Lezione 04

Simulazione Stocastica

Si considerano qui alcuni metodi di simulazione di fenomeni stocastici.

Il metodo Monte Carlo in particolare è utilizzato per stimare l'incertezza di un *sistema complesso,* quando è nota la sua dinamica.

Il metodo Bootstrap invece è una tecnica di simulazione condizionata che prevede l'uso di osservazioni reale come base delle simulazioni.

Entrambi trovano utilizzo, fra l'altro, nell'analisi di sensitività.

Metodo Monte Carlo

Origine del metodo: calcolo di un integrale

Devo calcolare

$$A = \int_0^1 h(u)du$$

dove h ha una forma complessa per cui non conosco la sua primitiva ma h è computabile per ogni u.

Allora posso usare un campione simulato di m variabili casuali $R(0,1)$:

$$U_1, \ldots, U_m \;\; iid \;\; R(0,1)$$

e stimare l'integrale A tramite

$$\bar{h}_m = \frac{1}{m} \sum_{j=1}^{m} h(U_j)$$

Dalla teoria dell'inferenza statistica abbiamo che lo stimatore è corretto

$$E(\bar{h}_m) = E(h) = A$$

e consistente

$$\bar{h}_m \to A \text{ per } m \to \infty$$

con errore

$$\sigma(\bar{h}) = \frac{\sigma_h}{\sqrt{m}}$$

dove

$$\sigma_h^2 = Var(h(U)).$$

E' il metodo "migliore" ? vedi anche MC modificato.

MC in generale

In numerosi problemi si dispone di un vettore stocastico, y con distribuzione $F(y)$, in cui la funzione di ripartizione F è nota, e di un funzione ”complessa”

$$z = h(y)$$

nota solo per via numerica o computabile ma non (facilmente) conoscibile per via analitica (es: computer models).

Obiettivo

Interessa la distribuzione di z, diciamo $G(z)$.

- In linea di principio nel calcolo delle probabilità questo problema è risolto con formule del tipo

$$G(t) = P(h(y) \leq t) = P(y \leq h^{-1}(t)) = F(h^{-1}(t)),$$

- se la funzione $h(\quad)$ è complessa questo approccio diventa impercorribile.
- L'idea è allora quella di

$$\boxed{\textit{simulare} \text{ un grande numero } m \text{ di volte}}$$

diciamo $m = 1000$, il vettore dei dati y_i e ripetere il calcolo della corrispondente $z_i = h(y_i),\ i = 1,\ldots,m.$ Si può, poi, studiare la distribuzione (media, variabilità etc.) di questo *grande campione* $z_1,\ldots,z_m$.

Si può quindi *stimare*

- la funzione di ripartizione G con la distribuzione empirica delle z

$$\hat{G}(t) = \frac{1}{m}\sum_{i=1}^{m} I(z_i \leq t)$$

con $I(\quad)$ funzione indicatrice

- la funzione di densità g (istogramma, Kernell smoothing, etc.)
- La media

$$E(z) \cong \bar{z} = \frac{1}{m}\sum_{i=1}^{m} z_i$$

- La varianza

$$Var(z) \cong s_z^2 = \frac{1}{m-1}\sum_{i=1}^{m}(z_i - \bar{z})^2$$

e così via.

Metodo MC in Statistica

In particolare interessa ora il caso in cui $z(\quad)$ è uno stimatore

$$\hat{\theta} = \hat{\theta}(y) = h(y)$$

di cui non si conosce la distribuzione. Consideriamo il caso di

$$y \text{ con distribuzione } F(y;\xi)$$

con $\xi = \xi^\circ$ prefissato e diciamo

$$y_j^* = (y_{j1}^*, \ldots, y_{jN}^*)$$

il campione $N - dim$ **simulato** tramite un generatore di numeri casuali indipendenti con distribuzione $F(y;\xi)$.

Indichiamo con

$$\hat{\theta}_j^* = \hat{\theta}(y_j^*)$$

la stima basata su y_j^*.

NB: θ e ξ possono essere lo stesso parametro oppure possono essere diversi. ξ identifica F, θ invece è solo oggetto di interesse.

Ripetendo le simulazioni per $j = 1, \ldots, m$ ed ordinando i risultati in senso crescente si ottengono le stime:

$$\hat{\theta}_1^* \leq \ldots \leq \hat{\theta}_m^*.$$

Da queste si può approssimare l'incognita distribuzione di $\hat{\theta}$ usando l'istogramma delle stime Monte Carlo $\hat{\theta}^*$ (o tecniche più sofisticate tipo Kernel-smoothing) e si può approssimare media e varianza di $\hat{\theta}$ con

$$E_\xi\left(\hat{\theta}\right) \cong \bar{\theta}^* = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j^*$$

e

$$Var_\xi\left(\hat{\theta}\right) \cong s_*^2 = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\theta}_j^* - \bar{\theta}^*\right)^2.$$

Metodo MC in Risk Analysis

Esempio

Si intende stimare il "rischio"

$$\theta = P(y_{\max} > 15)$$

dove

$$y_{\max} = \max(y_1, \ldots, y_{25})$$

e

$$y_i = \alpha y_{i-1} + \varepsilon_i$$

$$\varepsilon_i \equiv iid \ \frac{\sigma}{\sqrt{2}} t_4$$

NB: $\xi = (\alpha, \sigma)$ e $var(t_n) = \frac{n}{n-2} \Rightarrow var(t_4) = 2.$

Soluzione con Matlab

```
nstar=25;       alfa = 0.8;      sigma=1; m=10000;
innovaz=trnd(4,nstar,m)*sigma/2;
y=filter(1,[1, -alfa],innovaz);
ymax=max(y)';
hist(ymax,30)
sum(ymax>15)/m
```

Esercizi su MC con Matlab

1. Analisi Montecarlo del Teorema Limite Centrale sulla media
   a. Media e Varianza finita
   b. Media e/o Varianza infinita.

   A tal fine considerare campioni dalla distribuzione t di Student con gradi di libertà $g = 1, 2, 5, 10$.

   Per ciascuna distribuzione valutare il comportamento simulato della media campionaria in termini di:
   a. Valore atteso
   b. Varianza

   Per ampiezze campionarie $N = 10, 100, 1000, 10000$ ed un opportuno numero di replicazioni Monte Carlo m.

   Giustificare in particolare il comportamento per $g = 1$ e 2.

2. Distribuzione MC del coeff. di correlazione r_n per piccoli campioni gaussiani.

A tal fine considerare campioni dalla distribuzione normale bivariaiata standardizzata $N_2((0,0),(1,1),\rho)$ con coefficienti di correlazione $\rho = 0, 0.5, \pm 0.95$.

Per ciascuna distribuzione valutare il comportamento simulato della media campionaria in termini di:

a. Valore atteso

b. Varianza

c. Quando $\rho = 0$ valutare l'approssimazione $Var(r_n) = \frac{1}{n}$

Per ampiezze campionarie $N = 5, 10, 100$ ed un opportuno numero di replicazioni Monte Carlo m.

3. Metodo MC per misture
Calcolare in via approssimata la varianza delle stime di $\theta = (\mu, \sigma^2, \pi)$ usando i dati di Daphne o altra immagine e $k = 4$.
A tal fine

a. si usino come valori di θ i valori stimati su Daphne (Bootstrap parametrico) e

b. si usino $m = 100$ replicazioni casualizzate dei valori iniziali ottenuti da Gaussiane con deviazione standard pari al 3% del valore iniziale stesso.

4. Metodo MC nella sorveglianza (La sorveglianza non è stata fatta nel 2014)
Con riferimento alla lezione 08 "Carte di controllo e sorveglianza" si considerino le carte di controllo con bande a 3σ

- **a.** **i.** di Shewhart
  - **ii.** ed EWMA con $\lambda = 0.01$
- **b.** Dopo aver calcolato il Tempo Medio fra Falsi Allarmi
- **c.** confrontare il Tempo Medio di Ritardo

Si ottengano i risultati a. e b. considerando i seguenti tipi di dati con $N = 3000$ ed un opportuno numero di repicazioni m.

- **a.** $iid: \quad y_t \equiv N(0,1)$
- **b.** $AR(1): \quad y_i = 0.8 y_{i-1} + \varepsilon_i$ e $\varepsilon_i \equiv iid \ N(0,1)$

Il Metodo Bootstrap

Spesso la distribuzione di y non è nota o prefissare $\theta = \theta^{\circ}$ può essere limitativo. In questi casi una strada percorribile per capire la variabilità delle stime consiste nel costruire delle simulazioni basate sulla distribuzione empirica **del particolare vettore osservato**

$$y = (y_1, \ldots, y_n)$$

qui supposto per semplicità iid.

Il Bootstrap Parametrico

Quando la distribuzione delle componenti di y, $F(y_j|\xi)$ è nota nella forma ma ξ è ignoto, si può stimare $\hat{\xi}$ sui dati osservati e applicare poi il metodo Monte Carlo a $\xi = \hat{\xi}$.

Il Bootstrap Nonparametrico

Se F non è nota, l'idea è allora quella di stimarla tramite la funzione di ripartizione empirica del campione:

$$\hat{F}_n(x) = \frac{\sum I(y_t \leq x)}{n}$$

e applicare poi il metodo Monte Carlo a tale distribuzione.

In pratica ciascun elemento del campione simulato y_j^*, diciamo

$$y_{jt}^* \qquad t = 1, \ldots, N$$

con $N \lesseqgtr n$, viene estratto a caso (con rimessa) dal campione originario e si ottiene generando un numero casuale equidistribuito sugli interi da 1 ad n. Indicato con r_t tale numero casuale, si pone

$$y_{jt}^* = y_{r_t} \qquad t = 1, \ldots N.$$

Si ha così una stima $\hat{\theta}_{N,j}^*$. Iterando m volte come nel metodo MC si arriva ad una valutazione della distribuzione e delle proprietà dello stimatore $\hat{\theta}_N$.

Proprietà asintotiche

Ci interessa la distribuzione bootstrap di $\hat{\theta}_N$ al crescere di n ed $m,$ per fissato N.

Tale distribuzione è data da

$$\hat{G}_{n,m}^{N}(t) \cong G^{N}(t) = P\left(\hat{\theta}_N \leq t\right)$$

che calcoliamo come

$$\hat{G}_{n,m}^{N}(t) = \frac{1}{m}\sum_{j=1}^{m} I\left(\hat{\theta}_{N,j}^{*} \leq t\right)$$

Poiché le simulazioni sono iid, per la legge forte dei grandi numeri abbiamo che

$$\lim_{m\to\infty} \hat{G}^N_{n,m}(t) = G^N_n(t) = P\left(\hat{\theta}^*_{N,j} \leq t | y \equiv \hat{F}_n\right)$$

Inoltre, applicando ancora la legge forte dei grandi numeri ma al campione di partenza, abbiamo che

$$\lim_{n\to\infty} \hat{F}_n(x) = F(x)$$

Perciò si conclude che

$$\lim_{n,m\to\infty} \hat{G}_{n,m}(t) = G^N(t).$$

Esempio con Matlab

```
n=5; m=3;
% dati osservati
y=[3.75 4 -1 7 0.33]';
N=7;
% matrice N × m delle simulazioni:
ystar=y(unidrnd(n,N,m))
```

Uso del Bootstrap

Un semplice intervallo di confidenza bootstrap a livello $1-\alpha$ approssimato per θ si può basare sui percentili di $\hat{\theta}_j^*$ ed avrà la forma:

$$\hat{\theta}^*_{\left[\frac{\alpha}{2}m\right]} \div \hat{\theta}^*_{\left[\left(1-\frac{\alpha}{2}\right)m\right]}$$

dove $[x]$ è la parte intera di x.

Un miglioramento dell'approssimazione si ottiene studentizzando le stime bootstrap. Se $s^2\left(\hat{\theta}\right)$ è una stima della varianza di $\hat{\theta}$, si costruiscono le replicazioni bootstrap studentizzate

$$t_j^* = \frac{\hat{\theta}_j^* - \hat{\theta}}{s\left(\hat{\theta}_j^*\right)}$$

che vengono poi ordinate

$$t_1^* \leq \ldots \leq t_m^*$$

e l'intervallo di confidenza bootstrap studentizzato è dato da

$$\hat{\theta} + t^*_{\left[\frac{\alpha}{2}m\right]} s\left(\hat{\theta}\right) \div \hat{\theta} + t^*_{\left[\left(1-\frac{\alpha}{2}\right)m\right]} s\left(\hat{\theta}\right).$$

Bootstrap per serie storiche

Premessa

Consideriamo l'esempio della simulazione di un $AR(1)$.

Se ha parametri $\xi = (\alpha, \sigma^2)$ noti allora simuliamo y_t tramite la ricorsione:

$$y_t = \alpha y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \ iid \ N(0, \sigma^2)$$

Bootstrap parametrico

Si stima $\hat{\xi} = (\hat{\alpha}, \hat{\sigma}^2)$ noti allora si simula $y_j^* = (y_{j1}^*, \ldots, y_{jN}^*)$ tramite:

$$y_t^* = \hat{\alpha} y_{t-1}^* + \varepsilon_t^*$$

$$\varepsilon_t^* \ iid \ N(0, \hat{\sigma}^2)$$

Bootstrap semi-parametrico lineare

Si rinuncia alla normalità delle innovazioni, perciò stimato ξ e i residui $e = (e_1, \ldots, e_n)$, si simula $y_j^* = (y_{j1}^*, \ldots, y_{jN}^*)$ tramite

$$\begin{aligned} y_t^* &= \hat{\alpha} y_{t-1}^* + e_t^* \\ e_t^* &= \text{bootstrap } \mathit{IID} \text{ da } (e_1, \ldots, e_n) \end{aligned}$$

Bootstrap semi-parametrico nonlineare

Si rinuncia all'ipotesi iid per gli errori, ammettendo che abbiano una struttura di dipendenza nonlineare di carattere generale all'interno della stazionarietà.

Perciò stimato ξ e i residui $e = (e_1, \ldots, e_n)$, si simula $y_j^* = (y_{j1}^*, \ldots, y_{jN}^*)$ tramite

$$
\begin{aligned}
y_t^* &= \hat{\alpha} y_{t-1}^* + e_t^* \\
e_t^* &= \text{bootstrap a blocchi da } (e_1, \ldots, e_n)
\end{aligned}
$$

dove il bootstrap non-parametrico a blocchi è descritto nel lucido successivo.

Bootstrap non-parametrico a blocchi

Nel Block-stationary-bootstrap l'idea è quella di costruire il campione simulato $y_j^* = (y_{j1}^*, \ldots, y_{jN}^*)$ come unione di un certo numero k_j di blocchi di lunghezza variabile estratti dai dati osservati $y = (y_1, \ldots, y_n)$, con $n \lesseqgtr N$. In pratica si definisce

$$y_j^* = \left(\vec{y}_{j1}^*, \ldots, \vec{y}_{jk_j}^*\right)$$

dve ciascun blocco

$$\vec{y}_{ji}^* = (y_{t_i}, \ldots, y_{t_i+l_i})$$

è scelto con inizio casuale

$$t_i = Uniforme(1, n)$$

e lunghezza casuale l_i,

NB: occorre verificare che

$$t_i + l_i \leq n$$

Inoltre k_j è tale che

$$\sum_{i}^{k_j - 1} (l_i + 1) = a_j < N$$

infine l'ultimo blocco è troncato: $l_{k_j} = N - a_j$ in modo da garantire la dimensione corretta a y_j^*.

Esempi:

La lunghezza dei blocchi può essere:

1. $l_i = Poisson(L)$
2. $l_i = geom(1/L)$
3. $l_i = Uniforme(1, 2L)$ => crea anomalie sul correlogramma

Matlab:

la function **`stationaryBB()`** esegue il bootstrap a blocchi.

Nell'esempio **`ese_2018_block_bootstrap.m`** si ricampiona un modello

$$y^*_{j,t} = \hat{m}_t + e^*_{j,t}$$

dove m_t è modellato tramite una spline mentre e^*_t è ottenuto col bootstrap a blocchi applicato ai residui della spline:

$$e_t = y_t - \hat{m}_t.$$

Metodi Monte Carlo modificati

Consideriamo dapprima il campionamento di una vc unidimensionale Y con pdf $f(\quad)$ e cdf $F(\quad)$.

Indichiamo con $h(x)$ la risposta del modello di simulazione sottoposto ad input casuali Y, e τ, il suo valore atteso, è la grandezza da stimare mediante simulazione. Cioè

$$\tau = E(h(Y)) = \int_{-\infty}^{+\infty} h(y)f(y)dy$$

e

$$T = \frac{1}{n}\sum h(y_i)$$

Premessa

Ricordiamo che se $R_i \equiv R(0,1)$ e $Y \equiv F$, allora

$$Y_i = F^{-1}(R_i) \equiv Y$$

MC Stratificato

Interessa avere un campione $y_1, \ldots, y_n$ da F in modo che siano rappresentate *tutte le zone* della distribuzione F. In pratica dividiamo l'intervallo $(0, 1)$ in n intervalli di pari ampiezza, diciamo $c_i = (\frac{i-1}{n}, \frac{i}{n})$ e

$$C_i = \{Y : F(Y) \in c_i\}$$

è la partizione di Ω con elementi equiprobabili

$$P(C_i) = \frac{1}{n}.$$

Estraiamo a caso un valore da ciascuno di essi usando la distribuzione condizionata $f(y|C_i)$.

Questo equivale a scrivere

$$y_i = F^{-1}\left(\frac{i - 1 + R_i}{n}\right)$$

con $R_i \equiv R(0, 1)$ e $\quad F(x_i) \in (\frac{i-1}{n}, \frac{i}{n})$.

Proprietà di SMC

Indichiamo con $T_S = \frac{1}{n}\sum h(y_i)$ la stima di τ basata sul campionamento stratificato e con T_{MC} la corrispondente stima basato sul metodo Monte Carlo standard.

Nondistorsione

$$ET_S = \tau$$

Dimostrazione

Notiamo innanzitutto che visto il tipo di campionamento, la x_i simulata ha distribuzione $F(y|C_i)$ e

$$E(h(y_i)) = E(h(Y)|C_i)$$

Usando la media a due stadi, segue che

$$ET_S = \frac{1}{n}\sum_{i=1}^{n} E(h(y_i)) = E_C\{E(h(Y)|C)\} = E(h) = \tau.$$

Ottimalità

Premessa: ricordiamo che, data una vc casuale doppia (h, C) vale la seg. scomposizione della varianza:

$$Var(h) = E_C(Var(h|C)) + Var_C(E(h|C))$$

Traccia dimostrazione

Ricordando che $Var(h) = E(h - \tau)^2$, aggiungendo $\pm E(h|C)$ ed applicando la media a due stadi, si deduce la scomposizione della varianza.

Ottimalità

Andiamo ora a verificare che il campionamento proporzionale fornisce stime migliori del campionamento casuale semplice nel senso che

$$Var(T_S) \leq Var(T_{MC})$$

Dimostrazione

Ricordiamo innanzitutto che, poichè nel MC le x_i sono iid, si ha:

$$Var(T_{MC}) = \frac{1}{n} Var(h)$$

Poichè le x_i di SMC sono indipendenti, abbiamo che

$$Var(T_S) = \frac{1}{n^2} \sum Var(h(Y)|C_i) = \frac{1}{n} E_C Var(h|C).$$

Usando la scomposizione della varianza di cui sopra abbiamo quindi il risultato.

Esercizio: campionamento proporzionale

Dalla dimostrazione della pagina precedente abbiamo che il guadagno del campionamento stratificato è dato da:

$$Var(T_{MC}) - Var(T_S) = Var_C(E(h|C))$$

Commentare questo risultato

Esercizio: campionamento sistematico

Nell'approccio visto, diviso in modo sitematico il dominio in intervalli equiprobabili, si prende un valore a caso da ciascun intervallo:

$$y_i = F^{-1}\left(\frac{i-1+R_i}{n}\right), \; R_i \equiv R(0,1).$$

Si consideri ora la strategia per cui invece di y_i si usa la media della classe C_i

$$\hat{y}_i = E(Y|C_i)$$

e poi si stima T di conseguenza:

$$\hat{T} = \frac{1}{n}\sum h(\hat{y}_i)$$

Dire, in particolare, cosa succede alle proprietà di nondistorsione ed efficienza.

Esercizio: campionamento sistematico (segue)

Dire, inoltre, cosa succede alle proprietà di nondistorsione ed efficienza se invece di $\hat{y}_i$ si prende il valore centrale di C_i :

$$\check{y}_i = \frac{F^{-1}(\frac{i-1}{n}) + F^{-1}(\frac{i}{n})}{2}$$

e poi, come sopra,

$$\check{T} = \frac{1}{n} \sum h(\check{y}_i).$$

Importance MC

L'idea del campionamento per importanza è di usare un campione stratificato da una cdf $Q(y) \neq F(y)$ che assegna probabilità più alta a quei valori della y dove $|h|$ è più grande.

Perciò l'i^o campione è dato da

$$y_i' = Q^{-1}\left(\frac{i-1+R_i}{n}\right)$$

dove $R_i \equiv R(0,1)$ come sopra.

La stima di τ è ora data dalla media ponderata

$$T_I = \sum h(y_i)\frac{f(y_i)}{q(y_i)}.$$

E’ facile vedere che se

$$q(y) = n\frac{h(y)f(y)}{\tau}$$

allora, identicamente,

$$T_I = \tau.$$

Perciò, se si hanno idee a priori sul modello di simulazione e su $h(y)$ la stima di τ può essere migliorata usando un opportuno campionamento.

Tuttavia si nota come in quest’approccio sia importante la ponderazione per evitare la distorsione che si avrebbe ad usare la media aritmetica semplice $\bar{h}$.

Ipercubi Latini

Questo metodo, acronimizzato *LHS*, è una generalizzazione multivariata dello *SMC*.

Supponiamo ora che, come accade in pratica, y sia k –dimensionale

$$y = (y_1, \ldots, y_k)$$

con distribuzione mutlipla

$$F = \Pi_{j=1}^{k} F_j$$

e sia

$$y_{1j}^0, \ldots, y_{nj}^0$$

il campione *SMC* di n termini dalla marginale F_j.

Interessa stimare

$$\tau = E(h(Y)) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} h(y) f(y_1, \dots, y_k) dx_1 \dots dx_k$$

tramite la statistica

$$T = \frac{1}{n} \sum_{i=1}^{n} h(y_{i1}, \dots, y_{ik})$$

L'idea è quella di costruire un campione ad n elementi di $y \in R^k$ che rispetti le condizioni di SMC e di indipendenza delle marginali combinando le diverse componenti in modo casuale.

Per esempio se $k = 2$ in figura si riporta una scelta delle combinazioni degli intervalli equiprobabili per y_1 ed y_2 in corrispondenza ad $n = 5$.

		X		
	X			
			X	
				X
X				

Per garantire l'indipendenza tali abbinamenti sono effettuati casualmente.

A tal fine sia

$$p = (p_1, \ldots, p_n)$$

una permutazione casuale degli interi $1, \ldots, n$.

NB:

l'insieme delle possibili permutazioni p ha $n!$ elementi.

Campione *LHS*

Definiamo campione *LHS* quel campione con termine generico $y_{i,j}$ dato da

$$y_{i,1} = y^0_{i,1} \quad i = 1, \ldots, n$$

$$y_{ij} = y^0_{p(j)_i, j} \quad i = 1, \ldots, n, j = 2, .., k$$

Correlazione

Interessa generare dei campioni con marginali SMC da una F normale multivariata con una matrice di covarianza Σ. A tal fine si usa la trasformazione in ranghi.

Ranghi

Si dice rango dello scalare y_i fra $y = (y_1, \ldots, y_n)$ la sua posizione nella sequenza ordinata di $y_1, \ldots, y_n$. Indicheremo tale rango con

$$r(y_i) = r(y_i | y).$$

Tornando al problema di LHS per dati correlati, si generano gli n campioni normali $k-$ variati

$$y_i^0 \equiv NID_k(\mu, \Sigma)$$

che mettiamo nella matrice $n \times k$

$$Y^0 = (y_{i,j}^0)$$

e si generano poi i ranghi delle sue colonne Y_j^0

$$r_{ij} = r(y_{i,j}^0 | Y_j^0).$$

Si retrotrasforma poi, ancora con la gaussiana

$$y_{i,j} = \Phi^{-1}\left(\frac{r_{i,j} - R}{n}\right)\sqrt{\sigma_{jj}} + \mu_j.$$

Note

La matrice Y così ottenuta ha marginali SMC e correlazione $\cong \Sigma$.

In particolare il coefficiente di correlazione dei ranghi di Spearman è invariato:

$$corr(r(Y_i), r(Y_j)) = corr(r(Y_i^0), r(Y_j^0))$$

Matlab

Vedi randperm(), lhsdesign(), lhsnorm(), mvnrnd(), ranghi() ed lhsnorm.rank().

Ottimalità *LHS*

Se

$$h(y_1, \ldots, y_k)$$

è monotona in almeno $k-1$ componenti allora

$$Var(T_{LHS}) \leq Var(T_{MC})$$

Note:

- Se non vale la condizione di monotonicità il guadagno può essere nullo.
- Se si usa *LHS* correlato la nodistorsione è solo asintotica

Esercizi con Matlab

1. Confrontare MC ed LHS per la stima di $\mu = E(Y)$.
2. Confrontare MC ed LHS per la stima di $\rho(X, Y)$ e $\mu_{11} = E(XY)$ nella normale bivariata.

Bibliografia

Approfondimenti metodologici:

1. **Monte Carlo e Bootstrap**
2. Davison A.C., Hinkley D.V. (1997) *Bootstrap methods and their application*. Cambridge University Press.
3. Hjorth U. (1994) *Computer Intensive Statistical Methods: Validation, Model Selection and Bootstrap.* Chapman & Hall
4. Ripley B.D. (1987) *Stochastic Simulation*. Wiley.