UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA
IN STATISTICA E TECNOLOGIE INFORMATICHE

RELAZIONE FINALE

DISEGNO DI UNA CARTA EWMA CON PARAMETRI DEL PROCESSO INCOGNITI

RELATORE: CH.MA PROF.SSA GIOVANNA CAPIZZI

LAUREANDA: ERIKA BELLIN

ANNO ACCADEMICO 2003-04

A
Tai Gon.

Indice

Introduzione

La carta di controllo Ewma (Roberts, 1959) viene tipicamente implementata assumendo che i parametri del processo siano noti. In realtà solo raramente i parametri sono conosciuti e quindi la carta di controllo viene disegnata sostituendo ai parametri delle loro stime, con una conseguente diminuzione della efficienza della carta.

Il seguente studio tratterà la stima del valore h di una carta di controllo Ewma con parametri del processo (μ, σ) incogniti e di come la grandezza campionaria può influire su di essa. Nel primo capitolo vi sarà una descrizione generale della carta Ewma; nel secondo capitolo verrà esposto il problema della stima del parametro h e proposto, come metodo di stima, l'utilizzo delle approssimazioni stocastiche, nel terzo capitolo verranno esposti i risultati e le conclusioni.

Capitolo 1

Concetti generali

1.1 La variabilità nei processi produttivi

In ogni processo produttivo si possono distinguere due tipi di variabilità. Vi è una variabilità minima, intrinseca al processo, che non può essere eliminata in quanto risultato di fattori casuali di scarsa influeza, che nel loro complesso determinano un comportamento del processo da considerarsi tipico. Essa viene chiamata variabilità naturale ed un processo che presenta solo questa variabilità viene detto ***sotto controllo*** (*in control*, IC). Esistono poi fonti di variabilità che possono invece incidere sul livello di qualità del prodotto e principalmente sono riconducibili a guasti dei macchinari, errori degli operatori e materiali difettosi. La variabilità prodotta da questi fattori è molto più evidente di quella precedente e può portare a lotti di produzione difettosi. Un processo che presenta una variabilità, riconducibile a fattori specifici, viene detto ***fuori controllo*** (*out of control*, OC). Obiettivo principale del controllo della qualità è quindi di monitorare la produzione, attraverso la misurazione delle sue caratteristiche e di alcune statistiche ad esse associate, per individuare eventuali cambiamenti, di ampiezza δ, nella statistica di controllo. Lo strumento maggiormente utilizzato, ed affidabile, per stabilire se un processo abbia subito un cambiamento nella caratteristica monitorata, è la *carta di controllo*.

1.2 Le carte di controllo

La carta di controllo consiste in un grafico nel quale vengono riportati i valori della statistica di controllo, in funzione del tempo, t. Essa è costituita da:

- una statistica di controllo, w_t, corrispondente ad una funzione dei dati osservati, che rappresentano misurazioni della caratteristica del processo;

- CL (Central Line), una linea centrale che rappresenta il valore nominale della statistica per il processo in controllo;

- UCL (Upper Central Line), una linea superiore che rappresenta il valore più grande che la statistica può assumere affinchè il processo risulti in controllo;

- LCL (Lower Central Line), una linea inferiore che rappresenta il valore più piccolo che la statistica può assumere affinchè il processo risulti in controllo.

I valori UCL e LCL sono detti *limiti di controllo*. Essi rappresentano quegli indici con cui la statistica di controllo viene confrontata ad ogni istante t. La carta di controllo non è altro che una sequenza di test di verifica d'ipotesi per saggiare se il processo risulti in controllo o meno.
Data una statistica di controllo w_t che misuri la caratteristica X:

$$w_t = f(\underline{x}_1 \cdots + \underline{x}_n)$$

si considera dunque il seguente problema di verica d'ipotesi :

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

non appena una statistica di controllo cade al di fuori dei limiti di controllo viene lanciato l'allarme di una possibile anomalia nel processo. Esistono diversi tipi di carte di controllo ciascuna applicabile in determinate situazioni.

Una prima distinzione viene effettuata tra carte di controllo per *variabili* nel caso in cui si trattino variabili continue, e carte di controllo per *attributi*, nel caso si trattino di variabili discrete. Successivamente le carte per variabili sono distinte in carte di controllo con memoria o senza memoria. Le prime utilizzano le informazioni provenienti dai campioni precedenti in aggiunta a quella del campione attuale e permettono di cogliere cambiamenti di piccole dimensioni ($0 < \delta < 1.5$), mentre le seconde utilizzano solamente l'informazione derivante dal campione corrente e sono adatte ad individuare $\delta > 1.5$.

1.3 La carta di controllo EWMA

1.3.1 Applicabilità

La carta di controllo EWMA, *Exponentially Weighted Moving Average*, venne introdotta da Roberts (1959) per individuare piccoli cambiamenti nella media nei processi produttivi. Essa può essere utilizzata in due diversi modi, (1) come strumento per monitorare il processo produttivo, (2) come previsore per le successive osservazioni del processo.Tipicamente per una corretta applicazione della carta EWMA viene assunto che le osservazioni del processo siano distribuite come una Normale con parametri, media e deviazione standard noti. Nel seguente studio vengono imposte altre due assunzioni fondamentali: le osservazioni del processo siano indipendenti ed identicamente distribuite.

1.3.2 Calcolo della carta EWMA

Si assuma che le osservazioni della carattestica di qualità di interesse, denotata con X_t, siano indipendenti ed identicamente distribuite secondo una normale di media μ_t e varianza σ_X, dove il valore μ_t indica che la media delle osservazioni può subire dei cambiamenti nel tempo.

Il valore della statistica EWMA al tempo t, che indicheremo come Q_t, è definita quindi come

$$Q_t = (1-\lambda)Q_{t-1} + \lambda Y_t \tag{1.1}$$

dove λ, pari alla costante di lisciamento esponenziale appartenente all'intervallo $0 < \lambda \leq 1$, determina quanto peso assegnare alle osservazioni passate nel calcolo di Q_t. Più il suo valore sarà piccolo più memoria avrà la statistica. L'equazione (1.1) può essere scritta come

$$Q_t = \lambda \sum_{i=0}^{t-1} (1-\lambda)^i + (1-\lambda)^t Q_0 \tag{1.2}$$

dove Q_0 è pari al valore target della statistica.
Assumendo che $Q_0 = \mu$, è facile dimostrare che dall'equazione (1.2) risulta $E[Q_t] = \mu$, mentre la varianza è pari a

$$\sigma^2_{Q_t} = \sigma^2_X \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]. \tag{1.3}$$

I limiti di controllo risultano pari a:

$$UCL_t, LCL_t = \mu \pm h\sigma_X \sqrt{\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]} \tag{1.4}$$

dove h è una costante che determina l'ampiezza dei limiti di controllo.
Dal momento che la quantità $1-(1-\lambda)^{2t} \to 1$, per $t \to \infty$ si ottiene la varianza asintotica

$$\sigma^2_{Q_t} = \sigma^2_X \left(\frac{\lambda}{2-\lambda}\right) \tag{1.5}$$

ed i limiti di controllo asintotici risultano pari a

$$UCL_t, LCL_t = \mu \pm h\sigma_X \sqrt{\left(\frac{\lambda}{2-\lambda}\right)}. \tag{1.6}$$

1.3.3 Interpretazione

Gli andamenti non casuali all'interno del processo monitorato vengono segnalati usualmente con la fuoriuscita della statistica dai limiti di controllo. Quando si utilizza una carta EWMA andamenti non casuali possono essere individuati anche studiando l'andamento della statistica. Infatti i valori della statistica EWMA fluttuano attorno alla linea centrale, quando il processo risulta in controllo, mentre quando vi è un cambiamento essa si sposta lentamente verso il nuovo livello. L'individuazione di cambiamenti nel processo può avvenire evidenziando la presenza di andamenti crescenti o decrescenti della statistica di controllo.

1.3.4 Valutazione di una carta EWMA

Per ottenere una carta di controllo EWMA che sia in grado di raggiungere determinate *performance*, è necessario individuare la coppia di parametri, λ e h, ottimali. I criteri per valutare le capacità di una carta sono basati maggiormente sulla distribuzione della *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 EWMA fuoriesce dai limiti di controllo. Questo evento è segnale di una potenziale condizione di fuori controllo. In realtà la distinzione tra un falso allarme ed un vero fuori controllo non è sempre individuabile, si cerca così di costruire una carta che segnali il minor numero di falsi allarmi e che riesca e individuare un vero cambiamento il prima possibile. La più comune misura per valutare le capacità di una carta di controllo è l'ARL, *Average Run Length*, pari al valore atteso di T. Per un processo in controllo é desiderabile ottenere un ARL il più alto possibile per ottenere una percentuale di falsi allarmi bassa; quando il processo risulta fuori controllo l'ARL deve risultare molto basso così da cogliere il prima possibile il cambiamento avvenuto. Per effettuare un confronto tra diverse carte di controllo usualmente l'ARL in controllo è fissato uguale per tutte le carte come anche il cambiamento da individuare, e la carta con ARL fuori controllo minore risulta essere la migliore.

1.3.5 Procedura per il disegno di una carta di controllo EWMA

Il disegno statistico di una carta di controllo consiste nella determinazione dei limiti di controllo che devono essere ottimali in termini di ARL, in controllo e fuori controllo, nella scelta della dimensione dei campioni e nel determinare la frequenza di campionamento.
Per un ARL in controllo fissato, i limiti di controllo vengono determinati in modo che l'ARL fuori controllo risulti il più piccolo possibile per un cambiamento di grandezza fissata.
Per una carta EWMA ciò consiste nella determinazione delle costanti (λ, h).
Il numero di combinazioni di λ e h per ottenere una carta EWMA con un determinato ARL in controllo sono infinite, ma è necessario individuare quella combinazione in grado di produrre il minor ARL fuori controllo quando un cambiamento è effettivamente avvenuto. L'approccio generale per il disegno di una carta EWMA è il seguente:

- stabilire l'ARL in controllo della carta;
- determinare il cambiamento minimo che la carta deve individuare;
- trovare la costante di lisciamento, λ^*, utilizzando le tavole appropriate di Crowder (1987b, 1989) o di Lucas e Saccucci (1990);
- dato λ, trovare il valore di h che soddisfa il vincolo sull'ARL in controllo;

La combinazione ottimale di parametri, (λ^*, h^*), è dunque tale da produrre per un dato ARL in controllo, l'ARL fuori controllo minimo per un cambiamento prefissato.

Capitolo 2

Stima del valore h

2.1 Il problema dei parametri ignoti

Una delle assunzioni fondamentali per una corretta applicazione della carta EWMA è la conoscenza dei parametri, μ_0 e σ_0, del processo. In realtà nella maggior parte dei casi, la media e la deviazione standard del processo risultano sconosciuti e vengono quindi sostituiti con delle loro stime, calcolate da un processo in controllo. Questo però provoca un'alterazione nelle capacità della carta di individuare i cambiamenti nel processo. Studi precedenti hanno dimostrato che la sostituzione delle stime ai parametri noti porta nella carta EWMA ad un incremento nella segnalazione di falsi allarmi ed ad una diminuzione nella sensibilità di cogliere cambiamenti nel processo; ciò è dovuto al fatto che la distribuzione della carta non riesce a tener conto della variabilità degli stimatori e questo viene maggiormente amplificato quando si hanno a disposizione pochi campioni.

Ad esempio una carta EWMA con parametri $\lambda = 0.2$ e h=2.636, utilizzando i veri parametri del processo, fornisce un ARL in controllo pari a 200. Con i parametri stimati a partire da m=20 campioni di ampiezza n=5, risulta invece un ARL in controllo pari a 144. Anche l'ARL fuori controllo subisce delle conseguenze, sempre per la stessa carta ipotizzando un valore del cambiamento $\delta = 0.5$, l'ARL fuori controllo utilizzando i veri parametri risulta

27, mentre con i parametri stimati risulta pari a 45 (Jones, Champ e Ridgon 2001).
Obiettivo principale di questo lavoro è di valutare come la stima dei parametri del processo, $(\hat{\mu},\ \hat{\sigma})$, influenzi la stima del valore h. Inoltre si vuole effettuare un confronto tra il metodi di stima di h descritto nel seguente studio e quello proposto da Jones (2002) il quale, utilizzando il metodo della quadratura Gaussiana per risolvere l'equazione integrale che fornisce l'ARL in controllo per una carta EWMA con parametri stimati, ottiene una stima esatta della costante h.

2.2 La stima dei parametri

Lo stimatore comunemente usato per il parametro μ_0 è

$$\hat{\mu}_0 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij} \tag{2.1}$$

dove X_{ij} è la *j-esima* osservazione dal sottogruppo i.
Per quanto riguarda la stima del parametro σ_0 vi sono tre possibili stimatori non distorti dati da

$$\hat{\sigma}_0 = \frac{\overline{R}}{d_2}, \hat{\sigma}_0 = \frac{\overline{S}}{c_4} \tag{2.2}$$

e

$$\hat{\sigma}_0 = \frac{S_p}{c_{4,m}} \tag{2.3}$$

dove $\overline{R}$ e $\overline{S}$ sono rispettivamente la media degli m range campionari e la media degli m scarti tipo, mentre S_p e $c_{4,m}$ sono pari a

$$S_p = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij} - \overline{X}_i\right)}{m(n-1)}} \tag{2.4}$$

e

$$c_{4,m} = \frac{\sqrt{2}\Gamma\left(\frac{m(n-1)+1}{2}\right)}{\sqrt{m(n-1)}\Gamma\left(\frac{m(n-1)}{2}\right)}. \tag{2.5}$$

I valori d_2, c_4, e $c_{4,m}$ sono funzioni della dimensione campionaria n, ($c_{4,m}$ è anche funzione di m). Gli stimatori dati nella (2.2) e (2.3) sono tali che $E[\overline{R}] = d_2\sigma$, $E[\overline{S}] = c_4\sigma$ e $E[S_p] = c_{4,m}\sigma$. Derman e Ross (1995) mostrano che lo stimatore S_p ha errore medio quadratico più piccolo rispetto allo stimatore usuale $\overline{S}/c_4$, dove $c_4 = c_{4,1}$ quando il campione casuale proviene da una distribuzione normale. Infatti è possibile dimostrare che

$$var[S_p/c_{4,m}] \leq var[\overline{S}/c_4] \leq var[\overline{R}/d_2] \tag{2.6}$$

per una ragionevole grandezza campionaria e con campioni provenienti da una distribuzione normale.

In questo studio verranno utilizzati gli stimatori descritti nelle formule (2.1) e (2.3).

2.3 Scelta dei parametri λ e h

Fissato l'ARL in controllo ed il valore del cambiamento δ che la carta deve individuare, attraverso i normogrammi di Crowder (Crowder, 1987b, 1989) viene scelto il valore ottimale di λ^*.

Per quanto riguarda il parametro h si utilizza qui una particolare funzione che calcola per ciascun insieme di campioni, in numero pari a m, ciascuno di ampiezza n, una stima della soglia h, utilizzando un metodo basato sull'approssimazione stocastica (Capizzi, Masarotto 1999).

2.3.1 Uso dell'approssimazione Stocastica

Il concetto di approssimazione stocastica fu introdotto da Robbins e Monro (1951) per lo studio del seguente problema. Si supponga di poter osservare, per ciascun punto $h \in \Re$, una variabile casuale Y, la cui distribuzione F dipende da h ed il cui valore atteso

$$M(h) = \int_{-\infty}^{+\infty} y dF_h(y) \tag{2.7}$$

esista, anche se su esso non vi è fatta alcuna assunzione parametrica.
L'obiettivo è quello di stimare quel valore di h, diciamo h^*, per il quale il valore atteso (2.7) risulti uguale ad un valore prefissato α che supporremo pari a zero. In particolare si osserva Y nei punti $h_1, h_2, h_3, \ldots$, in modo che y_n abbia distribuzione F_{h_n}, indipendente da $h_1, \ldots, h_{n-1}$
Nel cosiddetto *processo di Robbins e Monro*, i punti h_n, $n = 2, 3, \ldots$, vengono generati utilizzando la formula ricorsiva

$$h_{n+1} = h_n - a_n(y_n - \alpha) \tag{2.8}$$

con h_1 pari ad un opportuno valore iniziale. I coefficienti $\{a_n\}$ costituiscono una succesione di costanti positive che convergono a zero quando $n \to \infty$.
La procedura così costruita fa sì che, sotto opportune condizioni di regolarità su F_h, h_n converga in probabilità a h^*, quando $n \to \infty$ (*Robbin e Monro*, 1951).
Nel caso oggetto di studio si tratta di applicare tale approccio al problema del calcolo del parametro h a cui corrisponda un ARL in controllo pari a B.
Si tratta dunque di risolvere:

$$ARL(f^{(0)}, h) - B = 0 \tag{2.9}$$

dove $ARL(f^{(0)}, h)$ corrisponde all'ARL di un dato schema in cui non vi è stato alcun cambiamento nella distribuzione. Nei casi in cui dato $f^{(0)}$ ed h, sia possibile simulare la *Run Lenght*, T, di un certo schema, ossia il numero delle osservazioni sino alla prima volta in cui la statistica di controllo eccede il limite di controllo, si possono calcolare i valori di h, utilizzando la (2.8), e ponendo $Y = (h - B)/B$. In questo caso, $M(h) = [ARL(f^{(0)}, h) - B]/B$ e quindi il valore h^*, a cui converge il processo di Robbins e Monro coincide con la soluzione della (2.9). La differenza *(h-B)* è stata divisa per B per migliorare la velocità di convergenza dell'algoritmo. Per quanto riguarda a_n sono state considerate solamente successioni del tipo $a_n = \frac{A}{n}$.

2.3.2 Regola d'arresto

Al fine di interrompere la procedura descritta precedentemente non appena h_n risulta sufficientemente vicino a h^* viene usata la regola d'arresto ideata da *Stroup, Braun* (1982). Essa consiste nel simulare in ciascun passo, per uno stesso livello di h, due osservazioni indipendenti della *Run Lenght* $(y_n^{'}, y_n^{''})$, in modo che risulti

$$E[y_n^{'}|h_n] = E[y_n^{''}|h_n] = M(h_n).$$

Il processo di Robbins e Monro, per l'aggiornamento di h, viene modificato sostituendo a y_n, la media $\bar{y}_n = (y_n^{'} + y_n^{''})/2$. Il processo iterativo viene arrestato non appena le ultime k medie della *run length* risultano sufficientemente vicine a B e quindi

$$N = inf\{n \geq k : u_n^{(1)} = \sum_{i=n-k+1}^{n} \frac{\bar{y}_i^2}{ks_n^2} < b\} \tag{2.10}$$

dove k è un appropriato intero mentre

$$s_n^2 = \sum_{i=1}^{n} \frac{e_i}{n}$$

con

$$e_i = (y_i^{(1)} - \bar{y}_i)^2 + (y_i^{(2)} - \bar{y}_i)^2$$

costituisce uno stimatore corretto della varianza.

Per n sufficientemente grande, la statistica $2ku_n$ si distribuisce come una variabile X_k^2, quindi *2kb* può essere posto pari al percentile p-simo di un X_k^2, per vari valori di p.

Capitolo 3

I risultati

3.1 Introduzione

Il problema dell'applicabilità di una carta EWMA ad un processo di cui risulatano ignoti i parametri è già stato oggetto di svariati studi. Tra le più recenti trattazioni, (Jones, L. Allison, 2002), viene proprio considerato il disegno di una carta EWMA per processi con parametri ignoti. In tale articolo le stime del valore h vengono ottenute in modo esatto risolvendo, ripetto ad h, l'equazione integrale che fornisce l'ARL in controllo per una carta EWMA che utilizza parametri stimati. La risoluzione di tale problema risulta difficile, quindi viene effettuata un'approssimazione basata sulla quadratura Gaussiana, simile a quella descritta in Crowder(1987b). Un metodo numerico viene dunque utilizzato per trovare l'appropriata costante che faccia raggiungere alla carta l'ARL in controllo desiderato. Gli autori prendono in considerazione processi di diverse dimensioni, più precisamente composti da un numero di campioni m pari a 30, 50, 100 e 200, di diversa ampiezza (n pari a 1, 5, 7 e 10) ed in grado di ottenere differenti valori dell'ARL in controllo (pari a 100, 370, 500 e 1000).
Tutte le stime del valore h presentate nell'articolo sono maggiori ripetto ai valori della letteratura tradizionale, ma con l'aumentare del numero m di campioni tendono ad avvicinasi. Viene effettuato inoltre uno studio sull'ARL

fuori controllo: vengono confrontate le curve dell'ARL fuori controllo per un numero di campioni pari a 30, 100, 200 tutti di ampiezza n=5 con la curva dell'ARL del processo con i parametri noti. La carta con i veri parametri risulta la più veloce nel cogliere i cambiamenti nel processo ma aumentando il numero dei campioni si incrementa la sensibilità delle carte che utilizzano i parametri stimati.
Inoltre in tale articolo viene evidenziato che le carte costruite con tale metodo hanno un ARL fuori controllo leggermente più grande in quanto nei casi di un effettivo fuori controllo le statistiche, w_t, si posizionano sul nuovo valore del processo più lentamente. Questa riduzione di sensibilità è dovuta all'ampliamento dei limiti di controllo necessario per ottenere l'ARL in controllo stabilito.
In entrambi gli studi vengono considerate diverse dimensioni del campione da cui vengono stimati i parametri, in quanto si vuole valutare come all'aumentare di queste variabili cambi la stima del valore di $\bar{h}$. Con diminuzione del campione qui si intende il numero m di campioni, ed il numero n che va ad indicare la numerosità di ciascun campione.

3.2 La procedura per la stima di h

La procedura qui applicata è composta da tre parti fondamentali:

1. una prima parte genera un campione di dimensione m ed n, dal quale vengono stimati i parametri $\hat{\mu}$, $\hat{\sigma}$ (lobiettivo è quello di ricavare un insieme di dati del nostro processo in controllo di cui si ipotizza di non conoscere i parametri);

2. una seconda parte simula le osservazioni della *Run Length*: viene generata un osservazione con media pari a $\hat{\mu}$ e deviazione standard $\hat{\sigma}$ e la si confronta con il limite di controllo calcolato con l'$\hat{h}$ stimato. Il primo istante t in cui l'osservazione generata eccederà il limite di controllo sarà pari all'osservazione della *Run Length* ;

3. nella terza parte viene implementata la formula ricorsiva di Robbins e Monro e la regola d'arresto: dato un valore inziale di h si generano due osservazioni indipendenti della *Run Length* richiamando la (2), e con la media delle due osservazioni si effettua una successiva stima di h utilizzando la formula (2.8) fin tanto che la condizione d'arresto viene verificata.

Le tre parti vengono ripetute per un numero N di iterazioni ed al termine della sua esecuzione la procedura fornisce come output la lista degli N $\hat{h}$, il numero di osservazioni necessarie per giungere a quella stima, $\bar{h}$ e σ_h rispettivamente pari al valore medio degli $\hat{h}$ stimati e alla loro deviazione standard.
La procedura viene riportata interamente in Appendice ed è stata sviluppata utilizzando il software GAUSS.

3.3 I risultati

Vengono in particolare effettuate delle simulazioni per ottenere la stima delle costanti h per una carta EWMA in grado di cogliere un cambiamento pari a δ=1 e con un ARL in controllo pari a 500. I campioni di ampiezze m ed n sono stati simulati da una $N(0, 1)$ ed i valori delle variabili utilizzate nella procedura sono: h1=1, A=1.5 , k=250, p=0.5, δ=1, λ=0.13 e B=500 e con un numero di iterazioni pari a N=150. Le dimensioni dei processi generati in questo studio sono m 30, 50, 100 e 200 campioni, ciascuno a sua volta di ampiezza n 1, 5, 7 e 10; inoltre vengono affiancati i risultati ottenuti in questo studio con quelli riportati nell'articolo di Jones L. A. (2002).
I risultati sono riportati nella Tab. 3.1.
Nella letteratura tradizionale i parametri ottimali per una carta EWMA con le caratteristiche indicate precedentemente sono λ=0.13 e h=2.88. Come si può notare le stime sono tutte maggiori di 2.88 ma all'aumentare del numero m dei campioni queste tendono ad avvicinarsi a tale valore, mentre aumentando l'ampiezza del campione non vi è un andamento costante delle

Tabella 3.1:

m	**n**	$\bar{h}$ **o** μ_h	σ_h	h *Articolo*
30	1	3.17	0.77267	/
	5	3.16	0.72015	3.03
	7	3.13	0.74036	3.05
	10	3.20	0.75363	3.06
50	1	3.04	0.53892	2.88
	5	3.10	0.60406	3.00
	7	3.18	0.67200	3.02
	10	3.05	0.55756	3.02
100	1	2.98	0.41094	2.91
	5	2.99	0.40182	2.96
	7	3.04	0.44538	2.97
	10	2.97	0.38393	2.97
200	1	2.96	0.27415	2.91
	5	2.99	0.40182	2.93
	7	2.95	0.28709	2.94
	10	2.95	0.28844	2.94

stime. Per processi di piccole dimensioni si ottiene una stima dei parametri, $(\hat{\mu}, \hat{\sigma})$, meno esatta in quanto disponendo di pochi dati la stima dei parametri risulta meno precisa. Tenere conto della maggiore variabilità delle stime dei parametri porta ed un aumento del valore di h e dunque a dei limiti di controllo più ampi con una conseguente dimunuzione del numero dei falsi allarmi. All'aumentare delle dimensioni, (m, n) del campione, la stima di h diminuisce, in quanto le stime dei parametri del processo risultano più precise, con conseguente restringimento dei limiti di controllo. Si nota inoltre come la deviazione standard delle stime di h diminuisca all'aumentare della dimensione del processo. Questo ci porta a concludere che aumentando sia il numero di campioni a disposizione sia la loro ampiezza si ottiene una stima del valore h più precisa.

Le stime ottenute risultano leggermente più grandi rispetto alle stime dell'articolo ma la differenza diventa minima all'aumentare della dimensione dell'intero campione . Questo scostamento può essere imputato ad una minore efficienza del metodo proposto per numerosità campionarie piccole.

3.4 Conclusioni

Possiamo concludere che l'utilizzo dell'approssimazione stocastica come metodo per il disegno di una carta EWMA con parametri del processo incogniti fornisce buoni risultati, utilizzando un procedimento più semplice in termini di implementazione. Inoltre si ribadisce l'importanza della dimensione dell'intero campione da cui si stimano i parametri e quindi si raccomanda un utilizzo più accorto di questo metodo in quei casi in cui si voglia disegnare lo schema EWMA con poche osservazioni a disposizione.

Appendice A

Appendice

Di seguito verranno riportate le procedure utilizzate per ottenere i risultati presentati in questo studio ed un elenco dei parametri di input necessari.

lambda: valore di λ scelto per la carta Ewma
h1: valore inziale per il calcolo di h
A: costante per il calcolo del coefficiente a_n nel processo di Monroe
seq: corrisponde al valore k
w: pari al valore p
B: corrisponde all'ARL IC
nSim: numero di simulazioni
m: numero di campioni
n: dimensione dei campioni
mu: valore della media del processo IC
s0: valore della deviazione standard del processo IC

Procedura ValoreH

```
output file=NomeFile.txt  ;
output on;

proc (2)= par(m,n,mu,s0);
local  dati,s,sp,c4,sqns,media,medie,i,j,a,b;
dati=rndn(m,n)*s0+mu;
if (n==1);
media=meanc(dati);
s=0;
i=1;
do while i <= m;
     j=1;
     do while j <= n;
          s=s+(dati[i,j]-media)^2;
j=j+1;
endo;
i=i+1;
endo;
sqns=sqrt(s/(m-1));
else;
medie=meanc(dati');
media=meanc(medie);
s=0;
i=1;
do while i <= m;
     j=1;
     do while j <= n;
          s=s+(dati[i,j]-medie[i,1])^2;
j=j+1;
endo;
```

```
i=i+1;
endo;
sp=sqrt(s/(m*(n-1)));
a=(m*(n-1)+1)/2;
b=(m*(n-1))/2;
c4=(sqrt(2)/sqrt(m*(n-1)))*exp(lnfact(a-1)-lnfact(b-1));
sqns=sp/c4;
endif;
retp(media,sqns);
endp;

proc runl(media,lambda,sqns,h,m,n);
local y,hl,ns1,x;
y=0;
x=0;
hl=media+(h*sqrt(lambda/(2-lambda))*sqns/sqrt(n));
do while abs(x) <=hl;
ns1=media+rndn(1,1)*(sqns/sqrt(n));
x=(1-lambda)*x+lambda*ns1;
y=y+1;
if y >= 10000;
        break;
endif;
endo;
retp(y,);
endp;

proc (2)= hNss(lambda,h1,A,seq,w,B,m,n,media,sqns);
local u2,u1,u,j,h, sume,ybar,ybar2,y1,y2,v;
j=1;
h=h1;
```

```
sume=10.0e-255;
ybar2=zeros(seq,1);
u=2;
do while u > 1;
y1=(runl(media,lambda,sqns,h,m,n)-B)/B;
y2=(runl(media,lambda,sqns,h,m,n)-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));
if j > seq;
u=u2;
endif;
j=j+1;
endo;
retp(h,j-1);
endp;

proc prova(lambda,h1,A,seq,w,B,nSim,m,n,mu,s0);
local h,i,x,p1,p2,str,medie,std,c;
h=zeros(nSim,2);
i=1;
```

```
c=ones(nSim,2);
do while i <= nSim;
{c[i,1],c[i,2]} =par(m,n,mu,s0);
{h[i,1],h[i,2]}=hnss(lambda,h1,A,seq,w,B,m,n,c[i,1],c[i,2] ) ;
i=i+1;
endo;
output on;
locate 10,1;
x=seqa(1,1,nSim);
format /m1/rdn 12,2;
medie=ones(1,1)|meanc(h);
std=ones(1,1)|stdc(h);
p1=x~h;
p2=p1|(medie')|(std');
print c;
print "lambda" lambda  "h1" h1  "A" A   "seq" seq  "w" w;
print "B" B  "nSim" nSim  "m" m  "n" n  ;
str={ NSim, h,  j };
$str';
format /m1/rdn 12,5;
print p2;
output off;
retp(h);
endp;

{h}=prova(lambda,h1,A,seq,w,B,nSim,m,n,mu,s0);
output off;
```

Bibliografia

[1] 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.

[2] CROWDER, S. W. (1987b), 'A simple method for studying run-length distribution of exponentially weighted moving average charts', *Techometris* 29.

[3] CROWDER, S. W. (1989), 'Designing exponentially weighted moving average schemes', *Journal of Quality Tecnology* 21.

[4] DERMAN, C., e ROSS, S. (1995), 'An Improved Estimator of σ in Quality Control', *Probability in the Engineering and Informational Sciences*, 9, pp. 411-415.

[5] LUCAS, J. M. e SACCUCCI, M. S. (1990), 'Exponentially weighted moving average control schemes: Properties and enhancements', *Techometris* 32.

[6] JONES, L. ALLISON (2002). "The Statistical Design of EWMA Control Charts with Estimated Parameters ". *Techometris*, 34, pp.277-288.

[7] JONES, L. ALLISON; CHAMP, C. W.; and RIGDON, S. E. (2001). "The Performance of Exponentially Weighted Moving Everage Charts with Estimated Parameters ". *Techometris*, 43, pp.156-167.

[8] ROBERTS, S. W. (1959). "Control Charts Tests Based on Geometric Moving Averages ". *Techometris*, 1, pp.239-250.

[9] RONCALLI, Par Thierry (1995). *Introduction á la programmation sous GAUSS Volume 1- Mèthodes Numèriques en Mathematiques et Statisques*, Ritme Informatique, Parigi.

Errata Corriege

Bellin Erika

Disegno di una carta EWMA
con parametri del processo incogniti

- La formula (2.4) a pagina 10 è pari a

$$S_p = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_i\right)^2}{m(n-1)}}$$

e non

$$S_p = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_i\right)}{m(n-1)}}.$$

- A pagina 16, riga 15, la frase
"Con dimunzione del campione qui si intende...." è errata va sostituita con
"Con dimensione del campione qui si intende...".