Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica e Tecnologie Informatiche

relazione finale

LA METODOLOGIA SELF-STARTING IN AMBITO UNIVARIATO: CONFRONTO FRA UNA CARTA PARAMETRICA E UNA NON PARAMETRICA

Relatore: Prof. Giovanna Capizzi
Dipartimento di: Scienze Statistiche

Laureando: Leonardo Gulino
Matricola: 1010813

Anno Accademico 2012/2013

Indice

Introduzione

Le aziende moderne puntano ad essere competitive economicamente a livello mondiale, cercando di ottenere la massima efficienza con dei costi piuttosto bassi; il fine principale è quello di raggiungere alti livelli di *qualità*, utilizzando tecniche statistiche adeguate per sorvegliare i processi produttivi e verificarne l'effettiva regolarità, analizzando non soltanto il risultato finale, ma ogni singolo stadio della produzione.

Il controllo statistico della qualità consiste nello studiare la stabilità dei parametri che caratterizzano la distribuzione di una caratteristica di qualità. Lo strumento statistico che consente di monitorare il processo produttivo è la carta di controllo, attraverso la quale viene segnalata la presenza di fonti di variazione diverse da quelle comuni.

Le classiche carte di controllo sono disegnate assumendo che i parametri del processo siano noti o debbano essere stimati utilizzando un gran numero di campioni di FASE I (Quesenberry, 1993), ovvero quei campioni ottenuti con un primo ciclo di produzione, assunto in uno stato di controllo statistico, ovvero influenzato dalle sole fonti di variazione comune.

Sono presenti, però, alcuni processi dove non è possibile raccogliere un gran numero di campioni per ottenere delle stime accurate dei parametri in controllo. Nelle applicazioni produttive di *start-up* e *short-run* non si dispone di un numero elevato di misurazioni per ricavare delle stime affidabili. In particolare, con *start-up* si intende il periodo nel quale viene avviata un'impresa, mentre con *short-run* vengono identificati i processi in cui si avvia una produzione limitata da effettuare nel modo più rapido possibile, analizzando quindi un numero piccolo di osservazioni. Per ovviare a quanto detto, sono state create delle carte di tipo *self-starting* (Hawkins, 1987 e Quesenberry,

1991) che permettono di aggiornare i parametri stimati, verificando simultaneamente le condizioni di fuori controllo del processo. Partendo quindi da un numero minimo di 3 osservazioni, vengono stimati i parametri e ad ogni nuova misurazione rilevata vengono aggiornate le stime, controllando che il processo non risulti fuori controllo.

In varie applicazioni industriali di *start-up* e *short-run*, però, la distribuzione sottostante del processo non è nota, per mancanza di un adeguato numero di osservazioni; ciò compromette le proprietà statistiche delle carte di controllo self-starting parametriche. In questi casi risulta utile implementare carte di controllo di tipo non parametrico che presentano il vantaggio di non assumere un'ipotesi distributiva di partenza.

Questa tesi si propone, in particolare, di presentare due carte di tipo self-starting, una parametrica introdotta da Capizzi e Masarotto (2010) e una non parametrica, proposta da Liu et al. (2012), e di mostrarne un confronto tramite l'uso di profili ARL. Nel capitolo 1, dopo una breve introduzione ai principali aspetti dell' SPC, viene esposta una breve rassegna delle carte self-starting univariate di Hawkins (1987) e Quesenberry (1991), introducendo la distribuzione della Run Length (RL) di queste carte esposta da Zantek (2004). Verranno trattate, in particolare, le carte Q-Shewhart, Q-Cusum e Q-Ewma. Il capitolo 2 presenta una carta self-starting univariata parametrica, l'Adaptive CUSCORE (ACUSCORE), che combina una semplice carta CUSCORE (Box e Ramirez, 1992) con una Adaptive EWMA, suggerita da Capizzi e Masarotto (2012). Verrano poi esposti gli aspetti pratici per il disegno di questa carta, con la scelta delle varie costanti, presentando i valori dei limiti di controllo a seconda di un'ARL in controllo prefissata.

Nel capitolo 3 viene esposto il disegno di una carta self-starting univariata non parametrica (NAE), basata sui ranghi standardizzati, introdotta da Liu et al. (2013); verranno presentati gli aspetti teorici della carta e gli aspetti pratici per il suo disegno. Nel quarto e ultimo capitolo, viene effettuato un confronto fra le due carte simulando gli ARL fuori controllo di 3 diverse distribuzioni (Normale, Gamma e t di Student). La statistica Acuscore è una carta disegnata assumendo che le osservazioni siano normali indipendenti ed identicamente distribuite (i.i.d.); tramite questo confronto si vuole verificare quanto la sua performance venga alterata dalle violazioni degli assunti distributivi, e quanto sia meglio utilizzare, in questi casi, una carta self-starting non parametrica.

Per questo confronto sono state implementate e utilizzate delle funzioni col software statistico *R*, riportate in Appendice.

Capitolo 1

Il Self-starting univariato

1.1 Il controllo statistico di processo (SPC)

Alla base di ogni processo ci sono delle caratteristiche di qualità, delle particolari proprietà che permettono di valutarne il rendimento. Il **controllo statistico del processo** (SPC, dall'inglese *Statistical Process Control*) rappresenta un insieme di tecniche statistiche applicate durante il processo di lavorazione per ottenere informazioni sullo stato del processo, in modo tale da controllare se la caratteristica di qualità assuma andamenti indesiderati non imputabili alla sola variabilità casuale (ad esempio difetti della macchina o errori degli operai). Durante la produzione deve quindi essere sorvegliata la distribuzione di queste caratteristiche, facendo sì che i parametri di ciascuna di esse non si discostino troppo da dei parametri *target*, ovvero dei valori ideali che corrispondono ad un desiderato rendimento del processo. L'SPC ci permette, quindi, di migliorare la qualità del processo riducendone la variazione rispetto al target. Variazioni casuali/comuni, comportano solitamente piccoli scostamenti dal target.

Quando i valori dei parametri delle caratteristiche prese in considerazione sono prossimi a quelli dei target, il processo viene definito *in controllo* (IC); altrimenti, il processo è detto *fuori controllo* (FC), e ci indica la presenza di cause "speciali" che creano una variabilità non imputabile a cause accidentali. La presenza di tali fonti di variazione altera la distribuzione delle caratteristiche di qualità e non dipende dal funzionamento naturale del processo.

Lo strumento base dell'SPC che permette di mantenere sotto controllo la variazione del processo è la **carta di controllo**, introdotta per la prima volta da Shewhart (1924). Nella letteratura sono presenti vari tipi di carte di controllo:

- carte per variabili o per attributi, a seconda che la caratteristica di qualità sia di tipo continuo o discreto;

- carte che sorvegliano media, varianza o altri parametri della distribuzione della caratteristica di qualità;

- carte di tipo parametrico e non parametrico, a seconda che siano noti o meno gli assunti distributivi della caratteristica di qualità.

- carte senza memoria (ad es. Shewhart) e con memoria (ad es. CUSUM e EWMA), a seconda che si valuti la qualità del processo basandosi soltanto sull'ultimo campione estratto, oppure considerando anche i precedenti.

Una carta di controllo consiste, alla fine, in una serie consecutiva di test di verifica del seguente sistema di ipotesi:

$$\begin{cases} H_0: & \text{processo in controllo} \\ H_1: & \text{processo fuori controllo} \end{cases} \tag{1.1}$$

Se ad esempio stiamo monitorando la media di un processo, questo sistema di ipotesi si traduce in:

$$\begin{cases} H_0: \mu = \mu_0 \quad \forall t \\ H_1: \begin{cases} \mu = \mu_0 & t \leq \tau \\ \mu = \mu_1 & t > \tau \end{cases} \end{cases} \tag{1.2}$$

dove τ rappresenta l'istante di tempo in cui la media ha subito un cambiamento da μ_0 a μ_1. Per verificare tali ipotesi, vengono calcolati un **limite di**

controllo inferiore (LCL, dall'inglese *Lower Control Limit*), un **limite di controllo superiore** (UCL, dall'inglese *Upper Control Limit*) e una **statistica di controllo**, che varia a seconda della carta scelta e che tiene conto del parametro che si vuole sorvegliare. L'ipotesi nulla viene accettata se il valore della statistica di controllo rientra nell'intervallo *decisionale* dato dai limiti di controllo, in alternativa viene segnalata la presenza di cause speciali di variazione.

Siamo quindi in presenza di un approccio sequenziale, in cui si continua a campionare fintanto che l'ipotesi nulla è vera, fermandosi non appena la statistica risulta fuori dall'intervallo, indicandoci così la presenza di un campione fuori controllo. L'istante di tempo in cui viene indentificato il primo fuori controllo viene detto **Run Length** (RL); la RL identifica il numero di osservazioni campionate fino alla segnalazione di un allarme. In particolar modo, indicando con t l'istante di tempo e w_t la statistica di controllo, $RL = \inf\{t : w_t \notin (LCL, UCL)\}$.[1]. Possiamo, inoltre, distinguere fra:

- $RL_0 = \inf\{t : w_t \notin (LCL, UCL)|H_0\}$, ovvero la run length in controllo, che rappresenta il primo istante di tempo in cui la statistica non appartiene all'intervallo decisionale quando H_0 vera. RL_0 ci indica quindi il tempo di attesa fino alla segnalazione di un falso allarme.[2]

- $RL_1 = \inf\{t : w_t \notin (LCL, UCL)|\bar{H}_0\}$, ovvero la run length fuori controllo, che indica il tempo di attesa fino alla segnalazione di un vero allarme.

In generale, la RL rappresenta una misura di efficienza della nostra carta di controllo, permettendoci di valutarne l'effettiva funzionalità.

Con **Average Run Length** (ARL) si intende il tempo medio di attesa per segnalare un allarme e coincide con il valore atteso della RL. Anche in questo caso possiamo distinguere tra ARL_0 e ARL_1, che rappresentano quindi i tempi medi di attesa per la segnalazione rispettivamente di un falso e di un vero allarme. Una carta di controllo è *efficiente* se segnala raramente

[1] Si considera la RL di una carta Shewhart

[2] Infatti se H_0 è vera, vuol dire che in realtà il processo è in controllo, e quindi la nostra statistica di controllo appartiene all'intervallo decisionale.

un falso allarme, avvisando il più presto possibile la presenza di un effettivo fuori controllo.

1.2 Statistiche Q e carta Q-Shewhart

Si assuma che, quando il processo è in controllo, le osservazioni $X_1, X_2, \ldots, X_i$ del processo siano variabili casuali normali indipendenti con media μ e varianza σ^2, quindi

$$X \sim N(\mu, \sigma^2) \tag{1.3}$$

Consideriamo, per il momento, i parametri noti e definiamo la statistica standardizzata

$$Z_i = \frac{(x_i - \mu)}{\sigma} \tag{1.4}$$

Con la classica carta di controllo di Shewhart, viene segnalato un cambiamento in media quando $Z_i > UCL$ o $Z_i < LCL$.

Supponiamo ora che i parametri siano non noti e che, a partire da un istante di tempo τ, il processo subisca una deviazione dalla situazione di controllo nella forma di uno shift di μ. Definiamo:

1. m_i : media campionaria delle prime i osservazioni, ottenuta dalla formula ricorsiva

$$m_i = m_{i-1} + \frac{1}{i}(x_i - m_{i-1}) \tag{1.5}$$

2. s_i^2 : varianza campionaria delle prime i osservazioni, ottenuta dalla formula ricorsiva

$$s_i^2 = s_{i-1}^2 + \frac{1}{i}\left((x_i - m_{i-1})^2 - \frac{i}{i-1}s_{i-1}^2\right) \tag{1.6}$$

dove $m_2 = (x_1 + x_2)/2$ e $s_2^2 = (x_1 - x_2)^2/2$. Tutto ciò vale se si hanno almeno 3 osservazioni di partenza, e quindi se $i \geq 3$. Ponendo $a_i = \sqrt{\frac{i-1}{i}}$, possiamo definire la statistica:

$$T_i = a_i \frac{(x_i - m_{i-1})}{s_{i-1}} \sim t_{i-2} \tag{1.7}$$

La statistica 1.4 ci permette di verificare l'ipotesi nulla che l'osservazione corrente x_i e il campione precedente $(x_1, x_2, \ldots, x_{i-1})$ provengano dalla stessa distribuzione di media ignota. Ciascuna T_i risulta statisticamente indipendente dalle precedenti e segue una distribuzione t di student con $i-2$ gradi di libertà. Questa statistica ci permette, quindi, di verificare se x_i ha la stessa media non nota dei precedenti $i-1$ campioni.

Si può notare che i gradi di libertà della statistica sopra descritta dipendono dall'istante di tempo i a cui ci si riferisce; ciò implica che le T_i non sono identicamente distribuite. Per ovviare a questo problema, Quesenberry (1991) e Hawkins (1987) introducono le statistiche Q, delle trasformazioni delle originali osservazioni in una nuova sequenza di osservazioni indipendenti ed identicamente distribuite (i.i.d.) come una normale standard.

Indichiamo con:

- G_v la funzione di ripartizione di una variabile casuale t di Student con v gradi di libertà;
- ϕ^{-1} la funzione quantile di una normale standard.

Possiamo ora definire la quantità:

$$Q_i = \Phi^{-1}[G_{i-2}(T_i)], \qquad i \geq 3 \tag{1.8}$$

Poichè, quando il processo è in controllo, le Q_i sono variabili casuali i.i.d provenienti da $N(0,1)$, possono essere utilizzate tutte le tradizionali carte di controllo, basate sulle Q_i, per identificare gli shift della media del processo da un valore μ ad un valore $\mu + \delta\sigma$. La carta di controllo Q-Shewhart è una carta Shewhart basata sulle statistiche Q_i; perciò, la carta segnala un fuori controllo quando $Q_i >$ UCL o $Q_i <$ LCL.

Si noti che la statistica Q_i è definita per $i \geq 3$; così, invece di attendere un campione sufficientemente grande, come quello di FASE I, Q_i può essere calcolata e rappresentata tramite una Shewhart non appena sono disponibili tre osservazioni. E' inoltre interessante evidenziare come la carta faccia uso di ciascuna nuova osservazione. In particolare, la carta calcola la statistica T_i dell'equazione (1.4), che misura la compatibilità della corrente i-esima osservazione con le precedenti, trasformandola nella statistica Q_i, da porre a confronto con i limiti di controllo. Se la statistica sta all'interno dell'intervallo decisionale, l'i-esima osservazione viene utilizzata per aggiornare la stima della media e della deviazione standard.

1.3 Probabilità di segnalare fuori controllo

Con probabilità di segnalare un fuori controllo si intende la probabilità che la statistica Q_i non appartenga all'intervallo decisionale, ovvero:

$$1 - Pr(LCL < Q_i < UCL) = 1 - F_i(UCL) + F_i(LCL) \tag{1.9}$$

dove $F_i(y)$ corrisponde alla funzione di ripartizione di Q_i nel punto y, non condizionata alle precedenti statistiche (Q_{i-1}, Q_{i-2},...).

Dalla (1.6), possiamo scrivere la funzione di ripartizione di Q_i come:

$$\begin{aligned} F_i(y) = Pr[\Phi^{-1}[G_{i-2}(T_i)] \leq y] = \\ = Pr[T_i \leq G_{i-2}^{-1}[\Phi(y)]] = \\ = \tilde{F}_i[G_{i-2}^{-1}[\Phi(y)]] \end{aligned} \tag{1.10}$$

dove $\tilde{F}_i(\cdot)$ è la funzione di ripartizione di T_i, $G_{i-2}^{-1}(\cdot)$ è la funzione quantile di una t di Student e $\Phi(\cdot)$ è la funzione di ripartizione di una Normale standard.

Possiamo quindi riscrivere la (1.10) come

$$Pr_i = 1 - \tilde{F}_i[G_{i-2}^{-1}[\Phi(UCL)]] + \tilde{F}_i[G_{i-2}^{-1}[\Phi(LCL)]] \tag{1.11}$$

Si può notare che, se la media del processo non ha subito variazioni dal suo valore iniziale μ, la funzione $\tilde{F}_i(\cdot)$ coincide semplicemente con la funzione di ripartizione di una t di Student con $i-2$ gradi di libertà, e la probabilità (1.11) si riduce a $Pr_i = 1 - \Phi(UCL) + \Phi(LCL)$, ovvero la stessa probabilità della classica carta di controllo della Shewhart.

Supponiamo ora che la media delle X_i passi da μ a $\mu + \delta$ a partire dalla $v - esima$ osservazione. Come risultato del cambiamento, il valore atteso del numeratore di $T_v = a_v \frac{(x_v - m_{v-1})}{s_{v-1}}$ passa da 0 a $a_v\delta$ è la distribuzione di T_v diventa una t di Student non centrale con $v-2$ gradi di libertà e parametro di non-centralità pari a $a_v\delta/\sigma$ (Hawkins, 1987); ciò ci permette di calcolare la probabilità (1.11) utilizzando la funzione di ripartizione di una t di Student non centrale al posto di $\tilde{F}_i(\cdot)$, ottenendo quindi una Pr_v che dipende soltanto da v, ovvero dall'istante in cui si ha avuto il cambiamento, e da δ/σ. Nel caso di limiti di controllo simmetrici (LCL= -UCL), Pr_v aumenta all'aumentare di v o dell'ampiezza di cambiamento $|\delta/\sigma|$; la media campionaria del numeratore delle statistiche successive all'istante in cui si ha avuto il cambiamento (T_{v+1}, $T_{v+2} \dots$), è aggiornata con osservazioni ottenute dopo lo shift.

Vediamo in particolare quanto vale il valore atteso del numeratore di T_i, con $i = v, v+1 \dots$:

$$\begin{aligned} E(a_i(x_i - \bar{x}_{i-1}) &= a_i[\mu + \delta - \mu - (i-v)\delta/(i-1)] \\ &= \delta(v-1)/\sqrt{i(i-1)} \\ &= \sigma\theta_i \end{aligned} \tag{1.12}$$

dove $\theta_i = \delta(v-1)/\sigma\sqrt{i(i-1)}$. Si noti che $a_i(x_i - \bar{x}_{i-1})$ è normalmente distribuito con media $\sigma\theta_i$ e varianza σ^2, in quanto combinazione lineare di variabili normali.[3] Questo valore atteso si avvicina a 0 al crescere di i, e ciò porta le statistiche T_i e Q_i a tendere a 0, diminuendo la probabilità Pr_i di segnalare un fuori controllo; questa riduzione del valore di Pr_i è inoltre dovuto al fatto che il denominatore di T_i, lo stimatore della deviazione standard, viene anch'esso aggiornato dalle nuove osservazioni prese dopo il cambiamento, e questo aggiornamento tende a incrementare il valore di s_{i-1}, diminuendo di conseguenza quello delle T_i.

Consideriamo adesso la distribuzione del denominatore delle T_i, e quindi della varianza campionaria; utilizzando la formula ricorsiva proposta nella (1.7), possiamo scrivere

$$(v-1)s_v^2 = (v-2)s_{v-1}^2 + a_v^2(x_v - \bar{x}_{v-1})^2 \tag{1.13}$$

[3]Si ricorda che le X_i sono variabili inizialmente distribuite come una $N(\mu, \sigma^2)$ e, dopo lo shift, come una $N(\mu + \delta, \sigma^2)$.

dove s_{v-1}^2 è la varianza campionaria relativa alle osservazioni rilevate prima del cambiamento in media e, poichè $x_1, x_2, \ldots, x_{v-1}$ sono variabili normali indipendenti e identicamente distribuite, segue che $(v-2)s_{v-1}^2 \sim \sigma^2\chi_{v-2}^2$; il termine $a_v^2(x_v - \bar{x}_{v-1})^2$ si distribuisce invece come σ^2 volte un chi-quadro non centrale con parametro di non-centralità pari a θ_v^2 (infatti $a_v(x_v - \bar{x}_{v-1})$ è una $N(\sigma\theta_v, \sigma^2)$). Essendo i due termini stocasticamente indipendenti possiamo affermare che $(v-1)s_v^2 \sim \sigma^2\chi_{v-1}^2(\theta_v^2)$. Ripetendo quanto detto per le successive osservazioni, troviamo che

$$(i-1)s_i^2 \sim \sigma^2\chi_{i-1}^2(\textstyle\sum_{j=v}^{i}\theta_j^2)$$

La statistica T_i può essere riscritta come il rapporto di due variabili indipendenti $(a_i(x_i - \bar{x}_{i-1}))/\sigma$ e $\sqrt{(i-2)s_{i-1}^2/(i-2)\sigma^2}$. Segue che il numeratore è normalmente distribuito con media θ_i e varianza pari a 1; il denominatore è distribuito come la radice quadrata di un chi-quadro non centrale con $i-2$ gradi di libertà e parametro di non-centralità pari a $\lambda_i = \sum_{j=v}^{i-1}\theta_j^2$. La statistica T_i ha quindi una distribuzione t di Student doppiamente non centrale con $i-2$ gradi di libertà e parametri di non-centralità θ_i e λ_i, per $i = v+1, v+2, \ldots$; in particolare, il secondo parametro di non-centralità riflette l'influenza dello stimatore della deviazione standard dovuto al cambiamento in media del processo. Questi risultati distributivi ci permettono di calcolare esattamente le probabilità $Pr_{v+1}, Pr_{v+2}, \ldots$ della (1.9), utilizzando la funzione di ripartizione di una t di student doppiamente non centrale. Per calcolarla, si ripropone il procedimento proposto da Kocherlakota e Kocherlakota nel 1991. Indichiamo con $H[u; m, \theta_i]$ la funzione di ripartizione di una t di student non-centrale con m gradi di libertà e parametro di non centralità θ_i; possiamo allora esprimere la funzione di ripartizione di T_i come:

$$\tilde{F}_i(\omega) = \sum_{\alpha=0}^{\infty} \tau_\alpha(\lambda_i/2)H(u; i+2(\alpha-1), \theta_i) \qquad i = v+1, v+2, \ldots \quad (1.14)$$

dove

$$\tau_\alpha(\beta) = \tfrac{\beta^\alpha e^{-\beta}}{\alpha!} \qquad u = \omega\sqrt{\tfrac{1+2\alpha}{i-2}}$$

Tramite l'equazione (1.11) riusciamo quindi a calcolare la funzione di ripartizione di una t di Student doppiamente non centrale utilizzando quella per una singola t di Student non centrale.

1.4 La distribuzione della Run Length

Come abbiamo anticipato nell'introduzione, quando il processo è in controllo, la RL rappresenta il numero di osservazioni prima che la carta Q-Shewhart segnali un falso allarme. Poichè le statistiche Q_i sono variabili indipendenti normali standard quando il processo è stabile, la RL ha distribuzione geometrica di parametro $\pi = 1 - \Phi(UCL) + \Phi(LCL)$ (come nel caso di una semplice carta Shewhart con μ e σ^2 noti). Quando il processo subisce un cambiamento in media, la RL coincide con il numero di osservazioni precedenti allo shift. La prima statistica dopo il cambiamento , T_v, è stocasticamente indipendente da ciascuna delle precedenti statistiche $T_3, T_4, \ldots, T_{v-1}$; la probabilità esatta che la RL sia uno è quindi $Pr(RL = 1) = Pr_v$, dove Pr_v corrisponde alla probabilità definita nella (1.9).

Consideriamo ora l'evento che la statistica Q_i sia fuori dai limiti di controllo, e indichiamo tale evento con A_i. Il fatto che le statistiche $Q_v, Q_{v+1}, \ldots$ non siano variabili casuali indipendenti implica che A_i e A_j non sono indipendenti fra loro, per ogni $i, j \geq v$; questa dipendenza rende più complicato trovare l'esatta espressione delle probabilità che la RL assuma valori maggiori di uno, in quanto ciascuna di queste probabilità è la probabilità congiunta che la statistica Q_i sia fuori dai limiti di controllo e che le precedente statistiche Q siano dentro i limiti di controllo. Tuttavia, le statistiche $Q_v, Q_{v+1}, \ldots$ sono approssimativamente indipendenti, e possiamo quindi ricavare un'approssimazione accurata per le probabilità delle RL trattando le statistiche Q_i come se fossero indipendenti. Assumendo l'indipendenza, la distribuzione della RL è definita dall'equazione (1.9); ricaviamo quindi che

$$Pr(RL = r) = \begin{cases} Pr_v & r = 1 \\ Pr_{v+r-1} \prod_{i=v}^{v+r-2} (1 - Pr_i) & r = 2, 3, \ldots \end{cases} \tag{1.15}$$

La probabilità espressa nella (1.15) è esatta per $r = 1$ e approssimata per $r = 2, 3, \ldots$.

1.5 Carte Q-CUSUM e Q-EWMA

Finora abbiamo descritto una carta Q di tipo Shewhart; essa è una carta senza memoria che risulta molto utile in presenza di cambiamenti in media molto grandi. Non tenendo conto delle osservazioni precedenti, la Q-Shewhart ha bisogno, infatti, che l'osservazione sia molto grande per segnalare l'allarme. Se, invece, si presentano piccoli cambiamenti, la carta continua a considerare il processo in controllo lanciando mediamente in ritardo un segnale di cambiamento..

Quando si è in presenza di processi in cui ci si aspettano piccoli cambiamenti, risulta più utile utilizzare carte con memoria come la CUSUM (*Cumulative Sum control chart*) o la EWMA (*Exponentially-Weighted Moving Average*) (Page, 1954). In particolare la EWMA ha la peculiarità di rilevare in modo efficiente il fuori controllo sia nel caso di shift piccoli che nel caso di shift grandi, grazie ad un'opportuna scelta della costante di lisciamento.

La statistica di una classica carta di controllo EWMA è infatti così definita:

$$Z_i = \lambda g(x_i) + (1 - \lambda) Z_{i-1} \tag{1.16}$$

dove λ è una peso che viene assegnato all'osservazione e corrisponde ad un valore compreso fra 0 e 1. Nel dettaglio, se $\lambda \to 1$, vuol dire che si dà maggior peso alle ultime osservazioni, e quindi la EWMA tende a comportarsi come una Shewhart, risultando utile nel caso ci si aspettino grandi cambiamenti in media; se invece $\lambda \to 0$, il peso viene distribuito in tutte le osservazioni e la EWMA si comporta come una CUSUM, permettendo di segnalare subito il fuori controllo anche nel caso di piccoli cambiamenti.

Quando si è in situazioni di *short-runs* o start-up e si reputa necessario utilizzare delle carte per rilevare piccoli cambiamenti, può tornare utile usare una carta Q-CUSUM o una carta Q-EWMA. In particolare, una carta self-starting CUSUM delle Q_i (Hawkins, 1987 et al.) è definita, sempre per $i \geq 3$

tramite le seguenti statistiche:

$$\begin{cases} C_i^L = \min\{0, C_{i-1}^L + Q_i + k\} \\ C_i^U = \max\{0, C_{i-1}^U + Q_i - k\} \end{cases} \tag{1.17}$$

con $C_2^L = C_2^U = 0$ e k che rappresenta il valore di riferimento della CUSUM, scelto in funzione di un'ampiezza del cambiamento che si vuole determinare.

Una carta self-starting EWMA delle Q_t è data da:

$$Z_i = Z_{i-1} + \lambda(Q_i - Z_{i-1}) \tag{1.18}$$

dove $Z_2 = 0$ e $0 < \lambda \leq 1$.

Capitolo 2

La carta parametrica Adaptive CUSCORE

I tradizionali schemi di controllo, come la CUSUM o la EWMA, sono disegnati assumendo che il cambiamento nelle osservazioni originali risulti un cambiamento significativo nelle statistiche Q_i basate su di esse. In ogni caso, dopo il momento dell'avvenuto cambiamento, la media di Q_i risulta essere, come già anticipato, una funzione dipendente dal tempo; in particolare, è possibile dimostrare che il modello che descrive il cambiamento della media è sostanzialmente influenzato dal valore assunto da δ (l'ampiezza del cambiamento) e da τ (l'istante di tempo in cui si ha avuto lo shift). Quando si ha un cambiamento della media del processo, c'è un immediato e rapido cambiamento del valore della media di Q_i; questo valore atteso, comunque, tende a zero non appena le osservazioni successive allo shift vengono utilizzate per aggiornare la stima della media, e quindi al crescere di i. Così, una carta di controllo che non riesce a segnalare entro poche osservazioni successive al cambiamento, presenta perfomance insoddisfacenti nel segnalare un fuori controllo.

Appare quindi auspicabile disegnare una carta di controllo self-starting utilizzando non soltanto il livello di Q_i, ma anche la *fault signature* del cambiamento della media (ovvero la correlazione fra lo shift e la media dei residui), e in particolare l'informazione che si ha dalla correlazione fra le Q_i

e il suo valore atteso. Quando la *fault signature* è nota, questo problema può essere risolto utilizzando una carta di controllo CUSCORE (*Cumulative SCORE chart*), come proposto da Box e Ramirez (1992); sfortunatamente, il valore atteso delle Q_i dipende dai valori ignoti di δ e τ. Per ovviare a ciò, Capizzi e Masarotto (2012) hanno proposto una rivisitazione della carta CUSCORE che, partendo da $i = 3$, viene così implementata:

1. Stimare il valore corrente della media di Q_i, utilizzando la seguente Adaptive EWMA (AEWMA)

$$f_i = (1 - w_i)f_{i-1} + w_i Q_i \tag{2.1}$$

dove $f_2 = 0$ e w_i è una funzione peso il cui valore è dato da:

$$w_i = \begin{cases} \lambda & \text{se } |Q_i - f_{i-1}| \leq \gamma \\ 1 - (1-\lambda)\frac{\gamma}{|Q_i - f_{i-1}|} & \text{se } |Q_i - f_{i-1}| > \gamma \end{cases} \tag{2.2}$$

con $\gamma \geq 0$ e $0 < \lambda \leq 1$.

2. Monitorare il processo utilizzando la seguente carta di controllo di tipo CUSCORE:

$$\begin{cases} AC_i^L = \min\{0, AC_{i-1}^L + |f_i|(Q_i + |f_i|/2)\} \\ AC_i^U = \max\{0, AC_{i-1}^U + |f_i|(Q_i - |f_i|/2)\} \end{cases} \tag{2.3}$$

con $AC_2^L = AC_2^U = 0$ e f_i data dalla (2.1)

Questa carta segnala quando $AC_i^L < -h$ o $AC_i^U > h$, dove h e $-h$ rappresentano i limiti di controllo della carta (sul valore di h si discuterà successivamente). Chiameremo questa carta Adaptive CUSCORE (ACUSCORE).

Non appena sono disponibili due osservazioni, il metodo self-starting fa uso di ciascuna nuova osservazione per aggiornare la media del campione m_i e la varianza campionaria s_i^2 e per calcolare le statistiche T_i e Q_i attraverso le equazioni mostrate nel capitolo precedente. Il campionamento e l'aggiornamento della stima dei parametri continua finchè l'ipotesi che il processo sia in controllo non viene rifiutata.

Le statistiche di controllo AC_i^L e AC_i^U sono simili ad una CUSUM bilaterale. La differenza sta nel fatto che il rilevamento è basato sulla somma cumulata del prodotto fra $|f_i|$ e x_i, e non sulla somma cumulata delle sole x_i; tenuto conto di questa somiglianza, l'interpretazione successiva al rilevamento di un fuori controllo tramite le statistiche di controllo ACUSCORE può essere svolta come nella CUSUM. In particolare:

- La direzione del cambiamento può essere determinata da quale delle due statistiche AC_i^L e AC_i^U segnala il fuori controllo;
- L'istante di tempo in cui si ha il cambiamento può essere stimato utilizzando l'ultima volta in cui la statistica che ha segnalato il FC era 0.

2.1 Disegno della carta

Per disegnare la carta di controllo ACUSCORE dobbiamo conoscere i valori di γ e λ che garantiscono buone prestazioni della carta in una varietà di condizioni fuori controllo. Capizzi e Masarotto (2010) hanno effettuato un intensivo studio di simulazione per capire l'effetto di λ e γ sulla performance dell'ACUSCORE. In particolare, per diversi valori di λ, γ, δ e τ, hanno studiato i ritardi attesi nel segnalare un cambiamento in media di ampiezza δ alla $\tau - esima$ osservazione, ovvero:

$$d_\delta = E_{OC}(RL - \tau - 1|RL \geq \tau) \tag{2.4}$$

I risultati di questi studi possono sono riassunti nella tabella 2.1.

Sulla base di questi risultati, Capizzi e Masarotto (2010) raccomandano di utilizzare $\boldsymbol{\lambda = 0.15}$ e $\boldsymbol{\gamma = 3}$, perchè garantiscono le migliori perfomance della carta in svariate condizioni di fuori controllo. Il disegno dell'ACUSCORE si riduce quindi alla scelta dei limiti di controllo h, la cui scelta verrà discussa in seguito.

E' interessante notare che questo pratico vantaggio dello schema suggerito è dovuto all'utilizzo di una statistica *adaptive EWMA* per stimare la *fault*

Performance della carta per diversi valori di λ e γ			
γ<6		$\forall\lambda$	
$\lambda \in [0.05, 0.25]$	$\lambda \in [0.30, 0.60]$	Valori piccoli di γ	Valori grandi di γ
Buona performance per piccoli e grandi cambiamenti in media	Performance ridotta per piccoli cambiamenti in media	Performance migliore contro grandi cambiamenti in media	Performance migliore contro piccoli cambiamenti in media

Tabella 2.1: Performance della carta per diversi valori di γ e λ

signature; infatti se l'AEWMA fosse sostituita da una standard EWMA, ovvero se avessimo stimato la fault signature utilizzando

$$f_i = \psi f_{i-1} + (1 - \psi)Q_i \qquad 0 < \psi \leq 1 \tag{2.5}$$

con ψ fissato, la performance della carta dipenderebbe crucialmente dal valore di ψ; in particolare, gli autori evidenziano come piccoli valori di ψ portano ad una veloce rilevazione dei piccoli cambiamenti, mentre grandi valori di ψ permettono una performance migliore in caso di grandi cambiamenti. L'ACUSCORE basata sull'adaptive EWMA, invece, assicura un'ottima performance nel rilevare sia piccoli che grandi cambiamenti.

2.1.1 Scelta dei limiti di controllo

La scelta dell'intervallo decisionale $(-h; +h)$ deve essere svolta bilanciando la performance della carta di controllo sia nelle condizioni in controllo, sia in quelle di fuori controllo, verificando che, fissato un determinato ARL in controllo, si abbia un ARL fuori controllo minimo.

Capizzi e Masarotto (2009) hanno determinato, via approssimazione stocastica, i valori dei limiti di controllo utili ad ottenere un desiderato valore della ARL in controllo (si veda tabella 2.2).

ARL_0	h
50	2.698
100	4.196
200	6.033
370.4	7.970
500	8.977
1000	11.558

Tabella 2.2: Valori di h in funzione dell'ARL_0

Capitolo 3

La carta non parametrica Adaptive EWMA

L'SPC veniva inizialmente utilizzato solo in ambito industriale, mentre ora questa tecnologia è stata adottata anche in altri campi, come la biologia, la medicina, la finanza e molte altre aree. La maggior parte delle carte di controllo assume che i dati provengano da una qualche distribuzione nota, più comunemente dalla distribuzione normale. In molte situazioni reali, però, la distribuzione di una caratteristica di qualità è non normale o sconosciuta; in questi casi risulta indispensabile usare una carta di controllo non parametrica, che non necessita di conoscere la distribuzione della caratteristica di qualità e per la quale le distribuzioni delle RL in controllo sono le stesse per ogni distribuzione continua. In questo capitolo verrà presentata una carta AEWMA sequenziale, suggerita da Liu et al. (2013), basata sui ranghi standardizzati per rilevare cambiamenti persistenti nel parametro di posizione.

Come anticipato nel capitolo 1, la letteratura sulle carte EWMA ci dice che esse possono essere costruite per avere una buona *performance* sia nel caso di piccoli cambiamenti che nel caso di grandi cambiamenti, a seconda dei valori di λ che vengono utilizzati; non è però possibile che una singola carta EWMA abbia un minimo ARL in entrambi i casi. Ciò che si propongono gli autori è di rilevare piccoli e grandi cambiamenti con una singola carta che non richieda nessuna conoscenza preliminare sulla distribuzione.

3.1 Lo schema proposto

Si supponga che le osservazioni del processo che vengono raccolte nel corso del tempo provengano dal seguente modello con istante di cambiamento τ:

$$x_i \sim \begin{cases} F(x, \mu_0) & \text{per } i = 1, 2, \ldots, \tau \\ F(x, \mu_1) & \text{per } i = \tau + 1, \tau + 2, \ldots \end{cases} \tag{3.1}$$

dove μ_0 e μ_1 sono, rispettivamente, i parametri di posizione in controllo e fuori controllo e $F(\cdot)$ è una distribuzione continua sconosciuta.

Indichiamo con R_n l'$n - esimo$ rango sequenziale definito come

$$R_n = \sum_{j=1}^{n} I(x_n \geq x_j) \tag{3.2}$$

dove $I(\cdot)$ è una funzione indicatrice che assume valore 1 se $x_n \geq x_j$, 0 altrimenti. La distribuzione di R_n varia al crescere di n; in particolare, il rango ha una distribuzione uniforme che, quando il processo è in controllo, ha valore atteso pari a $E[R_n] = (n+1)/2$ e varianza pari a $V[R_n] = (n+1)(n+2)/12$. Consideriamo adesso il rango standardizzato:

$$R_n^* = \frac{R_n - E[R_n]}{\sqrt{V[R_n]}} \quad n \geq 2 \tag{3.3}$$

Una possibile carta EWMA non parametrica basata sui ranghi standardizzati è quindi data da $Z_n = (1 - \lambda)Z_{n-1} + \lambda R_n^*$, con $Z_0 = 0$ e $0 \leq \lambda \leq 1$ (carta NE). Per disegnare, però, una carta di controllo che sia robusta per diverse grandezze dei cambiamenti, gli autori propongono la seguente statistica EWMA:

$$Z_n = (1 - \eta)Z_{n-1} + \eta R_n^* \tag{3.4}$$

con $Z_0 = 0$ e dove il peso η è dato da

$$\eta = 1 - \frac{1 - \lambda}{\max\{1, |\bar{R}_{n,k}^*|/\omega\}} \tag{3.5}$$

dove $\bar{R^*_{n,k}} = k^{-1}\sum_{i=0}^{k-1}$ rappresenta la media degli ultimi k ranghi standardizzati e k, ω, λ sono costanti predefinite; è importante notare che se $\bar{R^*_{n,k}}$ è più piccolo di ω, allora η è uguale a λ. In ogni caso, η tende maggiormente a 1 man mano che la media degli ultimi k valori di R^*_n tende ad infinito.

Questa carta può essere vista come una combinazione di una EWMA con una Shewhart, e può essere implementata solo quando $n \geq 2$.

3.2 Disegno della carta

Il disegno di questa carta prevede la scelta di tre costanti: λ, k e ω. Per una completa comprensione della performance della carta NAE, Liu et al. (2013) illustrano i risultati per diversi valori di queste costanti.

Analizziamo per primo il valore della costante λ. Lucas e Saccucci (1990) suggeriscono che valori piccoli di λ possono essere utilizzati per rilevare velocemente piccoli cambiamenti nel parametro di posizione, mentre grandi valori di λ devono essere utilizzati se si è intenzionati a rilevare velocemente grandi cambiamenti. Gli autori, per avere il valore iniziale di η, hanno scelto un piccolo valore di λ; quando il processo è OC il valore di η aumenta velocemente. Viene raccomandato quindi di scegliere $\boldsymbol{\lambda = 0.03}$.[1].

I parametri k e ω possono essere scelti utilizzando l'indice *RMI*, un indice che permette di valutare la performance complessiva di una carta rispetto a un range di possibili valori del cambiamento δ. L'RMI di una carta di controllo, suggerito da Han e Tsung (2006), è definito come:

$$RMI = \frac{1}{N}\sum_{l=1}^{N}\frac{ARL_{\delta_l} - MARL_{\delta_l}}{MARL_{\delta_l}} \tag{3.6}$$

dove N è il numero totale di cambiamenti δ considerati, ARL_{δ_l} è l'ARL_{OC} della carta di controllo presa in considerazione quando rileva un cambiamento di δ_1, e $MARL_{\delta_l}$ è il più piccolo valore dell'ARL_{OC} tra tutti quelli delle altre carte poste a confronto quando si vuole individuare un cambiamento δ_1. In questo modo, $(ARL_{\delta_l} - MARL_{\delta_l})/MARL_{\delta_l}$ può essere considerato

[1]Ovviamente possono essere scelti altri piccoli valori di λ

una misura di efficienza della carta considerata, confrontata con la carta migliore nel rilevare cambiamenti di ampiezza δ_1, mentre l'RMI è la media di tutte le differenze. Una carta di controllo con il più piccolo valore di RMI è considerata la migliore rispetto alle altre scelte per il confronto. I valori di δ considerati dagli autori variano in un range da 0.25 a 4.

Gli autori hanno studiato l'indice RMI in diverse condizioni, ponendo $k = 1, 2, \ldots, 10$ e $\omega = 0.8, 0.9, \ldots, 1.6$, notando che l'RMI più basso si ha nel caso di $\boldsymbol{k = 5}$ e $\boldsymbol{\omega = 1.2}$, che possono essere quindi considerati i parametri ottimali della carta. In ogni caso, se si utilizzano valori diversi di δ, si possono ottenere dei parametri ottimali differenti.

Un'altra misura che si può utilizzare per verificare l'efficienza della NAE per diversi tipi di shifts è la *ratio of change* (ROC) di η, così definita

$$ROC = E[N(\eta_c)/N(\eta)] \tag{3.7}$$

dove $N(\eta_c)$ si riferisce al numero di volte che η è cambiato e $N(\eta_c)$ rappresenta il numero totale di simulazioni. In particolare, si ha $ROC = 0$ quando il processo è in controllo, mentre si ha $ROC = 1$ quando la grandezza dello shift tende a infinito. Quando $k = 5, \omega = 1.2, \lambda = 0.03$ e il processo è in controllo, si ottiene $ROC = 0.00038$.

Gli autori hanno confrontato (tramite profili ARL e indice ROC) la loro carta con la non parametrica EWMA (NE) con un peso λ pari a 0.03, con un cambiamento dello shift che avviene sotto distribuzione normale standard. Sono state utilizzate 4 osservazioni per inizializzare la carta NAE e la carta NE. I risultati di questi confronti, attuati sia per $\tau = 100$ che per $\tau = 1000$, sono riassunti nella tabella 3.1:

δ	ROC	ARL
$0.25 \leq \delta \leq 0.75$	ROC molto basso ma non uguale a 0	NE ha una performance leggermente migliore di NAE.
$0.75 < \delta \leq 4$	ROC cresce velocemente	La performance della NAE è nettamente migliore di quella della NAE

Tabella 3.1: Confronto tra carta NAE e NE

Questi risultati valgono sia per $\tau = 100$ che per $\tau = 1000$. Il motivo per cui per piccoli δ la carta NE ha una performance migliore della NAE è che, anche quando il processo è IC ($\delta = 0$), l'indice ROC non è uguale a 0, e η ha un valore leggermente più grande dello 0.03 di λ. Perciò il limite di controllo della carta NAE è un pò più grande di quello della NE. Dai risultati si evince che la carta di controllo NAE basata su un'Adaptive Ewma è robusta rispetto a diverse ampiezze di cambiamento; inoltre, il fattore di adattamento η risulta robusto sia nel caso di $\tau = 100$ che nel caso di $\tau = 1000$; questo ci fa notare che utilizzare un'Adaptive EWMA risulta vantaggioso rispetto ad una semplice EWMA, in quanto il fattore di adattamento η si presenta robusto sia per τ grandi che per τ piccoli.

E' stato poi effettuato un confronto sulla relazione fra gli ARL_{OC} e i valori di k e ω. Sulla base di questo confronto si può affermare che grandi valori di k e ω sono più robusti e efficienti per piccoli shift, ma non sono sensibili a grandi valori di δ; piccoli valori di k e ω portano, invece, al risultato opposto. I valori ottimali dei due parametri risultano quindi, come già anticipato, $k = 5$ e $\omega = 1.2$, che garantiscono un'ottima performance della carta sia con piccoli che con grandi cambiamenti.

Capitolo 4

Performance Comparisons

In questo capitolo confronteremo la performance della carta parametrica *self-starting* ACUSCORE con quella non parametrica NAE basata sui ranghi standardizzati, utilizzando i profili ARL.

Per questo confronto sono state simulate osservazioni da tre distribuzioni differenti: Normale standard, Gamma e t di Student. In ogni simulazione vengono generate 50 osservazioni "in controllo", e dalla $51-esima$ osservazione (τ=50) in poi viene modificato il parametro di posizione della distribuzione introducendo lo shift, ovvero δ. In particolare si seguono i seguenti sistemi :

$$\begin{cases} X_i \sim N(0,1) & i = 1,2,\dots,\tau \\ X_i \sim N(\delta,1) & i = \tau+1, \tau+2, \dots \end{cases}$$
$$\begin{cases} X_i \sim G(3,1) & i = 1,2,\dots,\tau \\ X_i \sim G(3,1)+\delta & i = \tau+1, \tau+2, \dots \end{cases} \tag{4.1}$$
$$\begin{cases} X_i \sim T(4) & i = 1,2,\dots,\tau \\ X_i \sim T(4)+\delta & i = \tau+1, \tau+2, \dots \end{cases}$$

con $\delta \in (0.25, 3)$.

Per ogni valore di δ, il procedimento per la simulazione degli ARL fuori controllo consiste nel calcolare per ogni nuova osservazione la relativa statistica di controllo, nel confrontarla con i limiti e, nel caso la statistica risulti fuori controllo, conservando la corrispondente RL. Ripetendo per 3000 volte questa procedura, si ottengono 3000 run lengths, delle quali viene poi fatta la media, ottenendo l'ARL_{OC} simulato relativo a ciascun δ.

Nel caso dell' ACUSCORE si era già a disposizione dei limiti di controllo (vedi Capizzi e Masarotto, 2012). Per quanto riguarda, invece, la carta non parametrica NAE, i limiti sono stati calcolati via simulazione.

E' noto che i limiti di controllo asintotici di una carta EWMA, sotto l'ipotesi che il processo sia in controllo, sono dati da

$$E_{Z_n} \pm L\sigma_{Z_n}\sqrt{\frac{\lambda}{2-\lambda}} \tag{4.2}$$

dove λ è la costante di lisciamento che assume valori in $(0,1)$ e L è una costante determinata tramite il valore di λ e quello dell'ARL_{IC}. La carta di controllo presa in considerazione si basa sui ranghi standardizzati, che hanno distribuzione uniforme di media 0 e varianza 1; il valore atteso e la varianza della statistica di controllo Z_n (3.4) coincidono con quelli dei ranghi standardizzati e quindi, utilizzando il valore ottimale di λ proposto dagli autori, troviamo che i limiti di controllo della carta sono $\pm L\sqrt{0.03/1.97}$. L'unica componente ignota di questi limiti è la costante L, da trovare tramite simulazione; in particolare viene simulato l'ARL_{IC} utilizzando come primo valore di L quello che fornisce un ARL in controllo pari a 500 nel caso normale, per un valore di λ pari a 0.03. Se l'ARL simulato è più basso di 500, viene diminuito il valore di L, altrimenti viene aumentato, e si procede simulando finchè non si ottiene un ARL molto vicino a 500. Da queste simulazioni è risultato che il valore di L che dà un ARL_{IC} vicino a 500 è 2.663. [1]

Successivamente sono stati simulati i profili ARL delle due carte, come spiegato precedentemente. I risultati ottenuti sono riassunti nella tabella 4.1.

La carta parametrica ACUSCORE è una carta disegnata assumendo che le osservazioni siano Normali (i.i.d.). In questi casi, il valore di h che garantisce un'ARL in controllo pari a 500 è 8.977. I risultati mostrano, però, che la performance della carta viene alterata notevolmente quando si ha una violazione dell'assunto distributivo normale. In particolare, distribuzioni asimmetriche o con code più pesanti (come la Gamma e la t di Student), non garantiscono, come nel caso Normale, un ARL_0 pari a 500; nel caso

[1] La funzione utilizzata per trovare questo valore è mostrata nell'Appendice A.

della distribuzione Gamma, ad esempio, con $h = 8.977$, si avrebbe un ARL in controllo pari a 209, mentre con la t di Student si avrebbe un'ARL_0 pari a 177. Si nota che le due ARL in controllo sono notevolmente più basse di 500. Ciò implica che violazioni dall'ipotesi distributiva di normalità inducono la carta a segnalare prima anche quando il processo è in controllo. La carta non parametrica NAE, invece, non richiedendo un'ipotesi distributiva di partenza, garantisce un'ARL in controllo pari a 500 per tutte e tre le distribuzioni, e risulta preferibile alla carta parametrica quando si hanno dati che non supportano l'ipotesi di normalità.

Dai risultati si evince, inoltre, una preferenza della carta parametrica ACUSCORE quando si vogliono rilevare cambiamenti di piccola ampiezza ($\delta \in [0.25 : 0.75]$), mentre la carta non parametrica NAE dimostra una performance migliore nel caso di grandi cambiamenti ($\delta \in [1 : 3]$). Possiamo vedere che, nel caso di distribuzione Normale, la carta non parametrica presenta un ritardo nel segnalare un allarme simile a quello della parametrica già con un cambiamento di ampiezza pari a 0.75, mostrando una performance migliore per i successivi valori di δ.

Profili ARL_δ					
ACUSCORE			NAE		
$N(\delta,1)$	$Gamma(3,1)+\delta$	$T(4)+delta$	$N(\delta,1)$	$Gamma(3,1)+\delta$	$T(4)+delta$
$ARL_{0.25} = 281.51$	$ARL_{0.25} = 106.05$	$ARL_{0.25} = 141.40$	$ARL_{0.25} = 335.68$	$ARL_{0.25} = 331.13$	$ARL_{0.25} = 297.12$
$ARL_{0.5} = 80.84$	$ARL_{0.5} = 51.74$	$ARL_{0.5} = 71.44$	$ARL_{0.5} = 128.69$	$ARL_{0.5} = 104.56$	$ARL_{0.5} = 71.43$
$ARL_{0.75} = 36.71$	$ARL_{0.75} = 30.53$	$ARL_{0.75} = 36.77$	$ARL_{0.75} = 35.63$	$ARL_{0.75} = 28.53$	$ARL_{0.75} = 21.99$
$ARL_1 = 23.12$	$ARL_1 = 20.80$	$ARL_1 = 22.27$	$ARL_1 = 15.91$	$ARL_1 = 15.82$	$ARL_1 = 12.10$
$ARL_{1.5} = 12.27$	$ARL_{1.5} = 11.07$	$ARL_{1.5} = 12.08$	$ARL_{1.5} = 8.29$	$ARL_{1.5} = 8.21$	$ARL_{1.5} = 6.69$
$ARL_2 = 8.24$	$ARL_2 = 6.98$	$ARL_2 = 7.60$	$ARL_2 = 5.52$	$ARL_2 = 5.43$	$ARL_2 = 5.14$
$ARL_{2.5} = 5.81$	$ARL_{2.5} = 5.15$	$ARL_{2.5} = 5.42$	$ARL_{2.5} = 4.71$	$ARL_{2.5} = 4.59$	$ARL_{2.5} = 4.58$
$ARL_3 = 4.28$	$ARL_3 = 3.95$	$ARL_3 = 4.12$	$ARL_3 = 4.35$	$ARL_3 = 4.37$	$ARL_3 = 4.42$

Tabella 4.1: ARL_δ per la carta ACUSCORE e NAE. ARL in controllo pari a 500, $\delta \in [0.25 : 3]$

Conclusioni

In questa tesi si è cercato di presentare dei metodi *self-starting* in ambito univariato. Si è voluto, in particolare, mostrare gli aspetti teorici e pratici della carta parametrica ACUSCORE, proposta da Capizzi e Masarotto (2012), e della carta non parametrica NAE, proposta da Liu et al. (2013). E' stato poi mostrato un confronto fra queste due carte basato sui profili ARL.

La metodologia *self-starting* in ambito univariato, che trova le sue fondamenta negli approcci proposti da Quesenberry (1991) e Hawkins (1987), si basa principalmente sulla trasformazione dei dati originali in statistiche distribuite come una normale standard, chiamate statistiche Q. Il vantaggio di questa metodologia è quello di essere molto utile in applicazioni industriali come *short-run* e *start-up*, dove non si dispone di un numero elevato di misurazioni e dove, spesso, i parametri della distribuzione della caratteristica di qualità risultano ignoti. Avendo un numero minimo di 3 osservazioni, possono essere calcolate queste statistiche e, ad ogni nuova osservazione campionata, vengono aggiornate le stime dei parametri e si verifica se il processo è ancora in controllo.

Una carta di controllo di tipo parametrico che può essere utilizzata nell'ambito univariato del *self-starting* è la carta ACUSCORE (Capizzi e Masarotto, 2012), che unisce l'efficienza di una carta CUSCORE con quella di una EWMA adattiva. La carta EWMA adattiva risulta utile per garantire una migliore performance della carta nel rilevare sia piccoli che grandi cambiamenti.

Quando non è nota la distribuzione della caratteristica di qualità, risulta conveniente utilizzare una carta non parametrica. In questa tesi è stata analizzata la carta non parametrica Adaptive EWMA, proposta da Liu et al. (2013), basata sui ranghi standardizzati. La particolarità di questa carta sta nel peso utilizzato per le statistiche precedenti; gli autori hanno infatti proposto un valore che dipende anche dalla media degli ultimi k ranghi standardizzati rilevati, dove k è una costante che gli autori suggeriscono di porre uguale a 5 (valore ottimale).

Il confronto fra queste due carte è stato attuato tramite l'uso dei profili ARL, ponendo un ARL in controllo pari a 500. Il risultato di questo confronto ha mostrato come la performance della carta parametrica venga alterata in caso di violazioni dalla normalità, non garantendo un ARL in controllo di partenza pari a 500, come avviene invece nel caso Normale. La carta non parametrica, invece, si mostra più efficiente nel caso i dati non si distribuiscano normalmente, in quanto è indipendente dagli assunti distributivi.

Il confronto suggerisce, inoltre, una preferenza della carta parametrica nel rilevare cambiamenti di piccola ampiezza e della carta non parametrica nel rilevare cambiamenti di grande ampiezza.

Appendice A

Codici R utilizzati

1. **Codice simulazione ARL_δ della carta Acuscore** ($ARL_{IC} = 500$)

```
ARL=function(delta)
{
rl=rep(0,3000)
l=0.15
h=8.977
m=rep(0,2);a=rep(0,2);s2=rep(0,2);t=rep(0,2)
q=rep(0,2);f=rep(0,2);w=rep(0,2);acl=rep(0,2);acu=rep(0,2)
for(j in 1:length(rl))
{
x=rnorm(2,0,1)
m[2]=(x[1]+x[2])/2
s2[2]=((x[1]-x[2])^2)/2
for(i in 3:50)
{
x=rnorm(1,0,1)
a[i]=x-m[i-1]
m[i]=m[i-1]+a[i]/i
s2[i]=s2[i-1]+(1/i)*((a[i]^2)-i*s2[i-1]/(i-1))
t[i]=sqrt((i-1)/i)*a[i]/sqrt(s2[i-1])
q[i]=qnorm(pt(t[i],i-2))
```

```
if(abs(q[i]-f[i-1])<=3)
w[i]=l
else
w[i]=1-(1-l)*3/(abs(q[i]-f[i-1]))
f[i]=(1-w[i])*f[i-1]+w[i]*q[i]
acl[i]=min(0,(acl[i-1]+abs(f[i])*(q[i]+abs(f[i])/2)))
acu[i]=max(0,(acu[i-1]+abs(f[i])*(q[i]-abs(f[i])/2)))
}
for(i in 51:2000000)
{
x=rnorm(1,0,1)+delta
a[i]=x-m[i-1]
m[i]=m[i-1]+a[i]/i
s2[i]=s2[i-1]+(1/i)*((a[i]^2)-i*s2[i-1]/(i-1))
t[i]=sqrt((i-1)/i)*a[i]/sqrt(s2[i-1])
q[i]=qnorm(pt(t[i],i-2))
if(abs(q[i]-f[i-1])<=3)
w[i]=l
else
w[i]=1-(1-l)*3/(abs(q[i]-f[i-1]))
f[i]=(1-w[i])*f[i-1]+w[i]*q[i]
acl[i]=min(0,(acl[i-1]+abs(f[i])*(q[i]+abs(f[i])/2)))
acu[i]=max(0,(acu[i-1]+abs(f[i])*(q[i]-abs(f[i])/2)))
if((acl[i]<(-h))||(acu[i]>h))
{
rl[j]=i-50
break}}}
mean(rl)
}
```

[1]

[1] Questa funzione è stata usata anche per la distribuzione Gamma e t di Student, sostituendo a *rnorm* sia *rgamma* che *rt* e aggiungendo la relativa *sd* nell'aggiunta dello shift al parametro di posizione delle osservazioni fuori controllo.

2. **Codice simulazione ARL_{IC} per trovare i limiti di controllo della NAE**

```
L.NAE=function(L,k,omega,lambda){
rl=rep(0,10000)
rkst=rep(0,4)
z=rep(0,4)
h=L*sqrt(lambda/(2-lambda))
for(j in 1:length(rl)){
x=rnorm(1)
for(i in 2:4){
x=c(x,rnorm(1))
rk=rank(x)
n=length(x)
E=(n+1)/2
V=((n+1)*(n-1))/12
rkst[i]=(rk[i]-E)/sqrt(V)}
for(i in 5:200000){
x=c(x,rnorm(1))
rk=rank(x)
n=length(x)
E=(n+1)/2
V=((n+1)*(n-1))/12
rkst[i]=(rk[i]-E)/sqrt(V)
rm=rkst[(n-k+1):n]
Rm=sum(rm)
rkm=(k^(-1))*Rm
eta=1-((1-lambda)/max(1,(abs(rkm)/omega)))
z[i]=(1-eta)*z[i-1]+eta*rkst[i]
if((z[i]<(-h))||(z[i]>h)){
rl[j]=i
break}}}
mean(rl)}
```

3. **Codice simulazione ARL_δ della carta NAE ($ARL_{IC} = 500$)**

```
ARL.NAE.N=function(delta,k,omega,lambda)
{
rl=rep(0,3000)
rkst=rep(0,4)
z=rep(0,4)
h=2.663*sqrt(lambda/(2-lambda))
for(j in 1:length(rl))
{
x=rnorm(1)
for(i in 2:4)
{
x=c(x,rnorm(1))
rk=rank(x)
n=length(x)
E=(n+1)/2
V=((n+1)*(n-1))/12
rkst[i]=(rk[i]-E)/sqrt(V)
}
for(i in 5:50)
{
x=c(x,rnorm(1))
rk=rank(x)
n=length(x)
E=(n+1)/2
V=((n+1)*(n-1))/12
rkst[i]=(rk[i]-E)/sqrt(V)
rm=rkst[(n-k+1):n]
Rm=sum(rm)
rkm=(k^(-1))*Rm
eta=1-((1-lambda)/max(1,(abs(rkm)/omega)))
z[i]=(1-eta)*z[i-1]+eta*rkst[i]
```

```
}
for(i in 51:200000)
{
x=c(x,rnorm(1)+delta)
rk=rank(x)
n=length(x)
E=(n+1)/2
V=((n+1)*(n-1))/12
rkst[i]=(rk[i]-E)/sqrt(V)
rm=rkst[(n-k+1):n]
Rm=sum(rm)
rkm=(k^(-1))*Rm
eta=1-((1-lambda)/max(1,(abs(rkm)/omega)))
z[i]=(1-eta)*z[i-1]+eta*rkst[i]
if((z[i]<(-h))||(z[i]>h))
{
rl[j]=i-50
break
}
}
}
mean(rl)
}
```

[2]

[2]Questa funzione è stata usata anche per la distribuzione Gamma e t di Student, sostituendo a *rnorm* sia *rgamma* che *rt* e aggiungendo la relativa *sd* nell'aggiunta dello shift al parametro di posizione delle osservazioni fuori controllo.

Bibliografia

[1] Box, G. E. P. and Ramirez, J. G. (1992). Cumulative score charts. *Quality and Reliability Engineering International*, 8, 17-27.

[2] Capizzi, G. and Masarotto, G. (2012). An enhanced control chart for start-up processes and short runs. *Quality Technology and Quantitative Management*, 9, 189-202.

[3] Capizzi, G. and Masarotto, G. (2003). An adaptive exponentially weighted moving average control chart. *Technometrics*, 45, 199-207.

[4] Han, D., Tsung, F. (2006). A reference-free Cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *Journal of the American Statistical Association*, 101,368-386.

[5] Hawkins, D. M. (1987). Self-Starting cusum Charts for Location and Scale. *The Statisti-cian*, 36, 299-315.

[6] Kocherlakota, K. and Kocherlakota, S. (1991) On the doubly noncentral t distribution. *Communications in Statistics, Series B-Simulation and Computation*, 20, 23-31.

[7] Liu L., Xuemin Z., Jian, Z. and Zhaojun, W. (2013). A sequential rank-based nonparametric adaptive EWMA control chart. *Communications in Statistics, Series B-Simulation and Computation*, 42, 841-859.

[8] Lucas, J. M., Saccucci, M. S. (1990). Exponentially weighted moving average control schemes properties and enhancements. *Technometrics*, 32,1-29.

[9] Quesenberry, C. P. (1991). SPC Q charts for start-up processes and short or long Runs. *Journal of Quality Technology*, 23, 213-224.

[10] Zantek, P. F. (2005). Run-length distributions of Q-chart schemes. *IIE transactions*, 37, 1037-1045.

Ringraziamenti

Desidero anzitutto ringraziare la Prof.ssa Giovanna Capizzi, relatrice della mia tesi, per l'aiuto e la disponibilità fornitami durante la stesura dell'elaborato.

Un ringraziamento particolare a mia mamma, che da sempre mi sostiene, mi sta vicino e che mi ha permesso di portare a termine questa esperienza.

Un grazie a tutti i miei colleghi universitari con cui ho condiviso momenti belli e brutti di questi tre anni.

Vorrei, infine, ringraziare tutte le persone che mi sono state accanto e i miei amici di Bassano del Grappa.