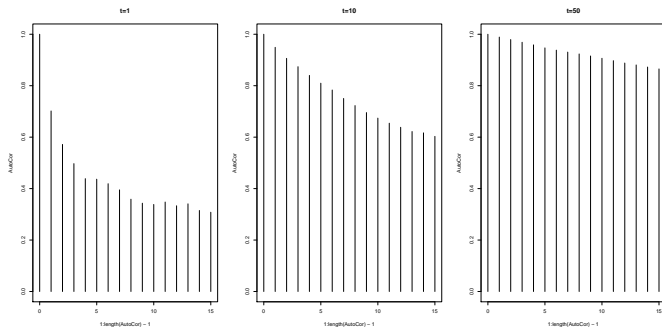


Figure: Stima MC dell'autocorrelazione per $t=1$ (sinistra), $t=10$ (centro) e $t=50$ (destra)

