Testo di riferimento:

**Steven Lavenberg** "Computer Performance Modeling Handbook" Academic Press,

Stime puntuali e Intervalli di confidenza

- Output della simulazione è costituito da sequenze di variabili casuali
  - stimare media, varianza, mediana, etc.
  - la stima consisterà in una v.c.
  - una quantità deterministica verrà stimata mediante una v.c.

Stima di caratteristiche transienti (terminating simulation)

Eseguiamo M osservazioni ( M run di simulazione) della v.c di interesse X:

Campione di dati i.i.d

$X_1, X_2, X_3, \ldots X_M$

F(x) = Prob(X<x) : distribuzione di probabilità

Media: $\mu = E[X]$;

Varianza: $\sigma^2 = E[(X- \mu)^2] = Var[X]$

Utilizziamo il campione per stimare $\mu$.

$$\hat{\mu} = \overline{X} = \frac{1}{M}\sum_{m=1}^{M} X_m$$

$\hat{\mu}$ è una variabile casuale; $E[\hat{\mu}] = \mu$.

La distribuzione di $\hat{\mu}$ deve essere centrata intorno a $\mu$

L'accuratezza della stima dipende dalla differenza

$$\hat{\mu} - \mu$$

necessità di ulteriori informazioni sulla stima.

$\sigma^2 = Var[X]$

$Var[\hat{\mu}] = \sigma^2/M$

Utilizziamo il campione per stimare anche $\sigma^2$

$$s^2 = \hat{\sigma}^2 = \frac{1}{M-1}\sum (X_m - \overline{X})^2$$

$$\hat{\mu} = \overline{X} = \frac{1}{M}\sum_{m=1}^{M} X_m$$

La quantità $s^2/M$ è una stima della varianza di $\hat{\mu}$

Inoltre la variabile casuale:

$$\frac{\hat{\mu} - \mu}{s / M^{1/2}}$$

ha una distribuzione di probabilità **t-student** con M-1 gradi di libertà.

Fissato un livello di confidenza $\alpha$ e in base ai valori tabulati della **t-student** ($t_{M-1}(\alpha/2) = t_{M-1}(1-\alpha/2)$)

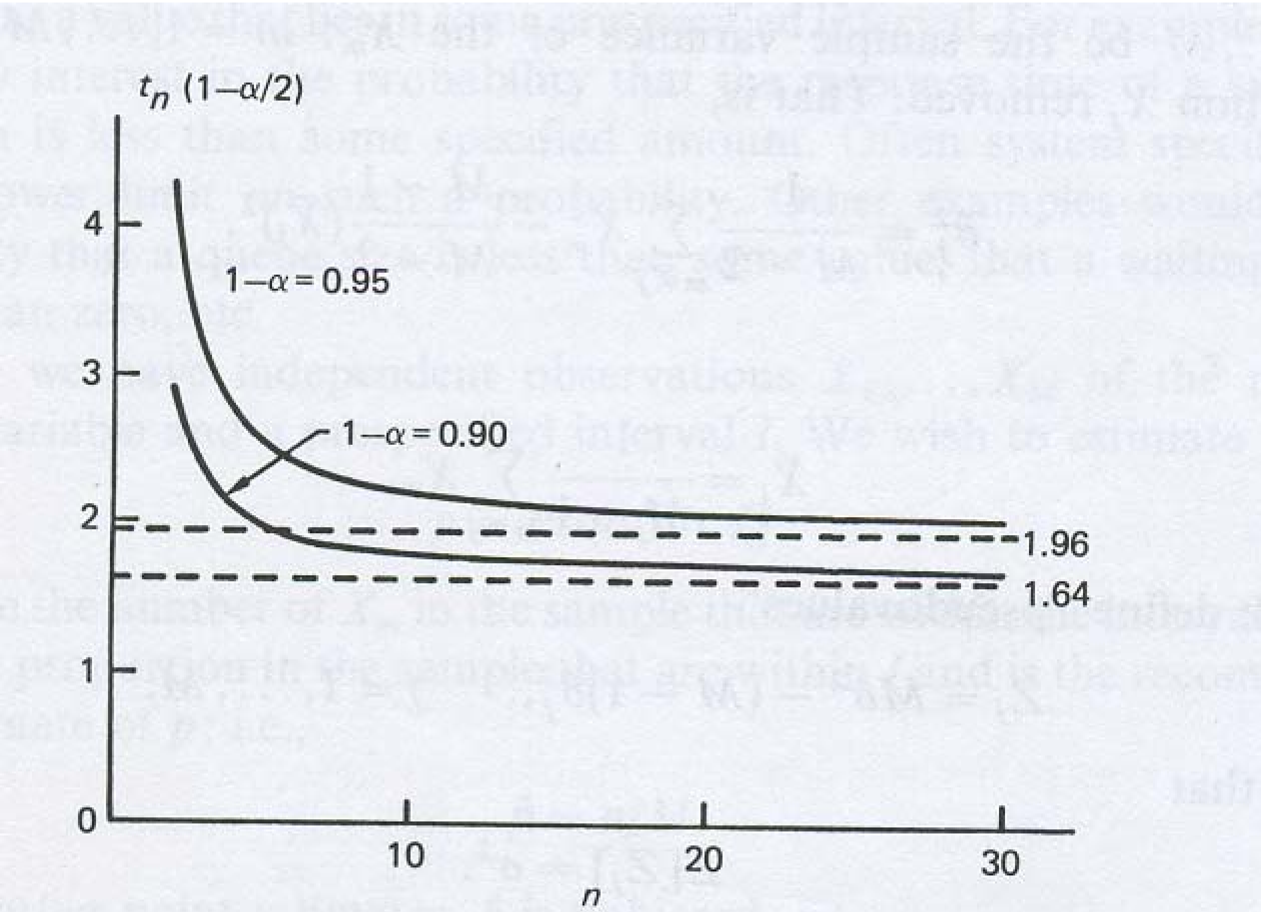
siamo in grado di calcolare intervalli di confidenza:

$$\hat{\mu} - t_{M-1}(1-\alpha/2)s/M^{1/2} \le \mu \le \hat{\mu} + t_{M-1}(1-\alpha/2)s/M^{1/2}$$

Con la proprietà che:

$$\text{Prob}\{\hat{\mu} - t_{M-1}(1-\alpha/2)s/M^{1/2} \leqslant \mu \leqslant \hat{\mu} + t_{M-1}(1-\alpha/2)s/M^{1/2}\} \approx 1-\alpha.$$

Per valori di (1-α) pari a 0.95 e 0.90 i valori di $t_{n-1}(1-\alpha/2)$ sono:




Per valori di n > 30 i valori dei moltiplicatori diventano: 1.64 [(1-α) = 0.90] e 1.96 [(1-α) = 0.95]

$$\hat{\mu} - 1.64(s / M^{1/2}) \leq \mu \leq \hat{\mu} + 1.64(s / M^{1/2})$$

$$\hat{\mu} - 1.96(s / M^{1/2}) \leq \mu \leq \hat{\mu} + 1.96(s / M^{1/2})$$

## Stima di caratteristiche stazionarie

- la sequenza di output converge

$\{V_n : n = 1,2,\dots\}$

$$\lim_{n \to \infty} \Pr ob\langle V_n \leq x \rangle = F(x)$$

Due tipi di problemi:

a) determinare la fase transiente;

b) eliminare la correlazione.

Fase transiente

1) osservare $F_n(x)$ e stabilire (mediante istogrammi) quando la fase transiente è terminata; costo elevato).

2) Esaminiamo la convergenza della media della distribuzione:

$$\mu_n = E[V_n]$$

Supponiamo di ripetere la simulazione M volte (M run di simulazione) e otteniamo N osservazioni per ogni run

- Run (1): $V_{1,1}, V_{1,2}, V_{1,3}, V_{1,4}, V_{1,5}, \ldots.. V_{1,N}$
- Run (2): $V_{2,1}, V_{2,2}, V_{2,3}, V_{2,4}, V_{2,5}, \ldots.. V_{2,N}$
- Run (3): $V_{3,1}, V_{3,2}, V_{3,3}, V_{3,4}, V_{3,5}, \ldots.. V_{3,N}$
-
-
-
- Run (M): $V_{M,1}, V_{M,2}, V_{M,3}, V_{M,4}, V_{M,5},.., V_{M,N}$

$$\hat{\mu}_n = \overline{V_n} = \frac{1}{M}\sum_{m=1}^{M} V_{n,m} \qquad n = 1,2,\ldots N$$

Sequenza random che converge:

$E[\hat{\mu}_n] = \mu_n$

Esempio:

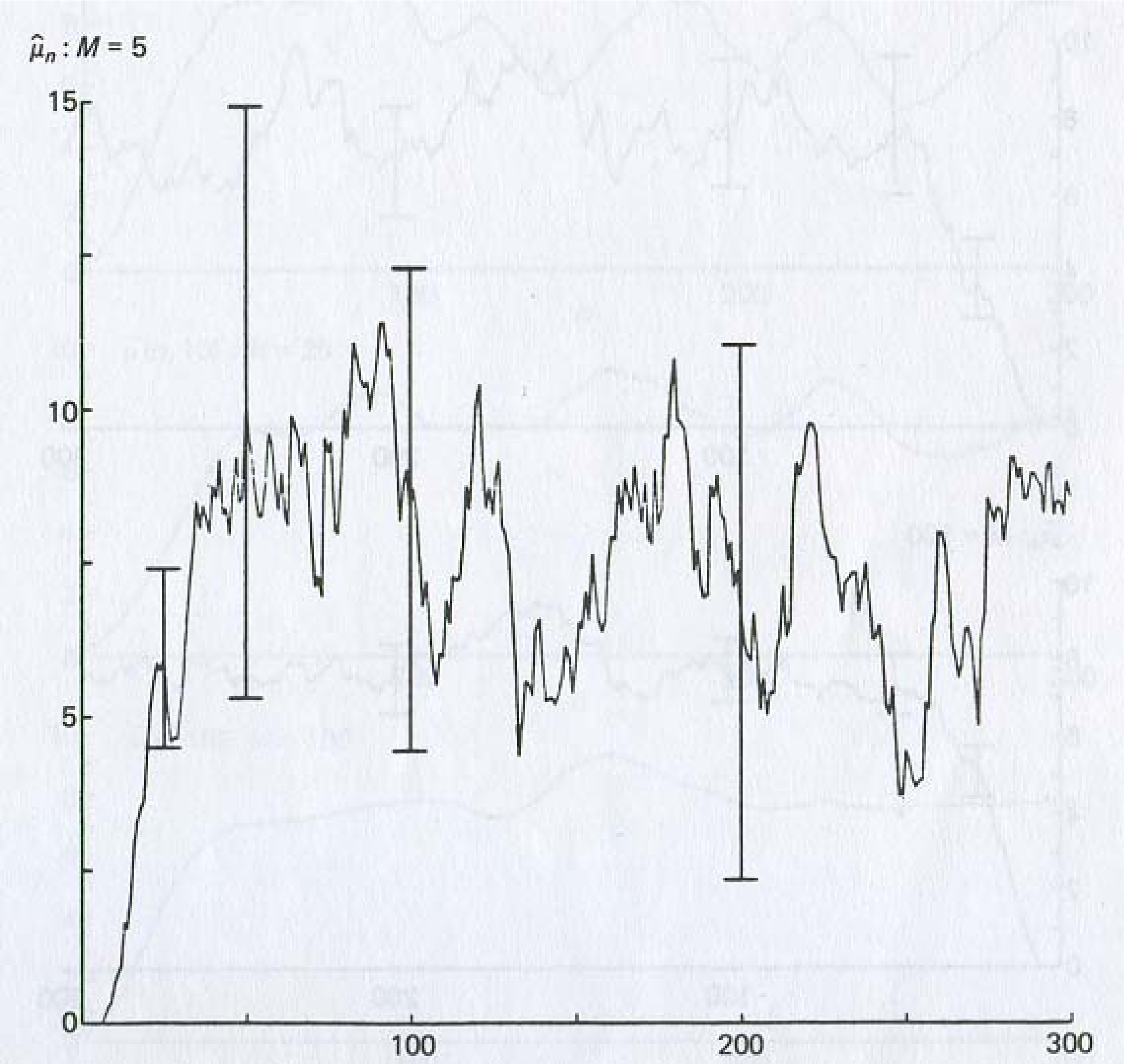
$\boldsymbol{\mu_2}$: Tempo medio di permanenza nella CODA 2

$\mathbf{W_{2,n}}$ : Tempo di attesa dell'n-ennesimo utente nella CODA 2.

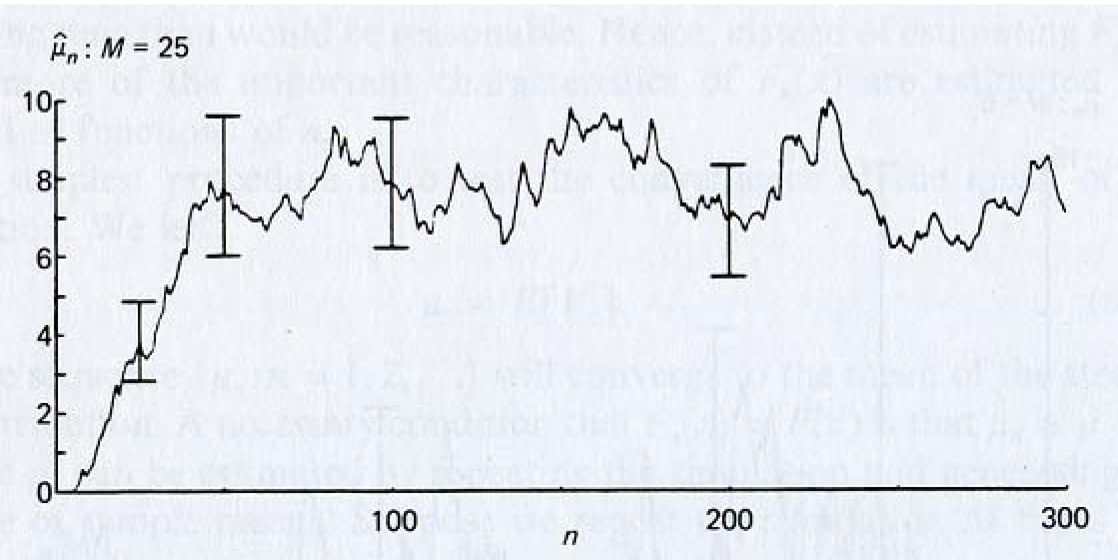
Consideriamo $V_{m,n}$ valore assunto da $W_{2,n}$ durante l'm-esimo run di simulazione;

m = 1,2…,M;
n = 1,2,…..N.

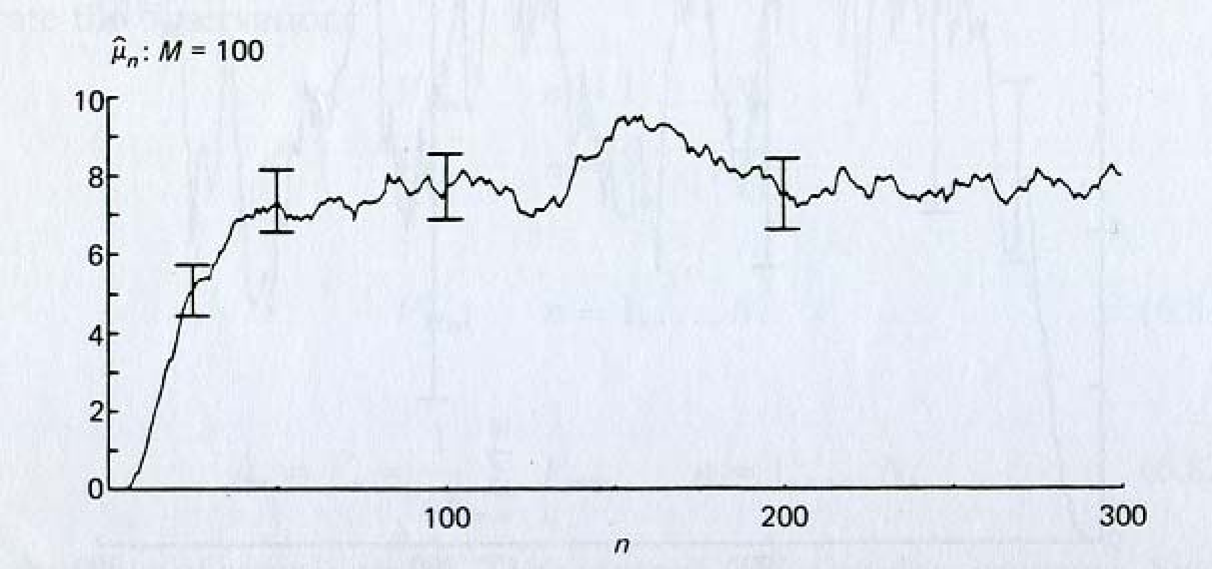
M = 5; N = 300.




M = 25; N = 300.

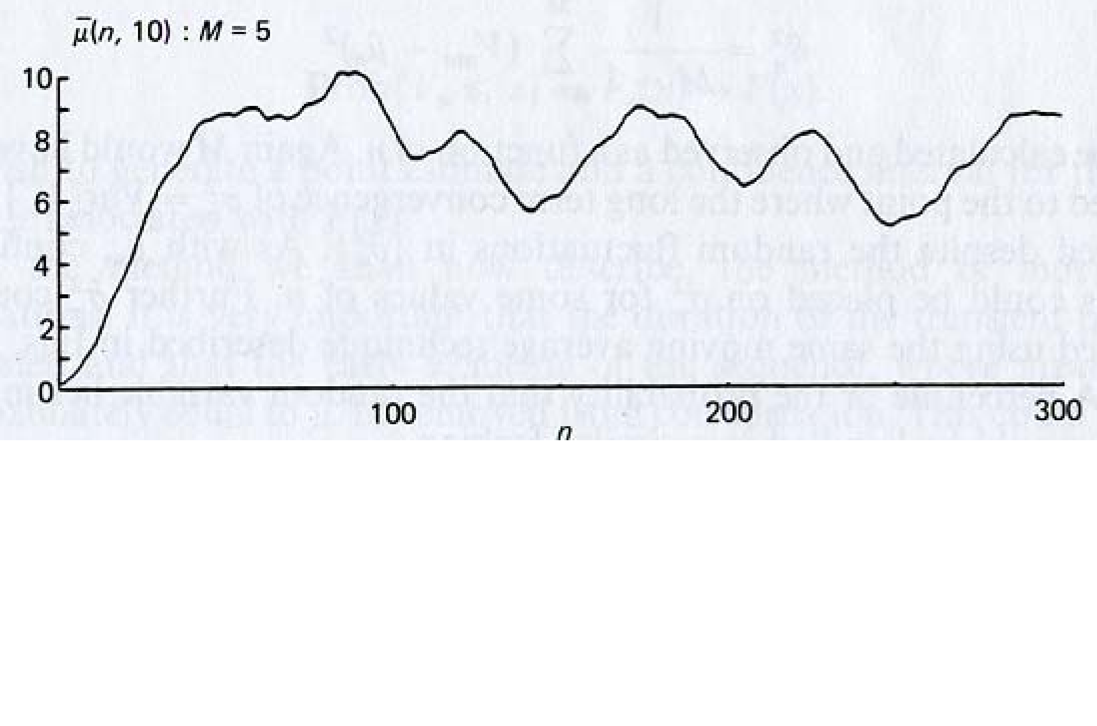


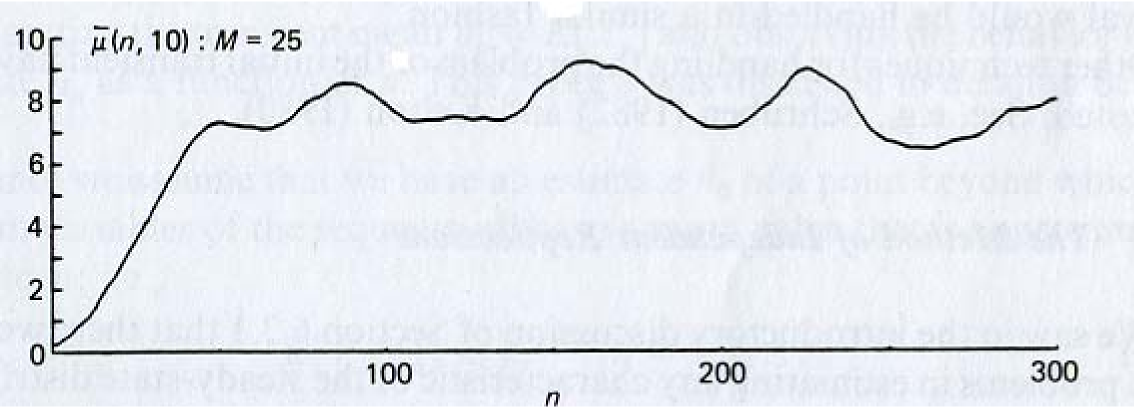
M = 100; N = 300.

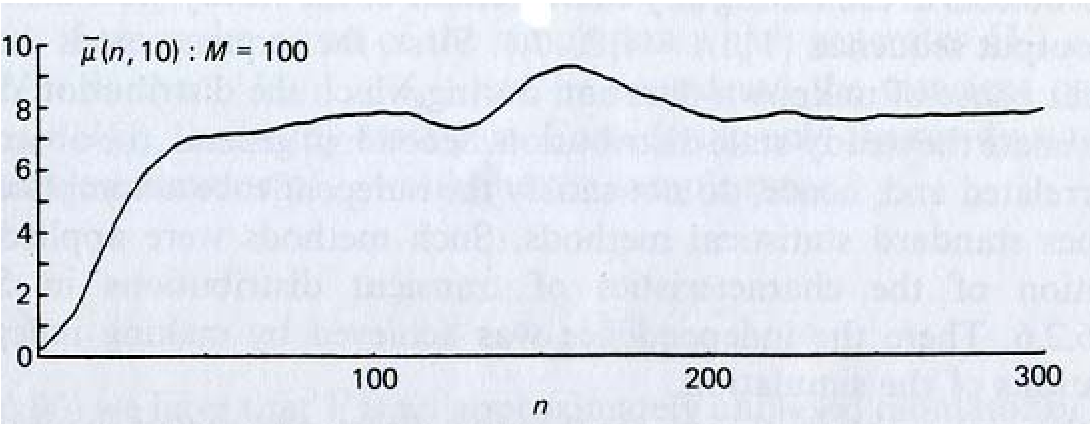



utilizziamo un metodo per ottenere grafici che attenuino il contributo delle fluttuazioni a breve termine:

$$\bar{\mu}(n; K) = \begin{cases} (2K+1)^{-1} \sum_{k=-K}^{K} \hat{\mu}_{n+k} & \text{if} \quad n \geqslant K+1, \\ (2n-1)^{-1} \sum_{k=-(n-1)}^{n-1} \hat{\mu}_{n+k}. & \text{if} \quad n < K+1. \end{cases}$$






3) stima delle caratteristiche steady state mediante il metodo delle prove ripetute.

   a) suddividere i dati raccolti in ogni run in fase transiente e fase stazionaria;

   b) per ogni singolo run si ottiene, utilizzando i dati relativi alla fase stazionaria, una stima puntuale della caratteristica di interesse;

   c) la stima puntuale finale si ottiene come media delle stime puntuali ottenute;

d) si calcolano gli intervalli di confidenza.

Stima della media di una v.c. stazionaria

$\{V_n : n = 1,2,\ldots\}$

$$\mathrm{Prob}\{V_n \leqslant x\} = F_n(x) \rightarrow F(x).$$

Si vuole stimare la media μ della associata ad F(x).

a) eliminazione del transiente:

è necessario calcolare il valore $\hat{n}_0$ tale che:

$$E[V_{\hat{n}0+n}] \approx \mu, \qquad n = 1, 2, \ldots .$$

I dati raccolti:

$\{V_1, \ldots, V_N\}$

vengono suddivisi in due insiemi:

Fase transiente $\{V_1, \ldots, V_{\hat{n}0}\}$

Fase stazionaria $\{V_{\hat{n}0+1}, \ldots, V_N\}$

b) stima puntuale della media (singolo run) indichiamo con $V_{m,n}$ l'n-esimo elemento del run m-esimo; m = 1,2,… M; n = 1,2..N

Stimiamo per ogni run la media:

$$\hat{\mu}_m = \frac{1}{N - \hat{n}_0} \sum_{n=\hat{n}_0+1}^{N} V_{mn}.$$

$$\hat{\mu}_m, \; m = 1, \ldots, M,$$

utilizzando questi ultimi valori siamo in grado di ottenere una stima puntuale:

$$\hat{\mu} = \frac{1}{M} \sum_{m=1}^{M} \hat{\mu}_m.$$

Stimiamo la varianza campionaria:

$$s^2(\hat{\mu}_m) = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\mu}_m - \hat{\mu})^2 = \frac{1}{M-1} \sum_{m=1}^{M} \hat{\mu}_m^2 - \frac{M}{M-1} \hat{\mu}^2$$

Sappiamo che la v.c.

$$\frac{\hat{\mu} - \mu}{s(\hat{\mu}_m)/M^{1/2}}$$

ha una distribuzione t-student ed è quindi possibile calcolare, fissato un livello di confidenza α, gli intervalli di confidenza:

$$\text{Prob}\{\hat{\mu} - t_{M-1}(1-\alpha/2)s(\hat{\mu}_m)/M^{1/2} \leqslant \mu \leqslant \hat{\mu} + t_{M-1}(1-\alpha/2)s(\hat{\mu}_m)/M^{1/2}\} \approx 1-\alpha$$

il metodo Batch è usato frequentemente

viene effettuato un unico (lungo) run di simulazione

assumiamo che la lunghezza di tale run sia $M$

assumiamo di osservare la variabile $X$:

$$X_1, X_2, X_3, \ldots, X_M$$

siamo interessati a stimare $\mu = E[X]$

eliminiamo il transiente iniziale (costituito da $K$ osservazioni)

i dati su cui effettuare statistiche risulta di lunghezza (M-K)

i dati vengono suddivisi in $N$ batch; in ciascun batch sono quindi presenti: $n = \frac{M-K}{N}$ osservazioni all'interno del batch *i* calcoliamo la media:

$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$

la stima della media:

$\hat{\mu} = \hat{\mu_N} = \frac{1}{N}\sum_{i=1}^{N} \bar{X}i = \frac{1}{nN}\sum_{i=1}^{N}\sum_{j=1}^{n} X_{i,j}$

$\hat{\mu}$ è la stima puntuale della media

suddivisione in batch influenza solo l'intervallo di confidenza assumendo batch abbastanza lunghi, le medie puntuali $\bar{X}_i$ possono essere considerate indipendenti

la varianza campionaria è: $S^2 = \frac{1}{(N-1)} \sum_i^N (\bar{X}_i - \hat{\mu_N})^2$

intervallo di confidenza ( con livello di confidenza $1 - \alpha$):
$\hat{\mu_N} \pm z_{1-\alpha/2} \frac{S}{\sqrt{N}}$

un solo periodo di warm-up

necessari 30-40 batch per una stima affidabile della varianza

i batch devono essere molto più lunghi della fase di warm-up (3-4 volte)

se vi è dipendenza, la correlazione è positiva

la dipendenza non degrada del tutto la stima fatta
la precisione della stima è inferiore