

Probabilità e Processi Stocastici (455AA)

Lezione 22

Dario Trevisan

11/12/2023

Section 1

Legge dei grandi numeri generale

Verso una legge dei grandi numeri generale

Estendiamo a situazioni in cui l'esito di ciascun "esperimento" sia una variabile aleatoria X_i a valori reali.

- la frequenza relativa dei successi è sostituita dalla media empirica

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Verso una legge dei grandi numeri generale

Estendiamo a situazioni in cui l'esito di ciascun "esperimento" sia una variabile aleatoria X_i a valori reali.

- la frequenza relativa dei successi è sostituita dalla media empirica

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

- Per dedurre la convergenza in media quadratica delle frequenze relative, avevamo usato che la legge della somma $\sum_{i=1}^n X_i$, ossia il numero di successi, ha densità discreta binomiale di parametri (n, p) .

Verso una legge dei grandi numeri generale

Estendiamo a situazioni in cui l'esito di ciascun "esperimento" sia una variabile aleatoria X_i a valori reali.

- la frequenza relativa dei successi è sostituita dalla media empirica

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

- Per dedurre la convergenza in media quadratica delle frequenze relative, avevamo usato che la legge della somma $\sum_{i=1}^n X_i$, ossia il numero di successi, ha densità discreta binomiale di parametri (n, p) .
- Tuttavia ripercorrendo l'argomento, basta conoscere molto meno: infatti è sufficiente che il valor medio $\mathbb{E}[\bar{X}_n]$ converga a una costante m e la varianza $\text{Var}(\bar{X}_n)$ sia infinitesima per $n \rightarrow \infty$.

Teorema

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie non correlate, tutte con lo stesso valor medio e varianza

$$\mathbb{E}[X_n] = m, \quad \text{Var}(X_n) = \sigma^2 < \infty.$$

Allora, si ha la convergenza in media quadratica (e quindi in probabilità)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = m.$$

Dimostrazione

Dalla dimostrazione segue che la deviazione standard della variabile \bar{X}_n è

$$\sigma_{\bar{X}_n} = \sqrt{\text{Var}(\bar{X}_n)} = \frac{\sigma}{\sqrt{n}},$$

e quindi informalmente possiamo scrivere

$$\bar{X}_n = m \pm \frac{\sigma}{\sqrt{n}}.$$

Applicazione: convergenza della varianza campionaria

Possiamo usare la legge dei grandi numeri anche per mostrare la convergenza della varianza campionaria

$$\bar{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

- Supponendo ad esempio che le $(X_i)_i$ siano tutte indipendenti, tutte con la stessa legge e dotate di momento quarto finito, quindi in particolare i momenti sono tutti uguali:

$$m_1 = \mathbb{E}[X_i], \quad m_2 = \mathbb{E}[X_i^2],$$

vale la convergenza in media quadratica e in probabilità

$$\lim_{n \rightarrow \infty} \bar{\sigma}_n^2 = \sigma^2,$$

Dimostrazione

Sull'ipotesi di indipendenza

Senza l'ipotesi di indipendenza (o non correlazione), la varianza della media campionaria è in generale la somma di n^2 termini

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i,j=1}^n \text{Cov}(X_i, X_j),$$

e quindi non segue necessariamente che sia infinitesima, anche tenendo in conto del denominatore n^2 .

- Può accadere tuttavia che $\text{Cov}(X_i, X_j)$ sia piccolo per “molte” coppie di indici, ad esempio per opportuni processi stocastici.

Media temporale di un processo

Dato un processo stocastico reale $(X_t)_{t=1}^{\infty}$, la media sui primi T tempi è $\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$.

- Se l'insieme degli stati E del processo non è un sottoinsieme di \mathbb{R} , si può considerare una qualsiasi *osservabile* $g : E \rightarrow \mathbb{R}$ e considerare la media

$$\overline{g(X)}_T = \frac{1}{T} \sum_{t=1}^T g(X_t).$$

Media temporale di un processo

Dato un processo stocastico reale $(X_t)_{t=1}^{\infty}$, la media sui primi T tempi è $\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$.

- Se l'insieme degli stati E del processo non è un sottoinsieme di \mathbb{R} , si può considerare una qualsiasi *osservabile* $g : E \rightarrow \mathbb{R}$ e considerare la media

$$\overline{g(X)}_T = \frac{1}{T} \sum_{t=1}^T g(X_t).$$

- Se g è la funzione indicatrice di un qualsiasi stato $x_0 \in E$,

$$g(x) = \begin{cases} 1 & \text{se } x = x_0 \\ 0 & \text{se } x \neq x_0. \end{cases}$$

Allora $\overline{g(X)}_T$ è la frazione di tempo trascorsa dal processo sullo stato x_0 , dal tempo $t = 1$ al tempo $t = T$.

Teorema ergodico

Se il processo $(X_t)_t$ è stazionario, si vuole indentificare il limite (se esiste)

$$\lim_{T \rightarrow \infty} \overline{g(X)}_T.$$

come il valor medio di g rispetto alla legge marginale in un qualsiasi istante, ad esempio nel caso di E discreto

$$\mathbb{E}[g(X_i)] = \sum_{x \in E} g(x)P(X_i = x).$$

- Se g è l'indicatrice di uno stato x_0 , il valor medio è la probabilità $\mathbb{E}[g(X_i)] = P(X_i = x_0)$.

Teorema ergodico

Se il processo $(X_t)_t$ è stazionario, si vuole indentificare il limite (se esiste)

$$\lim_{T \rightarrow \infty} \overline{g(X)}_T.$$

come il valor medio di g rispetto alla legge marginale in un qualsiasi istante, ad esempio nel caso di E discreto

$$\mathbb{E}[g(X_i)] = \sum_{x \in E} g(x)P(X_i = x).$$

- Se g è l'indicatrice di uno stato x_0 , il valor medio è la probabilità $\mathbb{E}[g(X_i)] = P(X_i = x_0)$.
- Risultati che garantiscono tale identificazione sono storicamente detti *teoremi ergodici*.

Enunciato del teorema ergodico

Sia $(X_t)_{t=0}^{\infty}$ un processo stazionario sull'insieme degli stati E e sia $g : E \rightarrow \mathbb{R}$ una osservabile. Se

$$\lim_{t \rightarrow \infty} \text{Cov}(g(X_0), g(X_t)) = 0,$$

allora vale la convergenza in media quadratica e in probabilità

$$\lim_{T \rightarrow \infty} \overline{g(X)}_T = \mathbb{E}[g(X_0)].$$

(per semplicità abbiamo specificato X_0 , ma un qualsiasi altro tempo X_t sarebbe lo stesso, essendo il processo stazionario).

Dimostrazione

Applicazioni: catene di Markov

Sia $(X_t)_t$ una catena di Markov stazionaria e irriducibile su un insieme di stati finito E . Allora per ogni $g : E \rightarrow \mathbb{R}$ vale la convergenza

$$\lim_{T \rightarrow \infty} \overline{g(X)}_T = \sum_{i \in E} g(i) \pi_i,$$

dove $\pi = (\pi_i)_{i \in E}$ è l'unica distribuzione invariante per la catena.

- In particolare, la frazione di tempo trascorsa dalla catena su uno stato è, nel limite, pari alla probabilità che la catena si trovi su quello stato.

Applicazioni: catene di Markov

Sia $(X_t)_t$ una catena di Markov stazionaria e irriducibile su un insieme di stati finito E . Allora per ogni $g : E \rightarrow \mathbb{R}$ vale la convergenza

$$\lim_{T \rightarrow \infty} \overline{g(X)}_T = \sum_{i \in E} g(i) \pi_i,$$

dove $\pi = (\pi_i)_{i \in E}$ è l'unica distribuzione invariante per la catena.

- In particolare, la frazione di tempo trascorsa dalla catena su uno stato è, nel limite, pari alla probabilità che la catena si trovi su quello stato.
- Un teorema analogo vale per processi di Markov a salti, dove la media nel tempo è intesa come integrale.

Applicazioni: processi ARIMA

Sia $(X_t)_t$ un processo ARIMA($p, 0, q$) stazionario. Allora

$$\lim_{T \rightarrow \infty} \bar{X}_T = 0,$$

e per ogni $t \in \mathbb{N}$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T X_s X_{s+t} = C(0, t).$$

- La funzione di autocovarianza empirica converge quindi a quella teorica. Questo fatto vale per processi stazionari anche più generali degli ARIMA.

Section 2

Il teorema limite centrale

Un raffinamento della legge dei grandi numeri

Torniamo all'esempio delle estrazioni con rimpiazzo da un'urna contenente una frazione $r \in (0, 1)$ di palline rosse.

- Posto R_n il numero di palline rosse estratte, la legge dei grandi numeri afferma informalmente che

$$\frac{R_n}{n} \approx r \pm \sqrt{\frac{r(1-r)}{n}}.$$

Un raffinamento della legge dei grandi numeri

Torniamo all'esempio delle estrazioni con rimpiazzo da un'urna contenente una frazione $r \in (0, 1)$ di palline rosse.

- Posto R_n il numero di palline rosse estratte, la legge dei grandi numeri afferma informalmente che

$$\frac{R_n}{n} \approx r \pm \sqrt{\frac{r(1-r)}{n}}.$$

- Il *teorema limite centrale* rende più preciso il simbolo \pm , mostrando che per una variabile gaussiana Z standard, ossia $\mathcal{N}(0, 1)$, vale

$$\frac{R_n}{n} \approx r + Z \sqrt{\frac{r(1-r)}{n}},$$

dove l'approssimazione è nel senso della convergenza in legge.

Un raffinamento della legge dei grandi numeri

Torniamo all'esempio delle estrazioni con rimpiazzo da un'urna contenente una frazione $r \in (0, 1)$ di palline rosse.

- Posto R_n il numero di palline rosse estratte, la legge dei grandi numeri afferma informalmente che

$$\frac{R_n}{n} \approx r \pm \sqrt{\frac{r(1-r)}{n}}.$$

- Il *teorema limite centrale* rende più preciso il simbolo \pm , mostrando che per una variabile gaussiana Z standard, ossia $\mathcal{N}(0, 1)$, vale

$$\frac{R_n}{n} \approx r + Z \sqrt{\frac{r(1-r)}{n}},$$

dove l'approssimazione è nel senso della convergenza in legge.

- In altre parole la variabile binomiale R_n si approssima con una variabile gaussiana avente la stessa media nr e varianza $nr(1-r)$.

Confronto grafico tra le densità

Possiamo visualizzare questo risultato graficamente confrontando la densità discreta binomiale e la gaussiana corrispondente.

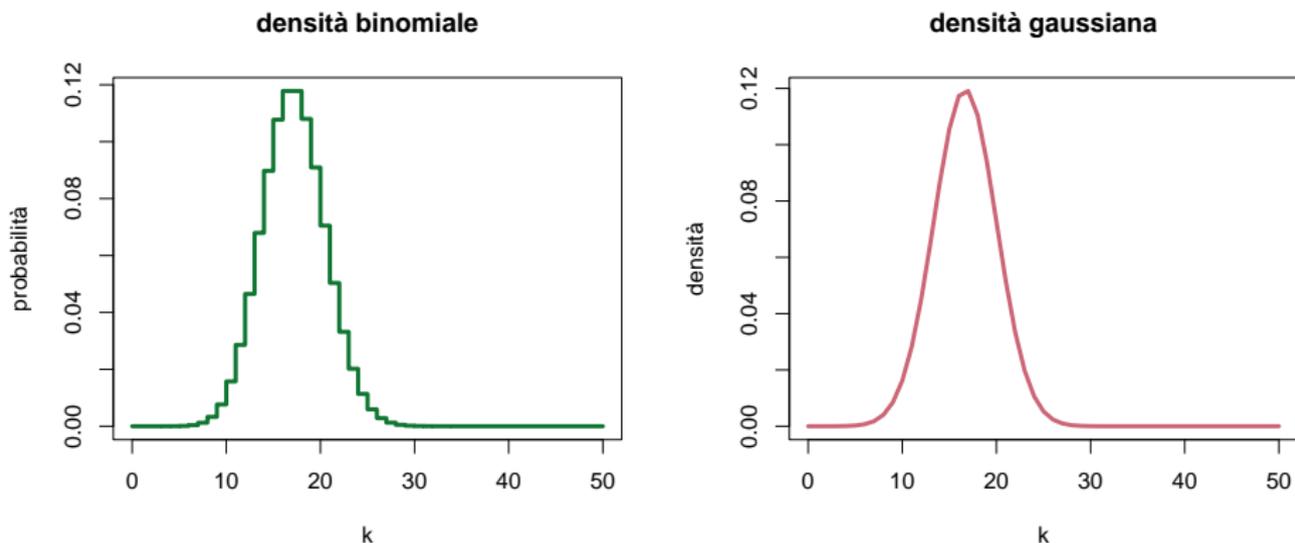


Figure 1: confronto tra densità binomiale di parametri $n = 50$, $r = 1/3$ e la densità gaussiana con medesima media nr e varianza $nr(1 - r)$.

Confronto tra le CDF

L'approssimazione vale nel senso della convergenza in legge: si confrontano le CDF piuttosto che le densità.

- Rigorosamente: dati $a \leq b \in \mathbb{R}$, vale

$$\lim_{n \rightarrow \infty} P \left(a \sqrt{\frac{r(1-r)}{n}} \leq \frac{R_n}{n} - r \leq b \sqrt{\frac{r(1-r)}{n}} \right) = \int_a^b \exp \left(-\frac{z^2}{2} \right) \frac{dz}{\sqrt{2\pi}}$$

Confronto tra le CDF

L'approssimazione vale nel senso della convergenza in legge: si confrontano le CDF piuttosto che le densità.

- Rigorosamente: dati $a \leq b \in \mathbb{R}$, vale

$$\lim_{n \rightarrow \infty} P \left(a \sqrt{\frac{r(1-r)}{n}} \leq \frac{R_n}{n} - r \leq b \sqrt{\frac{r(1-r)}{n}} \right) = \int_a^b \exp \left(-\frac{z^2}{2} \right) \frac{dz}{\sqrt{2\pi}}$$

- Conviene introdurre le variabili standardizzate delle R_n/n , ossia

$$Z_n = \left(\frac{R_n}{n} - r \right) \sqrt{\frac{n}{r(1-r)}},$$

in modo che la convergenza in legge sia

$$\text{CDF}_{Z_n}(t) \rightarrow \text{CDF}_Z(t) = \int_{-\infty}^t e^{-z^2/2} \frac{dz}{\sqrt{2\pi}}.$$

Possiamo verificare numericamente l'approssimazione scrivendo CDF_{Z_n} in termini della CDF_{R_n} , che è binomiale.

- usiamo l'identità, per $a > 0$, $b \in \mathbb{R}$,

$$CDF_{aX+b}(t) = CDF_X((t - b)/a).$$

Possiamo verificare numericamente l'approssimazione scrivendo CDF_{Z_n} in termini della CDF_{R_n} , che è binomiale.

- usiamo l'identità, per $a > 0$, $b \in \mathbb{R}$,

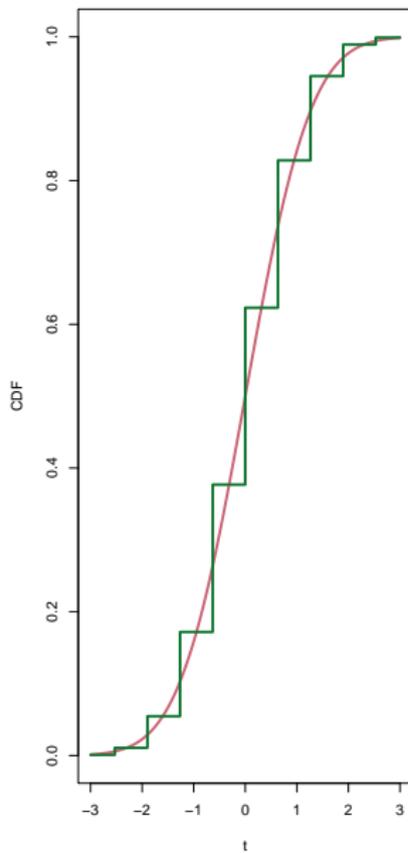
$$CDF_{aX+b}(t) = CDF_X((t - b)/a).$$

- Scriviamo

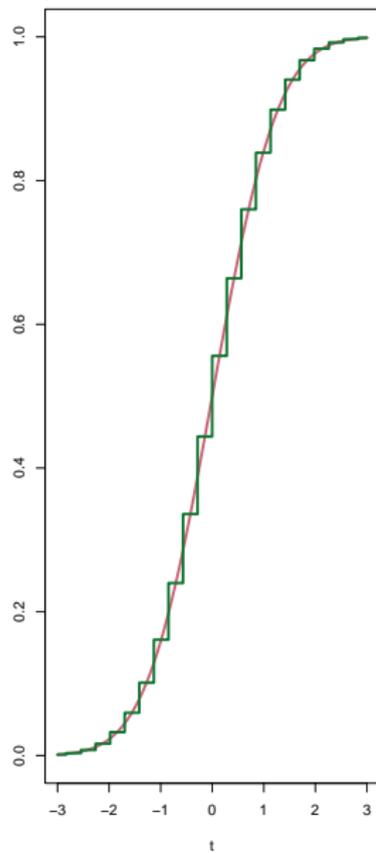
$$CDF_{Z_n}(t) = CDF_{R_n} \left(nr + t\sqrt{nr(1-r)} \right),$$

e visualizziamo graficamente in R.

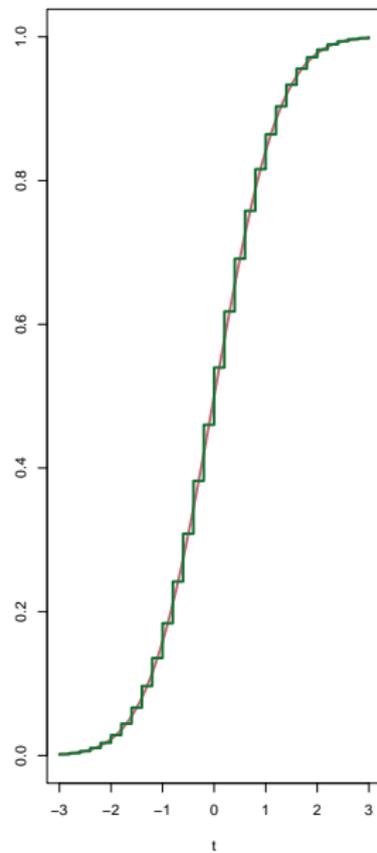
n=10



n=50



n=100



Il caso generale

Affrontiamo direttamente una dimostrazione del risultato generale per l'approssimazione gaussiana delle medie campionarie di variabili indipendenti.

Ricordiamo che informalmente, la legge dei grandi numeri dava

$$\bar{X}_n \approx m \pm \frac{\sigma}{\sqrt{n}}.$$

- Possiamo rendere più preciso il simbolo \pm introducendo una variabile gaussiana Z standard $\mathcal{N}(0, 1)$. L'approssimazione è nel senso della convergenza in legge.

Enunciato del teorema limite centrale

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie reali indipendenti, tutte con la stessa legge e quindi

$$m = \mathbb{E}[X_n] \quad \text{e} \quad \sigma^2 = \text{Var}(X_n) \in (0, \infty).$$

Allora, posta $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ e

$$Z_n = (\bar{X}_n - m) \frac{\sqrt{n}}{\sigma}$$

la sua standardizzata, si ha la convergenza in legge

$$\lim_{n \rightarrow \infty} Z_n = Z,$$

dove Z è gaussiana standard.

- Per ogni $a \leq b$, vale

$$\lim_{n \rightarrow \infty} P\left(a\sigma/\sqrt{n} \leq \bar{X}_n - m \leq b\sigma/\sqrt{n}\right) = \int_a^b e^{-z^2/2} \frac{dz}{\sqrt{2\pi}}.$$

Dimostrazione

Cenni al caso vettoriale

Se le variabili aleatorie sono a valori vettoriali indipendenti, a valori in \mathbb{R}^d , tutte con la stessa legge, quindi

$$\mathbb{E}[X_i] = m, \quad \Sigma_{X_i} = \Sigma,$$

allora si ha la convergenza in legge

$$\left(\bar{X}_n - m\right) \sqrt{n} \rightarrow Z$$

dove Z è una variabile gaussiana vettoriale con densità $\mathcal{N}(0, \Sigma)$.

- Oltre alle convergenze delle marginali si ha quindi anche convergenza delle variabili congiunte.

Section 3

Cenni ai metodi Monte Carlo

Simulare variabili aleatorie

Una delle applicazioni principali dei teoremi limite è di combinarli alle tecniche di *generazione* di numeri pseudo-casuali mediante opportuni algoritmi (che non descriviamo nel dettaglio).

- Tramite semplici comandi in R è possibile *simulare* variabili aleatorie con densità comuni (uniformi, binomiali, poisson, esponenziali, gaussiane ecc.).

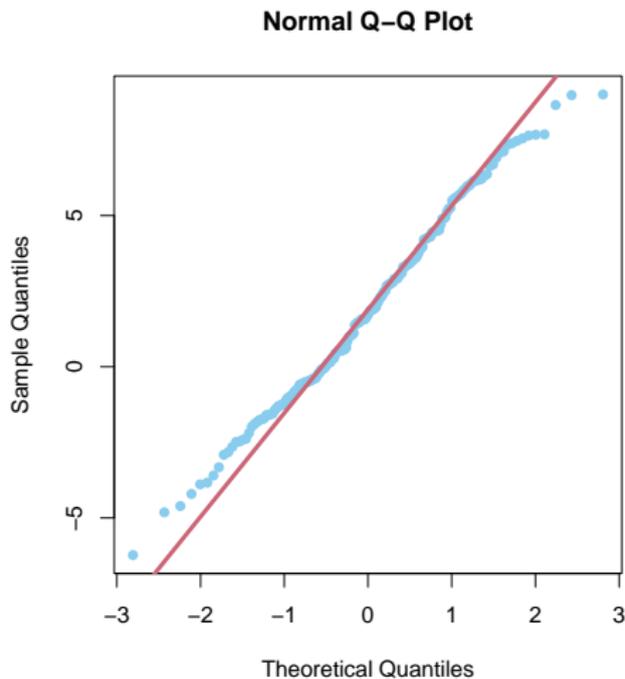
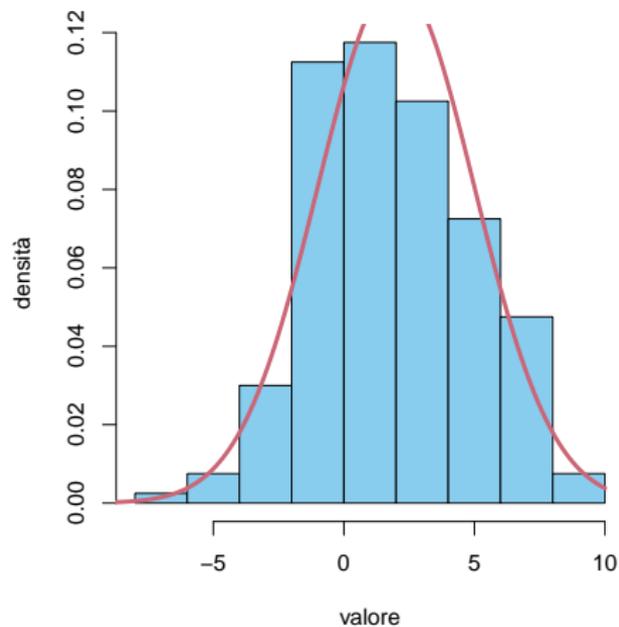


Figure 3: istogramma e qqplot di 200 variabili gaussiane indipendenti

Monte Carlo per l'integrazione numerica

I metodi *Monte Carlo* sono algoritmi che usano un gran numero di simulazioni di variabili per risolvere dei problemi numericamente, anche non legati alla probabilità in apparenza.

- Ad esempio, per calcolare l'integrale

$$\int_{\mathbb{R}^d} g(x)p(X = x)dx,$$

si ricorre alla simulazione di n variabili indipendenti con densità $p(X = \cdot)$ e quindi

$$\int_{\mathbb{R}^d} g(x)p(X = x)dx \approx \frac{1}{n} \sum_{i=1}^n g(X_i),$$

Monte Carlo per l'integrazione numerica

I metodi *Monte Carlo* sono algoritmi che usano un gran numero di simulazioni di variabili per risolvere dei problemi numericamente, anche non legati alla probabilità in apparenza.

- Ad esempio, per calcolare l'integrale

$$\int_{\mathbb{R}^d} g(x)p(X = x)dx,$$

si ricorre alla simulazione di n variabili indipendenti con densità $p(X = \cdot)$ e quindi

$$\int_{\mathbb{R}^d} g(x)p(X = x)dx \approx \frac{1}{n} \sum_{i=1}^n g(X_i),$$

- Il teorema limite centrale garantisce oscillazioni gaussiane:

$$\int_{\mathbb{R}^d} g(x)p(X = x)dx - \frac{1}{n} \sum_{i=1}^n g(X_i) \approx \frac{Z\sigma}{\sqrt{n}}.$$

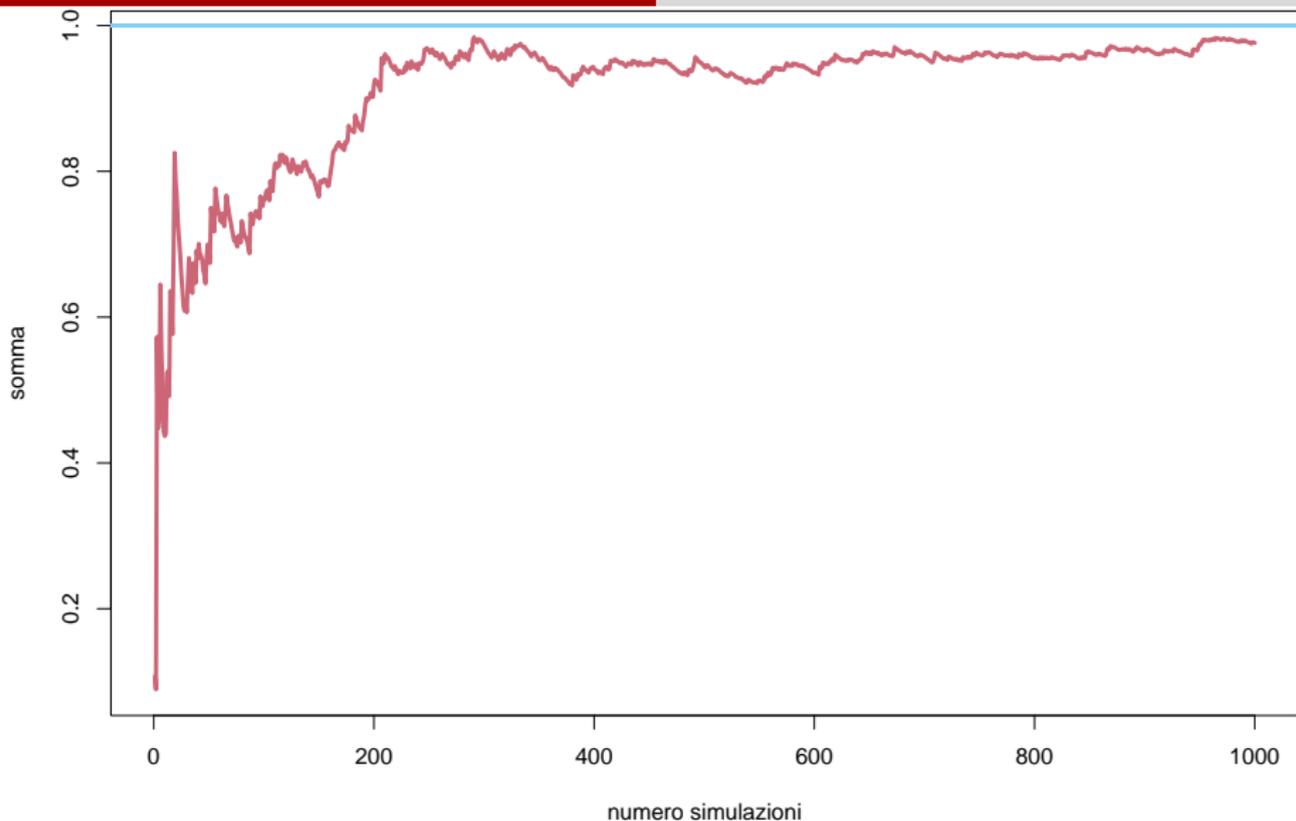


Figure 4: Calcolo del momento secondo di una gaussiana standard tramite media empirica di un campione indipendente

Section 4

Cenni agli eventi estremi

Massimo e minimo di variabili

È molto rilevante ai fini pratici lo studio delle caratteristiche *estreme* di una famiglia di variabili aleatorie.

- Invece della media campionaria, studiamo il massimo

$$M_n = \max_{i=1, \dots, n} X_i,$$

o il minimo

$$m_n = \min_{i=1, \dots, n} X_i.$$

Massimo e minimo di variabili

È molto rilevante ai fini pratici lo studio delle caratteristiche *estreme* di una famiglia di variabili aleatorie.

- Invece della media campionaria, studiamo il massimo

$$M_n = \max_{i=1, \dots, n} X_i,$$

o il minimo

$$m_n = \min_{i=1, \dots, n} X_i.$$

- Supponendo che le variabili $(X_n)_{n=1}^{\infty}$ siano indipendenti e tutte con la stessa legge, si può investigare il limite al tendere di $n \rightarrow \infty$ delle due variabili.

Per comprendere M_n è utile considerarne la CDF:

$$\begin{aligned} \text{CDF}_{M_n}(t) &= P\left(\max_{i=1,\dots,n} X_i \leq t\right) \\ &= P(X_1 \leq t, X_2 \leq t, \dots, X_n \leq t) \\ &= P(X_1 \leq t)P(X_2 \leq t) \dots P(X_n \leq t). \end{aligned}$$

- Nel caso di densità continue delle X_i , derivando questa identità si trova la densità di M_n .

Per comprendere M_n è utile considerarne la CDF:

$$\begin{aligned} \text{CDF}_{M_n}(t) &= P(\max_{i=1,\dots,n} X_i \leq t) \\ &= P(X_1 \leq t, X_2 \leq t, \dots, X_n \leq t) \\ &= P(X_1 \leq t)P(X_2 \leq t) \dots P(X_n \leq t). \end{aligned}$$

- Nel caso di densità continue delle X_i , derivando questa identità si trova la densità di M_n .
- Supponiamo che tutte le X_i abbiano la stessa legge (e quindi la stessa CDF) e mostriamo un risultato analogo alla legge dei grandi numeri.

Un teorema generale

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie a valori reali, indipendenti e tutte con la stessa legge.

- Se $t \in \mathbb{R}$ è tale che $\text{CDF}_{X_1}(t) < 1$, ossia

$$P(X_1 > t) > 0,$$

allora

$$\lim_{n \rightarrow \infty} P(M_n > t) = 1.$$

Un teorema generale

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie a valori reali, indipendenti e tutte con la stessa legge.

- Se $t \in \mathbb{R}$ è tale che $\text{CDF}_{X_1}(t) < 1$, ossia

$$P(X_1 > t) > 0,$$

allora

$$\lim_{n \rightarrow \infty} P(M_n > t) = 1.$$

- Se X_1 assume con probabilità positiva (anche piccola) valori arbitrariamente grandi, allora M_n converge verso $+\infty$.

Un teorema generale

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie a valori reali, indipendenti e tutte con la stessa legge.

- Se $t \in \mathbb{R}$ è tale che $\text{CDF}_{X_1}(t) < 1$, ossia

$$P(X_1 > t) > 0,$$

allora

$$\lim_{n \rightarrow \infty} P(M_n > t) = 1.$$

- Se X_1 assume con probabilità positiva (anche piccola) valori arbitrariamente grandi, allora M_n converge verso $+\infty$.
- Informalmente è perché se un evento ha probabilità positiva, prima o poi si realizzerà (ma in un tempo tipico inversamente proporzionale alla probabilità).

Dimostrazione

Un teorema più preciso

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie indipendenti, tutte con densità esponenziale del medesimo parametro $\lambda > 0$. Allora si ha la convergenza in legge

$$\lim_{n \rightarrow \infty} M_n - \frac{\log n}{\lambda} = G,$$

dove G è una variabile con distribuzione di Gumbel, ossia con funzione di ripartizione, per $t \in \mathbb{R}$,

$$\text{CDF}_G(t) = \exp(-e^{-t}).$$

- La densità di G si ottiene derivando:

$$\frac{d}{dt} \text{CDF}_G(t) = \exp(-t - e^{-t}).$$

Dimostrazione