UNIVERTÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA TRIENNALE IN
STATISTICA E TECNOLOGIE INFORMATICHE

Stima del parametro "h" di una carta di controllo CUSUM mediante approssimazione stocastica

RELATORE: CH.MA PROF.SSA GIOVANNA CAPIZZI

LAUREANDO: ALBERTO SABBION

ANNO ACCADEMICO 2004-2005

Alla mia famiglia, a mia zia, ai miei amici, ai compagni d'Università, ai professori, a tutti coloro che hanno contribuito al raggiungimento di questo importante traguardo.

Indice

Capitolo 1

Concetti generali

1.1 La variabilità nei processi produttivi

Ogni processo produttivo è caratterizzato da una certa variabilità. Questa variabilità può essere vista come somma di due altre variabilità distinte. La prima, chiamata variabilità naturale, è la variabilità intrinseca al processo, che non può essere eliminata in quanto risultato di fattori casuali sui quali è impossibile intervenire e che nel loro complesso determinano un comportamento detto "tipico" del processo stesso. Si dice quindi che un processo che presenta solo questa variabilità è *sotto controllo* (*in control*, IC).

La seconda è l'oggetto di studio del controllo della qualità ed è da attribuirsi a fonti di variabilità "esterne" che incidono sul livello qualitativo del prodotto in esame; i fattori a cui principalmente si riconducono sono: guasti o starature dei macchinari, errori degli operatori in fase di controllo o setup, alla difettosità del materiali utilizzati,ecc.

Quest'ultima variabilità, introdotta quindi da fattori esterni, è di norma molto più elevata in termini quantitativi della variabilità "naturale" e può portare facilmente la linea produttiva a realizzare partite difettose.

Si dice quindi che un processo avente variabilità riconducibile a fattori specifici, è *fuori controllo* (*out of control*, OC).
L'obiettivo primario del controllo statistico della qualità è di controllare costantemente la linea di produzione attraverso la misurazione delle caratteristiche d'interesse del prodotto e attraverso alcune statistiche ad esse associate individuare quanto prima eventuali anomalie. Lo strumento più utilizzato ed affidabile, per stabilire se un processo di produzione sia sotto controllo o fuori controllo secondo una cerca caratteristica del prodotto è la *carta di controllo*.

1.2 Le carte di controllo

La stesura di una carta di controllo consiste nella costruzione di un grafico che riporta i valori della statistica di controllo, funzione delle misurazioni d'interesse, affiancata a:

- una Linea Centrale (LC) che rappresenta il valore nominale della statistica di controllo quando il processo in esame è sotto controllo;
- una Linea Superiore di Controllo (UCL), sovrastante a LC, che rappresenta il valore accettabile più elevato che la statistica può assumere se il processo è sotto controllo;
- una Linea Inferiore di Controllo (LCL), sottostante a CL, che rappresenta il valore accettabile minore che la statistica può assumere se il processo è sotto controllo.

La carta di controllo consiste nella rappresentazione grafica dei valori della statistica di controllo e delle tre linee CL, UCL e LCL. UCL e LCL rappresentano i *limiti di controllo*, più precisamente costituiscono i valori critici con cui la statistica viene confrontata per decidere se il processo è sotto controllo *(IC)* o fuori controllo *(OC)*.

Indicata con w_t la statistica di controllo nell'istante t, funzione delle x_i misurazioni della caratteristica X:

$$w_t = f(x_1, \ldots x_n)$$

dove n è la numerosità campionaria, si tratta di verificare una sequenza d'ipotesi del tipo:

$$\begin{cases} H_0 : w_t \in (LCL, UCL) \\ H_1 : w_t \notin (LCL, UCL) \end{cases}$$

Finché si accetta l'ipotesi nulla H_0, il processo che si osserva è sotto controllo; alla prima accettazione dell'ipotesi alternativa H_1, viene lanciato un allarme sul possibile fuori controllo del processo.
Vi sono vari tipi di carte di controllo in funzione delle caratteristiche attese del prodotto e del tipo di produzione. La prima grossa differenza è nel tipo di variabile che si vuole monitorare, vi sino infatti carte di controllo *per variabili* applicabili a variabili di tipo quantitativo e carte di controllo *per attributi* applicabili a variabili di tipo qualitativo.
Un'ulteriore differenziazione suddivide le carte di controllo per variabili in: carte *con* o *senza memoria* a seconda che la statistica di controllo tenga conto o meno dei precedenti valori della stessa. Quest'ultima scelta è in funzione del tipo di variazione sulla variabilità della caratteristica in osservazione che si desidera cogliere. Per identificare grandi variazioni (1,5σ) sono efficienti carte di controllo senza memoria, mentre se si desidera osservare variazioni di piccola dimensione si deve necessariamente ricorrervi.

1.3 La Carte di Controllo CUSUM

1.3.1 Introduzione

Le Carte CUSUM (Cumulative SUM) o "*somme cumulate*" sono state inizialmente proposte da Page *(1954)* e successivamente studiate da diversi autori, di particolare rilevanza i lavori di Ewan *(1963)*, dello stesso Page *(1961)*, Gan *(1991)*, Lucas *(1976)*, Hawkins *(1981,1993a)* e Woodall e Adams *(1993)*. Esse sono un ottimo strumento per il controllo dei processi dove si incontrano frequentemente campioni di ampiezza unitaria.

Le carte CUSUM sono carte con memoria preferibilmente utilizzabili per individuare piccole variazioni (inferiori a 1,5σ) della produzione.

1.3.2 La statistica vera e propria CUSUM

Le carte CUSUM si basano sull'osservazione degli scarti dalla media, più precisamente si rappresentano graficamente le somme cumulate degli scarti tra le osservazioni ed un opportuno valore di riferimento, di solito la media del processo in controllo. Definendo come μ_0 il valore obiettivo per la media generale del processo e considerando $\overline{x_j}$ la media del *j*-esimo campione (con numerosità campionaria $n \geq 1$), la carta di controllo a somme cumulate CUSUM per *i*-esimo campione è:

$$C_i = \sum_{j=1}^{i} \left(\overline{x_j} - \mu_0 \right)$$

È evidente che quando il processo è sotto controllo la media è pari a μ_0. La somma cumulata appena definita si comporta dunque come un processo si tipo *random-walk*, di media nulla. Se invece la media dovesse aumentare, raggiungendo ad esempio un valore $\mu_1 \geq \mu_0$, allora la somma cumulata C_i verrebbe a presentare una cosiddetta deriva positiva. Viceversa, nel caso di un decremento nella media del processo ad un generico valore $\mu_1 \leq \mu_0$, si avrebbe per C_i un comportamento opposto. Pertanto lo svilupparsi di un trend crescente o decrescente, nell'ambito delle somme cumulate, può essere interpretato come evidenza empirica di un avvenuto salto nel livello medio nel processo, rendendo così necessaria l'attivazione di una procedura di ricerca delle possibili cause di sregolazione.

La C_i può anche essere scritta per semplificarne i calcoli come:

$$C_i = \sum_{j=1}^{i}\left(\overline{x_j} - \mu_0\right) = \left(\overline{x_i} - \mu_0\right) + \sum_{j=1}^{i-1}\left(\overline{x_j} - \mu_0\right) = \left(\overline{x_i} - \mu_0\right) + C_{i-1}$$

risulta ora evidente che come valore iniziale della somma cumulata si considera $C_0 = 0$.

1.3.3 *Forma tabulare o algoritmica delle CUSUM*

La forma più utilizzata delle carte CUSUM per il controllo delle media di un processo produttivo è la tabulare.

Essa presuppone, come le altre, che quando il processo è sotto controllo, $\overline{x_j}$ ha una distribuzione normale di media μ_0 e scarto quadratico medio σ, supponiamo per ora che quest'ultimo sia noto o che se ne disponga di una stima altamente attendibile.

La carta CUSUM tabulare opera accumulando in una statistica, indicata con C^+, le deviazioni da μ_0 che risultano superiori al valore obiettivo, e quelle inferiori a μ_0 in una statistica indicata C^-. Le statistiche C^+ e C^- vengono dette rispettivamente *CUSUM unilaterale superiore* e *CUSUM unilaterale inferiore* e sono così definite:

$$C_i^+ = \max\left[0, \overline{x_j} - (\mu_0 - K) + C_{i-1}^+\right]$$

$$C_i^- = \max\left[0, (\mu_0 - K) - \overline{x_j} + C_{i-1}^-\right]$$

Dove i valori iniziali sono $C_0^+ = 0$ e $C_0^- = 0$.

La quantità $\mathrm{K} = \mathrm{k}\sigma$ è solitamente detta valore di tolleranza ed è di norma pari alla metà dello scostamento tra il valore obiettivo μ_0 e quel valore μ_1 assunto fuori controllo.

Quindi se si esprimono i salti di livello della media del processo in funzione di σ, ovvero $\mu_1 = \mu_0 + \delta\sigma$ (e quindi $+\delta = |\mu_1 - \mu_0|/\sigma$), K corrisponde alla semiampiezza del salto cioè:

$$K = \frac{\delta}{2}\sigma = \frac{|\mu_1 - \mu_0|}{2}$$

È chiaro quindi che le statistiche C^+ e C^- accumulano solo le deviazioni del valore obiettivo μ_0 di ampiezza superiore a K; inoltre, entrambe si riportano a 0 non appena il loro valore risulta negativo.

A questo punto restano da definire i limiti di controllo (UCL,LCL) che per le carte CUSUM sono delle rette orizzontali poste in corrispondenza di : $\mathrm{UCL} = -\mathrm{LCL} = \mathrm{H} = \mathrm{h}\sigma$.

In pratica la scelta dei due parametri H e K è basilare per le prestazioni che la carta di controllo fornisce.

1.3.4 Procedura di avviamento a risposta accelerata

La procedura *Fast Initial Response* (*FIR*) è stata elaborata da Lucas e Crosier *(1982)* per migliorare la sensibilità della carta CUSUM nella fase di avvio di un processo. Ciò risulta utile quando l'azione correttiva non comporta il riportare la media esattamente al valore di riferimento. La procedura di *avviamento a risposta accelerata* (o *headstarter*) fissa i valori iniziali di C_0^+ e C_0^- a

opportune costanti non nulle, tipicamente $H/2$, accelerando così la risposta del 50%.
Quando il processo si avvia in uno stato di controllo la statistica si porta rapidamente a zero e la fase di avvio quindi non interferisce minimamente. Se invece il processo parte già da un livello differente da quello nominale, allora l'avvio accelerato fa sì che la carta CUSUM lo scopra più velocemente.

1.3.5 La Carta CUSUM Bilaterale Standardizzata

Le statistiche C^+ e C^- se usate singolarmente danno origine a carte CUSUM unilaterali; il loro impiego trova una giustificazione nel fatto di sapere a priori che la media del processo in esame in caso di fuori controllo della linea produttiva tenderà certamente ad aumentare o certamente a diminuire, da qui il tralasciare della statistica opposta.
In pratica però a priori non si è in grado di sapere in quale direzione si muove la media del processo in caso di un fuori controllo, da qui la necessità di utilizzare contemporaneamente le statistiche C^+ e C^- per poter così individuare sia scostamenti positivi che negativi.
Molti utilizzatori della carta CUSUM preferiscono standardizzare le variabili x_i prima di procedere con i calcoli delle statistiche:

$$y_i = \frac{x_i - \mu_0}{\sigma}$$

Le CUSUM standardizzare sono quindi definite da:

$$C_i^+ = \max\left[0, y_i - K + C_{i-1}^+\right]$$

$$C_i^- = \max\left[0, -K - y_i + C_{i-1}^-\right]$$

Dalla standardizzazione derivano essenzialmente due vantaggi: il primo è che con gli stessi valori di H e K possono così essere descritte diverse carte e quindi la scelta di questi parametri può farsi in maniera indipendente dalla scala di misura (cioè da σ); il secondo è che una carta CUSUM standardizzata aiuta nella costruzione di carte per il controllo della variabilità.
Si ricordi che nel caso di osservazioni non individuali (quindi $n \geq 1$) è sufficiente sostituire nelle precedenti formule: x_i con $\overline{x_i}$ e σ con $\sigma_x = \sigma / \sqrt{n}$.

1.3.6 Valutazione di una carta CUSUM

La carta CUSUM tabulare viene costruita scegliendo opportunamente il valore di riferimento K e l'intervallo di decisione H. Solitamente, viene suggerito di stabilire questi parametri in modo da ottenere buone prestazioni in termini di *ARL* (*Average Run Length*).
La *Run Length*, *T*, di una carta di controllo è una variabile casuale che rappresenta il primo istante temporale in corrispondenza del quale la statistica di controllo, nel nostro caso la CUSUM, fuoriesce dai limiti di controllo LCL e UCL, nel nostro caso $\pm \mathrm{H}$. In questo istante viene lanciato un allarme di un possibile fuori controllo.
In pratica si cerca di disegnare carte di controllo che segnalino falsi allarmi molto raramente ed invece segnalino veri allarmi di fuori controllo il più tempestivamente possibile.
L'*ARL* che indica quindi il tempo medio di attesa per un segnale d'allarme deve essere il più grande possibile per i falsi allarmi e minore possibile per quelli veri.
Nella costruzione di una carta l'ARL in controllo è posto pari a predeterminarti valori, di conseguenza verranno scelti i

parametri della carta in modo da ottenere il minor valore dell’ARL fuori controllo.

Capitolo 2

Stima del valore h

2.1 Il problema dei parametri ignoti

Una classica applicazione di una Carta CUSUM per monitorare la media di un processo produttivo richiede l'assunzione che la media e la varianza del processo in controllo siano note. Tuttavia, nella maggior parte delle applicazioni la media e la varianza devono essere stimate a partire dalle misurazioni campionarie del processo in controllo.
L'utilizzo di questa stima porta ad una peggioramento della capacità della carta di individuare cambiamenti nel processo in esame.

2.2 La stima dei parametri

Lo stimatore comunemente utilizzato per stimare la media μ_0 è quindi:

$$\hat{\mu}_0 = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}$$

dove X_{ij} è la j-esima osservazione dal sottogruppo i.

Per quanto riguarda invece la varianza del processo produttivo in controllo σ_0 lo stimatore più utilizzato è:

$$\hat{\sigma}_0 = \frac{S_p}{c_{4,m}}$$

dove:

$$S_p = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij}-X_{i.}\right)^2}{m(n-1)}} \quad \text{e} \quad c_{4,m} = \frac{\sqrt{2}\,\Gamma\left(\frac{m(n-1)+1}{2}\right)}{\sqrt{m(n-1)}\,\Gamma\left(\frac{m(n-1)}{2}\right)}$$

Il valore $c_{4,m}$ è funzione del numero di campioni e della loro ampiezza, Derman e Ross *(1995)* mostrano che lo stimatore S_p ha una varianza minore rispetto agli altri possibili stimatori soprattutto se i campioni provengono da una distribuzione normale.

2.3 Scelta del parametro "*h*"

Una volta fissato l'*ARL* in controllo resta da scegliere δ ossia lo shift dalla media che si desidera individuare.
Per quanto riguarda invece il parametro h, si utilizza una procedura di stima basata sull'approssimazione stocastica *(Capizzi-Masarotto,1999)*.

2.3.1 L'approssimazione stocastica

2.3.1.2 Il Problema

Si supponga che, in corrispondenza di ogni valore di un parametro, diciamo $h \in R$, sia possibile ottenere delle determinazioni indipendenti di una variabile casuale reale, Y_h, con funzione di distribuzione $F(y;h)$. A riguardo di quest'ultima si assume solamente che

$$M(h) = E\{Y_h\} = \int_R y\, dF(y;h)$$

esista per qualsivoglia h, sia monotona crescente ed inoltre che esista e sia unico un valore di h, indichiamolo con $\bar{h}$, per cui

$M(\bar{h})=0$. Il problema, affrontato dagli schemi di approssimazione stocastica, consiste nel determinare sequenzialmente una successione $h_1, h_2, \ldots$ convergente in qualche senso a h.

2.3.1.4 *Lo schema di Robbins e Monro*

Una prima soluzione al problema è stata suggerita da Robbins e Monro *(1951)*. Essi suggeriscono di generare la successione $h_1, h_2, \ldots$ utilizzando la relazione iterativa

$$h_n = h_{n-1} - \frac{A}{n} y_n$$

dove y_n è un'osservazione distribuita in accordo a $F(.; h_{n-1})$ mentre A è una opportuna costante positiva. Per quanto riguarda il valore di h_0, questo può essere scelto in modo arbitrario. Ovviamente, la convergenza sarà tanto più veloce quanto più h_0 risulta vicino al vero valore h.

È possibile fare vedere *(Ruppert, 1991)* che h_n converge quasi certamente ad h ed inoltre che, se esiste, oltre al valore atteso, anche la varianza di Y_n, allora $\sqrt{n}(h_n - \bar{h})$ converge in distribuzione ad una normale di media zero e varianza $A^2\sigma^2/(2AM'(\bar{h})-1)$, dove σ^2 indica la varianza di $Y_{\bar{h}}$.

2.3.1.4 *Regole d'arresto*

Lo schema produce una successioni convergente. Nelle applicazioni concrete è però necessario troncare le iterazioni in corrispondenza di un valore di n, diciamo N. La stima effettivamente generata dallo schema sarà quindi h_N.

Un possibile approccio per la scelta di N nel caso dello schema di Robbins e Monro è stato suggerito da Stroup e Braun *(1982)*.

Questi autori suggeriscono di ottenere per ogni n due valori y_{n1} e y_{n2} tali che $E\{y_{n1}\} = E\{y_{n2}\} = M(h_{n-1})$. Il processo di Robbins e Monro viene modificato semplicemente sostituendo nella ad y_n con $l_n = (y_{n1} + y_{n2})/2$ ed il processo viene arrestato in corrispondenza di

$$N = \inf\left\{n > k : u_n = \left(\sum_{i=n-k+1}^{n} l_i^2 /k\right)\Bigg/\left(\sum_{i=1}^{n} e_i^2 /n\right) < w\right\}$$

dove $e_n^2 = (y_{n1} + y_{n2})^2/2$, mentre k è un intero appropriato.
In sostanza, Stroup e Braun propongono di fermare la procedura nel caso in cui negli ultimi k passi si siano ottenute osservazioni sufficientemente vicine allo zero. Per quanto riguarda la scelta della soglia w, Stroup e Braun mostrano che la statistica u_n converge in distribuzione a $(k\sigma^2)^{-1}$ la distribuzione di $\lim_{n\to+\infty} \sum_{i=n-k+1}^{n} (l_i - E\{l_i\})^2$ Così se, ad esempio, le risposte y_{ni} sono gaussiane allora $2ku_n$ si distribuisce, per n sufficientemente grande, come una variabile casuale χ^2 con n gradi di libertà e, possiamo quindi pensare di scegliere $2kw$ uguale ad un percentile appropriato di questa distribuzione. Più in generale, dal momento che a convergenza ottenuta il valore atteso di un risulta pari ad $1/2$ possiamo pensare a delle soglie del tipo $\gamma/2$, con $\gamma > 0$.

2.3.1.4 Applicazione alla calibrazione di una carta di controllo

L'applicazione di quanto descritto nel paragrafo precedente al problema della scelta della soglia h di una carta di controllo risulta quindi:
la *run length* di uno schema definibile:

$$R_h = \inf\{t > 0 : s_t > h\};$$

la soglia h viene comunemente scelta risolvendo un'equazione del tipo:

$$E\{R_h\} = ARL(h) = B$$

dove B rappresenta un conveniente valore per il tempo medio intercorrente tra due falsi allarmi. Il valore atteso $E\{R_h\}$ è calcolato con riferimento alla distribuzione in controllo di $(z_1, z_2, \ldots)$.

Nei casi in cui $ARL(h)$ non sia né ottenibile in forma chiusa né approssimabile per via numerica, ma viceversa risulti possibile simulare il processo in controllo, possiamo ottenerlo risolvendo $Y_h = (R_h - B)/B$.

Si osservi la presenza di B al denominatore. Anche se non strettamente necessaria per la convergenza, si rivelata molto utile per accelerare la convergenza. In realtà, poiché la varianza della R_h è praticamente sempre proporzionale a $ARL^2(h)$, si tratta sostanzialmente di una standardizzazione.

Capitolo 3

I Risultati

3.1 La procedura per la stima di *"h"*

La procedura che si è implementata è composta da due parti principali:

- una prima funzione simula le osservazioni della *Run Length*
  - si generano m campioni di ampiezza n da una normale di media nulla e varianza pari a 0.21;
  - si procede al calcolo delle stime $\hat{\mu}$ e $\hat{\sigma}$;
  - si confronta la statistica CUSUM unilaterale superiore calcolata sul campione con la soglia $\hat{h}$.

Il primo istante in cui la statistica di controllo eccederà tale limite sarà pari alla *Run Length*.

- una seconda funzione aggiorna, mediante la formula ricorsiva di *Robbins e Monro,* la stima della soglia $\hat{h}$ fino al verificarsi della condizione della regola d'arresto.
  - Per un valore iniziale di h si generano due osservazioni indipendenti della *Run Length*, utilizzando la prima funzione, per poi passare ad una nuova stima di $\hat{h}$ fino all'arresto della ricorsione.

- Questa seconda funzione viene fatta girare per `nSim` volte affinché si ottengano `nSim` valori di $\hat{h}$. Alla fine si calcola il valor medio e lo scarto quadratico medio dei valori di $\hat{h}$, indicati con $\bar{h}$ e $\sigma_{\hat{h}}$.

Le procedure sono state implementate in GAUSS, il listato è riportato in Appendice.

3.2 La distribuzione dell'*ARL* con *h* stimato

Si è poi implementata una seconda procedura per la stima della distribuzione dell'*ARL* sulla base delle stime $\hat{h}$ ottenute. I risultati sono stati confrontati con quelli ottenuti da *Jones, Champ e Rigdon (2004)* sullo studio delle distribuzioni della *RL* di una CUSUM con parametri stimati.

Scelto come valore di soglia della statistica CUSUM il valore $\bar{h}$ ottenuto per mezzo dell'approssimazione stocastica, si procede al calcolo del profilo dell'*ARL* per nove diversi valori dello shift δ, a partire da $\delta = 0$, $\delta = 0{,}25$, ..., fino a $\delta = 2$.

La stima dell'*ARL* è ottenuta sulla base di campioni simulati (m campioni di ampiezza n) ottenendo per ciascun valore dello shift δ e coppia m ,n, 3000 simulazioni della *Run Length*.

La media di questi 3000 risulti sarà la nostra stima dell'*ARL*.

Anche in questo caso la procedura è stata implementata in GAUSS ed il listato è riportato anch'esso in Appendice.

3.3 I risultati

Il primo passo è stato quindi quello di stimare il valore della soglia h per una carta CUSUM unilaterale disegnata per individuare uno scostamento della media pari a σ, quindi $\delta = 1$ e $\mathrm{k} = 0{,}5$. Fissato l'*ARL* in controllo (B) pari a 250, per ciascuna coppia (m, n), si sceglie `nSim` pari a 100, in modo da ottenere 100 valori di $\hat{h}$. Per quanto riguarda i parametri h_0 ed A, per la formula ricorsiva di *Robbins e Monro*, si è posto arbitrariamente $h_0 = 1$ ed $A = 1{,}5$.

Per quanto riguarda invece m ed n: si scelgono i valori $\mathrm{m} = 30, 100$, $\mathrm{n} = 1, 5$; ottenendo così quattro stime per h.

Nel caso in cui i parametri del processo siano noti, la letteratura suggerisce un valore di h è pari a 3.716. I nostri risultati sono:

m	n	$\hat{h}$	$\sigma_{\hat{h}}$
30	1	3.7166898	0.04313423
	5	3.7113611	0.04578681
100	1	3.7114881	0.04713989
	5	3.7035758	0.04475102

I risultati ottenuti sono abbastanza buoni anche se non si vede chiaramente un miglioramento nell'aumentare la numerosità dei campioni m e/o della loro ampiezza n. Resta il fatto che approssimando alla seconda cifra decimale i risultati sono più che accettabili alla luce anche della bassa varianza della stima. Eccetto la prima stima, le altre tre, sottostimano la soglia h il che porta a concludere che carte di controllo basate su questa procedura tendono ad avere in *ARL* in controllo leggermente più basso.

Passiamo quindi ad analizzare i simulati sul comportamento dell'*ARL:*

Per $m = 30$, $n = 1$

Shift (δ)	*ARL stimato*	*ARL articolo*
0.00	251.6547	/
0.25	66.77267	/
0.50	22.879	/
0.75	12.38833	/
1.00	7.622333	/
1.25	5.712333	/
1.50	4.437667	/
1.75	3.634667	/
2.00	3.140667	/

Per $m = 100$, $n = 1$

Shift (δ)	*ARL stimato*	*ARL articolo*
0.00	246.9973	368.79
0.25	62.37133	81.47
0.50	23.329	26.96
0.75	12.126	12.99
1.00	7.769333	8.06
1.25	5.747667	5.78
1.50	4.469	4.51
1.75	3.657	3.71
2.00	3.163333	3.15

Per $m = 30$, $n = 5$

Shift (δ)	*ARL stimato*	*ARL articolo*
0.00	257.7337	658.75
0.25	63.37133	124.36
0.50	24.09267	34.55
0.75	12.099	14.66
1.00	7.724	8.54
1.25	5.635	5.96
1.50	4.484	4.59
1.75	3.718667	3.75
2.00	3.170667	3.18

Per $m = 100$, $n = 5$

Shift (δ)	*ARL stimato*	*ARL articolo*
0.00	245.9737	353.20
0.25	64.45167	76.07
0.50	23.48033	26.16
0.75	12.33967	12.82
1.00	7.781667	8.01
1.25	5.680333	5.76
1.50	4.441	4.50
1.75	3.665	3.70
2.00	3.118667	3.15

La prima cosa da notare è la nostra procedura consente di mantenere l'*ARL* in controllo ad un valore prossimo a quello prefissato, laddove quello ottenuto da *Jones, Champ e Rigdon (2004)* risulta molto più alto. Inoltre, osservando i valori dell'ARL per valori crescenti dello shift la nostra stima si avvicina sensibilmente al valore ottenuto da *Jones, Champ e Rigdon (2004)*, mediante un approccio probabilistico.

3.4 Conclusioni

La stima quindi del parametro h mediante un metodo di approssimazione stocastica, risulta abbastanza buona sia tenendo conto dei risultati ottenuti sia dal punto di vista della complessità dell'implementazione dell'algoritmo necessario.

Appendice

1 – Procedura per la stima di “h”

```
/*
CUSUM UNILATERALE - Processo Robbins-Monro
                 Stima di h
*/

df=hsec;
{h}=prova(0,0.5,5,sqrt(0.0212),1,1.5,200,0.5,250,100);
100*(hsec-df)/(100*60*60);
save cuss11=h;



proc runl(media,k,sizeg,sqns,h);

  local i,y,ns,z,t,hl,kl,zc,sqnsm,tmp;
  y=0; t=0; zc=0;
  ns=zeros(30,1);

  sqnsm=sqns/sqrt(sizeg);
  hl=h*sqnsm;
  kl=k*sqnsm;

  locate 1,1;print "hl" hl "h" h;
  do while (t<=hl);
    locate 2,1; print "t" t "y" y "zc" zc;
    i=1;
    do while (i<=30);
      ns[i]=meanc(rndn(sizeg,1)*(sqns));
      i=i+1;
    endo;
    z=media+ns;
    zc=meanc(z);
```

```
    t=maxc(0|(t+zc-kl));
    y=y+1;
    if y >= 10000;
      break;
    endif;
  endo;

  retp(y);

endp;

proc (2)= hNss(media,k,sizeg,sqns,h1,A,seq,w,B);

  local u2,u1,u,j,h,hlc, sume,ybar,ybar2,y1,y2,aa;
  j=1; h=h1; u=2;
  sume=10.0e-255;
  ybar2=zeros(seq,1);

  do while u > 1;
    y1=(runl(media,k,sizeg,sqns,h)-B)/B;
    y2=(runl(media,k,sizeg,sqns,h)-B)/B;
    ybar=(y1+y2)/2;
    if seq==1;
      ybar2=ybar;
    else;
      ybar2=ybar2[2:seq]|ybar;
    endif;
    sume=sume+((y1-y2)^2)/2;
    h=h-(A/j)*ybar;
    if h<0;
      h=0;
    endif;
    u1=2*sumc(ybar2^2)/(sume/j);
    u2=u1/(2*(seq*w));
    locate 4,1;
    hlc=4.39*sqns;
```

```
    locate 5,1;
    print "j=" j "seq=" seq "h=" h "hlc" hlc;
    if j > seq;
      u=u2;
      locate 6,1;
      print "u" u;
    endif;
    j=j+1;
  endo;

  retp(h,j);

endp;

proc prova(media,k,sizeg,sqns,h1,A,seq,w,B,nSim);

  local h,i;
  h=zeros(nSim,2);
  i=1;
  locate 12,1;
  do while i <= nSim;
    print "nSim" i;

{h[i,1],h[i,2]}=hnss(media,k,sizeg,sqns,h1,A,seq,w,B);
    i=i+1;
  endo;
  locate 10,1; print "h" h;

  retp(h);

endp;
```

2 – Procedura per la stima dell’ARL fuori controllo

```
/*
  ARL fuori controllo per il cusum
*/

proc cusum1(shift,k,sqns,h,sizeg);

  local y,ns,z,t,kt,hl,media,sqnsm,i;
  y=0; i=0; t=0; zc=0; i=1;
  ns=zeros(30,1);

  sqnsm=sqns/sqrt(sizeg);
  hl=h*sqnsm;
  kt=k*sqnsm;
  do while (t<=hl);
    locate 2,1;
    print "t" t "y" y;
    do while (i<=30);
      ns[i]=meanc(rndn(sizeg,1)*(sqns));
      i=i+1;
    endo;
    z=media+ns;
    zc=meanc(z);
    t=maxc(0|(t+zc-kl));
    print "y=" y  "t" t;
    y=y+1;
    if y >= 10000;
      break;
    endif;
  endo;

  retp(y);

endp;
```

```
proc cusumout(shift,k,sqns,h,nSim,sizeg);

  local riss,i;
  riss=zeros(nsim,1);
  i=1;

  do while i <= nSim;
    {riss[i,1]}=cusum1(shift,k,sqns,h,sizeg);
    i=i+1;
    locate 3,1;
    print "riss" riss;
  endo;

  retp(riss);

endp;



let shift= 0 0.25 0.5 0.75 1 1.25 1.5 1.75 2;
kk=1;
nome="cusum";
do while kk<=9;
  nomeout=nome$+ftocv(kk,1,0)$+".dat";
  output file=^nomeout;
  ii=1;
  do while ii<=3000;
    {riss}=cusumout(shift[kk],0.5,sqrt(0.0212),3.71767,
    1,1);
    output on;
    riss;
    output off;
    ii=ii+1;
    locate 10,1;
```

```
    print "ii" ii "kk" kk "cusum fuori controllo-
           shift";
  endo;
  kk=kk+1;
endo;
```

Bibliografia

- **Capizzi G.,Masarotto G. (1999)** *"Calibrazione di una carta di controllo mediante approssimazione stocastica"* - Atti dello XXXIX Riunione Scientifica della Società Italiana di Statistica, in supplemento alla Rivista di Scritti di Statistica Economica, pp.809-815
- **L. Allison Jones, Charles W. Champ, Steven E. Rigdon (2004)** *"The Run Length Distribution of the CUSUM with estimate Parameters"* - Journal of Quality Tecnology, Vol. 36, No. 1, Jenuary 2004.
- **Page E. S. (1954)** *"Continius ispaction schemes"* – Biometrics 61.
- **Ewan (1963)** *"When and how to use Cu-Sum charts"* - Tecnometrics 5.
- **Page E. S. (1961)** *"Cumulative sum control charts"* - Tecnometrics 3.
- **Gann F.F. (1991)** ***"An optional design of CUSUM quality control charts"* -** Journal of Quality Tecnology, Vol. 23.
- **Lucas J.M. (1976)** ***"The design and use of cumulative sum quality control schemes"* -** Journal of Quality Tecnology, Vol. 14.
- **Hawkins D.M. (1981)** ***"A CUSUM for a scale parameter"* -** Journal of Quality Tecnology, Vol. 13.
- **Hawkins D.M. (1993a)** ***"Cumulative sumcontrol charting"* -** Quality engineering, 5.
- **Woodall W.H., Adams B.M. (1993)** ***"the statistical design of Cusum charts"* -** Quality engineering, 5.

- **Lucas J.M., Crosier R.B. (1982) *"The initial response dor CUSUM quality control schemes"* –** Tecnometrics 24.
- **Derman C., Ross S. (1995) *"An improved estimator of σ in quality contro"* –** Probability in the engeering in informatical sciences 9, pp 411-415.
- **Montgomery D.C.(1999)**, *"Il controllo statistico della qualità"*, McGraw-Hill 1999.
- **Kuan-Pin Lin (2001)**, *"Computational Econometrics – Gauss Programmino for Econometricians and Financial Analysis"* – ETEXRT Los Angeles 2001
- **http://www.trigconsulting.co.uk/gauss/manual.html (2002)**