



Università degli Studi di Padova
Facoltà di Scienze Statistiche
Corso di laurea in Statistica e Tecnologie Informatiche

TESI DI LAUREA:

**CARTA NON PARAMETRICA EWMA
BASATA SUL RAPPORTO DI
VEROSIMIGLIANZA**

Relatore: Prof.ssa GIOVANNA CAPIZZI

Laureando: MARCO CATTAPAN

Anno Accademico 2010/2011

INDICE

Introduzione	I
Capitolo 1: Controllo statistico di processo	1
1.1 Cos'è l'SPC	1
1.2 Cause di variazione	2
1.3 Componenti essenziali dell'SPC	3
1.4 Carte di controllo	5
1.5 Limiti di specifica e limiti di controllo	7
1.6 Classificazione dei processi	8
1.7 Fasi del monitoraggio	9
Capitolo 2: Carte di controllo non parametriche	11
2.1 Introduzione	11
2.2 La carta NLE	13
2.2.1 Distribuzione in controllo $F_0(x)$ nota	14
2.2.2 Distribuzione in controllo $F_0(x)$ incognita	16
2.3 Osservazioni raggruppate	18
Capitolo 3: Progettazione e implementazione della carta NLE	19
3.1 Determinazione dei limiti di controllo	19
3.2 Scelta del parametro λ	20
3.3 On post-signal diagnostic	21
3.4 Altre carte di controllo non parametriche	23
3.4.1 Wilcoxon rank sum test	24
3.4.2 Fligner-Killen test	25
Capitolo 4: Valutazione della performance numerica	27
4.1 L'ARL per il confronto tra carte	27
4.2 Confronto tra NLE e le convenzionali carte parametriche	28
4.3 Confronto tra NLE e le carte alternative non parametriche	31
4.4 Diagnostic Performance Analysis	33
Capitolo 5: Carta EWMA basata sui ranghi standardizzati	35
5.1 Ranghi standardizzati	35
5.2 Statistica test Z_t	36
5.3 ARL per processi IC	37
5.4 ARL per processi OC	38
Capitolo 6: Esempio pratico	39
Conclusioni	43
Appendice	47
Bibliografia	51

Introduzione

La prima carta di controllo fu proposta nel 1924 da Walter A. Shewhart che scrisse, accompagnandola: "*il modello di rapporto allegato è stato progettato per indicare se le variazioni osservate nella percentuale di apparati difettosi siano o no significative, a indicare, cioè, se il processo sia soddisfacente*". Iniziò così l'era del controllo statistico della qualità.

L'ISO (International Organization for Standardization) fa riferimento allo standard ISO 9001 per definire la qualità come una caratteristica che un prodotto o servizio deve avere per soddisfare i bisogni e le aspettative dei consumatori.

Le organizzazioni riconoscono che, per essere competitivi nell'economia globale di oggi, devono sforzarsi di raggiungere elevati livelli di qualità per cui è nel loro interesse porre un'estrema attenzione sui metodi statistici per la sorveglianza e il mantenimento della qualità affinché la scarsa qualità possa essere rilevata precocemente.

E' noto che per garantire la qualità all'interno di un'azienda, non è sufficiente controllare il prodotto finito, ma sono necessarie continue analisi durante tutte le fasi del processo produttivo che permettono di verificare la conformità dei prodotti alle esigenze e alle specifiche di progettazione.

Il controllo della qualità consiste nel realizzare una serie di ispezioni e misurazioni per determinare se gli standard di qualità vengano rispettati e nel caso in cui questo non si verificasse, per il raggiungimento e il mantenimento della conformità.

Lo strumento statistico che consente di monitorare il processo produttivo e di stabilire se è o meno in controllo, è la carta di controllo che consente di individuare eventuali anomalie presenti nel processo causati da fonti di variabilità anomale.

Le proprietà statistiche delle carte di controllo tradizionali, valgono al meglio solo se è soddisfatta l'assunzione della normalità della distribuzione delle osservazioni del processo sorvegliato.

E' ben risaputo tuttavia, che in molte applicazioni industriali di *start-up* e *short-run*, la distribuzione sottostante del processo non è sufficientemente nota per la mancanza di un adeguato numero di misurazioni o ha spesso distribuzione diversa da quella normale

così da compromettere le proprietà statistiche e quindi l'interpretazione delle carte di controllo parametriche.

Si utilizza il termine *start-up* per identificare l'operazione ed il periodo durante il quale si avvia un'impresa appena costituita, nella quale ci sono ancora processi organizzativi in corso.

Le situazioni di *short-run* si verificano quando volumi bassi di produzione vengono eseguiti in tempi sempre più rapidi, e di conseguenza vengono analizzate un numero di osservazioni insufficiente per una stima accurata dei parametri del processo e per il disegno dei limiti di controllo necessari per sorvegliare la stabilità del processo.

Per questi motivi, sono state progettate le carte di controllo non parametriche che non dipendono in generale, da alcuna assunzione parametrica sulla distribuzione. In altre parole, i limiti di controllo di queste carte possono essere scelti in modo da ottenere una desiderata distribuzione della run-length indipendentemente dalla distribuzione sottostante della caratteristica sorvegliata.

Questa tesi si propone in particolare di presentare una carta EWMA non parametrica, introdotta da C. Zou e F. Tsung (2010)¹, che si fonda sull'integrazione di un potente test non parametrico sulla bontà di adattamento (GOF), basato sul rapporto di verosimiglianza, e su una carta EWMA standard. Con tale carta, denominata NLE, si studia la Fase II di monitoraggio di un metodo per sorvegliare i cambiamenti in una distribuzione. La carta, facendo parte della categoria *self-starting*, per definizione gestisce quasi simultaneamente l'aggiornamento delle stime dei parametri e la verifica delle condizioni di fuori controllo che possono presentarsi.

Queste carte *self-starting* vengono utilizzate quando non si hanno a disposizione campioni sufficientemente grandi per stimare i parametri.

L'idea principale delle carte *self-starting* è quella di non considerare il problema dei parametri sconosciuti, producendo dai dati con parametri ignoti, un corrispondente flusso di osservazioni indipendenti e identicamente distribuite come una normale standard.

La carta NLE è in grado di rilevare uno spostamento in entrambe le direzioni cioè sia un incremento che un decremento in posizione, scala e forma nei parametri. Oltre a rilevare



¹C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pp. 174-196

velocemente un cambiamento nel processo, è importante diagnosticare anche il cambiamento e identificare se c'è stato uno spostamento in posizione, scala, o in entrambi dopo che si è verificato un segnale di allarme. Per far ciò, oltre che presentare la carta sono descritti due metodi basati su test non parametrici, il *Wilcoxon rank sum test* e l'*Aligned rank scale test*, che confrontando i dati raccolti prima e dopo lo stimatore del change-point, riescono a stabilire quali parametri sono stati interessati da qualche modifica.

Il primo capitolo tratterà il Controllo Statistico di Processo riassumendo quali tecniche vengono impiegate per il monitoraggio della qualità nelle realtà produttive, focalizzandosi maggiormente sul concetto del funzionamento delle carte di controllo.

Nel secondo capitolo verrà presentata la carta non parametrica NLE nei suoi aspetti teorici descrivendo il test non parametrico sulla bontà di adattamento su cui si basa e la sua statistica test. Nel terzo capitolo verranno esposti gli aspetti pratici legati al disegno della carta come la determinazione dei limiti di controllo, la scelta ottimale del parametro di liscio λ e i due test non parametrici che servono ad identificare quali parametri sono stati modificati che sono il *Wilcoxon rank sum test*² per la posizione e l'*Aligned rank scale test*³ per la varianza.

L' Average Run Length (ARL) è la più comune misura per valutare le prestazioni di una carta di controllo. Per avere una carta efficiente, si desidera un elevato valore dell'ARL quando il processo è in controllo, per ottenere una percentuale bassa di falsi allarmi, mentre quando è fuori controllo è consigliabile avere un'ARL molto bassa così da cogliere il prima possibile il cambiamento avvenuto.

Nel quarto capitolo verrà comparata in termini di ARL la performance della NLE con alcune convenzionali carte parametriche disegnate sotto l'assunzione di normalità, e con altri metodi non parametrici alternativi.

Nel quinto capitolo, sarà introdotta la carta non parametrica EWMA basata sui ranghi standardizzati, suggerita da Hackl e Ledolter, 1991, che verrà usata nell'esempio pratico presente nel sesto capitolo, per avere un confronto della sua prestazione con la carta non



² Chakraborti, Van de Wiel (2008) – “A Nonparametric Control Chart Based on the Mann-Whitney Statistic”. *IMS Collections. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, 1, pp. 156-172

³ Fligner, Killeen (1976) - “Distribution-Free Two-Sample Test for Scale”. *Journal of the American Statistical Association*, 71, pp. 210–213.

parametrica NLE, analizzando i risultati ottenuti con le funzioni create con il software R disponibili nell'appendice.

3

CAPITOLO 1

Controllo Statistico di Processo

1

1

1.1 Cos'è l'SPC

La qualità di un processo può essere adeguatamente rappresentata da alcune sue caratteristiche. Il Controllo Statistico del Processo (SPC) è un insieme di strumenti statistici e di gestione che, attraverso la sorveglianza della distribuzione delle caratteristiche di qualità, permettono ai processi di funzionare in modo più prevedibile e più redditizio.

I processi controllati producono meno sprechi e maggiori rendimenti per quanto riguarda sia la produzione che le performance dei macchinari. Infatti, i processi tenuti sotto controllo, realizzano prodotti di maggior qualità essendo i valori dei parametri vicini a target desiderati. Un parametro target è il valore ideale per il parametro che caratterizza la distribuzione della caratteristica di qualità. Tale valore target viene definito sulla base di esperimenti, simulazioni ed esperienza.

Le operazioni di produzione, l'usura degli impianti, la difettosità dei materiali acquistati e tra l'altro anche i difetti degli operatori, possono provocare scarsa qualità in output. L'obiettivo è di stabilire se il processo può essere lasciato continuare o se dovrebbe essere aggiustato o corretto per portare l'output ad un accettabile livello di qualità.

In altre parole, l'SPC consiste in una procedura sistematica di sorveglianza e di intervento per migliorare la stabilità del processo riducendo la variazione rispetto al target e di conseguenza migliorando la qualità del prodotto.

Ci sono molti effetti benefici dell'SPC, ma tutti derivano dalla riduzione della variazione della caratteristica di qualità rispetto al target stabilito dall'azienda. Il beneficio più immediato dell'SPC è il miglioramento della qualità del prodotto dovuto principalmente alla riduzione della variazione dal target. I prodotti risultano infatti più uniformi ed affidabili e per il fatto che ci sono meno sprechi, i costi sono ridotti.

L'SPC fornisce una base per il miglioramento continuo del processo grazie all'analisi continua dei dati. Gli operatori che sorvegliano il processo, decidono quando le correzioni sono necessarie e fanno la maggior parte degli aggiustamenti.

I costi dell'SPC non sono banali poiché le persone devono essere addestrate ad implementare e usare gli strumenti statistici utili per la sorveglianza. I sistemi di controllo dei processi devono essere stabiliti, e l'apparecchiatura necessaria deve essere acquistata e mantenuta in buone condizioni. Tuttavia questi costi sono certamente giustificati se si confrontano con i benefici che si traggono. Infatti, i clienti possono prontamente rilevare una qualità superiore reagendo positivamente con un aumento degli acquisti e ciò comporta un vantaggio critico nel marketing.

1.2 Cause di variazione

Il successo dell'SPC poggia sugli studi pionieri di Shewhart ¹ che ha individuato due tipi fondamentali di variazione che hanno diverse cause sottostanti e richiedono differenti reazioni quando si manifestano.

L'SPC viene usato per rilevare questi tipi di variazione e quindi per prevenire variazioni dal target adottando le più opportune ed efficaci azioni.

Le cause "comuni" di variazione determinano la variazione intrinseca del processo, la quale è presente anche quando il processo funziona al suo meglio per il fatto che è parte del processo stesso. Le cause comuni producono spostamenti dal target che sono relativamente piccoli in grandezza e non presentano nel corso del tempo nessun pattern sistematico. Se nel processo intervengono solo queste cause, i valori della produzione tenderanno di norma a disporsi in una curva a campana contenuta in limiti ben precisi, formando la cosiddetta variabilità naturale, e l'ampiezza della dispersione dei dati risulta prevedibile, analizzando il trend precedente, e casuale.

Se la variazione nella qualità dell'output della produzione è dovuta a queste cause comuni che si manifestano attraverso variazioni casuali nei materiali, nella temperatura, nell'umidità e così via, il costruttore non può farci niente. Se è troppo grande rispetto alle esigenze, la variabilità dei dati può ridursi solo modificando il processo nel suo complesso cambiando aspetti fondamentali del processo come ad esempio le modalità di manutenzione, l'addestramento del personale, il fornitore di materie prime.



¹ W. A. Shewhart (1980) – "Economic control of quality of manufactured product". Edizione commemorativa. E' la sua opera più famosa e descrive in maniera estremamente completa i principi di base del controllo della qualità.

Un processo sotto queste condizioni di variazioni comuni si considera stabile ovvero “in-controllo”. Come in precedenza detto, un processo stabile ha la stessa distribuzione nel corso del tempo. Un processo stabile è già performante ed aggiustarlo spesso comporta uno spreco di tempo degli ingegneri ed un aumento della variazione del processo.

Oltre alle cause comuni, ci sono delle cause che non dipendono dal funzionamento naturale del processo, che possono di per sé creare una variabilità. Le cause “speciali” di variazione si riferiscono a cause assegnabili che comportano uno spostamento significativo dal target oppure la presenza di un pattern sistematico non prevedibile che disturba il funzionamento del processo. Queste cause alterano la variabilità naturale e inducono uno shift in media o in varianza o una modifica della forma della distribuzione della caratteristica di qualità.

Se la variazione nella qualità è dovuta alle cause assegnabili come macchinari in logoramento, setting scorretti delle macchine, scarsa qualità delle materie prime o errori degli operatori, il processo dovrebbe essere aggiustato o corretto il più presto possibile per ridurre o eliminare la causa.

Con le azioni correttive, il processo viene riportato ad un accettabile livello di qualità, ma se viene lasciato procedere influenzato dalle cause speciali, la variazione del processo dal target potrebbe continuare o aggravarsi. Se d'altra parte l'intervento della causa speciale di variazione ha portato invece ad un miglioramento della qualità, il processo deve essere modificato in modo da incorporarla al suo interno, facendola diventare così una causa comune nel nuovo processo.

Quando un processo viene disturbato da una o più cause speciali, si dice che è instabile o in stato di “fuori-controllo”.

1.3 Componenti essenziali dell'SPC

Le componenti essenziali dell'SPC sono cinque: misurazioni, calcoli, carte di controllo, trend rules e piani di interventi (o response plans).

La misurazione è fondamentale ed è la parte più dispendiosa sia in termini di tempo che in implementazione del sistema di sorveglianza, senza questa non sarebbe possibile conoscere la performance del processo. Durante tutto il corso del processo produttivo, le

misurazioni quantitative vengono effettuate perché possono essere indicatori della qualità.

I calcoli vengono svolti soprattutto per ottenere la statistica di controllo per le varie osservazioni del campione e quelle più semplici sono la media campionaria, il range e la deviazione standard.

Le carte di controllo sono grafici in cui vengono rappresentati in ordine di tempo, i valori della statistica di controllo che permettono di individuare, tramite i limiti di controllo, l'instabilità di un processo.

Le trend rules sono delle regole che aiutano l'interpretazione di ciascuna carta di controllo infatti la posizione ed il pattern dei punti in una carta di controllo consentono di determinare, con una piccola probabilità di errore, se un processo è in controllo statistico.

Ce ne sono molte ma la più semplice trend rule è quella che ci si trova in una situazione di fuori controllo se un punto, associato ad una statistica di controllo, non rientra nei limiti di controllo e in questo caso, azioni correttive dovrebbero essere prese il prima possibile.

In aggiunta a questi punti fuori dai limiti di controllo, ci sono certe sequenze di punti che, nonostante si trovino all'interno dei limiti di controllo, evidenziano che qualcosa di anomalo sta intervenendo nel processo, dovuto a delle cause che devono essere quindi rimosse. Un esempio è la "*eight on one side rule*", dove una causa speciale si sospetta ogni volta che otto punti in fila cadono su un lato della linea centrale.

Scegliere le giuste trend rules è importante: se il sistema causa troppi falsi allarmi, diventerà inefficace nel corso del tempo perché le persone ignoreranno questi allarmi ma se il sistema non riesce a rilevare importanti cause speciali, la qualità del prodotto ne risentirà, sia immediatamente sia più tardi se il processo continuerà ad allontanarsi dal target.

Un altro motivo di preoccupazione per l'interpretazione delle carte di controllo è l'andamento nel corso del tempo del trend che può constatare un eventuale spostamento graduale e nel caso di verificasse, si dovrebbero prendere azioni correttive per riportare indietro il processo in una situazione soddisfacente.

Le cause che possono influenzare il trend possono essere i macchinari deteriorati, i graduali cambiamenti in temperatura e umidità, l'accumulo di sporcizia, l'efficacia degli operatori.

Il piano di intervento porta l'operatore, una volta riscontrata la possibile instabilità del processo con la violazione di una trend rule, ad effettuare una serie di dettagliate e sistematiche istruzioni documentate per riscontrare la causa speciale responsabile dell'escursione, correggerla e per far ripartire il processo. In questo modo, l'operatore riporta il processo ad uno stato stabile il più velocemente possibile.

Un piano di intervento ben costruito accelera la risoluzione di problemi e raccoglie le informazioni per futuri miglioramenti di processo.

1.4 Carte di controllo

Se è vero che la variabilità di un processo è inevitabile, è anche vero che l'uomo ha bisogno di tenerla sotto controllo, in modo da mantenere livelli di qualità accettabili.

Una carta di controllo rappresenta un metodo grafico per decidere se la variazione nell'output è dovuta a cause comuni (processo IC) o a cause speciali (processo OC).

Le carte di controllo hanno avuto ampio successo per cinque ragioni:

- forniscono informazioni diagnostiche
- forniscono informazioni sulla stabilità nel tempo del processo
- inducono un miglioramento della produttività perché si riducono gli sprechi e la perdita di tempo
- sono efficaci nel prevenire la produzione di pezzi difettosi per la tempestività delle segnalazioni
- evitano di dover apportare inutili aggiustamenti al processo produttivo

La carta registra in ordine cronologico e graficamente i valori della statistica di controllo di un processo e li confronta con i limiti di controllo superiore (UCL) e inferiore (LCL) che tengono presente della variabilità naturale del processo in esame.

Ogni volta che una situazione fuori controllo viene rilevata, assestamenti o altre azioni correttive vengono prese per riportare il processo in una situazione di controllo.

Uno stato di fuori controllo può consistere cambiamenti o in media o in varianza. In tal caso le carte di controllo vengono disegnate in coppia: una che monitori la media del processo e l'altra la varianza. Perché un processo possa essere considerato sotto

controllo, è necessario che entrambe le carte non presentino valori esterni ai limiti di controllo.

Le procedure statistiche per il controllo di processo nelle carte di controllo si basano sulla metodologia di verifica d'ipotesi. Per sapere se il processo si comporta in maniera standard, ovvero la distribuzione della caratteristica in esame è stabile, ad ogni istante temporale t si va a verificare se la statistica di controllo rientra all'interno dei limiti di controllo. In tal caso si conclude, con un margine prefissato di errore, che il processo è stabile. L'ipotesi nulla H_0 è dunque: il processo produttivo è in controllo (IC). Mentre l'ipotesi alternativa H_1 è: il processo produttivo è fuori controllo (OC) o instabile.

Ad ogni istante t , se si accetta l'ipotesi H_0 , si continua a campionare finché un valore della statistica di controllo risulta non appartenere alla regione di accettazione delimitata dai limiti di controllo e così, si rifiuta l'ipotesi H_0 e si lancia l'allarme. Un valore fuori dai limiti di controllo fornisce una forte evidenza che il processo è fuori controllo per l'influenza sul processo di una o più cause speciali di variazione. In altre parole, le carte di controllo vedono l'insorgere di cause speciali di variabilità nel momento in cui si arriva ad una situazione a favore dell'ipotesi H_1 associata al processo fuori controllo.

Con questo procedimento si svolge un'analisi sequenziale. Si tratta infatti di inferenza a dimensione casuale non conoscendo il numero preciso di campioni da estrarre prima della segnalazione di un allarme.

Attraverso questo sistema di verifica d'ipotesi, si possono avere due tipi di errori nelle decisioni di far continuare un processo in controllo o di far aggiustare un processo fuori controllo. L'errore del I tipo, ovvero l'ipotesi nulla viene erroneamente rifiutata e l'errore di II tipo, ovvero l'ipotesi nulla viene erroneamente accettata. L'efficienza della carta di controllo è dunque misurata dalla probabilità di rifiutare correttamente l'ipotesi nulla quando questa è falsa, dunque di lanciare un allarme in presenza di cause assegnabili di variazione.

Un passo importante nell'uso delle carte di controllo è il loro disegno che avviene sulla base di considerazioni statistiche ed economiche. Il disegno di una carta consiste nella scelta:

□ dei limiti di controllo: questo è uno dei passaggi più critici poiché si deve trovare un trade-off tra l'individuazione di falsi allarmi e quella di difetti

□ della dimensione campionaria: la probabilità di individuare uno scostamento cresce all'aumentare della dimensione campionaria e per scegliere quella ottimale bisogna aver presente qual è lo scostamento del processo che si vuole individuare più velocemente

□ della frequenza di campionamento: la situazione ottimale sarebbe quella di poter esaminare grandi campioni di frequente, ma tale situazione è spesso poco praticabile dal punto di vista economico

Le carte di controllo possono essere classificate a seconda del tipo di parametro che si voglia sorvegliare (la media, la deviazione standard, la percentuale di unità difettose) e a seconda che la caratteristica misurata vari nel continuo (carte di controllo per variabili) o nel discreto (carte di controllo per attributi).

Le carte di controllo si distinguono anche in “parametriche”, se sono verificati alcuni assunti e quindi dipendono dai parametri della distribuzione e in “non parametriche”, se sono indipendenti dagli assunti distributivi.

Ciascuna di queste carte può infine essere “senza” memoria (esempio carta Shewhart) o “con” memoria (ad esempio CUSUM ed EWMA). Le carte senza memoria si basano solo sull'istante t e verificano, ogni volta che si estrae un campione, se vi è evidenza a favore dell'ipotesi H_0 non considerando la storia passata del processo. Tali carte sono efficienti nel segnalare subito se il processo è fuori controllo di fronte a consistenti valori anomali. Con le carte con memoria, ogni volta che si estrae un campione, si verifica se vi è evidenza a favore dell'ipotesi H_0 considerando non solo l'ultimo campione ma anche tutti i campioni precedenti.

1.5 Limiti di specifica e limiti di controllo

I limiti di specifica sono i limiti estremi di una caratteristica del prodotto, determinati esternamente, per esempio dai clienti o dai progettisti.

Il limite superiore (LSL, Lower Specification Limit) e quello inferiore (USL, Upper Specification Limit) si basano sulla qualità, sulle caratteristiche dei macchinari e su considerazioni economiche ma non hanno alcuna relazione con l'attuale performance del processo ovvero sono specifiche del disegno.

Nessun processo produttivo è perfetto, c'è sempre almeno una minima variazione in ogni cosa che viene fabbricata e qualsiasi spostamento dal target è dannosa per la

qualità del prodotto anche se si trova ancora all'interno dei limiti di specifica. Quando la variazione dal target è abbastanza considerevole, o il prodotto deve essere eliminato oppure il cliente è costretto a subire effetti negativi da tale variazione.

L'SPC rileva cause speciali con l'aiuto dei limiti di controllo che rappresentano il normale processo di funzionamento e dipendono solamente dall'attuale performance del processo, e non da quella desiderata. Il limite di controllo superiore (UCL, Upper Control Limit) e quello inferiore (LCL, Lower Control Limit) sono importanti nel decidere se un processo è in controllo o fuori controllo.

I limiti di controllo se sono stretti provocano falsi allarmi con elevata probabilità mentre se sono larghi, c'è il rischio di non individuare un difetto. In generale, i limiti di controllo sono posti a tre deviazioni standard dal valore target di un parametro. E' da notare che tra -3σ e $+3\sigma$ si definisce l'intervallo di variazione naturale dovuta alle cause comuni.

Spesso nelle carte di controllo, le linee sono disegnate anche ad uno e a due deviazioni standard per aiutare a verificare la presenza di cause speciali anche quando il processo procede all'interno dei suoi limiti di controllo.

Bisogna infine aggiungere che non esiste alcuna relazione matematica o statistica tra limiti di specifica e limiti di controllo.

1.6 Classificazione dei processi

Sulla base di quanto evidenziato dalle carte di controllo un qualsiasi processo può essere classificato in uno dei quattro seguenti stati in funzione della sua stabilità e della conformità alle specifiche dei materiali prodotti.

Il processo "ideale" è stabile (opera all'interno dei limiti di controllo LCL e UCL) e quasi sempre produce materiale all'interno dei limiti di specifica (LSL e USL). In questo caso, l'SPC viene usato per mantenere lo stato ideale grazie alla rilevazione delle possibili cause speciali.

Il processo "promettente" è stabile (opera all'interno di LCL e UCL), ma a volte produce materiale fuori dai limiti di specifica. Se il processo è fuori target, un semplice intervento può essere sufficiente per riportarlo in target e per produrre costantemente materiale di buona qualità. Tuttavia, se il processo è centrato nel target, allora la variazione intrinseca deve essere ridotta e ciò richiede ricerche ed esperimenti per

trovare la procedura di correzione per ridurre le cause di variazione che l'SPC di per sé non può correggere. In questa situazione, l'SPC è necessario per prevenire alterazioni e per mantenere la stabilità del processo in modo che questi esperimenti possano essere eseguiti su processi stabili.

Il processo “pericoloso” è instabile anche se produce di solito materiale all'interno dei limiti di specifica (LSL e USL). Questo stato è pericoloso perché l'instabilità del processo può eventualmente produrre materiale scadente destinato inconsapevolmente ai clienti. Dal momento che non viene prestata attenzione alla dinamica del processo, correggere il problema risulta difficile e l'SPC è necessario per questo tipo di processo, per ridurre la variazione e per conoscerlo meglio.

Il processo “turbolento” è instabile e spesso produce materiale fuori dalle specifiche. L'instabilità smaschera la vera natura del processo e così l'SPC viene usato per stabilizzare il processo prima che vengano provati dei possibili miglioramenti. E' da notare che un processo deve essere stabile prima che possa significativamente migliorarsi e gli sforzi sono orientati per avvicinarsi al target e per ridurre la variazione.

1.7 Fasi del monitoraggio

Il controllo di qualità svolto su un processo produttivo presuppone la conoscenza degli standard su ciascuna caratteristica del prodotto presa in esame, che viene trattata come una variabile casuale. La sorveglianza di un processo consiste:

- nel verificare se c'è stato un cambiamento
- nel dire quale parametro è cambiato
- nell'individuare l'istante temporale in cui è avvenuto il cambiamento e stimare la sua grandezza

Le operazioni che si svolgono nella Fase I della sorveglianza sono volte a caratterizzare la distribuzione in controllo della caratteristica in esame. In particolare consistono nella:

- fase esplorativa (EDA): per verificare l'indipendenza delle osservazioni, ed eventualmente individuare la struttura di correlazione nelle osservazioni, e la stabilità della media e della varianza
- stima dei parametri della distribuzione in controllo della caratteristica in esame
- stima dei limiti con cui confrontare il livello corrente del processo

Il controllo in Fase II ha lo scopo di verificare se ci sono scostamenti rilevanti dagli standard. In particolare, nella Fase II si sorveglia la produzione estraendo osservazioni e confrontandole con i limiti di controllo stimati nella Fase I.

CAPITOLO 2

Carte di controllo non parametriche

2.1 Introduzione

Le proprietà statistiche delle carte di controllo parametriche funzionano al meglio, solo se le soddisfatti l'assunzione della normalità della distribuzione sottostante del processo composto da osservazioni indipendenti e identicamente distribuite (i.i.d.).

Se questo requisito viene violato con deviazioni estreme, i risultati possono risultare falsati perché si può verificare la riduzione del tasso dell'errore di tipo I, ma a compagna la sostanza è la riduzione di potenza.

Si ritiene che l'uso delle carte di controllo non parametriche per distribuzioni libere (*non parametric distribution-free charts*) sia particolarmente utile nel caso in cui c'è la violazione dell'ipotesi di normalità in presenza di campane subitaneamente grandi.

Inoltre, in alcune applicazioni possono sorgere dei dubbi sulla adeguatezza delle assunzioni distributive e sul potenziale impatto sulla prestazione della carta per un'errata decisione della distribuzione.

Se le carte di controllo non parametriche vengono usate anche nelle situazioni in cui il teorema del limite centrale non può essere invocato, che si verifica quando si ha la distribuzione sottostante del processo che non è subitaneamente nota per il fatto che si ha una distribuzione molto piccola, o nel caso in cui esse hanno singole osservazioni.

D'altra parte, l'assunzione di normalità è ragionevole, e quindi la procedura di controllo basata sulla media del processo di quattro o cinque osservazioni è adatta in altre circostanze, per ragioni tecniche, economiche, o perché è richiesto solamente una singola osservazione ad ogni istante.

Queste carte di controllo non parametriche superano questi limiti concettuali, proprio per il fatto che non dipendono dalla normalità, o più in generale, da alcuna assunzione sulla distribuzione della caratteristica di qualità.

Per questo motivo, nella realtà industriale risultano più applicabili rispetto ai metodi parametrici, soprattutto nelle situazioni di start-up e short-run.

Infatti, i metodi non parametrici si usano nei contesti in cui le caratteristiche delle distribuzioni sono difficili da verificare o da soddisfare.

I vantaggi che derivano dall'uso di queste carte non parametriche possono essere così riassunti:

- flessibilità cioè la non necessità di ipotizzare una distribuzione di probabilità per il processo
- capacità di individuare cambiamenti nel processo quando la distribuzione dei dati non è normale, in particolare in presenza di code pesanti
- più potenza rispetto ai tradizionali test quando i campioni sono generati da distribuzioni altamente asimmetriche
- distribuzione per la run length in controllo uguale per tutte le distribuzioni continue
- robustezza contro i valori anomali

In questi ultimi anni, le carte di controllo non parametriche hanno attratto l'attenzione di molti ricercatori che si sono maggiormente focalizzati nel monitoraggio e nell'individuazione di cambiamenti nei parametri di posizione di un processo (media, mediana o percentili della distribuzione).

Altre caratteristiche sulla distribuzione, come la scala e la forma, vengono opportunamente visionate anche perché sono importanti indicatori di qualità, infatti se si riscontra un incremento della varianza, questo spesso può corrispondere ad un deterioramento del processo che richiede puntuali e corrette misure.

Data la numerosità di questi studi, saranno di seguito elencati soltanto alcuni approcci non parametrici che verranno considerati nel proseguo della tesi:

- Hackl e Ledolter (1991) hanno sviluppato una carta EWMA basata sui ranghi standardizzati per singole osservazioni, denominata HLE ²
- McDonald (1990) ha proposto una carta CUSUM basata sui ranghi sequenziali per singole osservazioni, denominata McE ³
- procedura self-starting della HLE, denominata SHL



²D. Hackl, J. Ledolter (1991) – “A Control Chart Based on Ranks”, *Journal of Quality Technology*, vol. 23, n. 2, pp. 117-124

³D. McDonald (1990) – “A CUSUM Procedure Based on Sequential Ranks”, *Naval Logistic Research*, vol. 37, pp. 627-646

La carta di controllo non parametrica che verrà discussa in questa tesi è la carta EWMA basata sul rapporto di verosimiglianza denominata NLE. Essendo distribution-free, per definizione non si richiede la conoscenza della distribuzione sottostante.

Questa carta si basa sull'integrazione di un test non parametrico sulla bontà di adattamento (goodness-of-fit test) e sulla carta di controllo EWMA (Lucas and Saccucci 1990).⁴

Si evita la raccolta di una lunga serie di dati in Fase I prima di usare la carta, anche se in genere è necessario e consigliabile raccogliere un campione di osservazioni preliminari stabili, tipo $m_0 \geq 25$.

La carta è computazionalmente veloce, conveniente nell'usarla, ed efficace nello scoprire spostamenti potenziali in posizione, scala e forma e per questo motivo, offre una robusta protezione contro la variazione in varie distribuzioni sottostanti.

Per dimostrare ciò, nel capitolo 4, verrà confrontata la performance della carta NLE sia con schemi parametrici, sia con le carte non parametriche precedentemente elencate, in particolare con gli schemi HLE e McE.

2.2 La carta NLE

Per il disegno della carta NLE, si considerano m_0 osservazioni storiche, indipendenti e identicamente distribuite (i.i.d.) X_{-m_0+1}, \dots, X_0 tratte dalla Fase I.

La distribuzione della t-esima futura osservazione, X_t , è sintetizzata dal seguente modello di change-point:

$$X_t \stackrel{\text{i.i.d.}}{\sim} \begin{cases} F_0(x), & \text{for } t = -m_0 + 1, \dots, 0, 1, \dots, \tau, \\ F_1(x), & \text{for } t = \tau + 1, \dots, \end{cases} \quad (1)$$

dove F_0 e F_1 , con il punto di cambiamento τ e le funzioni di distribuzione in controllo (IC) e fuori controllo (OC), tutti assunti incogniti.

La carta NLE è legata al test statistico non parametrico sulla bontà di adattamento (GOF) di Zhang (2002) basato sul rapporto di verosimiglianza sequenziale.

I test di bontà di adattamento permettono di valutare la compatibilità di un campione con una distribuzione teorica, in altre parole, determina quanto è probabile che questa



⁴ Lucas, Saccucci (1990) – “Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements”, *Technometrics*, vol. 32, pp. 1-294

distribuzione abbia generato il campione. In particolare, i test goodness-of-fit non parametrici non hanno assunzioni restrittive sulla forma della distribuzione sottostante. Zhang (2002) ha proposto un nuovo approccio non parametrico per parametrizzazione per costruire un test GOF generale basato sul log-rapporto di verosimiglianza, che non solo genera i tradizionali test (Kolmogorov - Smirnov, Anderson - Darling, e Cramér - von Mises) ma produce anche nuovi tipi di omnibus test che sono generalmente molto più potenti rispetto ai precedenti.

Di seguito il test viene descritto nei due casi in cui F_0 , la distribuzione in controllo, sia supposta nota o incognita.

2.2.1 Distribuzione in controllo $F_0(x)$ nota

Sia X_1, \dots, X_n un campione casuale da X , che è una variabile casuale continua con funzione di distribuzione $F(x)$. Si vuole saggiare l'ipotesi nulla $H_0: F(x) = F_0(x)$ per ogni $x \in (-\infty, \infty)$ contro $H_1: F(x) \neq F_0(x)$ per alcune $x \in (-\infty, \infty)$.

Ciò è equivalente a H_{0v} , che è $F(v) = F_0(v)$ contro H_{1v} , che è $F(v) \neq F_0(v)$ per ogni $v \in (-\infty, \infty)$.

L'approccio di Zhang consiste nel combinare tutti i log-rapporti di verosimiglianza valutati nelle osservazioni X_i , $i = 1, \dots, n$, incorporando un appropriato peso (α_i) per ogni X_i , per avere $Z = \sum \alpha_i G_{X_i}$.

Esso si basa sul seguente log-rapporto di verosimiglianza (Einmahl e McKeague, 2003):

$$G_u = n \left\{ F_n(u) \ln \left(\frac{F_n(u)}{F_0(u)} \right) + [1 - F_n(u)] \ln \left(\frac{1 - F_n(u)}{1 - F_0(u)} \right) \right\} \quad (2)$$

dove si considera un campione binario con $F_n(u)$ come probabilità di successo e $F_n(u)$ è la funzione di distribuzione empirica (e.d.f.) del campione $\{X_1, \dots, X_n\}$ ovvero

$$F_n(u) = n^{-1} \sum_{j=1}^n I\{X_j \leq u\}.$$

Usando in particolare $\alpha_i = [F_n(X_i)(1 - F_n(X_i))]^{-1}$, Zhang (2002) suggerisce il test seguente:

$$Z_A = \sum_{i=1}^n \omega_i \left\{ \frac{1}{1 - F_n(X_i)} \ln \left(\frac{F_n(X_i)}{F_0(X_i)} \right) + \frac{1}{F_n(X_i)} \ln \left(\frac{1 - F_n(X_i)}{1 - F_0(X_i)} \right) \right\} \quad (3)$$

con $\omega_i = 1$ per ogni i .

Valori elevati di Z_A portano al rifiuto dell'ipotesi nulla H_0 e corrispondono alle osservazioni più estreme, cioè lontane dalla mediana, associate a valori maggiori di α_i .

Un approccio alternativo sta nel considerare invece la seguente funzione di distribuzione empirica pesata in ciascun punto u :

$$F_n^{(\lambda)}(u) = a_{\lambda,n}^{-1} \sum_{j=1}^n (1-\lambda)^{n-j} I_{\{X_j \leq u\}} \quad \text{con} \quad a_{\lambda,n} = \sum_{j=1}^n (1-\lambda)^{n-j} \quad (4)$$

dove λ è un parametro di lisciamento analogo a quello che viene utilizzato nella carta EWMA.

Quando la t -esima futura osservazione X_t , viene raccolta, alla luce di $F_n^{(\lambda)}(u)$, si calcola la statistica di controllo Z_t , sostituendo in Z_A , $F_n(X_i)$ con $F_i^{(\lambda)}(X_i)$ e usando $\omega_i = \lambda(1-\lambda)^{t-1}$, ovvero:

$$Z_t = \sum_{i=1}^t \lambda(1-\lambda)^{t-i} \left\{ \frac{1}{1 - F_i^{(\lambda)}(X_i)} \ln \left(\frac{F_i^{(\lambda)}(X_i)}{F_0(X_i)} \right) + \frac{1}{F_i^{(\lambda)}(X_i)} \ln \left(\frac{1 - F_i^{(\lambda)}(X_i)}{1 - F_0(X_i)} \right) \right\}$$

che è equivalente a:

$$Z_t = (1-\lambda)Z_{t-1} + \lambda Y_t \quad t=1,2,\dots \quad (5) ,$$

con $Z_0=0$ dove:

$$Y_t = \frac{1}{1 - F_t^{(\lambda)}(X_t)} \ln \left(\frac{F_t^{(\lambda)}(X_t)}{F_0(X_t)} \right) + \frac{1}{F_t^{(\lambda)}(X_t)} \ln \left(\frac{1 - F_t^{(\lambda)}(X_t)}{1 - F_0(X_t)} \right)$$

La statistica Z_t nell'equazione (5), opera in modo analogo alla convenzionale carta EWMA parametrica per variabili normali (Lucas e Saccucci, 1990) per il fatto che fa uso di tutte le osservazioni disponibili fino al tempo corrente t cambiando esponenzialmente il peso assegnato alle osservazioni passate.

Fissando $\omega_i = \lambda(1-\lambda)^{t-1}$, si combinano tutte le Y_t in modo che le osservazioni più recenti possano indicare il cambiamento più accuratamente e facilmente. A differenza di quella parametrica, lo sforzo computazionale per calcolare le Y_t , cresce in modo sequenziale con il tempo t .

Quindi, la carta proposta combina il lisciamento esponenziale all'interno della funzione di distribuzione empirica delle osservazioni e aggiorna la stima della sconosciuta distribuzione IC.

Si osservi che conoscere $F_0(x)$, equivale ad assumere che m_0 è sufficientemente grande e in questo caso si potrebbe direttamente sostituire $F_0(X_t)$ nell'equazione (5) con la funzione di distribuzione empirica $F_{-m_0,0}(X_t) = m_0^{-1} \sum_{j=-m_0+1}^0 I_{\{X_j \leq X_t\}}$.

2.2.2 Distribuzione in controllo $F_0(x)$ incognita

Il disegno di una carta di controllo prevede l'uso dei parametri del processo IC che si presume siano conosciuti esattamente, ma in pratica raramente sono ben conosciuti e spesso sono sostituiti con delle stime calcolate su un campione IC, detto di Fase I.

Il concetto principale è che con grandi campioni, le stime sono vicine ai veri parametri. E' stato dimostrato tuttavia che inevitabilmente le prestazioni della carta di controllo diminuiscono sia in termini di run-length IC che in quelli di run-length OC.

Esistono contesti applicativi, quali le produzioni di basso volume, in cui prendere campioni sufficientemente grandi per poter stimare i parametri, è praticamente impossibile.

Per far fronte alla situazione di non avere a disposizione un dataset vasto, sono stati sviluppati metodi self-starting che gestiscono il controllo sequenziale aggiornando simultaneamente le stime dei parametri e verificando le condizioni OC che possono presentarsi (Hawkins e Quesenberry).

I metodi self-starting rimuovono la necessità del campione di Fase I, poiché il monitoraggio del processo inizia quasi subito e non smettono mai di aggiornare i parametri e di aumentare la precisione delle stime con l'aggregarsi dei dati.

La procedura suggerita da Zhang nel caso di F_0 incognita, consiste nel $F_{-m_0, t-1}(X_t) = (m_0 + t - 1)^{-1} \sum_{j=-m_0+1}^{t-1} I_{\{X_j \leq X_t\}}$ in sostituzione di $F_0(X_t)$ in Z_t per $t \geq 1$, ottenendo la seguente statistica con m_0 osservazioni di riferimento:

$$\begin{aligned} \tilde{Z}_t &= (1 - \lambda)\tilde{Z}_{t-1} + \lambda\tilde{Y}_t, \\ \tilde{Y}_t &= \frac{1}{1 - F_t^{(\lambda)}(X_t)} \ln \left(\frac{F_t^{(\lambda)}(X_t)}{F_{-m_0, t-1}(X_t)} \right) + \frac{1}{F_t^{(\lambda)}(X_t)} \ln \left(\frac{1 - F_t^{(\lambda)}(X_t)}{1 - F_{-m_0, t-1}(X_t)} \right) \quad (6) \end{aligned}$$

La corrispondente carta di controllo lancia un segnale se:

$$\tilde{Z}_t > L_t,$$

dove $L_t > 0$ è una sequenza di limiti di controllo dinamici scelta per ottenere una specifica distribuzione della run-length IC. Si usano i limiti dinamici perché se si usassero quelli fissi, le probabilità di falsi allarmi potrebbero significativamente aumentare in presenza di short-runs.

La carta definita dalla (5) o dalla (6), a seconda che F_0 sia nota o incognita, è denominata carta EWMA non parametrica basata sul rapporto di verosimiglianza, nel seguito denotata con NLE.

La carta NLE è in grado di rilevare uno spostamento in entrambe le direzioni, cioè sia un incremento sia un decremento, in posizione, scala e forma nei parametri.

Come si può vedere, la statistica di controllo \tilde{Z}_t è non parametrica perché la statistica \tilde{Y}_t usa solo l'informazione sul rango delle X_t ma non le loro dimensioni.

Infatti, dalla definizione di \tilde{Z}_t , si può vedere che da $t \geq 1$, la distribuzione della run-length IC della carta NLE è determinata dalle densità $f_{\tilde{Y}_1}, f_{\tilde{Y}_2|\tilde{Y}_1}, \dots, f_{\tilde{Y}_t|\tilde{Y}_{i < t}}$ dove $f_{x|y_1, \dots, y_k}$ denota la densità condizionata di x dato y_1, \dots, y_k .

Da notare che $\{\tilde{Y}_1, \dots, \tilde{Y}_t\}$ sono determinate dai ranghi di $\{X_1, \dots, X_t\}$ nel campione aggregato $\{X_{-m_0+1}, \dots, X_t\}$, che sono libere dalla distribuzione IC sottostante del processo essendo X_{-m_0+1}, \dots, X_t i.i.d. distribuite come F_0 . Per deduzione, si ricava che, $f_{\tilde{Y}_1}, f_{\tilde{Y}_2|\tilde{Y}_1}, \dots, f_{\tilde{Y}_t|\tilde{Y}_{i < t}}$ sono anche loro libere da $F_0(x)$, che produce il seguente risultato.

Pertanto una carta NLE è *distribution-free*, perché la distribuzione della run-length IC è la stessa per tutte le distribuzioni continue della caratteristica di interesse.

Questo è particolarmente utile nel determinare i limiti di controllo perché, per ogni distribuzione continua, i limiti L_t per ottenere la desiderata distribuzione della run-length IC sono gli stessi.

In altre parole, i limiti di controllo possono essere scelti indipendentemente da F_0 per ottenere una desiderata distribuzione della IC run-length.

In pratica, potrebbe essere più conveniente considerare la statistica normalizzata Z_t/L_t , e in tali casi, il limite di controllo normalizzato è una costante pari a 1.

2.3 Osservazioni raggruppate

Quando un gruppo di g osservazioni $\{X_{j1}, \dots, X_{jg}\}$, vengono prese sequenzialmente dal processo ad ogni t , la carta NLE può essere prontamente definita in modo simile all'equazione (6) usando la seguente funzione di distribuzione empirica pesata modificata:

$$F_{-m_0, k}^{(g)}(u) = \frac{1}{(k + m_0)g} \sum_{j=-m_0+1}^k \sum_{i=1}^g I_{\{X_{ji} \leq u\}}$$

$$F_k^{(\lambda, g)}(u) = a_{\lambda, k, g}^{-1} \sum_{j=1}^k (1 - \lambda)^{k-j} \sum_{i=1}^g I_{\{X_{ji} \leq u\}}$$

con:

$$a_{\lambda, k, g} = g \sum_{j=1}^k (1 - \lambda)^{k-j}$$

così la statistica della carta NLE per osservazioni raggruppate è data dall'equazione (6)

con:

$$\tilde{Y}_t^{(g)} = \sum_{i=1}^g \frac{1}{1 - F_t^{(\lambda, g)}(X_{ti})} \ln \left(\frac{F_t^{(\lambda, g)}(X_{ti})}{F_{-m_0, t-1}^{(g)}(X_{ti})} \right) + \frac{1}{F_t^{(\lambda, g)}(X_{ti})} \ln \left(\frac{1 - F_t^{(\lambda, g)}(X_{ti})}{1 - F_{-m_0, t-1}^{(g)}(X_{ti})} \right)$$

invece di \tilde{Y}_t , dove si segue ancora il principio che le osservazioni nello stesso gruppo hanno gli stessi pesi e i gruppi più recenti hanno maggior peso.

CAPITOLO 3

Progettazione e implementazione della carta NLE

3.1 Determinazione dei limiti di controllo

Come Z_A , la distribuzione nulla della statistica proposta, \tilde{Z}_t , è discreta per ogni t dato ma anche se si ottiene la completa distribuzione per ogni dato λ , t , e m_0 , la distribuzione della run-length IC della carta NLE è ancora sconosciuta perché dipende dalla distribuzione condizionata $f_{\tilde{Z}_t | \tilde{Z}_i, i \leq t-1}$.

Per una data probabilità di falso allarme, α , i limiti di controllo L_t possono essere approssimati da valori che soddisfano:

$$\begin{aligned} \Pr\left(\tilde{Z}_t > L_t \mid \tilde{Z}_i \leq L_i, 1 \leq i < t\right) &= \alpha, \quad \text{for } t > 1, \\ \Pr\left(\tilde{Z}_1 > L_1\right) &= \alpha. \end{aligned} \quad (7)$$

Ovviamente, queste probabilità dovrebbero essere interpretate come frequenze relative su molte simulazioni ripetute.

Da notare che, nel caso F_0 sia assunta nota, il procedimento sopra descritto è ancora applicabile salvo che si usi rispettivamente la formula (5) e Z_t invece della formula (6) e \tilde{Z}_t .

Anche se la carta NLE è uno schema self-starting e quindi può essere implementata allo start-up di un processo, si consiglia comunque di avere a disposizione prima del monitoraggio, tramite l'analisi di Fase I, un numero modesto di osservazioni storiche (almeno $m_0 \geq 25$) per ottenere almeno una prima verifica che il processo è stato effettivamente stabile, e in seguito procedere con il disegno della carta NLE.

Esaminando la definizione di $F_t^{(\lambda)}$ nell'equazione (4), i possibili valori di \tilde{Z}_t per t molto piccoli, ad esempio $t=1$ o 2 , sono molto limitati quando m_0 è piccolo, quindi si potrebbe non essere in grado di calcolare L_t per $t=1,2$ per ottenere un desiderato α (errore di I tipo).

Utilizzando i metodi self-starting, per la quale non si dispone della Fase I e della Fase II, servono almeno due osservazioni ossia $t \geq 2$.

In questo caso per risolvere il problema di trovare L_1 o L_2 , si consiglia di prendere le ultime due osservazioni storiche X_{-1} e X_0 delle m_0 osservazioni storiche trattandole come le prime due pseudo-future osservazioni. Successivamente, si calcolano le corrispondenti statistiche test \tilde{Z}_t per $t=-1,0$ ma si inizia a sorvegliare il processo dopo che la prima osservazione della Fase II è stata raccolta.

In pratica, i valori di \tilde{Z}_t relativi alle prime due pseudo-future osservazioni, vengono usati per calcolare i possibili valori di \tilde{Z}_1 e \tilde{Z}_2 sufficienti per ottenere i limiti di controllo.

Sulla base dei risultati empirici in Hawkins (2003), i limiti L_t convergono gradualmente ad una costante quando t aumenta.

Dato lo sforzo computazionale dovuto anche al procedimento self-starting, usando diverse combinazioni di valori di λ e della ARL_0 , Zou e Tsung (2010) forniscono i limiti di controllo della carta NLE per valori di α pari a 0.005, 0.0027 e 0.002, che corrispondono, ad una ARL_0 di 200, 370 e 500, quando F_0 è conosciuta (usando la formula (5))⁴, e, se m_0 è sufficientemente grande, i limiti di controllo della carta NLE (usando la formula (6)) per ARL_0 di 370 e 500, per varie combinazioni di (λ, m_0) .⁵

In particolare, sono presi in considerazione i valori di m_0 pari a 25, 50, 100, e 200 mentre per il parametro di lisciamiento, i valori di $\lambda=0.05$ e $\lambda=0.1$.

I limiti di controllo tabulati in queste tabelle possono essere usati per ciascuna distribuzione continua perché la carta NLE è distribution-free.

3.2 Scelta del parametro λ

In generale, la scelta di λ in una comune EWMA dipende dall' ampiezza del cambiamento che si desidera individuare nel più breve tempo possibile.

In particolare, un λ più piccolo porta ad aumentare la sensibilità della carta ai piccoli cambiamenti, mentre un λ più grande permette una rilevazione più rapida di grandi spostamenti (Lucas e Saccucci, 1990). Se λ viene scelto sufficientemente piccolo (max



⁴C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pag. 1794

⁵C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pp. 180-1815

0.05), si ottiene una carta EWMA robusta alle condizioni di normalità ma si deve comunque tener conto che diminuendo il suo valore, si perde sensibilità nel rilevare grandi spostamenti.

Queste affermazioni sono valide anche per la carta NLE e l'effetto di λ sulla performance della carta è rilevante.

Gli autori hanno fornito delle figure che mostrano le ARL_{δ} Steady-State (in scala logaritmica) della NLE sotto lo scenario $N(0,1)$ contro $N(\delta,1)$ per $\delta \neq 0$, con i valori di λ pari a 0.05, 0.1, 0.2 e 0.5 quando F_0 è nota e quando $m_0=100$, usando in entrambi i casi $\tau=50$, $ARL_0=370$ e 20.000 replicazioni.⁶

Si può constatare che in entrambi i casi, un λ più piccolo porta ad una rilevazione più rapida di piccoli spostamenti ed un λ più grande funziona meglio nel rilevare spostamenti più consistenti e ciò è coerente con le proprietà della convenzionale carta EWMA (Lucas e Saccucci, 1990) e la sua versione self-starting (Quesenberry, 1995).

Sulla base di risultati empirici, una scelta ragionevole consiste nel scegliere λ appartenente all'intervallo [0.05,0.2].

3.3 On post-signal diagnostic

Nella pratica del controllo di qualità, oltre a rilevare velocemente un cambiamento nel processo, è anche fondamentale diagnosticare il cambiamento e identificare se c'è stato uno spostamento in posizione, scala, o in entrambi dopo che si è verificato un segnale di OC.

Uno stimatore del change-point basato sul test non parametrico Z_A viene proposto da Zou e Tsung (2002) per contribuire all'aumento di efficienza della carta NLE.

Si assume che la carta segnali alla k -esima osservazione, che cioè ci siano m_0 osservazioni storiche IC e k future osservazioni, ed uno spostamento si sia verificato dopo il τ -esimo campione ($0 \leq \tau < k$) come illustrato dal modello (1).

⁶C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pag. 1826

Data una stima del change-point v , si può ricavare la seguente versione generalizzata di Z_A :

$$Z_A(v, k) = \sum_{i=v+1}^k \left\{ \frac{1}{1 - F_{v,k}(X_i)} \ln \left(\frac{F_{v,k}(X_i)}{F_{-m_0,v}(X_i)} \right) + \frac{1}{F_{v,k}(X_i)} \ln \left(\frac{1 - F_{v,k}(X_i)}{1 - F_{-m_0,v}(X_i)} \right) \right\} \quad (8)$$

dove $F_{k_1, k_2}(X_i)$ è definita, come prima, da:

$$(k_2 - k_1)^{-1} \sum_{j=k_1+1}^{k_2} I_{\{X_j \leq X_i\}}$$

La statistica $Z_A(v, k)$ può essere intuitivamente vista come un test statistico basato su Z_A per verificare che il campione $\{X_i, i=v+1, \dots, k\}$ ha la stessa distribuzione di $F_{-m_0, v}(\cdot)$ dove $F_{-m_0, v}$ può essere considerata come una stima di F_0 .

Allora lo stimatore suggerito del change-point τ , è dato da:

$$\hat{\tau} = \arg \max_{0 < v < k} \{Z_A(v, k)\} \quad (9)$$

che in questo contesto viene utilizzato in un'applicazione SPC on-line, ma in letteratura viene usato anche per la rilevazione off-line di change-point e inoltre risulta asintoticamente efficace.

Dopo aver ottenuto una stima dell'istante di cambiamento nel parametro di posizione, si hanno $(k - \hat{\tau})$ osservazioni trattate come OC.

Tra queste $(k - \hat{\tau})$ osservazioni, tuttavia alcune potrebbero effettivamente essere osservazioni IC.

Tuttavia, come Reynolds e Stoumbos (2005) hanno sottolineato, le carte di controllo utilizzate per la diagnostica, non necessariamente devono essere le stesse carte di controllo che sono utilizzate per individuare un cambiamento.

Per questo motivo, si suggerisce di usare metodi basati su test non parametrici come uno strumento ausiliario per determinare quali parametri hanno subito un cambiamento dopo che la carta ha lanciato il segnale.

3.4 Altre carte di controllo non parametriche

Qui di seguito vengono brevemente elencati alcuni test non parametrici sui quali si basano le carte di controllo non parametriche scelte per le comparazioni.

Parametric		Non parametric using ranks
Test on Pearson's $\rho = 0$	➔	Test on Kendall's $\tau = 0$ Test on Spearman's $\rho_S = 0$
t -test on two independent samples	➔	Wilcoxon-Mann-Whitney test
1-Way ANOVA	➔	Kruskal-Wallis test
2-Way ANOVA	➔	Friedman test

6

La maggior parte di questi si basa sui ranghi.

Il rango definisce la posizione di un'osservazione all'interno di un dataset di n dati messi in ordine crescente partendo dal rango 1 fino al rango n . Per esempio, se si considerano questi dati numerici 3.4, 5.1, 2.6, 7.3, i loro ranghi sono rispettivamente 2, 3, 1 e 4.

Se nel campione aggregato si trovano due o più valori uguali, il valore del rango è dato dalla media aritmetica delle posizioni che avrebbero nel dataset ordinato.

I ranghi godono della proprietà di rimanere immutati quando si verifica una trasformazione monotona della variabile numerica come la modifica della sua scala.

Nel seguito vengono considerati i seguenti due test non parametrici a due campioni basati sui ranghi: il Wilcoxon rank sum test (o chiamato Mann-Whitney-Wilcoxon test, Conover, 1999) per il parametro di posizione e l'aligned rank scale test (Fligner e Killeen, 1976), indicati rispettivamente come TW e TA.

In questo contesto, si considerano due campioni che corrispondono al campione di dati raccolto prima del change-point $\hat{\tau}$, ossia per $j = -m_0 + 1, \dots, \hat{\tau}$, e il campione di dati provenienti dopo il presunto cambiamento $\hat{\tau}$ cioè per $j = \hat{\tau} + 1, \dots, k$.

3.4.1 Wilcoxon rank sum test

Il test di Wilcoxon-Mann-Whitney (o test di Wilcoxon della somma dei ranghi) si applica nel caso in cui si chiede di confrontare le mediane di due campioni che non seguono una distribuzione normale, per rilevare uno spostamento in posizione.⁷

Il Wilcoxon rank sum test si applica nel caso si abbiano numerosi dati ($n \geq 20$ e $m \geq 20$), anche se i due campioni non hanno bisogno di avere lo stesso numero di osservazioni.

Il test viene usato per testare l'ipotesi nulla H_0 , secondo la quale i due campioni indipendenti provengono dalla stessa popolazione (o popolazioni identiche).

L'ipotesi nulla H_0 si riferisce solo all'uguaglianza delle mediane delle distribuzioni, e quindi richiede che le due distribuzioni, sebbene arbitrarie, abbiano la stessa forma.

Il sistema d'ipotesi per il Wilcoxon Rank Sum test con ipotesi alternativa bilaterale, risulta quindi:

H_0 : i due campioni hanno la stessa distribuzione di probabilità

H_1 : la distribuzione di probabilità per la popolazione 1 è spostata a sinistra/destra rispetto alla 2

Dato un set di n dati per la prima popolazione ed un set di m dati per la seconda $\{X_{1j}, \dots, X_{nj}; j=1,2\}$, tutti i dati vengono uniti e messi in ordine crescente a prescindere dalla popolazione di partenza e viene attribuito il rango agli $n+m$ valori partendo dal più piccolo fino al più grande.

Il passo successivo consiste nel sommare i ranghi per ciascun campione: $R_j = \sum_{i=1}^{n_j} r_{ij}$ per $j=1,2$ dove r_{ij} è il rango della osservazione X_{ij} .

Per ottenere il valore della statistica test che risulta essere diverso a seconda di quale campione si prende in considerazione, dopo aver ottenuto R_1 , a questo valore si sottrae il risultato che si ottiene dalla seguente formula: $N_1(N_1+1)/2$ (dove N_1 è la numerosità campionaria del gruppo 1).



⁷ S. Chakraborti, Van de Wiel (2008) – “A Nonparametric Control Chart Based on the Mann-Whitney Statistic”. *IMS Collections. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, 1, pp. 156-1727

Per grandi campioni, la statistica Wilcoxon Rank Sum $W=R_2$ può essere approssimata da una distribuzione normale. In particolare:

$$W^* = \frac{R_2 - [n_2(n_1 + n_2 + 1)/2]}{[n_1 n_2 (n_1 + n_2 + 1)/12]^{1/2}}$$

In questo modo la statistica W^* ha una distribuzione approssimata normale standard. Per verificare l'ipotesi nulla contro l'ipotesi alternativa bilaterale con livello di significatività α , si rifiuta l'ipotesi nulla se $|W^*| \geq z_{\alpha/2}$, dove il valore critico $z_{\alpha/2}$ può essere trovato nella tavola della distribuzione della normale standard.

3.4.2 Fligner-Killeen test

Il Fligner-Killeen test è stato considerato in uno studio, come uno dei test più robusti per l'omogeneità delle varianze tra campioni indipendenti, in presenza di variazioni dalla normalità.⁸

Il sistema d'ipotesi per il Fligner-Killeen Test consiste in:

H_0 : i campioni provengono da popolazioni con una comune distribuzione e quindi con stesse varianze

H_1 : le distribuzioni hanno la stessa media (o mediana) ma differente scala e quindi diverse varianze

Per verificare l'omogeneità delle varianze delle popolazioni, Fligner e Killeen (1976) hanno proposto di classificare i valori assoluti $|X_{ij}|$ e assegnare punteggi crescenti come $\alpha_{N,i} = i$, $\alpha_{N,i} = i^2$, o $\alpha_{N,i} = \Phi^{-1}((1+i/(N+1))/2)$ sulla base dei ranghi di tutte le osservazioni. Conover, Johnson e Johnson (1981) hanno modificato il Fligner-Killeen test e hanno suggerito di classificare congiuntamente $|X_{ij} - \tilde{X}_j|$ invece di $|X_{ij}|$, dove \tilde{X}_j è la mediana del campione della j -esima popolazione e di assegnare a questi, punteggi crescenti.

Questa versione del Fligner-Killeen test è chiamata median-centering Fligner-Killeen test.

Dopo aver assegnato i punteggi $\alpha_{N,i}$ ai $|X_{ij} - \tilde{X}_j|$, il Fligner-Killeen test modificato può essere calcolato come:

$$X^2 = \sum_{j=1}^k n_j (\bar{A}_j - \bar{a})^2 / V^2$$



⁸Fligner, Killeen (1976) - "Distribution-Free Two-Sample Test for Scale". *Journal of the American Statistical Association*, 71, pp. 210-213.

dove \bar{A}_j è il punteggio medio (*mean score*) per il campione j-esimo, \bar{a} è il punteggio medio globale (*overall mean score*) e V^2 è la varianza del campione di tutti i punteggi.

Precisamente:

$$N = \sum_{j=1}^k n_j$$

$$\bar{A}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} a_{N,m_i} \quad \text{dove } a_{N,m_i} \text{ è il rank score crescente per la } i\text{-esima osservazione nel } j\text{-esimo campione}$$

$$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_{N,i}, \text{ e } V^2 = \frac{1}{N-1} \sum_{i=1}^N (a_{N,i} - \bar{a})^2$$

Il Fligner-Killeen test modificato risulta molto robusto contro variazioni dalla normalità. Per campioni di grandi dimensioni, la statistica test Fligner-Killeen modificata X^2 ha una distribuzione asintotica χ^2 con (k-1) gradi di libertà.

CAPITOLO 4

Valutazione della performance numerica

In questo capitolo, verrà confrontata la performance della carta NLE con le altre carte parametriche e non parametriche prese in esame nello studio svolto dagli autori.

Quanto segue rappresenta in buona parte un riepilogo della struttura e del contenuto dell'articolo scritto da Zou e Tsung (2010).

4.1 L'ARL per il confronto tra carte

La più comune misura per valutare le prestazioni di una carta di controllo è l'Average Run Length (ARL) ovvero il tempo medio d'attesa per segnalare un allarme, o in altre parole, l'ARL è il numero di punti che, in media, vengono raffigurati nella carta di controllo prima che venga indicata una condizione di fuori controllo, come una statistica test che cade fuori dai limiti di controllo.

Per avere una carta efficiente, si desidera avere un elevato valore dell'ARL quando il processo è in controllo per ottenere una percentuale bassa di falsi allarmi, mentre quando è fuori controllo è consigliabile avere un'ARL molto bassa così da cogliere il prima possibile il cambiamento avvenuto. Questi obiettivi non possono essere perseguiti congiuntamente e quindi si deve trovare un trade-off. In particolare, una carta di controllo è efficiente quando ci impiega poco tempo per lanciare l'allarme, ovvero ha un breve ritardo nel segnalare l'avvenuto cambiamento causato dalla causa speciale di variazione, mentre non lancia un allarme quando non è avvenuto alcun cambiamento.

Per effettuare un confronto tra diverse carte di controllo, usualmente si fissa l'ARL in controllo (ARL_0) uguale per tutte le carte ad un valore elevato, come pure il cambiamento, δ , nella media del processo da individuare nel più breve tempo possibile.

I due autori per ottenere i risultati relativi alla performance numerica della procedura NLE, hanno impostato l'ARL in controllo a $ARL_0=370$, e hanno utilizzato nella maggior parte dei casi 20.000 replicazioni.

4.2 Confronto tra NLE e le convenzionali carte parametriche

In questo paragrafo, viene comparata la performance della NLE con alcune convenzionali carte parametriche prese in esame dagli autori, progettate sotto l'assunzione di normalità.

La questione della normalità univariata è stata studiata da Borror (1999) e da Stoumbos e Reynolds (2002), con la conclusione che la non normalità può significativamente compromettere la performance della carta Shewhart. Tuttavia, le carte EWMA e CUSUM possono essere disegnate per essere robuste, scegliendo un valore piccolo del parametro di lisciamiento λ .

Per il motivo che le carte CUSUM ed EWMA hanno simili capacità di rilevamento e di robustezza dimostrate in letteratura, nel seguito si farà uso solo della carta EWMA per fare il confronto. Vengono utilizzate le osservazioni standardizzate da $X_t^* = (X_t - \mu_0) / \sigma_0$, dove μ_0 e σ_0 sono rispettivamente la media e la deviazione standard che sono note a priori o stimate dalle m_0 osservazioni storiche.

Per il confronto con gli schemi parametrici, Zou e Tsung hanno considerato tre carte:

□ classica carta EWMA (EWM) per la rilevazione di un cambiamento nella media che è definita come: $w_t = (1 - \lambda) w_{t-1} + \lambda X_t^*$

□ carta tipo EWMA (EWW) per monitorare deviazioni, definite come:

$$v_t = (1 - \lambda) \max\{1, v_{t-1}\} + \lambda X_t^{*2}$$

Di solito, si considera maggiormente la rilevazione di un incremento in σ poiché ciò corrisponde ad una riduzione in qualità, mentre un decremento può risultare interessante in alcuni processi.

□ carta proposta da Reynolds e Stoumbos (2001) per sorvegliare congiuntamente sia la media che la varianza e frutto dell'implementazione congiunta delle due carte precedenti (CEW).

Nell'articolo gli autori hanno dato particolare attenzione al confronto tra la carta NLE e la carta CEW che fornisce una maggiore qualità di informazioni utili per eliminare le cause speciali di variazione che influenzano uno spostamento in media o in varianza oppure in entrambi i parametri.

Un approccio alternativo alla carta CEW, è quello di usare due carte di controllo, una per rilevare cambiamenti nella media □ e che di solito si basa sulle medie campionarie,

e l'altra per rilevare quelli nella deviazione standard σ e che spesso si basa su una misura della dispersione di breve periodo, come la varianza campionaria ed il range campionario.

Nell'articolo di Zou e Tsung (2010) si trovano i confronti della zero-state ARL_{δ} della carta NLE e quella della CEW sotto l'ipotesi di normalità della distribuzione.⁹

La zero-state ARL si riferisce alla ARL calcolata assumendo che il cambiamento nel processo si verifichi all'inizio del monitoraggio cioè in $t=0$.

Gli autori hanno considerato per lo studio di simulazione due scenari, $N(0,1)$ contro $N(\delta,1)$ per $\delta \neq 0$ e $N(0,1)$ contro $N(0,\delta^2)$ per $\delta \neq 1$, che corrispondono rispettivamente ai problemi di monitoraggio della media e della varianza. Per entrambe le carte, hanno utilizzato tre valori di λ ($\lambda = 0.05, 0.1, 0.2$) e due valori di m_0 ($m_0 = 20.000, 200$) che corrispondono rispettivamente ai casi in cui F_0 è approssimativamente nota e quando i parametri sono stimati.

I risultati ottenuti mostrano che la carta CEW ha una maggiore efficienza nel monitorare il cambiamento nella media.

Nonostante questo, la carta NLE offre prestazioni abbastanza soddisfacenti e la differenza con la CEW non è molto rilevante se si rilevano piccoli e moderati spostamenti. Infatti, la superiorità della CEW diventa più evidente quando $m_0 = 200$ e δ è abbastanza grande, tipo $\delta \geq 2$. Infatti, se m_0 non è grande abbastanza, la carta NLE che si basa sui ranghi invece che sulla dimensione dello spostamento, non è pienamente efficiente nel segnalare velocemente un cambiamento che si è effettivamente verificato anche quando questo è abbastanza grande, dal momento che i ranghi delle osservazioni non sono in grado di crescere in modo corrispondente. Quando però m_0 è molto grande, se si considera lo stesso valore di λ , la carta NLE funziona anche un po' meglio della carta CEW nella rilevazione di spostamenti molto grandi ($\delta \geq 3$).

Rispetto agli spostamenti nella media, per rilevare in modo più efficace il cambiamento nella dispersione, e quindi nella varianza, la carta NLE richiede più osservazioni in controllo prima della realizzazione del cambiamento.

Tuttavia, se il change-point τ è grande, la carta self-starting, avendo a disposizione un maggior numero di osservazioni in controllo, è in grado di aggiornare l'informazione



⁹C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pag. 185C

sulla distribuzione IC e a sua volta mostra un miglioramento nella sua capacità di rilevazione.

Un'altra tabella presentata nell'articolo, illustra i valori ARL_{δ} della carta NLE sempre con gli stessi scenari, con diversi valori di τ ($\tau = 0, 50, 100, 200$) quando $\lambda = 0.1$ e $m_0=200$, e si può constatare che c'è un abbassamento dell' ARL_{δ} all'aumentare di τ .¹⁰

Questo risultato è molto più evidente nel caso in cui si voglia rilevare un cambiamento piccolo o grande piuttosto di uno moderato.

In coerenza con l'analisi sulla robustezza di Stoumbos e Reynolds (2000), oltre alla normale standard, gli autori Zou e Tsung si sono soffermati anche sulle seguenti distribuzioni standardizzate in modo da avere media zero e varianza uno:

(i) $t(3)$ ossia la distribuzione t di Student con tre gradi di libertà

(ii) χ^2_3 ossia la distribuzione chi-quadro con tre gradi di libertà

I risultati mostrano che la carta NLE può essere più robusta rispetto alla carta CEW, in quanto quest'ultima è composta dalla carta EWV che può essere considerata non particolarmente robusta nel caso di violazione della normalità delle osservazioni.

Assumendo tali distribuzioni Zou e Tsung (2010) forniscono i valori dell' ARL_{δ} della carta NLE per due valori di λ e m_0 ($\lambda=0.05, 0.1$; $m_0= 20.000, 200$) e per $\tau=0$.¹¹

I limiti di controllo della carta EWM vengono ricavati dagli autori via simulazione ipotizzando una distribuzione normale standard per le osservazioni.

Dai confronti si può dedurre che le carte EWM e CEW sono molto più robuste quando $\lambda=0.05$. Inoltre, quando $m_0=200$, la NLE risulta più efficiente della carta EWM nella rilevazione di piccoli e moderati cambiamenti nella media mentre la EWM ha un certo vantaggio, come previsto, per quelli grandi.

Per sorvegliare la varianza, la superiorità della NLE sulla EWM è abbastanza evidente, essendo la carta EWM non sensibile ai cambiamenti nella dispersione.



¹⁰ C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pag. 1861

¹¹ C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pp. 187-1881

4.3 Confronto tra NLE e le carte alternative non parametriche

Zou e Tsung (2010) nell'articolo hanno successivamente comparato la performance della NLE con altri metodi non parametrici alternativi.

Per il confronto hanno considerato in particolare le due carte seguenti, basate su singole osservazioni:

□ carta EWMA basata sui ranghi standardizzati (HLE), proposta da Hackl e Ledolter (1991) e presentata nel capitolo 5 per l'esempio pratico. Tale carta è definita come:

$$w_t = (1 - \lambda) w_{t-1} + \lambda R_t$$

dove $R_i = 2(m_0 + 1)^{-1} [R_i^* - (m_0 + 2)/2]$ e R_i^* è il rango di X_i .

Gli autori considerano anche una versione self-starting della carta HLE (SHL) che consiste nel sostituire la funzione di distribuzione empirica, basata sul campione di riferimento, con il suo stimatore sequenziale. Più semplicemente, se la carta non segnala dopo la t-esima osservazione, allora $F_{-m_0,0}(X_t)$ usata nella HLE viene sostituita da $F_{-m_0,t-1}(X_t)$.

□ carta CUSUM basata sui ranghi sequenziali (McE), proposta da McDonald (1990) e definita come:

$$w_t = (1 - \lambda) w_{t-1} + \lambda(R_t/(t + 1) - 0.5)$$

dove $R_i = 1 + \sum_{j=1}^{i-1} I_{\{X_j \neq X_i\}}$ è il rango sequenziale di X_i

In tutte queste carte, compresa la NLE, si sceglie come parametro di liscio $\lambda=0.1$.

Per valutare la performance complessiva di una carta, rispetto a un range di possibili valori del cambiamento δ , si calcolano anche i loro valori dell'indice RMI. L'indice RMI di una carta di controllo è stato proposto da Han e Tsung (2006) ed è definito come:

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

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 uno spostamento δ_l , 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_l .

Una carta di controllo con il più piccolo valore di RMI è considerata la migliore rispetto alle altre scelte per il confronto.

In primis, si può dedurre che la carta NLE è più efficace rispetto alle altre nella rilevazione di moderati e grandi spostamenti nella media del processo, perché anche se δ aumenta, le ARL_{δ} delle altre tre carte difficilmente cambiano quando $\delta > 3$ ovvero si stabilizzano su valori che risultano più grandi di quelli relativi alla carta NLE.¹²

Le carte di controllo basate sulle trasformazioni lineari dei ranghi, quali le carte HLE, McE e SHL, condividono questo comportamento sulla performance poiché i ranghi non crescono in modo proporzionale rispetto all'ampiezza del cambiamento.

La carta NLE invece, poiché deriva dal test non parametrico del rapporto di verosimiglianza, è una funzione non lineare dei ranghi e quindi supera questo limite.

Si possono fare altre considerazioni focalizzandosi sui piccoli spostamenti, le altre tre carte EWMA offrono generalmente una più veloce rilevazione rispetto alla NLE, sebbene la differenza è relativamente piccola specialmente quando $\tau = 150$, questo perché la NLE è una carta omnibus, mentre le altre sono disegnate per essere sensibili ai cambiamenti nel parametro di posizione. Si può intuire che la carta NLE supera con un margine abbastanza ampio le prestazioni delle altre carte nella rilevazione dei cambiamenti nella varianza. Pertanto la carta NLE è più sensibile ai cambiamenti nel parametro di scala rispetto alle altre carte basate sui ranghi. La carta NLE sembra inoltre abbastanza efficiente anche nella rilevazione di una diminuzione nella varianza.

In termini di indice RMI, la carta NLE ha la performance migliore in assoluto. Si può aggiungere che la superiorità della NLE diventa più marcata quando m_0 e τ aumentano poiché si ha a disposizione un maggior numero di osservazioni, sia storiche che nuove, dalle quali trarre informazioni.

I valori dell' ARL_{δ} , relativi alla sorveglianza della media e della varianza, quando le distribuzioni sottostanti del processo sono $t(3)$ e χ^2_3 , mostrano una maggiore efficienza della carta NLE rispetto alle altre tre carte nella rilevazione di un cambiamento nella varianza e una performance perlomeno comparabile quando il cambiamento riguarda solo la media.¹³

Altre simulazioni mostrano che la carta NLE è in grado anche di rilevare altri cambiamenti nella distribuzione, come gli spostamenti in forma.



¹¹ C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pp. 189-190

¹² C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pag. 191

Pertanto la carta NLE sembra efficiente e robusta.

4.4 Diagnostic Performance Analysis

Zou e Tsung (2010) discutono anche l'accuratezza delle stime del punto di cambiamento ottenute, dopo che la carta NLE abbia segnalato, con il test di Wilcoxon e con quello di Fligner-Killen test.

Per quantificare la precisione dello stimatore $\hat{\tau}$ descritto nell'equazione (9), gli autori tabulano oltre la mediana (MED) delle stime del change-point, le probabilità $\Pr(|\hat{\tau} - \tau| = 0)$, $\Pr(|\hat{\tau} - \tau| \leq 1)$, e $\Pr(|\hat{\tau} - \tau| < 2)$ (indicate rispettivamente come P_0 , P_1 e P_2) e le stime delle probabilità $\Pr(T_w \text{ è significativo})$ per cambiamenti solo nella media, $\Pr(T_A \text{ è significativo})$ per quelli nella dispersione e $\Pr(\text{entrambi i test sono significativi})$ per entrambi i parametri, indicate rispettivamente come PT_w , PT_A e P_{Both} .¹⁴

I risultati mostrano che lo stimatore del change-point, descritto nell'equazione (9), funziona meglio dal punto di vista della mediana poiché si avvicina al suo valore per ciascuna dimensione del cambiamento δ , e $\hat{\tau}$ ha una migliore precisione quando la dimensione δ aumenta.

Si può vedere dalle probabilità calcolate PT_w e PT_A , che $\hat{\tau}$ funziona molto meglio per monitorare cambiamenti nella media rispetto ai cambiamenti nella dispersione perché i valori di PT_w sono più elevati soprattutto quando si è in presenza di un piccolo δ .

Per quanto riguarda la combinazione dei due test TW e TA per determinare quali parametri sono stati modificati dopo il segnale, è probabile che solo TW sarà significativo quando c'è un piccolo cambiamento nella media ma entrambi i test saranno significativi nel caso di un grande cambiamento nella media. Inoltre è probabile che solo TA sarà significativo quando si realizza una variazione nella dispersione.

Inoltre, i due autori sono giunti alla conclusione che l'uso dei nonparametric post-diagnostic test è almeno altrettanto affidabile della combinazione delle due carte EWMA che formano la carta CEW nell'identificare il cambiamento avvenuto nel parametro.

Risultati tratti da simulazioni aggiuntive indicano che la stessa conclusione può valere anche nel caso in cui la media e la varianza cambino entrambe allo stesso tempo.



¹⁴C. Zou, F. Tsung (2010) – “Likelihood Ratio-Based Distribution-Free EWMA Control Charts”, *Journal of Quality Technology*, vol. 42, n. 2, pag 1921

CAPITOLO 5

Carta EWMA basata sui ranghi standardizzati

In molte situazioni reali, la distribuzione di una caratteristica di qualità è non-normale e sconosciuta.

Come già è stato detto, tra le carte di controllo non parametriche proposte in letteratura c'è la carta di controllo EWMA basata sui ranghi (Hackl e Ledolter, 1991).

I metodi non parametrici basati sui ranghi sono indipendenti dalle ipotesi distributive e, rispetto alle tradizionali tecniche delle carte di controllo, risultano meno influenzati da distribuzioni contaminate o da singole osservazioni anomale.

Questa procedura, è dunque appropriata in situazioni dove le misurazioni sono ricavate da procedure analitiche soggette ad outliers occasionali.

Il rango di un'osservazione al tempo t può corrispondere alla distribuzione in controllo della caratteristica studiata o, se questa o i suoi parametri non sono specificati, a un campione sufficientemente grande di osservazioni dal processo in controllo.

5.1 Ranghi standardizzati

Sia X_t , $t=1,2,\dots$, una sequenza di variabili casuali indipendenti da una distribuzione di probabilità continua $F(x)$.

Il rango standardizzato R_t di un'osservazione X_t , proveniente da un processo in controllo ovvero tale che la sua distribuzione di probabilità sia $F(x) = F_0(x)$, è così definito:

$$R_t = 2\left[F_0(X_t) - \frac{1}{2}\right] \quad (1)$$

I ranghi standardizzati R_t sono indipendenti e uniformemente distribuiti tra -1 e 1 con media 0 e varianza $1/3$.

Se la distribuzione in controllo non è specificata, si può utilizzare un campione di dati di dimensione $g-1$ $\{Y_1, Y_2, \dots, Y_{g-1}\}$, estratto da un processo in controllo, e i ranghi standardizzati delle X_t sono definiti come:

$$R_t = \frac{2}{g} \left(R_t^* - \frac{g+1}{2} \right) \quad (2)$$

con R_t^* il rango di X_t , rispetto a $\{Y_1, Y_2, \dots, Y_{g-1}\}$, pari a:

$$R_t^* = 1 + \sum_{i=1}^{g-1} I(X_t > Y_i)$$

dove $I(X_t > Y_i)$ è la funzione indicatrice 1 se $X_t > Y_i$ e 0 altrimenti.

I ranghi R_t seguono una distribuzione uniforme g punti $\{1/g - 1, 3/g - 1, \dots, 1 - 1/g\}$ con media 0 e varianza $\left(\frac{g^2 - 1}{3}\right)g^2$.

La serie storica dei ranghi non è più indipendente, poiché i ranghi sono calcolati a partire dallo stesso campione di riferimento $\{Y_1, Y_2, \dots, Y_{g-1}\}$.

5.2 Statistica test Z_t

Per controllare la media del processo, Hack e Ledolter (1991) suggeriscono una carta EWMA basata sui ranghi standardizzati (1) o (2). La statistica di controllo è:

$$Z_t = \lambda R_t + (1 - \lambda)Z_{t-1}$$

calcolata per $t = 1, 2, \dots$

Il valore iniziale Z_0 viene di solito posto a 0 ed il parametro di liscio λ è un valore tra 0 e 1.

La statistica di controllo è una media pesata dei ranghi precedenti, dove i pesi dati ai ranghi diminuiscono con il tempo. Come per la carta EWMA standard, un piccolo valore di λ permette una rilevazione rapida di un piccolo spostamento, mentre si usa un elevato λ per i grandi spostamenti.

Hack e Ledolter (1991) suggeriscono di usare un λ compreso tra 0.1 e 0.3 per dare maggior peso ai dati più recenti (per esempio con $\lambda=0.3$, la statistica dà l'88% del peso alle ultime sei osservazioni).

Si rileva un segnale di fuori controllo al tempo t , in particolare una deviazione positiva da un valore target del processo, se $Z_t > h$, dove h è il limite di controllo scelto sulla base della desiderata ARL_0 . Nel caso bilaterale, che si verifica tenendo presente di entrambe le direzioni, $|Z_t| > h$ è considerato un segnale di fuori controllo al tempo t .

5.3 Average Run Length per processi IC

Per l'implementazione di questa procedura di controllo, occorre scegliere il parametro di lisciamento λ ed il limite di controllo h .

Le prestazioni dei vari approcci per il controllo del processo, vengono di solito valutate in base all'ARL e perciò questi parametri sono scelti in modo da avere un elevato valore dell'ARL quando il processo è stabile (IC) ed un piccolo valore dell'ARL fuori controllo (OC).

Per il calcolo dell'ARL sui ranghi standardizzati, Hackl e Ledolter (1991) basandosi sull'equazione di Crowder (1987)¹⁵ utilizzata per la carta EWMA parametrica, forniscono le coppie di parametri (λ, h) che sono in grado di garantire una determinata ARL in controllo.¹⁶

In particolare vengono forniti i parametri del disegno capaci di produrre una data ARL sia nel caso la procedura EWMA si basi sui ranghi definiti in (1), e quindi nel caso in cui si è a conoscenza della distribuzione in controllo, sia nel caso questa si basi sui ranghi standardizzati definiti in (2), ottenuti da un ampio campione di riferimento ($g \rightarrow \infty$).

Hackl e Ledolter (1991), riportano anche i valori dell'ARL in controllo per la carta EWMA basata sui ranghi standardizzati definiti in (2) e basata su un campione di riferimento di dimensione $g-1$.¹⁷



¹⁵ V. Crowder (1987) - "Average Run Lengths of Exponentially Weighted Moving Average Charts", *Journal of Quality Technology*, vol. 19, pp. 161-164

¹⁶ P. Hackl, J. Ledolter (1991) - "A Control Chart Based on Ranks", *Journal of Quality Technology*, vol. 23, n. 2, pag. 119

¹⁷ P. Hackl, J. Ledolter (1991) - "A Control Chart Based on Ranks", *Journal of Quality Technology*, vol. 23, n. 2, pag. 120

5.4 Average Run Length per processi OC

Hackl e Ledolter (1991) analizzano anche l'efficienza della carta EWMA, basata sui ranghi nel rilevare cambiamenti.

Dopo le g osservazioni iniziali in controllo, si verifica un cambiamento nel livello del processo, quando in un istante t la media del processo si sposta dal valore target ad un nuovo livello diverso dal precedente di δ deviazioni standard.

Confrontata con la carta EWMA standard e con la carta Shewhart nel caso in cui un cambiamento di dimensione $\delta = 0.3$ si realizzi nel parametro di posizione delle varie distribuzioni sottostanti (normale, uniforme e t di Student), si osserva che la procedura EWMA basata sui ranghi, disegnata per $g=30$ e 50 , è particolarmente vantaggiosa in presenza di distribuzioni con code pesanti e quando la dimensione del cambiamento è da piccola a moderata. La carta in questione perde però in efficienza, anche se di poco, rispetto a quella basata sulle osservazioni nel caso in cui la variabile di controllo è distribuita normalmente.¹⁸

Un altro vantaggio riscontrato è quello di essere meno sensibile alla presenza di valori anomali (outliers) e ciò implica un'abilità ridotta nel rilevare velocemente grandi spostamenti nella media. Per esempio, se $\lambda=0.3$, $h=0.5$, e $g \leq 50$, è raro che segnali in meno di tre osservazioni in media.

1
1
1
1
1
1
1
1



¹⁸ P. Hackl, J. Ledolter (1991) – “A Control Chart Based on Ranks”, *Journal of Quality Technology*, vol. 23, n. 2, pag. 1211

CAPITOLO 6

Esempio pratico

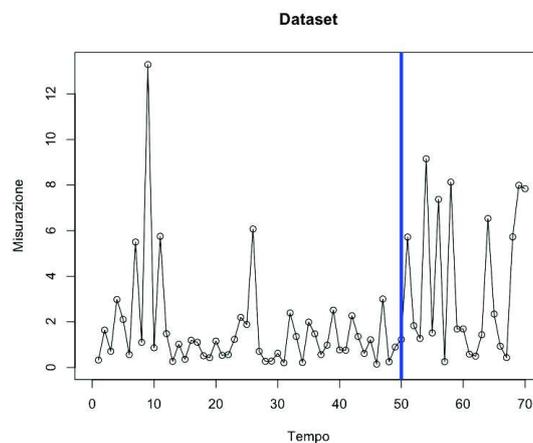
In questa sezione, confronteremo la performance della carta NLE con quella della carta EWMA basata sui ranghi standardizzati applicandole ad un dataset generato con il software R.

La distribuzione che è stata presa in esame per simulare il dataset è l'esponenziale con l'iniziale parametro di liscio λ fissato a $1/2$ che viene modificato introducendo lo shift nella 51esima osservazione.

Si hanno quindi 50 osservazioni di Fase I dalla distribuzione $\text{Exp}(\lambda=1/2)$ e le ultime 20 osservazioni dalla distribuzione $\text{Exp}(\lambda=1/4)$ che saranno utilizzate per il disegno delle carte.

Sapendo che la media di una variabile esponenziale è data da $1/\lambda$ e la varianza è $1/\lambda^2$, ciò significa che si è in presenza di un aumento nella distribuzione.

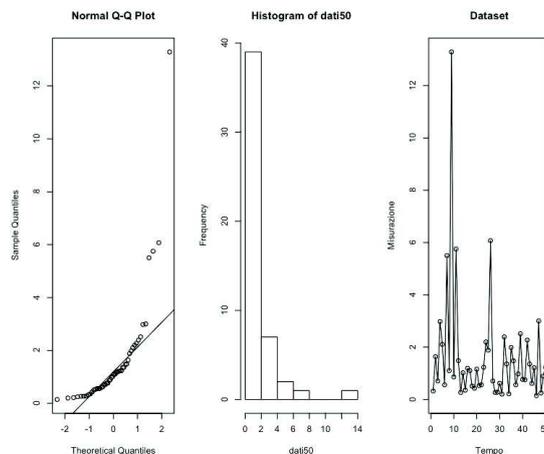
La seguente figura mostra la serie temporale delle 70 misurazioni del dataset che verrà presa in esame.



Un campione di questa dimensione può essere più piccolo rispetto all'ideale per identificare pienamente la distribuzione IC, ma è sufficiente per usare la carta proposta NLE per il fatto che è una carta distribution-free.

L'assunzione della normalità nella distribuzione sottostante al processo viene scartata, come suggerito dal grafico quantile-quantile (Q-Q plot) e dall'istogramma per le 50

osservazioni storiche di Fase I, infatti il test di Shapiro-Wilk risulta altamente significativo.



Tutti questi controlli suggeriscono che l'assunzione di normalità su F_0 non è valida e quindi gli schemi non parametrici saranno più robusti e potenti rispetto alle carte parametriche basate sulla normalità.

Per questa ragione, in questo dataset simulato è più ragionevole applicare questo tipo di carte di controllo ovvero quelle che non dipendono dalla normalità, o più in generale, da alcuna assunzione sulla distribuzione parametrica, che sono particolarmente utili nei processi quando c'è la mancanza o limitata conoscenza della distribuzione del processo.

Ora si passa al confronto tra le due carte non parametriche.

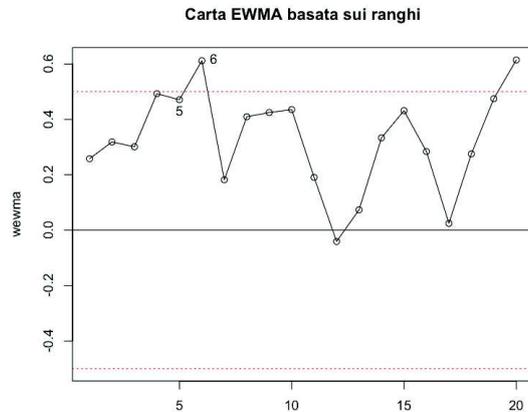
Avendo a disposizione 50 osservazioni raccolte in Fase I, è stato scelto il seguente scenario:

	m_0	λ	limiti	ARL_{IC}
Carta ranghi:	50	0.3	$h=0.5$	194
Carta NLE:	50	0.1	noti	370

Per l'implementazione della procedura EWMA basata sui ranghi standardizzati, sono stati scelti il parametro di liscio λ ed il limite di controllo h dalle stime di Monte Carlo delle Average Run Lengths che sono definite con $\lambda=0.3$ ed il limite di controllo impostato a $h=0.50$.

Avendo a disposizione un campione di dimensione $g=50$, si ha che la stima dell'ARL è pari a 194.

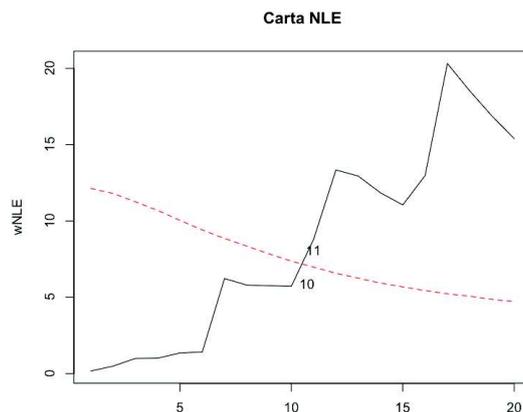
La carta EWMA basata sui ranghi standardizzati disegnata per questo esempio segnala il cambiamento alla sesta osservazione e quindi con un ritardo di cinque osservazioni avendo impostato il cambiamento in $t=5$.



Per quanto riguarda i limiti di controllo per la carta NLE, nell'articolo di Zou e Tsung è presente una tabella che fornisce la sequenza dei limiti dinamici L_t scelta per ottenere una specifica distribuzione della run-length IC, anche nel caso in cui il campione di riferimento sia composto proprio da 50 osservazioni raccolte in Fase I, come in questo esempio.

Per effettuare un congruo confronto con la carta EWMA basata sui ranghi, il parametro di liscio λ è stato impostato a $\lambda=0.1$.

La carta NLE segnala il cambiamento dopo la decima statistica di controllo e quindi con un ritardo di dieci osservazioni avendo impostato il cambiamento in $t=5$.



La carta EWMA basata sui ranghi standardizzati segnala più velocemente il cambiamento e la sua migliore efficienza viene confermata anche dai valori dei p-value relativi ai test TW e TA.

CONCLUSIONI

Lo scopo di questa tesi è di analizzare nei suoi aspetti teorici e pratici la carta non parametrica EWMA basata sul rapporto di verosimiglianza denominata NLE.

La carta NLE come tutte le altre carte non parametriche, è utile nei processi statistici di controllo nelle situazioni in cui c'è mancanza, o limitata conoscenza della distribuzione del processo oppure, quando la sua distribuzione sottostante non è normale.

Nei primi capitoli, abbiamo visto i principali aspetti teorici di questa carta parlando del test GOF non parametrico proposto da Zhang (2010), basato sul rapporto di verosimiglianza e della sua statistica test. La formula di questa statistica incorpora il liscio esponenziale, come nella comune carta EWMA, in modo che le osservazioni più recenti possano indicare il cambiamento accuratamente e facilmente.

Uno dei suoi principali vantaggi è che non richiede un lungo step per la raccolta di dati in Fase I per la conoscenza della distribuzione sottostante, essendo un metodo self-starting che studia la Fase II di monitoraggio di cambiamenti su distribuzione non parametrica che aggiorna l'informazione sulla distribuzione IC con l'aggregarsi delle osservazioni.

Usare le carte non parametriche comporta un'altra agevolazione che consiste nel mantenere sempre uguale per ogni distribuzione continua del processo la sequenza dei limiti di controllo L_t , per ottenere il desiderato comportamento della run-length IC.

Abbiamo constatato l'effetto significativo del parametro di liscio λ della carta NLE nel rilevare velocemente un cambiamento nel processo e gli autori nel citato articolo consigliano di sceglierlo tra 0.05 e 0.2.

Reynolds e Stoumbos (2005) hanno sostenuto la tesi che le carte di controllo utilizzate per la diagnostica non necessariamente devono essere le stesse per determinare quando segnalare.

Per questo, sono stati descritti due metodi non parametrici che fanno riferimento alla stima del change-point $\hat{\tau}$, applicabili anche nel software R, utili da usare dopo il segnale OC lanciato dalla carta, per individuare se c'è stato uno spostamento in posizione, scala o in entrambi.

Sotto l'aspetto pratico, la carta NLE risulta computazionalmente veloce, conveniente nell'usarla, ed efficace nello scoprire spostamenti potenziali che riguardino sia un

incremento che un decremento in posizione, scala e forma nei parametri e quindi offre una robusta protezione contro la variazione in varie distribuzioni sottostanti.

Nella prima parte del quarto capitolo, è stata comparata la performance della NLE in termini di ARL_{OC} con alcune carte progettate sotto l'assunzione di normalità, ma soprattutto con la carta CEW che essendo una combinazione di due schemi EWMA, è più agevolata nell'individuare un processo fuori controllo.

Il confronto dimostra che per i cambiamenti in media, la carta NLE offre prestazioni abbastanza soddisfacenti e la differenza con l'altra carta non è molto rilevante se si rilevano piccoli e moderati spostamenti.

La carta NLE risulta inferiore quando si hanno poche osservazioni stabili di Fase I con δ abbastanza grande ($\delta \geq 2$), ma quando m_0 è elevato, funziona anche un po' meglio nella rilevazione di spostamenti molto grandi ($\delta \geq 3$).

Focalizzandosi sui valori più bassi della ARL_{OC} , si può affermare che rispetto al cambiamento nella media, per rilevare lo spostamento nella varianza più efficacemente, la carta NLE richiede più osservazioni IC.

Nella seconda parte del capitolo, per il confronto sono state principalmente prese in considerazione due carte non parametriche, la carta EWMA basata sui ranghi standardizzati (HLE) e la carta CUSUM basata sui sequential ranks (McE).

Per valutare la performance complessiva, oltre alle ARL_{OC} , sono stati calcolati i valori dell'indice RMI, e in base a questo, la carta NLE risulta essere la migliore in assoluto.

Abbiamo così potuto notare analizzando l'andamento delle curve dei grafici, che la carta NLE risulta la più efficace nella rilevazione di moderati e grandi spostamenti nella media del processo, mentre per i piccoli spostamenti perde un po' di performance.

L'aspetto più importante è che la carta NLE supera di un bel ampio margine le prestazioni delle altre carte non parametriche nella rilevazione di spostamenti della varianza, sia in aumento che in diminuzione.

Quindi si può giungere alla conclusione che l'approccio alla carta NLE, non solo svolge altrettanto bene (in termini di ARL_{OC}) l'individuazione di spostamenti in posizione come negli schemi convenzionali, ma è anche notevolmente efficace nella rilevazione di cambiamenti in scala e forma in cui le carte non parametriche basate sui ranghi risultano insufficienti.

Si possono trarre diverse conclusioni analizzando i risultati riguardanti le performance della stima del change-point $\hat{\tau}$ e dei test TW e TA.

E' stato dimostrato che $\hat{\tau}$ ha una migliore precisione quando la dimensione dello spostamento δ aumenta e funziona molto meglio per monitorare cambiamenti nella media rispetto a quelli in dispersione.

Infine, l'uso di questi test è almeno altrettanto affidabile nell'identificare l'avvenuto cambiamento, come la combinazione delle due carte EWMA che formano la carta combinata CEW.

A P P E N D I C E

Wilcoxon Rank Sum test

Nel linguaggio R, il comando per il Wilcoxon Rank Sum test è *wilcox.test(x,y)* dove *x* è un vettore di dati numerico della prima popolazione e *y* è un vettore di dati numerico della seconda popolazione.

```
wilcox.test(a,b,correct=FALSE)
```

Con questa istruzione in R si confronta il campione *a* con il campione *b*, ma si ottiene la stessa conclusione invertendo l'ordine degli argomenti della funzione.

Valutando il p-value ottenuto dalla funzione, che risulta identico nei due casi, si accetta o si rifiuta l'ipotesi H_0 .

Fligner.Killeen test

Il Fligner.Killeen test implementato nel linguaggio R usa la mediana in ciascuno dei campioni e quindi utilizza la versione median-centering.

Il comando per questo test è *fligner.test(x,g)* dove *x* è un vettore di dati numerico e *g* è un vettore o un oggetto di fattori che danno il gruppo per i corrispondenti elementi di *x*.

Esempio:

```
gruppoa<-rep("A",55)
gruppob<-rep("B",15)
gruppi<-c(gruppoa,gruppob)
gruppi<-factor(gruppi)
dati=c(a,b)
fligner.test(dati,gruppi)
```

Valutando il p-value ottenuto dalla funzione, si accetta o si rifiuta l'ipotesi H_0 .

Funzioni in R per la carta EWMA basata sui ranghi

```
ranghi<-function(dati50,datitest)
```

```
{
m0<-length(dati50)
t<-length(datitest)
data<-c(dati50,datitest)
n<-length(data)
rstand<-c(rep(0,t))
for (i in 1:t)
{
dato=datitest[i]
cat("osservazione = ", dato, "\n")
r<-c(rep(0,(m0-1)))
for (j in 1:(m0-1))
{
if(dato > data[j])
r[j]<-r[j]+1
else
r[j]<-r[j]
}
somma=0
somma<-sum(r)+1
cat("R* = ", somma, "\n")
rt<-(2/m0)*(somma-((m0+1)/2))
cat("Rt = ", rt, "\n")
rstand[i]<-rt
}
cat("I ranghi standardizzati ottenuti sono:", "", "\n")
rstand
}
```

```
EWMAranghi<-function(rstand,lam)
```

```
{
k<-length(rstand)
wewma<-c(rep(0,k))
for (t in 1:k)
{
if (t==1)
wewma[t]<-lam*rstand[t]
else
wewma[t]<-((1-lam)*wewma[t-1])+(lam*rstand[t])
}
cat("I valori della statistica ottenuti sono:", "", "\n")
cat(wewma)
}
```

```
EWMAgrafico<-function(dati50,datitest,rstand,lam,h)
```

```
{
media<-0
ucl<-media+h
lcl<-media-h
rstand<-ranghi(dati50,datitest)
wewma<-EWMAranghi(rstand,lam)
matplot(wewma,ylim=c(min(lcl,min(wewma)),max(ucl,max(wewma))),type="l"
,main="EWMA basata sui ranghi")
points(wewma)
abline(a=media,b=0)
abline(h=ucl,lty=3,col=2)
```

```
abline(h=1cl, lty=3, col=2)
}
```

Funzioni in R per la carta NLE

```
fdistemp<-function(data,xt,lam)
```

```
{
data=sort(data)
n=length(data)
s=0
a=0
for (j in 1:n)
{
if(x[j]<=xt)
{
s=s+(1-lam)^(n-j)
}
a=a+(1-lam)^(n-j)
}
fxt=a^(-1)*s
cat("F(Xt) = ", fxt, "\n")
}
```

```
m0=length(dati50)
```

```
n=length(data)
```

```
fdistxt<-function(xt,m0,t,data,n)
```

```
{
fxt=((m0+t-1)^(-1))*(ecdf(data)(xt)*n)
cat("F(Xt) = ", fxt, "\n")
}
```

```
calcoloYt<-function(data,xt,lam,m0,t,n)
```

```
{
Yt=(1/(1-
fdistemp(data,xt,lam)))*log(fdistemp(data,xt,lam)/fdistxt(xt,m0,t,data
,n))+1/ fdistemp(data,xt,lam))*log((1-fdistemp(data,xt,lam))/(1-
fdistxt(xt,m0,t,data,n)))
}
```

```
calcoloZt<-function(data,xt,lam,m0,t,n,ztprec)
```

```
{
Yt<-calcoloYt(data,xt,lam,m0,t,n) //per ottenere il valore di Yt
Z=0
if (t==1)
{
Z<-lam*Yt
}
else
{
Z<-((1-lam)*ztprec)+(lam*Yt)
}
}
```

4

4

5

5

4

50

BIBLIOGRAFIA

- C. Zou, F. Tsung (2010) – “Likelihood ratio-based distribution-free EWMA control charts”, *Journal of quality technology*, vol. 42, n. 2, pp. 174-195
 - P. Hackl, J. Leodolter (1991) – “A control chart based on ranks”, *Journal of quality technology*, vol. 23, n. 2, pp. 117-124
 - D. McDonald (1990) – “A CUSUM procedure based on sequential ranks”, *Naval logistic research*, vol. 37, pp. 627-646
 - J. M. Lucas, M. S. Saccucci (1990) – “Exponentially weighted moving average control schemes: properties and enhancements”, *Technometrics*, vol. 32, pp. 1-29
 - S. Chakraborti, Van de Wiel (2008) – “A nonparametric control chart based on the Mann-Whitney statistic”. IMS Collections. *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, vol. 1, pp. 156-172
 - Fligner, Killeen (1976) – “Distribution-free two-sample test for scale”, *Journal of the American statistical association*, vol. 71, pp. 210-213
 - SV Crowder (1987) – “Average Run Lengths of Exponentially Weighted Moving Average charts”, *Journal of quality technology*, vol. 19, pp. 161-164
 - S. Chakraborti, P. Van Der Laan, S.T. Bakir (2001) – “Nonparametric control charts: an overview and some results”, *Journal of Quality Technology*, vol. 33, pp. 304-313
 - W. A. Shewhart (1980) – “Economic control of quality of manufactured product”.
 - D. Drain (1997) – “Statistical methods for industrial process control”, pp. 282-299, pp. 345-349
 - Anderson, Sweeney, Williams, Freeman, Shoesmith (2007) – “Statistics for business and economics”, pp. 757-772, pp. 732-737
 - Montgomery D.C (2000) – “Il controllo statistico della qualità”
 - A. Galgano (2004) – “I sette strumenti della qualità totale”, pp. 166-186
 - G. Capizzi – materiale del corso di “Metodi statistici per il controllo della qualità”
- 5
5
5
5
5
5

