

Probabilità e Processi Stocastici (455AA)

Lezione 21

Dario Trevisan

9/12/2024

Section 1

Processi a stati continui

Richiami

Dato un processo $(X_t)_{t \in \mathcal{T}}$ a stati continui (\mathbb{R}) , abbiamo introdotto

- La funzione di media $t \mapsto \mathbb{E}[X_t]$,

Richiami

Dato un processo $(X_t)_{t \in \mathcal{T}}$ a stati continui (\mathbb{R}) , abbiamo introdotto

- La funzione di media $t \mapsto \mathbb{E}[X_t]$,
- La funzione di autocovarianza $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$,

Richiami

Dato un processo $(X_t)_{t \in \mathcal{T}}$ a stati continui (\mathbb{R}) , abbiamo introdotto

- La funzione di media $t \mapsto \mathbb{E}[X_t]$,
- La funzione di autocovarianza $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$,
- La funzione di autocorrelazione $(s, t) \mapsto \text{ACF}(s, t) = \rho_{X_s, X_t}$

Richiami

Dato un processo $(X_t)_{t \in \mathcal{T}}$ a stati continui (\mathbb{R}) , abbiamo introdotto

- La funzione di media $t \mapsto \mathbb{E}[X_t]$,
- La funzione di autocovarianza $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$,
- La funzione di autocorrelazione $(s, t) \mapsto \text{ACF}(s, t) = \rho_{X_s, X_t}$
- Il concetto di *stazionarietà in senso lato*: media costante e $C(s, t) = C(0, |t - s|)$

Richiami

Dato un processo $(X_t)_{t \in \mathcal{T}}$ a stati continui (\mathbb{R}) , abbiamo introdotto

- La funzione di media $t \mapsto \mathbb{E}[X_t]$,
- La funzione di autocovarianza $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$,
- La funzione di autocorrelazione $(s, t) \mapsto \text{ACF}(s, t) = \rho_{X_s, X_t}$
- Il concetto di *stazionarietà in senso lato*: media costante e $C(s, t) = C(0, |t - s|)$
- La classe dei *processi gaussiani*

Section 2

Stima della funzione di autocovarianza

Stima della funzione di autocovarianza

Affrontiamo il problema generale di stimare la funzione di autocovarianza di un processo X a partire dall'osservazione dei valori $(X_t)_{t=0}^{n-1} = (x_t)_{t=0}^{n-1}$ (una *serie storica*).

- È una **generalizzazione** del problema di stimare **valor medio** e **varianza** di una famiglia di variabili aleatorie indipendenti, tutte con la stessa legge.

Stima della funzione di autocovarianza

Affrontiamo il problema generale di stimare la funzione di autocovarianza di un processo X a partire dall'osservazione dei valori $(X_t)_{t=0}^{n-1} = (x_t)_{t=0}^{n-1}$ (una *serie storica*).

- È una generalizzazione del problema di stimare valor medio e varianza di una famiglia di variabili aleatorie indipendenti, tutte con la stessa legge.
- Nel caso gaussiano avevamo ottenuto la media e la covarianza campionarie.

Verosimiglianza

Invece dell'indipendenza, supporremo la **stazionarietà del processo**, ossia la matrice di covarianza delle variabili $(X_t)_{t=0}^{n-1}$ è costante sulle diagonali:

$$(C(s, t))_{s, t=0}^{n-1} = (C(|t - s|))_{s, t=0}^{n-1}.$$

- Supponiamo pure che il processo X sia **gaussiano** e **centrato** (ossia la funzione di media è nota e costantemente nulla)

$$\left((C(s, t))_{s, t=0}^{n-1} \right) = \begin{pmatrix} C(0) & C(1) & C(2) & \dots \\ C(1) & & & \\ & \dots & & \\ & & & C(1) \\ & & & & C(0) \end{pmatrix}$$

Verosimiglianza

Invece dell'indipendenza, supporremo la *stazionarietà del processo*, ossia la matrice di covarianza delle variabili $(X_t)_{t=0}^{n-1}$ è costante sulle diagonali:

$$(C(s, t))_{s,t=0}^{n-1} = (C(|t - s|))_{s,t=0}^{n-1}.$$

- Supponiamo pure che il processo X sia *gaussiano* e **centrato** (ossia la funzione di media è nota e costantemente nulla)
- La verosimiglianza per la matrice di covarianza è

$$\begin{aligned} L(C; \mathbf{x}) &= p((X_t)_{t=0}^{n-1} = (\mathbf{x}_t)_{t=0}^{n-1} | C) \\ &\propto \exp\left(-\frac{1}{2} \mathbf{x}^T C^{-1} \mathbf{x}\right) \frac{1}{\sqrt{\det C}}, \end{aligned}$$

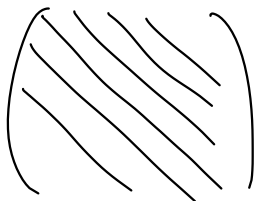
dove abbiamo posto, per alleggerire la notazione, $\mathbf{x} = (x_t)_{t=0}^{n-1}$.

Stima di massima verosimiglianza

Possiamo determinare la stima di massima verosimiglianza per C con i soliti passaggi: si tratta di minimizzare la funzione

$$C \mapsto x^T C^{-1} x + \log \det C.$$

- È difficile calcolare analiticamente C_{MLE} , ma si può ricorrere a metodi numerici.



Una ipotesi semplificativa

Per ottenere delle espressioni elementari per C_{MLE} introduciamo una ulteriore ipotesi nella struttura della matrice di covarianza.

- Oltre al fatto che $C(s, t)$ sia **simmetrica**, **semidefinita positiva** e **costante sulle diagonali** (per la stazionarietà), chiediamo che sia **circolante**:

$$C(k) = C(n - k) \quad \text{per ogni } k = 1, 2, \dots, n - 1.$$

$$\begin{pmatrix} C(0) & C(1) & \dots & C(n-1) \\ C(1) & C(0) & \dots & C(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ C(n-1) & C(n-2) & \dots & C(0) \end{pmatrix}$$

Una ipotesi semplificativa

Per ottenere delle espressioni elementari per C_{MLE} introduciamo una ulteriore ipotesi nella struttura della matrice di covarianza.

- Oltre al fatto che $C(s, t)$ sia simmetrica, semidefinita positiva e costante sulle diagonali (per la stazionarietà), chiediamo che sia *circolante*:

$$C(k) = C(n - k) \quad \text{per ogni } k = 1, 2, \dots, n - 1.$$

- È solo una ipotesi di comodo, una volta ottenute delle stime esplicite cerchiamo di rimuoverla.

Passaggio alla base delle frequenze

Tramite trasformata di Fourier a tempi finiti, C diventa diagonale.
Richiamiamo la notazione:

- Sia $F \in \mathbb{C}^{n \times n}$,

$$F_{\xi t} = e^{-2\pi i \xi t / n}, \quad \text{per } \xi, t = 0, 1, \dots, (n-1).$$

Passaggio alla base delle frequenze

Tramite trasformata di Fourier a tempi finiti, C diventa diagonale.
Richiamiamo la notazione:

- Sia $F \in \mathbb{C}^{n \times n}$,

$$F_{\xi t} = e^{-2\pi i \xi t / n}, \quad \text{per } \xi, t = 0, 1, \dots, (n-1).$$

- La trasformata di Fourier a tempi finiti di $(x_t)_{t=0}^{n-1}$ è

$$\hat{x}(\xi) = \sum_{t=0}^{n-1} x_t e^{-2\pi i \xi t / n} = \sum_{t=0}^{n-1} F_{\xi t} x_t,$$

ossia $\hat{x} = Fx$.

$$\Sigma_{AX+b} = A \Sigma_X \bar{A}^T \quad X \in \mathbb{C}^n \quad A \in \mathbb{C}^{m \times n}$$

Se il vettore x è l'osservazione del processo X , la matrice di covarianza del vettore aleatorio $\hat{X} = FX$ si calcola tramite la formula per le trasformazioni affini:

$$\Sigma_{\hat{X}} = \Sigma_{FX} = F \Sigma_X \bar{F}^T = FC\bar{F}^T,$$

- Questo cambio di coordinate: dalla base dei "tempi" a quella delle "frequenze" diagonalizza la matrice delle covarianze.

Se z v.v. a valori complessi $\text{Var}(z) = \mathbb{E}[|z - \mathbb{E}[z]|^2]$
 z, z' --- allora $\text{Cov}(z, z') = \mathbb{E}[(z - \mathbb{E}[z])(\overline{z' - \mathbb{E}[z']})]$
 z v.v. a valori \mathbb{C}^n $\Sigma_z = (\text{Cov}(z_i, z_j))_{i,j=1 \dots n}$
 Hermitiana e ≥ 0

Se il vettore x è l'osservazione del processo X , la matrice di covarianza del vettore aleatorio $\hat{X} = FX$ si calcola tramite la formula per le trasformazioni affini:

$$\Sigma_{FX} = F\Sigma_X\bar{F}^T = FC\bar{F}^T,$$

- Questo cambio di coordinate: dalla base dei “tempi” a quella delle “frequenze” diagonalizza la matrice delle covarianze.
- Introduciamo la notazione

$$\hat{C}(\xi) = \sum_{k=0}^{n-1} e^{-2\pi i \xi k/n} C(k). \quad \xi = 0, 1, \dots, n-1$$

$$(FC \bar{F}^T)_{\xi \xi'} = \sum_{s,t=0}^{n-1} F_{\xi s} C_{st} \bar{F}_{t \xi'}$$

$$= \sum_{s,t=0}^{n-1} e^{-2\pi i \xi s/n} C(|t-s|) e^{2\pi i \xi' t/n} =$$

$$= \sum_{s,t=0}^{n-1} C(|t-s|) e^{2\pi i \xi' (t-s)/n} e^{2\pi i (\xi' - \xi) s/n}$$

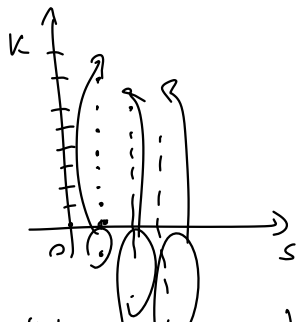
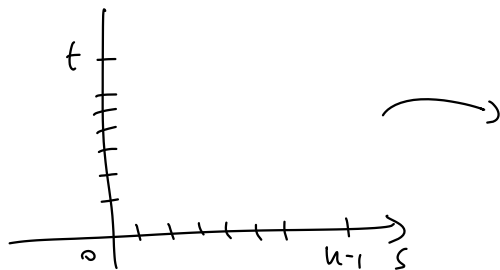
$$= \sum_{s=0}^{n-1} \sum_{k=0}^{n-1} \underbrace{C(k) e^{2\pi i \xi' k/n}}_{\text{periodica di periodo } n} \underbrace{e^{2\pi i (\xi' - \xi) s/n}}$$

$$\boxed{k = t - s}$$

$$s=0 \rightarrow k=0, 1, \dots, n-1$$

$$s>0 \rightarrow k = \underbrace{(-1, \dots, n-2)}_{n-1}$$

$$s>2 \rightarrow k = \underbrace{-2, -1, 0, \dots, n-2}$$



$$\Rightarrow (FCF^{-1})_{\xi\xi'} = \left(\sum_{k=0}^{n-1} c(k) e^{2\pi i \xi k/n} \right) \left(\sum_{s=0}^{n-1} \exp(2\pi i (\xi' - \xi) s/n) \right)$$

$$= \hat{C}(\xi) = \hat{C}(\xi')$$

$$\begin{cases} 0 & \text{se } \xi \neq \xi' \\ n & \text{se } \xi = \xi' \end{cases}$$

OSS

$$\hat{C}(\xi) = \sum_k c(k) e^{-2\pi i \xi k/n} = \sum_k c(k) e^{2\pi i \xi k/n} = \sum_k c(n-k) e^{2\pi i \xi k/n}$$

DSS $\hat{C}(\xi)$ sono Reali e ≥ 0

Stima di massima verosimiglianza

I numeri $n\hat{C}(\xi)$ sono quindi multipli degli autovalori della matrice di covarianza C e quindi sono tutti positivi.

- In queste nuove coordinate, le componenti del vettore FX sono non correlate e quindi, essendo gaussiane, indipendenti.

Stima di massima verosimiglianza

I numeri $n\hat{C}(\xi)$ sono quindi multipli degli autovalori della matrice di covarianza C e quindi sono tutti positivi.

- In queste nuove coordinate, le componenti del vettore FX sono non correlate e quindi, essendo gaussiane, indipendenti.
- La verosimiglianza assume quindi una espressione molto più trattabile:

$$L(C; \hat{x}) = p(FX = \hat{x} | C) \propto \exp\left(-\frac{1}{2n} \sum_{\xi=0}^{n-1} \frac{|\hat{X}(\xi)|^2}{\hat{C}(\xi)}\right) \frac{1}{\sqrt{\prod_{\xi=0}^{n-1} \hat{C}(\xi)}}$$

$$\boxed{\hat{C}_{MLE}(\xi) = \frac{|\hat{X}(\xi)|^2}{n}}$$

$$\begin{aligned} & \exp\left(-\frac{1}{2} \hat{x} \cdot \Sigma_X^{-1} \hat{x}\right) \frac{1}{\sqrt{\det \Sigma_X}} \\ &= \prod_{\xi=0}^{n-1} \exp\left(-\frac{1}{2n} \frac{|\hat{X}(\xi)|^2}{\hat{C}(\xi)}\right) \frac{1}{\sqrt{\hat{C}(\xi)}} \end{aligned}$$

Stima di massima verosimiglianza

I numeri $n\hat{C}(\xi)$ sono quindi multipli degli autovalori della matrice di covarianza C e quindi sono tutti positivi.

- In queste nuove coordinate, le componenti del vettore FX sono non correlate e quindi, essendo gaussiane, indipendenti.
- La verosimiglianza assume quindi una espressione molto più trattabile:

$$L(C; \hat{x}) = p(FX = \hat{x} | C) \propto \exp \left(-\frac{1}{2n} \sum_{\xi=0}^{n-1} \frac{|\hat{x}(\xi)|^2}{\hat{C}(\xi)} \right) \frac{1}{\sqrt{\prod_{\xi=0}^{n-1} \hat{C}(\xi)}}$$


- La stima di massima verosimiglianza si ottiene minimizzando la funzione

$$C \mapsto \sum_{\xi=1}^{n-1} \left[\frac{1}{n} \frac{|\hat{x}(\xi)|^2}{\hat{C}(\xi)} + \log \hat{C}(\xi) \right]$$

$$\left| \frac{\hat{x}(\xi)}{n} \right|^2 = \hat{C}_{MLE}(\xi) \Rightarrow \text{per trovare } C_{MLE} \\ \text{uso la trasformazione } \underline{\text{inversa}}$$

$k=0 \dots n-1$

$$\begin{aligned} C_{MLE}(k) &= \frac{1}{n} \sum_{\xi=0}^{n-1} \hat{C}_{MLE}(\xi) e^{2\pi i \xi k/n} \\ &= \frac{1}{n^2} \sum_{\xi=0}^{n-1} \left| \sum_{s=0}^{n-1} x_s e^{-2\pi i \xi s/n} \right|^2 e^{2\pi i \xi k/n} \\ &= \frac{1}{n^2} \sum_{\xi=0}^{n-1} \sum_{s,t=0}^{n-1} x_s x_t e^{-2\pi i \xi (s-t)/n} e^{2\pi i \xi k/n} \\ &= \frac{1}{n^2} \sum_{s,t=0}^{n-1} x_s x_t \left[\sum_{\xi=0}^{n-1} e^{-2\pi i \xi (s-t-k)/n} \right] \end{aligned}$$



Discutiamo l'espressione

$$\frac{1}{n^2} \sum_{s,t=0}^{n-1} X_t X_s \sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n}$$

- Se $t = s + k$, allora il termine

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} 1 = n,$$

Discutiamo l'espressione

$$\frac{1}{n^2} \sum_{s,t=0}^{n-1} X_t X_s \sum_{\xi=0}^{n-1} e^{-2\pi i \overbrace{(t-s-k)}^{=0 \text{ ok}} \xi/n}$$

$m \geq \text{ordine } \xi$

$$t-s-k = l \cdot n$$

$$l \in \mathbb{Z}$$

- Se $t = s + k$, allora il termine

$$\sum_{\xi=0}^{n-1} e^{-2\pi i (t-s-k) \xi/n} = \sum_{\xi=0}^{n-1} 1 = n,$$

- Se $t = s + k - n$ ugualmente si avrebbe

$$\sum_{\xi=0}^{n-1} e^{-2\pi i (t-s-k) \xi/n} = \sum_{\xi=0}^{n-1} e^{2\pi i \xi} = \sum_{\xi=0}^{n-1} 1 = n.$$

Discutiamo l'espressione

$$\frac{1}{n^2} \sum_{s,t=0}^{n-1} X_t X_s \sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n}$$

- Se $t = s + k$, allora il termine

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} 1 = n,$$

- Se $t = s + k - n$ ugualmente si avrebbe

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} e^{2\pi i\xi} = \sum_{\xi=0}^{n-1} 1 = n.$$

- In tutti gli altri casi possibili, si trova invece

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = 0,$$

Concludiamo che

$$C_{\text{MLE}}(k) = \frac{1}{n} \sum_{s=0}^{n-1-k} x_s x_{s+k} + \frac{1}{n} \sum_{s=n-k+1}^{n-1} x_s x_{s+k-n}.$$

- È la somma due contributi:

Concludiamo che

$$C_{\text{MLE}}(k) = \frac{1}{n} \sum_{s=0}^{n-1-k} X_s X_{s+k} + \frac{1}{n} \sum_{s=n-k+1}^{n-1} X_s X_{s+k-n}.$$

- È la somma due contributi:
 - 1 la covarianza campionaria tra X_s e il processo “traslato” avanti nel tempo di k istanti, X_{s+k}

Concludiamo che

$$C_{MLE}(k) = \frac{1}{n} \sum_{s=0}^{n-1-k} X_s X_{s+k} + \frac{1}{n} \sum_{s=n-k+1}^{n-1} X_s X_{s+k-n}.$$

- È la somma due contributi:
 - ① la covarianza campionaria tra X_s e il processo “traslato” avanti nel tempo di k istanti, X_{s+k}
 - ② la covarianza campionaria tra X_s e il traslato indietro di $n - k$ istanti, X_{s+k-n} .

Autocovarianza campionaria

I calcoli svolti suggeriscono di proporre come stima generale per C solo il primo contributo: la covarianza campionaria tra X_s e il traslato X_{s+k} .

- Definiamo la funzione di **autocovarianza campionaria** (o empirica) come

$$c(k) = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_s - \bar{x}_0)(x_{s+k} - \bar{x}_k),$$

dove le medie campionarie \bar{x}_0 , \bar{x}_k sono rispettivamente sui primi $n-k$ e sugli ultimi $n-k$ valori.

$$\bar{x}_0 = \frac{1}{n-k} \sum_{s=0}^{n-k-1} x_s \quad \bar{x}_k = \frac{1}{n-k} \sum_{s=k}^n x_s$$

Autocovarianza campionaria

I calcoli svolti suggeriscono di proporre come stima generale per C solo il primo contributo: la covarianza campionaria tra X_s e il traslato X_{s+k} .

- Definiamo la funzione di **autocovarianza campionaria** (o empirica) come

$$c(k) = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_s - \bar{x}_0)(x_{s+k} - \bar{x}_k),$$

dove le medie campionarie \bar{x}_0 , \bar{x}_k sono rispettivamente sui primi $n-k$ e sugli ultimi $n-k$ valori.

- $c(k)$ è la covarianza tra due variabili aleatorie così definite: si sceglie $S \in \{0, 1, \dots, n-1-k\}$ casuale uniforme e si considerano i valori x_S (prima variabile) e x_{S+k} (seconda variabile).

Funzione di autocorrelazione campionaria

Possiamo definire anche le varianze campionarie

$$\sigma_0^2 = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_s - \bar{x}_0)^2$$

e

$$\sigma_k^2 = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_{s+k} - \bar{x}_k)^2$$

e quindi la funzione di autocorrelazione campionaria, data da

$$\text{acf}(k) = \frac{c(k)}{\sigma_0 \sigma_k},$$

che assume sempre valori tra $[-1, 1]$

ACF in R

In R è l'autocorrelazione campionaria si calcola tramite la funzione `acf()`.

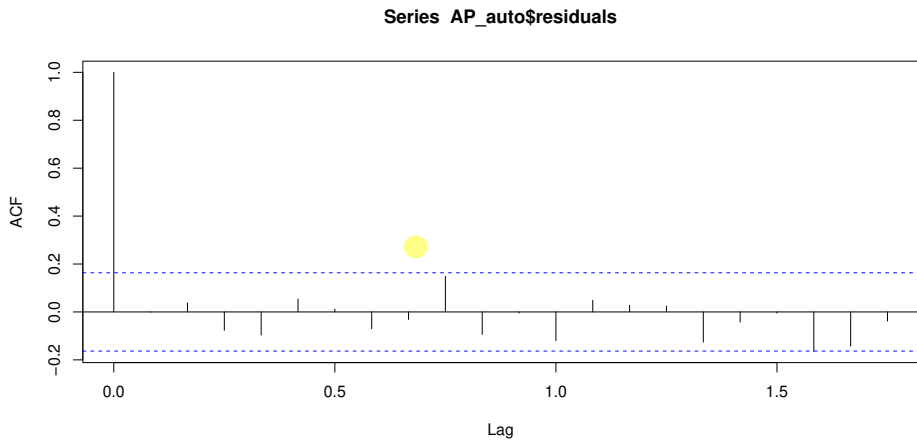


Figure 1: autocorrelazione empirica dei residui

Attenzione: la presenza di un “trend” può fare apparire la ACF uniformemente grande.

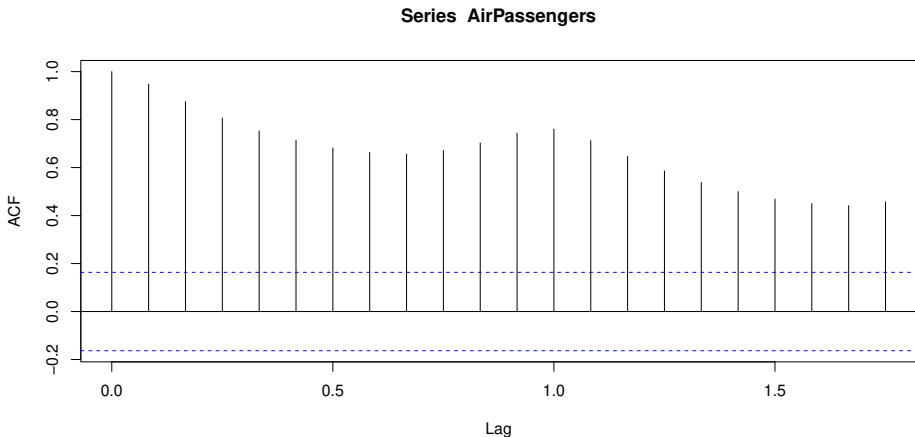


Figure 2: la funzione di autocorrelazione è uniformemente grande per via del trend

Per rimuovere questo effetto basta passare ad una derivata discreta tramite la funzione `diff()`.

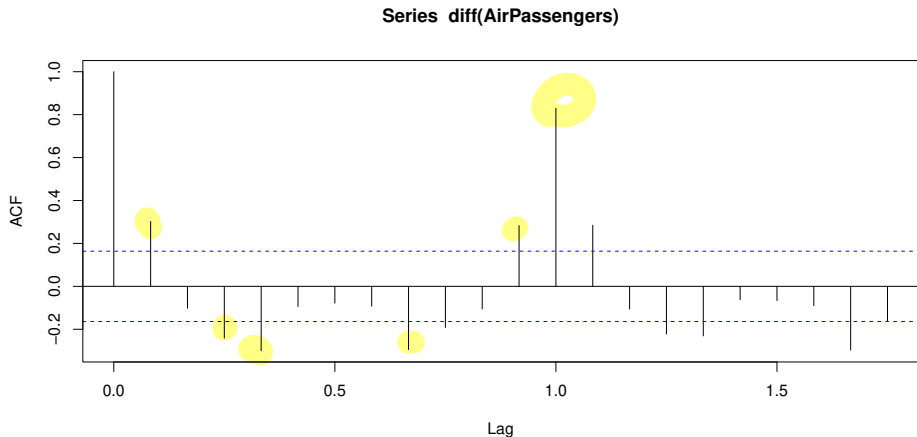


Figure 3: la derivata discreta rimuove il trend e la funzione di autocorrelazione evidenzia la stagionalità a 12 mesi

Densità spettrale di potenza

Torniamo alla base delle frequenze: cosa accade ai calcoli se non vale l'ipotesi semplificativa $C(k) = C(n-k)$?

- Se n è molto grande la stima per ξ fissato

$$\hat{C}(\xi) \approx \frac{|\hat{x}(\xi)|^2}{n} = \frac{1}{n} \left| \sum_{t=0}^{n-1} x_t e^{-2\pi i t \xi / n} \right|^2$$

è comunque una buona approssimazione.

Densità spettrale di potenza

Torniamo alla base delle frequenze: cosa accade ai calcoli se non vale l'ipotesi semplificativa $C(k) = C(n - k)$?

- Se n è molto grande la stima

Densità spettrale di potenza $\rightarrow \hat{C}(\xi) \approx \frac{|\hat{X}(\xi)|^2}{n} = \frac{1}{n} \left| \sum_{t=0}^{n-1} x_t e^{-2\pi i t \xi / n} \right|^2$

è comunque una buona approssimazione.

- Il membro a destra è l'energia associata alla frequenza ξ diviso il tempo n , quindi si interpreta come *potenza*.

Il Teorema di Wiener-Khinchin

Sia $(X_t)_{t=0}^{\infty}$ un processo a valori reali, stazionario in senso lato, con media nulla $\mathbb{E}[X_t] = 0$, e tale che

$$\sum_{k=0}^{\infty} |C(0, k)| < \infty. \quad \left(\text{OK, per ARMA stazionari} \right)$$

- Per ogni $\xi \in [0, 1]$, si definisce *densità spettrale di potenza* associata alla frequenza ξ la quantità

$$\hat{C}(\xi) = \sum_{k \in \mathbb{Z}} C(|k|) e^{-2\pi i k \xi}.$$

Il Teorema di Wiener-Khinchin

Sia $(X_t)_{t=0}^{\infty}$ un processo a valori reali, stazionario in senso lato, con media nulla $\mathbb{E}[X_t] = 0$, e tale che

$$\sum_{k=0}^{\infty} |C(0, k)| < \infty.$$

- Per ogni $\xi \in [0, 1]$, si definisce *densità spettrale di potenza* associata alla frequenza ξ la quantità

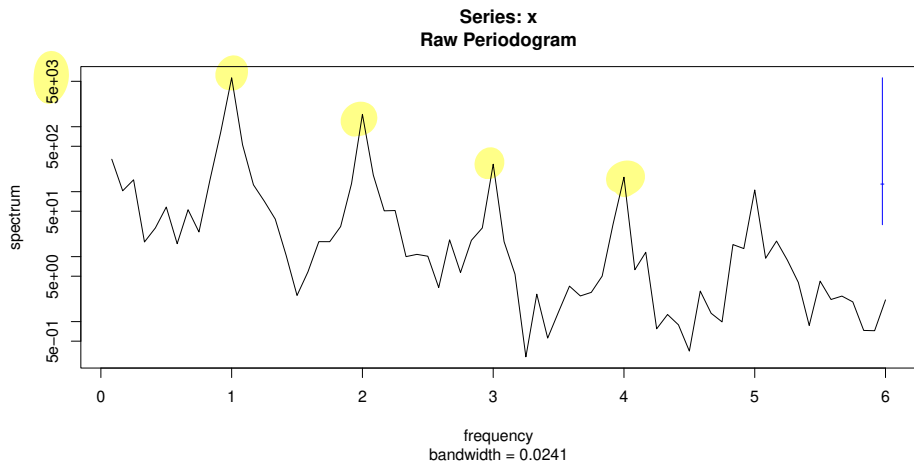
$$\hat{C}(\xi) = \sum_{k \in \mathbb{Z}} C(|k|) e^{-2\pi i k \xi}.$$

- Vale il limite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\left| \sum_{t=0}^{n-1} X_t e^{-2\pi i t \xi} \right|^2 \right] = \hat{C}(\xi).$$

Spettrogramma in R

In R si può utilizzare direttamente il comando `spectrum()`, che fornisce una stima della densità spettrale di potenza da una serie storica osservata.



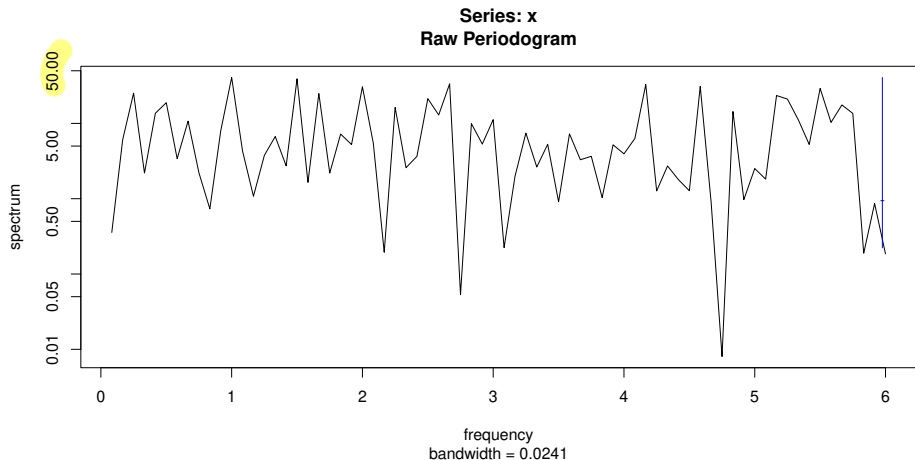


Figure 5: Stima della densità spettrale di potenza dei residui della serie AirPassengers dopo un fit con un modello ARIMA

Section 3

I teoremi limite

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.
- il **teorema limite centrale** fornisce una precisazione della legge dei grandi numeri facendo intervenire variabili **gaussiane**

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.
- il **teorema limite centrale** fornisce una precisazione della legge dei grandi numeri facendo intervenire variabili gaussiane
- Le applicazioni dei teoremi limite comprendono i metodi **Monte Carlo**, cui accenniamo.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.
- il **teorema limite centrale** fornisce una precisazione della legge dei grandi numeri facendo intervenire variabili gaussiane
- Le applicazioni dei teoremi limite comprendono i metodi Monte Carlo, cui accenniamo.
- Accenniamo infine ad altri teoremi limite, per gli eventi **estremi**.

Convergenza di variabili aleatorie

Dobbiamo precisare in che senso una successione di variabili aleatorie $(X_n)_{n=1}^{\infty}$ approssimi una variabile aleatoria limite X_{∞} , informalmente scriviamo

$$X_n \approx X_{\infty} \quad \text{o} \quad X_n \rightarrow X_{\infty}.$$

- Due approcci principali:

Convergenza di variabili aleatorie

Dobbiamo precisare in che senso una successione di variabili aleatorie $(X_n)_{n=1}^{\infty}$ approssimi una variabile aleatoria limite X_{∞} , informalmente scriviamo

$$X_n \approx X_{\infty} \quad \circ \quad X_n \rightarrow X_{\infty}.$$

- Due approcci principali:
 - 1 la distanza $|X_n - X_{\infty}|$ diventa piccola, con grande probabilità, al crescere di n ,

Convergenza di variabili aleatorie

Dobbiamo precisare in che senso una successione di variabili aleatorie $(X_n)_{n=1}^{\infty}$ approssimi una variabile aleatoria limite X_{∞} , informalmente scriviamo

$$X_n \approx X_{\infty} \quad \circ \quad X_n \rightarrow X_{\infty}.$$

- Due approcci principali:
 - 1 la distanza $|X_n - X_{\infty}|$ diventa piccola, con grande probabilità, al crescere di n ,
 - 2 le leggi di X_n convergono verso la legge di X_{∞} , ad esempio confrontandone le densità, le CDF, le MGF, i momenti, ecc.

usa la regola
 $\checkmark (X_n, X_{\infty})$

NON usa la regola
 congiunta

Convergenza in probabilità e in media quadratica

Siano $(X_n)_{n=1}^{\infty}$ e X_{∞} variabili aleatorie a valori in \mathbb{R}^d . Diciamo che X_n converge verso X_{∞}

- *in probabilità* se per ogni $\varepsilon > 0$, si ha

$$\lim_{n \rightarrow \infty} P(|X_n - X_{\infty}| \leq \varepsilon) = 1,$$

oppure, equivalentemente,

$$\lim_{n \rightarrow \infty} P(|X_n - X_{\infty}| > \varepsilon) = 0;$$

Convergenza in probabilità e in media quadratica

Siano $(X_n)_{n=1}^{\infty}$ e X_{∞} variabili aleatorie a valori in \mathbb{R}^d . Diciamo che X_n converge verso X_{∞}

- *in probabilità* se per ogni $\varepsilon > 0$, si ha

$$\lim_{n \rightarrow \infty} P(|X_n - X_{\infty}| \leq \varepsilon) = 1,$$

oppure, equivalentemente,

$$\lim_{n \rightarrow \infty} P(|X_n - X_{\infty}| > \varepsilon) = 0;$$

- *in media quadratica* se vale

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[|X_n - X_{\infty}|^2 \right] = 0.$$

La convergenza in media quadratica implica quella in probabilità:

$$P(|X_n - X_\infty| > \varepsilon) = P(|X_n - X_\infty|^2 > \varepsilon^2) \leq \frac{\mathbb{E}[|X_n - X_\infty|^2]}{\varepsilon^2},$$

$\rightarrow 0$
 $\text{se } n \rightarrow \infty$

\uparrow Markov

- se il membro a destra è infinitesimo per $n \rightarrow \infty$ anche quello a sinistra lo è.

Un criterio per la convergenza ad una costante

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie a valori in \mathbb{R}^d . Allora X_n converge verso una costante $c \in \mathbb{R}^d$ se e solo se

$$\mathbb{E}[X_n] \rightarrow c \quad \text{e} \quad \Sigma_{X_n} \rightarrow 0.$$

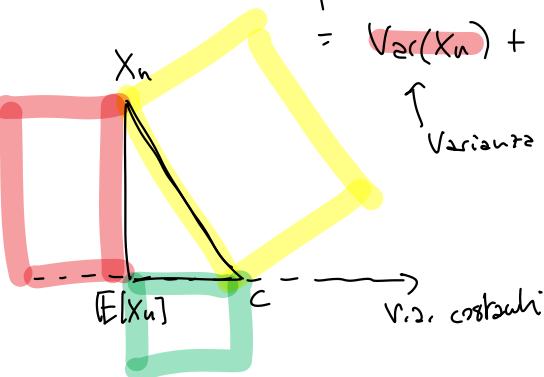
*in media
quadratica*

Dimostrazione

caso $d=1$ v.z. reali

$$\begin{aligned}
 E[|X_n - c|^2] &= E\left[\underbrace{|X_n - E[X_n]|}_A + \underbrace{|E[X_n] - c|}_B \right]^2 \\
 &= E[A^2] + E[B^2] + 2E[AB] \\
 &= \text{Var}(X_n) + |E[X_n] - c|^2 + \cancel{2B E[A]}
 \end{aligned}$$

\uparrow Varianza \uparrow bias



Dimostrazione

$$E[(X_n - c)^2] = \text{Var}(X_n) + |E[X_n] - c|^2$$

1) Se $X_n \rightarrow c$ in media quadratica

$$E[(X_n - c)^2] \rightarrow 0 \implies \text{Var}(X_n) \rightarrow 0$$

$$E[X_n] \rightarrow c$$

2) Viceversa se $\text{Var}(X_n) \rightarrow 0$ e $E[X_n] \rightarrow c \implies E[(X_n - c)^2] \rightarrow 0$

DSS Conv. in probabilità \implies Conv. in legge

Convergenza in legge

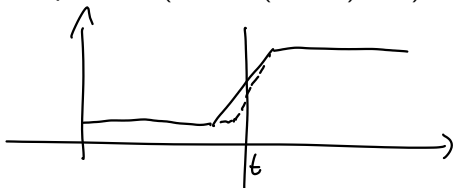
L'idea sarebbe di confrontare le densità delle X_n con la densità del limite X_∞ . Ma come fare se le X_n sono discrete e X_∞ è continua?

Siano $(X_n)_{n=1}^\infty$ e X_∞ variabili aleatorie a valori in \mathbb{R}^d . Diciamo che X_n converge verso X_∞ in legge se

- nel caso $d = 1$, si ha

$$\lim_{n \rightarrow \infty} \text{CDF}_{X_n}(t) = \text{CDF}_{X_\infty}(t)$$

[per ogni $t \in \mathbb{R}$ eccetto al più i punti t in cui $\text{CDF}_{X_\infty}(t)$ ha una discontinuità di tipo salto (ossia $P(X = t) > 0$)



Convergenza in legge

L'idea sarebbe di confrontare le densità delle X_n con la densità del limite X_∞ . Ma come fare se le X_n sono discrete e X_∞ è continua?

Siano $(X_n)_{n=1}^\infty$ e X_∞ variabili aleatorie a valori in \mathbb{R}^d . Diciamo che X_n converge verso X_∞ *in legge* se

- nel caso $d = 1$, si ha

$$\lim_{n \rightarrow \infty} \text{CDF}_{X_n}(t) = \text{CDF}_X(t)$$

per ogni $t \in \mathbb{R}$ eccetto al più i punti t in cui $\text{CDF}_X(t)$ ha una discontinuità di tipo salto (ossia $P(X = t) > 0$)

- nel caso generale $d \geq 1$, si ha

$$\lim_{n \rightarrow \infty} \text{MGF}_{X_n}(t) = \text{MGF}_{X_\infty}(t)$$

per ogni t in cui $\text{MGF}_{X_\infty}(t)$ sia finita, supponendo che $\text{MGF}_X(t)$ sia finita per t sufficientemente piccolo.

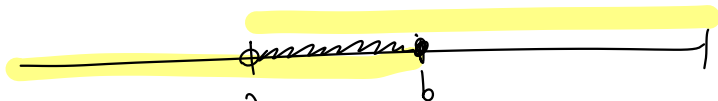
In alternativa, si può richiedere la convergenza delle funzioni caratteristiche per ogni $\omega \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(\omega) = \varphi_{X_\infty}(\omega).$$

- Se $d = 1$ e la variabile X_∞ ha densità continua, allora CDF_{X_∞} non ha salti. Possiamo allora chiedere che per ogni $a \leq b \in \mathbb{R}$,

$$\lim_n P(X_n \in (a, b]) = P(X_\infty \in (a, b]).$$

$$\text{CDF}_{X_n}(b) = P(X_n \leq b) \longrightarrow \text{CDF}_{X_\infty}(b) = P(X_\infty \leq b)$$



$$P(X_n \in (a, b]) = P(X_n \leq b) - P(X_n \leq a)$$

In alternativa, si può richiedere la convergenza delle funzioni caratteristiche per ogni $\omega \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(\omega) = \varphi_{X_\infty}(\omega).$$

- Se $d = 1$ e la variabile X_∞ ha densità continua, allora CDF_{X_∞} non ha salti. Possiamo allora chiedere che per ogni $a \leq b \in \mathbb{R}$,

$$\lim_n P(X_n \in (a, b]) = P(X_\infty \in (a, b]).$$

- La convergenza in legge è anche equivalente alla convergenza

$$\lim_{n \rightarrow \infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X_\infty)]$$

per ogni funzione g continua ovunque e uniformemente limitata (ossia esiste una costante c tale che $|g(x)| \leq c$ per ogni $x \in \mathbb{R}^d$).

Legge dei grandi numeri

La legge dei grandi numeri è un teorema che giustifica l'interpretazione della **probabilità** di una affermazione come **frequenza relativa** con si realizza in una successione di esperimenti ripetuti, sotto le stesse condizioni, **ma tutti indipendenti** tra loro.

- Ne diamo una dimostrazione usando la **convergenza in media quadratica** (e quindi in probabilità).

Legge dei grandi numeri

La legge dei grandi numeri è un teorema che giustifica l'interpretazione della **probabilità** di una affermazione come **frequenza relativa** con si realizza in una successione di esperimenti ripetuti, sotto le stesse condizioni, ma tutti indipendenti tra loro.

- Ne diamo una dimostrazione usando la convergenza in media quadratica (e quindi in probabilità).
- Cominciamo con il caso più semplice delle estrazioni con rimpiazzo da un'urna.

Modello dell'urna

La frazione delle palline rosse è $r \in [0, 1]$ e si effettuano n estrazioni con rimpiazzo.

- il numero R_n di palline rosse estratte ha densità binomiale di parametri (n, r) , con

$$\mathbb{E}[R_n] = n \cdot r \quad \text{e} \quad \text{Var}(R_n) = n r (1-r)$$

$$P(R_n = k) = \binom{n}{k} r^k (1-r)^{n-k}$$

$$\mathbb{E}\left[\frac{R_n}{n}\right] = \frac{\mathbb{E}[R_n]}{n} = \frac{n r}{n} = r \quad \left| \quad \text{Var}\left(\frac{R_n}{n}\right) = \frac{1}{n^2} n r (1-r) = \frac{1}{n} r (1-r)$$

Modello dell'urna

La frazione delle palline rosse è $r \in [0, 1]$ e si effettuano n estrazioni con rimpiazzo.

- il numero R_n di palline rosse estratte ha densità binomiale di parametri (n, r) , con

$$\mathbb{E}[R_n] = \quad \text{e} \quad \text{Var}(R_n) = \quad .$$

- Passando alla **frequenza relativa** R_n/n , possiamo scrivere informalmente:

$$\left[\frac{R_n}{n} \right] \approx r \pm \sigma\left(\frac{R_n}{n}\right) = r \pm \frac{\sqrt{r(1-r)}}{\sqrt{n}}$$

Modello dell'urna

La frazione delle palline rosse è $r \in [0, 1]$ e si effettuano n estrazioni con rimpiazzo.

- il numero R_n di palline rosse estratte ha densità binomiale di parametri (n, r) , con

$$\mathbb{E}[R_n] = \quad \text{e} \quad \text{Var}(R_n) = \quad .$$

- Passando alla **frequenza relativa** R_n/n , possiamo scrivere informalmente:

$$\frac{R_n}{n} \approx r \pm \frac{c(r)}{\sqrt{n}} \rightarrow r$$

- Al tendere di $n \rightarrow \infty$ la frazione $\boxed{R_n/n}$ converge verso r , la frazione di palline rosse sul totale.

Legge dei grandi numeri per esperimenti ripetuti

Partendo dal modello delle estrazioni dall'urna possiamo estendere ad una qualsiasi situazione in cui vi siano un grande numero, potenzialmente illimitato, di esperimenti ripetuti, tutti indipendenti tra loro, e ciascuno con probabilità di successo $p \in [0, 1]$.

- La frequenza relativa del numero di successi sul totale degli esperimenti converge quindi alla probabilità di successo di un singolo esperimento.

Legge dei grandi numeri per esperimenti ripetuti

Partendo dal modello delle estrazioni dall'urna possiamo estendere ad una qualsiasi situazione in cui vi siano un grande numero, potenzialmente illimitato, di esperimenti ripetuti, tutti indipendenti tra loro, e ciascuno con probabilità di successo $p \in [0, 1]$.

- La *frequenza relativa del numero di successi sul totale degli esperimenti* converge quindi alla *probabilità di successo di un singolo esperimento*.
- Attenzione: può essere limitante interpretare la **probabilità solo come frequenza**.

Applicazione: stima della probabilità

Supponiamo che inizialmente la frazione di **palline rosse** non sia nota, e si supponga una variabile aleatoria R a valori in $[0, 1]$ ad esempio uniforme.

Allora,

$$\begin{aligned}
 P(|R_n/n - R| \leq \varepsilon) &= \int_0^1 \boxed{P(|R_n/n - r| \leq \varepsilon | R = r)} dr \\
 &\geq 1 - \frac{\int_0^1 r(1-r) dr}{n\varepsilon^2} = 1 - \frac{1}{\boxed{n\varepsilon^2}} \rightarrow 1
 \end{aligned}$$

$\sqrt{\frac{r(1-r)}{n}}$
 $\varepsilon \approx \frac{1}{\sqrt{5n}}$
 $\sqrt{n} \approx \frac{1}{\varepsilon^2}$

Dis di Markov - che lo è

per $n \rightarrow \infty$, ossia con alta probabilità la frequenza relativa R_n/n è vicina alla variabile R (precisamente abbiamo mostrato la convergenza in probabilità).

- La probabilità calcolata sopra è rispetto all'informazione a priori, ossia prima di effettuare le estrazioni (o prima di essere informati dell'esito).

Istogrammi e densità teorica

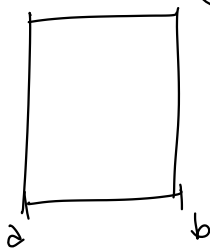
Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^n$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).

Istogrammi e densità teorica

Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^n$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).
- Consideriamo un rettangolo di base $a < b \in \mathbb{R}$:



Istogrammi e densità teorica

Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^n$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).
- Consideriamo un rettangolo di base $a < b \in \mathbb{R}$:
- ① l'istogramma delle frequenze (assolute) avrà altezza $H(a, b)$ pari al numero delle X_i tali che $a < X_i \leq b$,

Istogrammi e densità teorica

Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^n$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).
- Consideriamo un rettangolo di base $a < b \in \mathbb{R}$:
 - 1 l'istogramma delle frequenze (assolute) avrà altezza $H(a, b)$ pari al numero delle X_i tali che $a < X_i \leq b$,
 - 2 quello delle densità è ulteriormente diviso il numero delle osservazioni n e per la lunghezza della base $(b - a)$.

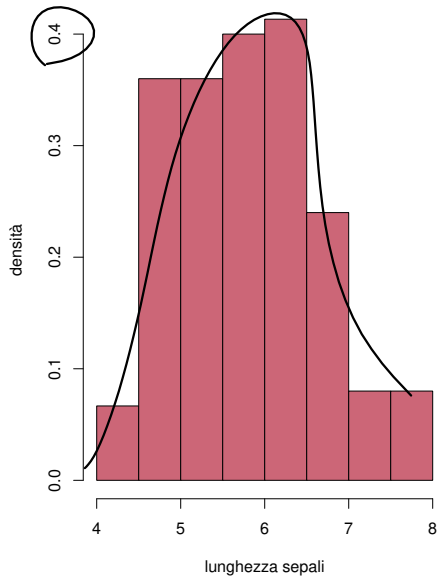
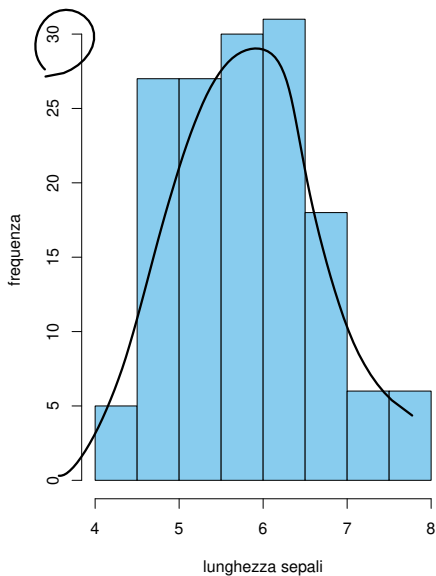


Figure 6: istogramma delle frequenze (a sinistra) e delle densità (a destra)

Per costruire l'istogramma, possiamo pensare ad un "successo" se $X_i \in (a, b]$, che avviene con probabilità

$$r = P(X_1 \in (a, b]) = \int_a^b p(X_1 = x) dx \approx p(X_1 = a)(b - a),$$

dove nell'ultima approssimazione supponiamo la densità regolare e $b - a$ piccolo.

- Considerando n esperimenti indipendenti si avrà

$$\frac{H(a, b)}{n} \approx r \pm \sqrt{\frac{r(1-r)}{n}} \approx p(X_1 = a)(b - a),$$

e quindi l'istogramma delle densità, che ha altezza $H(a, b)/(n(b - a))$, è, con alta probabilità vicino alla densità teorica.