Università degli Studi di Padova

DIPARTIMENTO DI SCIENZE STATISTICHE

Corso di Laurea Triennale in Statistica per le Tecnologie e le Scienze

Tesi di laurea Triennale

Un metodo di controllo statistico per la difesa da attacchi informatici volti alla manomissione di processi manifatturieri

Relatore:
Prof. **Guido Masarotto**
Dipartimento di Scienze Statistiche

Laureando:
Pietro Ferrazzi
Matricola 1193114

A.A. 2020/2021

INTRODUZIONE

La produzione industriale di ogni tipo necessita che le caratteristiche dei manufatti generati rispettino alcuni parametri. Per esempio, l'assemblaggio di orologi da polso è possibile soltanto se le lancette hanno una certa lunghezza, una certa larghezza, un certo spessore ed il foro terminale di un certo diametro: la produzione di queste deve essere dunque in linea con le qualità richieste per il prodotto finale. D'altra parte, ciascun processo è caratterizzato da una variabilità intrinseca: selezionate le specifiche desiderate e trasmesse ai macchinari responsabili della produzione, ogni articolo generato presenterà delle caratteristiche leggermente difformi rispetto ai precedenti. Poiché in genere si tratta di produzione di grandi quantità di articoli, sono state definite nel tempo delle metodologie di controllo statistico della qualità dei processi (di seguito QC) con lo scopo di tenere sotto osservazione i cicli di produzione e "avvertire" l'utilizzatore se questa non rispetti più, in maniera significativa, i canoni impostati. In sintesi, si tratta di verificare, tramite delle apposite carte di controllo, se la fabbricazione stia rispettando alcuni parametri desiderati. Una carta di controllo è composta da una statistica test costruita sui dati, confrontata ad ogni tempo con i limiti di controllo, superati i quali si considera emesso un allarme. Le metodologie esistenti di QC sono molteplici: si rimanda a Montgomery (2009) per eventuali approfondimenti. In particolare, per quanto segue si utilizza un approccio incentrato su una carta di controllo parametrica, composta in due fasi:

- Fase I: si stimano i parametri della distribuzione che ogni variabile di interesse segue quando è "in controllo", assumendone la normalità;
- Fase II: si osservano i dati via via che vengono generati dal processo e li si confronta con i limiti di controllo calcolati in modo che la probabilità di emettere un falso allarme sia pari ad una quantità desiderata (usualmente 0.05).

Poiché spesso le caratteristiche che è possibile considerare sono svariate, molte metodologie si basano sulla valutazione di un sottoinsieme di variabili maggiormente esplicative circa la qualità complessiva del prodotto, assumendo che sia possibile, per migliorare l'efficienza e contenere i costi, monitorare soltanto queste ultime. Rimanendo nell'esempio degli orologi, si potrebbe valutare di tenere sotto osservazione soltanto la lunghezza delle lancette ed il diametro del foro. All'interno di questo contesto, è possibile che un'azienda concorrente sia interessata

a manomettere la produzione delle lancette. Se tale azienda fosse a conoscenza del fatto che vengono osservate soltanto due tra le quattro variabili indicatrici dello stato di controllo del processo, potrebbe effettuare un attacco informatico ai software responsabili dello spessore e della larghezza delle lancette, compromettendo il risultato finale. Tutto ciò avverrebbe senza che la carta di controllo - costruita per monitorare lunghezza e diametro - emettesse alcun allarme.

Per definire e chiarificare la tipologia di attacchi ai quali si intende dare risposta sono riportati di seguito quattro esempi.

Autodistruzione di un generatore di corrente USA (2007)[1]

Al fine di testare la vulnerabilità delle strutture dei servizi energetici statunitensi il Dipartimento di Sicurezza Interna USA ha progettato un attacco informatico al software di gestione di un generatore di corrente, modificandone in maniera impropria il ciclo operativo, in modo da causare la rottura dello stesso. Il Dipartimento ne ha concluso che la struttura di controllo del sistema energetico non sia stata in grado di accorgersi per tempo della modifica del processo, al punto da iniziare una rivalutazione dell'intero sistema di controllo della rete energetica federale.

Stuxnet (2010)[2]

Nell'ambito delle tensioni tra Iran e Stati Uniti nel progetto nucleare del paese sciita, per rallentarne i progressi nella fase di arricchimento dell'uranio, il governo USA ha dato ordine di utilizzare un malware il quale, senza che il sistema di controllo iraniano emettesse alcun allarme, è stato in grado di neutralizzare circa il 20% delle centrifughe, essenziali per il processo di arricchimento, modificandone impropriamente la velocità in maniera scomposta, causando forti vibrazioni e distorsioni. Fonti governative americane hanno sostenuto che, in questo modo, il piano del governo di Ahmadinejad abbia subito una battuta di arresto compresa tra i 18 e i 24 mesi. Della vicenda si trova una spiegazione più dettagliata in D. Albright et al (2020).

Protesi animali

Negli studi veterinari è stato proposta da alcuni anni l'applicazione della stampa tridimensionale per ottenere protesi animali (Harrysson et al. 2015). È dunque stato suggerito

[1] https://edition.cnn.com/2007/US/09/27/power.at.risk/index.html, visitato in data 17/04/2020

[2] https://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html visitato in data 17/04/2020

di utilizzare lo stesso approccio per gli uomini (Zeltmann et al., 2017). Gli autori di questo studio hanno verificato come alterazioni nella direzione di stampa, difficilmente rilevabili dai sistemi di controllo proposti, inducano ad una importante diminuzione della qualità del prodotto. Risulta evidente come un attacco informatico che modifichi, senza essere rilevato, le caratteristiche della produzione di protesi possa risultare drammatico per i pazienti.

Fornace siderurgica (2014)

Nel 2014 il report annuale dell'Ufficio Federale per la Sicurezza Informatica tedesco ha reso nota la vicenda di un'acciaieria la quale, a seguito di un attacco tramite strategie di pishing di origine non identificata, non è stata in grado di mantenere in controllo una fornace, la quale ha così rovinato in maniera sostanziale la qualità della produzione. Per maggiori dettagli si veda Robert M. Lee et al (2014).

Le usuali metodologie di controllo della qualità non sono in ideate per individuare attacchi come quelli sopra esemplificati che possono, anzi, sfruttare le caratteristiche proprie del sistema di monitoraggio per non essere rilevati. Quelle, infatti, sono basate su ragionamenti ed assunzioni che non rimangono ragionevoli in caso di modifica malevola dei processi produttivi, come il monitoraggio basato su una sola caratteristica di base, il raggruppamento delle variabili in sottogruppi, il considerare soltanto variazioni costanti nel tempo delle caratteristiche di interesse. Inoltre, quando anche i metodi convenzionali rilevino la variazione occorsa, essi non tengono conto che possa essere stata causata da questi attacchi, con il conseguente aumento dei costi e dei tempi per la ricerca dell'origine.

Le pagine che seguono intendono sviluppare una metodologia per mettere in sicurezza i processi manifatturieri rispetto ad "avversari" che, conoscendo la struttura del sistema di controllo della qualità, siano in grado di disegnare attacchi intelligenti tali da non intaccare le caratteristiche monitorate, detti Passive Joint Attack (PJA). In particolare, si intende valutare un metodo di difesa da attacchi randomizzati che colpiscono, per ogni prodotto fabbricato, un sottoinsieme casuale di dimensione casuale di caratteristiche di interesse, in particolare quando il numero di parametri che si vuole tenere in controllo è elevato. In generale, attacchi di questo tipo sono discussi in DeSmit et al. (2017).

Evidentemente, non si richiede al QC che prevenga gli attacchi, ma che ne rilevi il prima possibile gli effetti.

Il presente lavoro illustra, nelle due sezioni, due metodologie rivolte a due diversi contesti rispetto a PJA:

1) Ad ogni ciclo di produzione alcune caratteristiche subiscono uno shift in media che viene riassorbito al ciclo successivo. Si analizza la proposta fornita da Ahmad E. Elhabashy et al. (2020).
2) Ad ogni ciclo di produzione alcune caratteristiche subiscono uno shift che viene riassorbito al ciclo successivo, mentre altre conservano la variazione occorsa in maniera permanente. Si propone una metodologia naïve basato su un approccio che tenga conto del "peso" delle osservazioni passate.

Poiché l'interesse è rivolto a variazione della media che sono in un caso reversibilli (durano cioè il tempo di un solo ciclo produttivo) e nell'altro anche non reversibili, si tenga conto che, è possibile utilizzare due diversi approcci di QC:

- *"alla Shewart"*: non considerare le osservazioni ottenute a tempi precedenti, basando la carta di controllo solo su quanto si è osservato al tempo presente;
- *"con memoria"*: cumulare le osservazioni ottenute a tempi precedenti in quanto possono già essere affette dalla variazione che si intende rilevare e dunque fornire informazione riguardo.

PRIMA SEZIONE

METODOLOGIA CORRENTE

Nei contesti di grande numerosità delle variabili all'interno di un processo è raro rilevare una variazione simultanea di tutte le caratteristiche di interesse. In genere, gli shift sono piuttosto causati da un sottoinsieme di variabili che fanno riferimento ad uno stesso fattore fisico latente che si trova in uno stato "fuori controllo". Per trovare il miglior equilibrio tra capacità di rilevamento delle alterazioni di interesse e il costo della procedura di monitoraggio bisogna compiere una scelta circa quante e quali variabili tenere sotto osservazione. Le possibilità a cui si può fare riferimento sono tre: 1) monitoraggio di uno specifico sottoinsieme di variabili, 2) monitoraggio di quante più variabili possibile e 3) selezione del sottoinsieme di variabili "più informativo" rispetto alla qualità compromessa dopo averle osservate tutte.

Il primo approccio è il più economico e consente di avere un'alta sensibilità rispetto a variazioni delle caratteristiche in controllo. Tuttavia, un metodo di questo tipo può essere facilmente aggirato se la sua struttura è nota a chi compie l'attacco malevolo, targetizzando le variabili che non sono osservate.

Il secondo approccio è meno vulnerabile del precedente, ma evidentemente più costoso e meno sensibile a variazioni che, spesso, possono non essere distinguibili dal rumore. Per esempio, si consideri una carta di controllo T^2di Hotelling, costruita su p_i variabili incorrelate, per cui la probabilità di avere un falso allarme sia pari ad s (cioè s pari alla probabilità che la carta segnali essere avvenuta una variazione delle caratteristiche di interesse quando invece questa non si sia verificata): la probabilità q_i di individuare una variazione di intensità unitaria sulla media della sola variabile x_1 - mentre le restanti $p_i - 1$ rimangono in controllo - decresce in maniera notevole con l'aumentare di p_i. In altre parole, questa carta di controllo non è efficace nel rilevare velocemente la variazione occorsa quando il numero delle variabili osservato è grande. Per questa ragione, le tradizionali carte di controllo - costruite sulla base di una statistica test e limiti di controllo numerici, superati i quali si considera che venga emesso un allarme- sono poco efficaci per la rilevazione rapida di variazioni di questo genere.

Nella Tabella 1 sono riportate le quantità di interesse al variare di p_i. I dati sono forniti da Kaibo Wang e Wei Jiang (2009).

Tabella 1: Probabilità di rilevare uno shift unitario di x_1 monitorando p_i variabili; la probabilità s di osservare un allarme quando il processo è in controllo pari a 0.002			
p_i	q_i $probabilità\ di\ rilevare\ lo\ shift$	p_i	q_i $probabilità\ di\ rilevare\ lo\ shift$
1	0.982	11	0.354
2	0.964	12	0.281
3	0.973	13	0.216
4	0.897	14	0.163
5	0.843	15	0.120
6	0.776	16	0.086
7	0.697	17	0.060
8	0.612	18	0.041
9	0.524	19	0.028
10	0.436	20	0.018

Il terzo metodo consiste nel selezionare, dopo averle ispezionate tutte, un sottoinsieme di variabili che sia il più informativo possibile circa l'eventuale perdita di qualità. Ci sono diversi approcci possibili riguardo l'implementazione di metodologie di controllo statistico basate sulla selezione di variabili (Variable Selection), di cui si evidenziano quelli basati sulla regolarizzazione via LASSO (Zou e Qiu, 2009), LAR (Capizzi e Masarotto, 2013) o selezione "forward" (Wang e Jian, 2009) del modello. In sintesi, viene predisposto in Fase I un modello, costruito considerando come esplicative tutte le variabili che è possibile osservare, che descriva la situazione in controllo tale per cui i coefficienti $\beta_k\ k = 1, \dots, p$ siano pari a 0 per tutte le p variabili rilevate. Nella Fase II, ad ogni rilevazione i viene valutato, tramite le tecniche di regolarizzazione illustrate (si veda Azzalini e Scarpa, 2009), quali siano le j variabili che più influiscono nel modello. Una volta individuate queste ultime, si procedere a testare l'ipotesi nulla che i coefficienti relativi alle j variabili selezionate siano uguali a 0. Se l'ipotesi nulla viene rifiutata, la carta di controllo emette un "errore" del quale è responsabile la variabile il quale coefficiente risulta significativamente diverso da 0.

Questa strada è costosa per via del grande numero di variabili da rilevare, ma poco vulnerabile e con alta sensibilità rispetto a variazioni della qualità complessiva.

Una rapida valutazione dei costi-benefici porta a concludere che il miglior approccio rispetto al problema in questione sia il terzo: più costoso, ma sensibile a piccoli shift e più difficile da aggirare tramite PJA.

APPROCCIO DI PARAGONE: VS-MSPC

Si è visto come il monitoraggio di più variabili e la selezione tra esse di un sottoinsieme dinamico di controllo sia la migliore strategia tra quelle fin qui proposte. Pertanto, la metodologia di Variable Selection è considerata, da qui in avanti, come termine di confronto per verificare l'ipotesi che la tecnica proposta dagli autori sia preferibile in termini di rapporto *costi - velocità di rilevamento shift* rispetto agli approcci esistenti. Tra i diversi metodi di Variable Selection esistenti, si è scelto di effettuare il confronto rispetto a quello proposto da Wang e Jiang (2009) di selezione "forward". Secondo gli autori, quest'ultimo bilancia in maniera valida la capacità di emettere un allarme in caso di shift e quella di individuare la variabile che ne è causa. Nello specifico, tale carta di controllo è costruita come segue.

Dato $y_t \sim N_p(\mu, \Sigma)$ con μ e Σ rispettivamente vettore di medie e matrice di covarianza di y_t, si assume senza perdita di generalità $\mu = 0$. Il test di ipotesi seguente è in grado di rilevare una variazione nella media del processo: $H_0: \mu = 0_p,\ H_1: \mu = \mu_1$.
Quando il processo è fuori controllo, la vera media di y_t è $\mu_1 = \delta d$ con δ scalare che rappresenta la grandezza della variazione, d vettore di lunghezza unitaria definita da $\| d \| = \sqrt{d^T \Sigma^{-1} d} = 1$. Si può verificare l'ipotesi di cui sopra tramite il log-rapporto di verosimiglianza:

$$l(y_t) = \log\left(L(y_t, 0_p) \Big/ \max\{ L(y_t, \mu_1) \} \right).$$

Sotto l'assunzione di normalità per y_t si ottiene:

$$l(y_t) = {y_t}^T \Sigma^{-1} \mathrm{y_t} - \min_{\mu \in \{\mu:\, \mu = \delta d,\ \delta > 0\}} \{ (\mathrm{y_t} - \mu)^T \Sigma^{-1} (\mathrm{y_t} - \mu) \}$$

dove la minimizzazione si ottiene in $\mu = \mu^*$, con $\mu^* = y_t$.

Poiché il metodo assume che solo q (fissato) delle p variabili siano fuori controllo, μ^* è un vettore $p \times 1$ con $p - q$ zeri. Tale soluzione matematica che identifica μ^*come stimatore di μ non tiene conto del fatto che gli elementi di y_t siano disturbati da del rumore, cioè che $p - q$ variabili siano in realtà "vicine" a 0 (e non esattamente pari a 0) e le restanti q "lontane". Per rendere evidente e quantificare ciò, si può aggiungere una penalizzazione al termine oggetto di minimizzazione in $l(y_t)$. In particolare, si minimizza $S^2 = [\,(y_t - \mu)^T \Sigma^{-1}(y_t - \mu) + \sum_{j=1}^{p} p_{\lambda_j}(|\mu_j|)\,]$ dove $|\mu_j|$ è il modulo della vera media della j-esima variabile. $p_{\lambda_j}()$ è una funzione di penalità che controlla la complessità del modello: se essa è scelta correttamente, la minimizzazione conduce a conteggiare medie piccole come pari a 0 e focalizzare l'attenzione soltanto sulle variabili che potrebbero essere fuori controllo.

Nello specifico si pone $p_{\lambda_j}(|\mu_j|) = \lambda I(|\mu_j| \neq 0)$ con $\mathrm{I}(|\mu_j| \neq 0)$ funzione indicatrice del fatto che la j-esima media sia diversa da 0.

Questo conduce a $S^2(\lambda) = [\,(y_t - \mu)^T \Sigma^{-1}(y_t - \mu) + \lambda q\,]$ con q numero di coefficienti diversi da zero nel modello finale. In questo modo si ottiene una limitazione per il numero di variabili potenzialmente fuori controllo e si stima la loro variazione tramite gli elementi non nulli di μ^*.

Tramite la decomposizione di Cholesky di Σ e procedure algoritmiche di selezione del modello è possibile calcolare $S^2(\lambda)$ tramite una minimizzazione dei minimi quadrati, dato il numero di predittori q. Sono così individuate le variabili che possono causare lo shift come quelle con coefficienti non nulli, cioè il sottoinsieme q-dimensionale che massimizza il criterio $F = \frac{(R^2_{k+1} - R^2_k)(n-k-1)}{1 - R^2_{k+1}}$, cioè ancora il gruppo di variabili che meglio descrive l'aumento di S^2.

Poiché la stima si basa sulla minimizzazione dei minimi quadrati ne consegue che il numero di variabili individuate come di interesse dal modello sia sempre pari al massimo possibile, cioè a q. Infatti, l'aggiunta di una concomitante al modello diminuisce sempre, anche se non significativamente, la devianza residua.

Una volta ottenuto μ^*, si ottiene la statistica di controllo all'osservazione t:

$$\Lambda(y_t) = 2{y_t}^T \Sigma^{-1} \mu^* - {\mu^*}^T \Sigma^{-1} \mu^* \quad [1]$$

La carta basata su tale statistica emette un allarme quando $\Lambda(y_t)$ non rientra nei limiti di controllo. Questi ultimi sono calcolati via simulazione in modo da fornire un ARL in controllo predefinito, cioè dopo aver valutato quale lasso temporale medio si sia disposti ad accettare, sapendo che il processo è rimasto in controllo per l'intera durata dell'indagine, tra l'inizio del monitoraggio e il primo falso allarme osservato.

Per riassumere, la procedura da seguire per definire la carta di controllo conseguente a questo ragionamento, detta Variable Selection for Multivariate Statistical Process Control (VS-MSPC), si compone in tre parti:

a) Selezione delle variabili: ogni volta che si ottiene un'osservazione si effettua la minimizzazione sopra indicata per identificare le variabili potenzialmente fuori controllo (con media diversa da 0) e il valore della loro variazione dallo stato in controllo.

b) Controllo del processo: le variabili con coefficienti diversi da zero vengono osservate tramite la carta di controllo definita da $\Lambda(y_t)$.

c) Diagnosi del segnale: se si rileva un allarme, il sottogruppo di variabili identificate in a) è responsabile dell'allarme.

Per applicare questa procedura è necessario assumere q, numero massimo di coefficienti non nulli, cioè di variabili possibili cause dell'allarme. Questo dipende dal contesto e va valutato di volta in volta.

APPROCCIO DI PARAGONE: STIMA DI VS-MSPC

L'approccio VS-MPSC è computazionalmente oneroso, in particolare se, come in questo caso, è applicato a simulazioni iterate in diversi contesti. Nello specifico, la regressione effettuata per trovare le q variabili con media diversa da zero tramite approcci di tipo stepwise richiede molto tempo. Per tale motivo, in questo frangente si è scelto di non affidarsi direttamente ai risultati forniti dalla VS-MPSC, quanto piuttosto di stimarne i risultati semplificando la parte che richiederebbe troppo tempo per essere calcolata direttamente. Invece che effettuare una regolarizzazione del modello per la selezione delle q variabili da monitorare, si propone di assumere che gli elementi non nulli di μ^* siano corrispondenti ai q elementi di y_t con maggiore distanza assoluta dalla loro media.

Per sostenere l'opportunità di questa scelta "semplificatrice" si è proceduto confrontando, in diversi scenari, i risultati dei due approcci in termine di ARL fuori controllo, cioè di tempo atteso tra l'intervento della causa speciale che modifica il processo e il rilevamento della variazione occorsa da parte della carta, dato un certo shift δ occorso. In particolare, si sono costruite simulazioni in cui M variabili subiscono una variazione e q (numero di variabili che si assume varino) $<<$ p, poiché per q che tende a p i risultati dei due approcci convergono.

Per rendere i risultati confrontabili con quelli ottenuti nell'articolo che propone l'approccio VS-MPSC, la grandezza della variazione complessiva osservata δ sulle M variabili è valutata come media delle variazioni individuali.

Nelle Tabelle 2 e 3 sono riportati i risultati delle simulazioni effettuate per diversi valori di p, q e M per diversi shift δ.

Tabella 2: paragone in termini di ARL tra il metodo VS-MSPC e la sua stima " naïve " per diversi valori di p, q = M = 2								
	p = 10		p = 20		p = 50		p = 100	
Dimensione shift (δ)	VS-MSPC	Naïve	VS-MSPC	Naïve	VS-MSPC	Naïve	VS-MSPC	Naïve
0	200.99	199.98	200.43	201.56	199.55	199.14	199.22	199.63
0.5	138.67	138.06	158.70	160.51	177.59	182.98	185.62	183.01
1	52.14	52.42	74.88	74.89	108.62	108.22	132.71	132.30
1.5	16.48	16.27	24.99	25.28	42.32	42.66	60.41	60.69
2	5.83	5.77	8.66	8.63	14.00	14.13	20.55	20.65
1.5	2.65	2.61	3.57	3.51	5.23	5.29	7.29	7.30
3	1.56	1.56	1.90	1.85	2.47	2.48	3.16	3.16
3.5	1.18	1.17	1.29	1.28	1.51	1.51	1.76	1.75
4	1.04	1.04	1.08	1.08	1.16	1.16	1.25	1.24

Tabella 3: paragone in termini di ARL tra il metodo VS-MSPC e la sua stima " naïve " per diversi valori di M, p=10, q=2												
	M = 1		M = 2		M = 3		M = 4		M = 5		M = 6	
Dimensione shift (δ)	VS-MSPC	Naïve	VS-MSPC	Naïve	VS-MSPC	Naïve	VS-MSPC	Naïve	VS-MSPC	Naïve	VS-MSPC	Naïve
0	199.68	199.9	200.99	201.5	200.96	199.1	200.36	199.6	199.78	200	200.05	200
0.5	145.70	142.7	138.67	138.3	134.39	140.7	133.10	133.7	132.73	134.8	133.09	132.0
1	59.23	58.93	52.14	52.85	50.45	50.69	49.74	49.90	49.75	50.86	50.53	50.11
1.5	19.82	19.18	16.48	16.49	15.92	16.12	15.92	16.09	16.18	16.65	16.49	16.55
2	7.37	3.27	5.83	5.77	5.74	5.68	5.88	5.89	6.01	6.1	6.16	6.05
1.5	3.32	3.28	2.65	2.64	2.64	2.60	2.71	2.67	2.78	2.77	2.85	2.83
3	1.87	1.85	1.56	1.55	1.56	1.56	1.60	1.60	1.63	1.64	1.66	1.66
3.5	1.32	1.31	1.28	1.18	1.18	1.18	1.20	1.19	1.21	1.21	1.23	1.23
4	1.10	1.10	1.04	1.05	1.04	1.05	1.05	1.06	1.05	1.06	1.01	1.06

Si noti che, per ogni scenario ricostruito, le prestazioni dei due approcci in termini di valore atteso dell'ARL sono pressoché equivalenti. Pertanto, è ragionevole utilizzare l'approccio naïve per stimare quello computazionalmente più oneroso.

APPROCCIO DEGLI AUTORI

L'idea che sta alla base del VS-MSPC è quella di osservare tutte le variabili possibili e scegliere tra di esse quelle che con maggior evidenza potrebbero causare un'uscita dalla condizione di controllo del processo, per tenerle sotto osservazione. Tuttavia, non è molto efficace in termini di difesa contro attacchi informatici malevoli volti alla modifica del processo stesso. Infatti, la scelta delle variabili sottoposte a controllo rimane deterministica, cioè prevedibile da chi voglia effettuare un attacco in maniera strutturata. Inoltre, a motivo della rilevazione sistematica di tutte le variabili, è costoso. L'alternativa proposta dagli autori intende risolvere questi due problemi tramite una scelta casuale di q_i variabili da monitorare tra le p osservabili, per tenerle sotto controllo tramite una carta T^2 di Hotelling con approccio alla Shewart. Inoltre, individuando in modo casuale il sottoinsieme è possibile attribuire alle variabili una probabilità di essere selezionate in base alla loro importanza nel processo: caratteristiche ritenute cruciali possono essere sorvegliate costantemente o quasi.

Le condizioni sotto le quali si identifica questa metodologia sono:

- il monitoraggio di ogni prodotto i fabbricato ad un tasso costante, cioè la possibilità di misurare $V_i = \{V_{ij}\} \forall\, i \geq 1\ e\ j = 1, \dots, p$ dove V_{ij} rappresenta la variabile j nel prodotto i,
- il fatto che le uniche variazione che interessi valutare siano quelle relative alla media,
- $V_i \sim N_p(\mu_{V_i}, \Sigma_{\mathrm{V_i}})$ con $\mu_{V_i}\ e\ \Sigma_{\mathrm{V_i}}$ vettore di medie e matrice di covarianza di cui si ottiene una stima in Fase 1, cioè sono stimate tramite dati raccolti prima della fase di monitoraggio, quando il processo viene assunto "in controllo".

Per scegliere la numerosità del sottoinsieme di q_i variabili da monitorare bisogna valutare a priori il numero massimo Q di caratteristiche che si ritiene possano aver subito variazioni, tale per cui $q_i \leq Q \leq p$. Una volta determinato q_i, si possono osservare le $y_i = \{\, y_{ik} \,\} \forall\, i \geq 1\ e\ k = 1, \dots, p$ dove y_{ik} rappresenta la realizzazione della k-esima variabile relativa al prodotto i. Si noti che q_i può variare nel tempo.

La selezione del sottoinsieme q_i-dimensionale da supervisionare può essere effettuata in tre diversi modi.

a) Sottoinsieme di grandezza fissa: $q_i = q$ costante, tutte le variabili hanno la stessa importanza.
b) Sottoinsiemi di grandezza variabile: si assume che q_i si comporti come una variabile casuale. In questo contesto si è optato per $q_i \sim U\{2, Q\}\ \forall\ i \geq 1$.
c) Sottoinsiemi di grandezza fissa con importanza variabile. Ogni variabile fornisce un contributo diverso alla qualità complessiva del prodotto. Si individuano due classi di importanza (alta e bassa) e si ispezionano sempre le q_{KQC} variabili considerate più rilevanti mentre le restanti $q - q_{KQC}$ vengono scelte in maniera casuale.

Una volta rilevate tramite uno dei tre metodi le variabili di interesse, esse sono monitorate secondo una carta di controllo T^2 di Hotelling basata sulla statistica $T_i^2 = n\left(y_i - \hat{\mu}_{V_i}\right)^T \left(S_{q_i}^2\right)^{-1} \left(y_i - \hat{\mu}_{V_i}\right)$ dove $\hat{\mu}_{V_i}$è il vettore delle medie stimato in Fase 1, $S_{q_i}^2$ è la stima della matrice di covarianza, ottenuta in Fase 1, basata sulla differenza tra osservazioni successive. Calcolata in questo modo, $S_{q_i}^2$risulta robusta rispetto a shift avvenuti nella stessa Fase I.
Viene generato un allarme quando T_i^2 supera i limiti di controllo definiti, via simulazione, a partire dal valore desiderato dell'$ARL_{in\ controllo}$.

CRITERIO DEL CONFRONTO

Per definire quale dei due metodi sia preferibile (VS-MSPC considerato tramite la sua stima, da qui in poi definita "naïve", o quello proposto dagli autori dell'articolo) è necessario individuare un criterio di scelta tramite un modello di valutazione economica dei costi. Di seguito è illustrata la proposta di valutazione dei costi per situazioni di uscita di controllo del processo fatta da Lorenzen e Vance (1986), semplificata secondo le necessità del presente confronto.

Assumendo che:

a) la produzione venga interrotta per la ricerca della causa una volta emesso un allarme,
b) vi sia una sola causa che determina un cambiamento nella media del processo,
c) le cause che possono verificarsi si distribuiscono come una variabile casuale di Poisson con intensità λ (occorrenze per unità di tempo), tale che $RL_{in\ controllo} \sim EsponenzialeNegativa(\ \mu = \frac{1}{\lambda}\ )$,
d) il processo possa ritornare in controllo solo a seguito di intervento esterno,
e) non siano considerate variazioni della varianza,

si può valutare il costo del sistema di QC tramite i seguenti due elementi:

a) costo di campionamento,
b) costo di produzione di non conformità mentre il processo è in corso.

Le equazioni di costo in situazioni fuori controllo per l'approccio di selezione casuale posso essere pensate come segue:

- $E(costo\ di\ campionamento) = n * a * \frac{E(T*q_i)}{h}$ dove n e la dimensione del campione, a è il costo di *campionamento e effettuazione del test* per ciclo di produzione, h è la frequenza di campionamento, T è il tempo trascorso in controllo più il tempo trascorso fuori controllo prima dell'allarme, q_i il numero di variabili monitorate;

- $E(costo\ di\ produzione\ non\ conformità) = C[\ h(ARL_{FC} - \tau)]$ dove C è il costo su unità prodotta legato alla produzione di non conformità e τ è il tempo atteso tra l'intervento della causa che genera non conformità e il momento in cui viene emesso l'allarme. Nel terzo tipo di campionamento casuale il valore di C viene raddoppiato per le variabili con importanza *alta.* Si noti che è possibile utilizzare h anche in questa equazione perché si è sotto l'assunzione che la frequenza di campionamento coincida con la frequenza di produzione.

Le medesime equazioni sono valide per l'approccio SV-MSPC naïve se si sostituisce q_i con p.

Per rendere confrontabili i costi tra il metodo SV-MSPC naïve e la Selezione Casuale delle Variabili si è proceduto introducendo un'equivalenza tra i costi in controllo (e non tra gli ARL in controllo!) dell'approccio usato come termine di confronto e di quello proposto. Per la sua costruzione, a parità di ARL il metodo SV-MSPC è migliore del secondo, poiché seleziona le variabili in maniera mirata, mentre quest'ultimo è preferibile a livello di spesa in controllo. Pertanto, il confronto viene effettuato a livello dei costi complessivi di due carte di controllo con medesime caratteristiche economiche nella fase in controllo:

$$\frac{c}{ARL_{in\ controllo}(SV-MSPC_{sempl})} + a * p = \frac{c}{ARL_{in\ controllo}(proposto)} + a * q_{medio}$$

dove q_{medio} e la numerosità media dei sottoinsiemi di variabili osservate (utile in particolare nel secondo piano di campionamento).

Per calcolare i costi dei due diversi approcci è necessario determinare a priori i valori di alcune costanti presenti nel modello: $n, T, h, \tau, \lambda, a$ (Tabella 4).

L'equazione che lega i costi in controllo permette di determinare, una volta impostato un $ARL_{IC}(SV - MSPC_{sempl})$ (di seguito sempre pari a 200), il valore di $ARL_{IC}(proposto)$ tale per cui i costi complessivi in controllo dei due approcci siano uguali.

Il miglior sistema di controllo è quello che minimizza il valore atteso dei costi complessivi della fase fuori controllo riscalato sul costo medio dei due approcci. In altre parole, la metodologia più economica è quella proposta dagli autori se la seguente differenza è positiva:

$$Differenza\ Costi = \frac{Costo_{(SV-MSPC)_{naïve}} - Costo_{(proposto)}}{\frac{1}{2} * \left(Costo_{(SV-MSPC)_{naïve}} + Costo_{(proposto)}\right)} \qquad [2]$$

Tabella 4: Valori scelti per i parametri del modello di valutazione dei costi

Parametro	n	T	h	τ	λ	a
Valore	1	1h	1h	1h	1h	\$5

CONFRONTO

Definito il criterio secondo il quale scegliere il miglior approccio, si procede al confronto via simulazione in diverse condizioni. Per ogni combinazione di situazioni possibili sono effettuate 10000 replicazioni basando la carta di controllo su stime di media e varianza ottenuta in fase 1, a sua volta basata su un campione di numerosità $m = 500$.

Gli scenari sono definiti dalle combinazioni delle seguenti variabili:

- numero di variabili $p = \{10, 20, 50, 100\}$,
- tipologia di campionamento (i tre precedentemente illustrati),
- numero massimo Q di caratteristiche che si ritiene possano essere variate (10%, 30%, 60%, 100% di p per i primi due piani di campionamento, 20%, 30%, 60%, 100% di p per il terzo piano di campionamento),

- il numero di variabili effettivamente attaccato, selezionato casualmente come $M \in \{\frac{p}{2}, \dots, p\}$,
- l'intensità della variazione complessiva sulle M caratteristiche attaccate, definita come $\delta = \delta^*/\sqrt{M}$ con δ^* shift medio di tutte le variabili individuali e $\delta \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$. Definire δ in questo modo rende confrontabili osservazioni successive in termini di intensità della variazione complessiva.

Ogni simulazione effettuata per ogni scenario possibile tra quelli individuati inizia con la generazione casuale dalla variabile aleatoria $V_i \sim N(0_p, I_p)$ con $i = 1, \dots, m$ utili per la Fase 1, dalle quali si ricavano $\hat{\mu}_{ic}$ e S_{ic}, stime in controllo della media e della varianza. La "vera" media è posta pari a zero senza perdita di generalità per le proprietà della distribuzione normale multivariata.
La Fase 2 viene simulata generando sequenzialmente (finché non viene emesso un allarme per entrambi gli approcci) le y_i, realizzazioni di $V_i \sim N(\mu_{fc}, S_{ic})$, dove μ_{fc} è la media fuori controllo con valori pari a 0 per le $p - M_i$ variabili che non subiscono variazioni, valori pari a δ^* per le M_i variabili che subiscono shift.
Si procede applicando i due approcci ad ogni passo i:

a) SV-MSPC naïve: definite le Q componenti di y_i da monitorare (sulla base del metodo precedentemente descritto) e la stima della media μ^*, si calcola il valore della statistica [1] $\Lambda(y_i) = 2{y_i}^T S_{ic}^{-1}\mu^* - {\mu^*}^T S_{ic}^{-1}\mu^*$ e vi si applica la relativa carta di controllo, ottenendo il valore della Run Length fuori controllo e valutandone il costo.

b) Selezione Casuale delle Variabili. Si campionano casualmente (secondo lo schema che si sta utilizzando nella specifica simulazione) le q_i variabili da osservare e vi si applica la carta di controllo T^2 di Hotelling basata sulla statistica $T_i^2 = n\,(y_i - \hat{\mu}_{ic})^T (S_{ic}^2)^{-1}(y_i - \hat{\mu}_{ic})$, ottenendo il valore della Run Length fuori controllo e valutandone il costo.

Una volta ottenuti i 10000 costi per entrambi gli approcci nello scenario corrente, se ne calcola la media aritmetica e si valuta la $Differenza\ Costi$ [2]. In questo modo si effettua un

confronto tra i costi medi dei due metodi per ogni possibile combinazione dei valori di $Q, p, M \; e \; \delta$.

Di seguito sono riportate alcuni dei risultati ottenuti:

Tabella 4: Differenza % di costo tra i due approcci con p=10, primo piano di campionamento e costo IC pari a 150$

	M = 5				**M = 10**			
δ	**Q=1**	**Q=3**	**Q=6**	**Q=10**	**Q=1**	**Q=3**	**Q=6**	**Q=10**
0.5	163.89	119.23	60.71	1.03	164.66	118.82	66.41	0.61
1.0	134.01	94.94	50.35	0.38	135.84	96.43	47.98	−1.47
1.5	87.39	60.66	33.63	−0.90	94.31	62.86	32.41	−1.97
2.0	33.54	26.02	14.86	−0.79	48.01	29.17	14.99	−0.61
2.5	−18.50	−5.61	1.32	−0.36	6.45	−3.50	−0.24	0.92
3.0	−51.11	−32.85	−9.12	−0.05	−32.31	−25.72	−13.51	−0.33
3.5	−78.10	−49.19	−17.88	−1.66	−55.29	43.05	−20.21	−0.36
4.0	−93.02	−59.85	−24.79	−0.02	−71.26	−55.05	22.79	−0.41

Tabella 5: Differenza % di costo tra i due approcci con p=100, primo piano di campionamento e costo IC pari a 600$

	M = 50				**M = 100**			
δ	**Q=10**	**Q=30**	**Q=60**	**Q=100**	**Q=10**	**Q=30**	**Q=60**	**Q=100**
0.5	192.57	174.42	133.19	0.49	192.56	174.94	133.46	0.33
1.0	190.88	172.61	130.78	−2.16	191.30	172.72	131.17	−2.09
1.5	188.86	168.42	128.18	−1.96	188.72	168.83	126.11	−3.97
2.0	184.73	162.37	121.19	−2.23	184.63	163.40	120.44	−1.61
2.5	177.67	155.15	114.59	1.18	178.66	155.86	114.58	−3.50
3.0	167.30	144.79	105.51	1.65	169.39	145.58	107.01	−0.29
3.5	151.19	131.88	95.31	0.17	156.45	132.77	94.96	−3.25
4.0	133.53	118.11	86.07	0.34	139.01	119.96	86.00	−2.73

Tabella 6: Differenza % di costo tra i due approcci con p=10, secondo piano di campionamento e costo IC pari a 150$								
	M = 5				**M = 10**			
δ	**Q=1**	**Q=3**	**Q=6**	**Q=10**	**Q=1**	**Q=3**	**Q=6**	**Q=10**
0.5	163.61	138.97	106.01	70.01	163.77	140.97	107.65	67.42
1.0	133.24	113.24	88.04	56.63	134.23	114.90	85.05	54.17
1.5	85.85	76.91	59.07	37.67	92.23	80.93	59.13	37.22
2.0	29.54	36.94	33.30	23.73	47.01	38.83	32.27	21.68
2.5	−14.04	−0.47	8.95	10.07	7.30	10.23	8.70	5.29
3.0	−53.06	−25.84	−9.52	21.67	−28.63	−15.10	−6.77	23.20
3.5	−76.32	−46.42	−24.48	212.03	−55.27	−39.28	−22.51	211.57
4.0	−94.53	−62.62	−35.40	217.42	−72.06	−52.36	−29.40	213.56
	$q_{avg} = 1$	$q_{avg} = 2$	$q_{avg} = 3.5$	$q_{avg} = 5.5$	$q_{avg} = 1$	$q_{avg} = 2$	$q_{avg} = 3.5$	$q_{avg} = 5.5$

Tabella 7: Differenza % di costo tra i due approcci con p=100, secondo piano di campionamento e costo IC pari a 600$								
	M = 50				**M = 100**			
δ	**Q=10**	**Q=30**	**Q=60**	**Q=100**	**Q=10**	**Q=30**	**Q=60**	**Q=100**
0.5	195.93	188.01	174.25	148.38	195.91	188.25	173.95	149.00
1.0	194.81	186.46	171.93	146.51	194.81	186.81	171.99	145.43
1.5	192.84	184.11	168.80	142.30	192.90	183.82	168.21	141.96
2.0	189.46	180.05	164.20	136.18	189.83	180.17	163.40	135.30
2.5	184.29	174.63	157.29	128.00	185.01	174.90	156.33	128.69
3.0	176.85	166.91	150.03	119.95	178.01	166.33	150.18	120.53
3.5	166.62	157.61	139.91	109.92	168.54	157.08	140.01	110.31
4.0	150.82	144.33	128.93	101.09	155.41	145.18	128.83	98.79
	$q_{avg} = 5.5$	$q_{avg} = 15.5$	$q_{avg} = 30.5$	$q_{avg} = 50.5$	$q_{avg} = 5.5$	$q_{avg} = 15.5$	$q_{avg} = 30.5$	$q_{avg} = 50.5$

Tabella 6: Differenza % di costo tra i due approcci con p=10, terzo piano di campionamento e costo IC pari a 160$								
	M = 5				**M = 10**			
δ	**Q=1**	**Q=3**	**Q=6**	**Q=10**	**Q=1**	**Q=3**	**Q=6**	**Q=10**
0.5	141.43	116.45	60.91	−0.38	140.14	117.20	62.67	−0.22
1.0	120.75	93.50	48.68	1.85	121.37	96.04	49.60	−0.40
1.5	96.36	64.15	29.04	−0.99	95.14	64.80	30.27	−3.12
2.0	76.28	31.09	12.11	1.06	79.17	36.32	12.61	0.10
2.5	60.97	5.77	−1.69	−1.48	61.99	14.07	−3.36	−1.13
3.0	47.54	−17.90	−11.44	0.89	48.41	−5.43	−11.25	−0.62
3.5	37.77	−32.43	−19.38	−2.33	41.93	−23.48	−22.63	0.60
4.0	32.95	−43.86	−24.71	0.10	33.32	−33.33	−23.11	0.20

Tabella 7: Differenza % di costo tra i due approcci con p=100, secondo piano di campionamento e costo IC pari a 700$								
	M = 50				**M = 100**			
δ	**Q=20**	**Q=30**	**Q=60**	**Q=100**	**Q=20**	**Q=30**	**Q=60**	**Q=100**
0.5	184.52	175.30	132.55	0.72	183.88	174.29	132.36	1.81
1.0	182.84	173.08	130.89	1.85	182.96	172.60	130.07	-2.11
1.5	181.68	169.27	125.30	-1.14	180.97	169.08	124.84	-1.63
2.0	178.40	163.28	121.85	0.01	178.01	163.64	120.36	-2.11
2.5	174.72	156.33	112.63	0.66	175.13	156.15	112.39	-1.21
3.0	170.47	145.52	103.13	1.41	169.99	146.48	105.37	0.27
3.5	164.65	132.16	91.99	2.83	165.99	136.39	94.54	0.89
4.0	157.44	117.38	84.79	2.45	159.85	123.57	84.49	3.13

CONCLUSIONE CONFRONTO

Per ogni piano di campionamento, si ottiene un risultato del confronto che indica la bontà dell'approccio proposto rispetto a quello VS-MSPC naïve:

1) Sottoinsieme di grandezza fissa: l'approccio proposto conduce ad un risparmio quando l'intensità delle variazioni non è molto elevata; la massima minimizzazione dei costi si ottiene quanto Q ha il suo minimo valore possibile; quando $Q = p$ non si notano differenze tra i due approcci; all'aumentare di p l'approccio proposto migliora rispetto al SV-MSPC naïve.
2) Sottoinsieme di grandezza variabile: i risultati sono simili al precedente, con maggior vantaggio dell'approccio proposto. La sola differenza notevole è che quando $Q = p$ i due metodi non sono più equivalenti.
3) Sottoinsieme di grandezza fissa con variabili di diversa importanza: i risultati sono analoghi ai precedenti.

Per riassumere, la metodologia proposta è particolarmente vantaggiosa quando l'intensità delle variazioni è bassa, quando la numerosità delle variabili che si possono monitorare è molto inferiore del numero delle variabili presenti e quando il numero totale delle variabili presenti è molto grande.

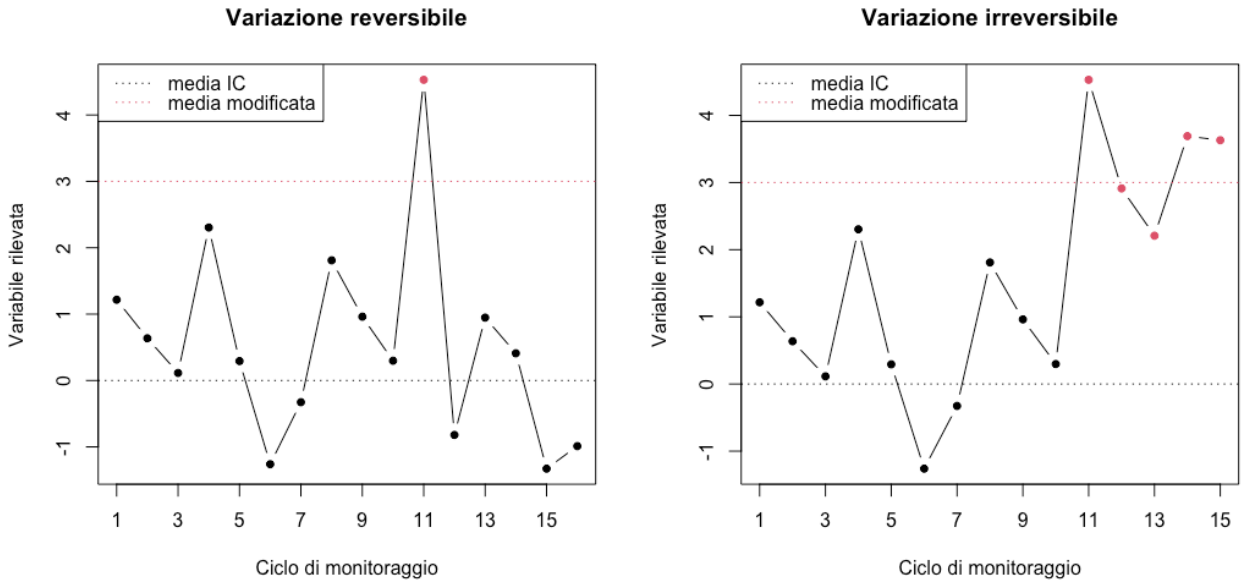
SECONDA SEZIONE

INTRODUZIONE

Nella sezione conclusiva dell'articolo vengono fatte alcune proposte riguardo possibili future direzioni di studio verso cui indirizzare la ricerca. In particolare, si evidenzia come possa essere di interesse approfondire approcci che tengano conto di una diversa struttura degli attacchi esterni rispetto a quella esclusivamente "randomizzata" considerata dagli autori, nella quale q_i variabili sono modificate ad ogni ciclo di produzione per ritornare al loro livello in controllo al tempo successivo. Nello specifico, si suggerisce di affrontare il caso in cui ad ogni ciclo alcune variabili vengano manomesse in modo randomizzato (per tornare al loro livello in controllo al tempo successivo all'attacco) e alcune, invece, una volta attaccate conservino lo shift occorso sulla loro media. I due casi sono esemplificati nella Figura 1. La seconda fattispecie può verificarsi:

- per peculiarità tecniche del processo generatore della caratteristica di controllo modificata malintenzionalmente,
- per uscita di controllo dovuta a cause non dovute all'attacco.

Figura 1



L'obiettivo è rilevare tempestivamente alterazioni della media delle caratteristiche di interesse, siano essa reversibile o meno.
Vista la presenza di variazioni che persistono costanti nel tempo, risulta ragionevole utilizzare una carta di controllo che abbia "memoria" di quanto è successo ai tempi precedenti, che cioè faccia rientrare nella statistica test utilizzata per monitorare il processo tutte le osservazioni che possano indicare le modificazioni della media occorse in tempi precedenti al ciclo in oggetto. Tra i vari modi di costruire carte di controllo che tengano conto di quanto detto, questa seconda sezione ha l'obiettivo di costruire un sistema statistico di QC che, pur mantenendo le caratteristiche di randomizzazione proposte da Elhabashy et al. (2020), sia basato su un approccio EWMA piuttosto che su quello alla Shewart fin qui utilizzato. Questa metodologia consente di:

- difendersi da attacchi malevoli intelligenti scegliendo in modo casuale, ad ogni ciclo, le variabili da osservare,
- mantenere bassi i costi di monitoraggio osservando, ad ogni ciclo, solo alcune variabili tra tutte quelle possibili,
- di rilevare più rapidamente di quanto proposto nella sezione 1 variazioni stabili delle caratteristiche di interesse.

Quanto segue giace sotto l'assunzione che i dati siano generati da una variabile casuale normale multivariata e che gli attacchi e i guasti interessino solamente la media. Per semplicità, la vera media e la vera varianza della distribuzione in controllo sono da ora in poi considerate note.

$MEWMA$

Di seguito vengono brevemente illustrate le caratteristiche di un sistema di monitoraggio per variabili multivariate basato su una carta di controllo Multivariate - $EWMA$.
Data $x_t \sim N_p(\mu_x, \Sigma_x)$ realizzazione p-dimensionale al tempo t delle variabili considerate nell'ambito del controllo della qualità di un processo, con x_{ti} realizzazione al tempo t dell'i-esima variabile monitorata, si può definire la statistica $W_t = \lambda x_t + (1-\lambda)W_{t-1}$, con $W_0 = 0$ e $\lambda \in [0,1]$. La varianza di W_t risulta pari a $\Sigma_{W_t} = \frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}] * \Sigma_{\mathrm{x}}$.

Si noti che al divergere di t $[1-(1-\lambda)^{2t}] \to 1$. La trasformazione $T^2 = W_t^T \Sigma_{W_t}^{-1} W_t$ può essere utilizzata come statistica di controllo se confrontata ad ogni tempo t con i limiti di controllo definiti sulla base dell'ARL_{IC} desiderato. Quest'ultimi possono essere individuati via simulazione: generate al tempo t, in maniera casuale, un numero elevato di statistiche T_{ti}^2 a partire dal modello in controllo, fissato $1-\alpha$ come probabilità di commettere un errore nell'emettere un allarme, si pone la probabilità che T_t^2 sia maggiore di L_t pari a α. È possibile approssimare $L_t \approx k \; tale \; che \; \frac{numero \; di \; T_{ti}^2 > k}{numero \; totale \; di \; T_{ti}^2} = \alpha$, cioè L_t è pari al quantile $1-\alpha$ della distribuzione di T_t^2 stimata tramite i T_{ti}^2. Oltre che via simulazione i limiti di controllo possono essere calcolati numericamente: per approfondimenti si veda la libreria *spc* di R.

SELEZIONE CASUALE DELLE VARIABILI IN MONITORAGGIO

Nel contesto di interesse affrontato in questa sezione, caratterizzato da grande numerosità delle variabili osservate ($p \gg 1$), per evitare di costruire la statistica di controllo su tutte le p variabili (con gli svantaggi conseguenti in termini di velocità nel rilevamento di eventuali variazioni illustrati nella prima sezione) si utilizza l'approccio di selezione casuale di $q < p$ variabili da monitorare. Notando che, per costruzione, la statistica test W_t "conserva" quello che è successo nei tempi precedenti, è ragionevole sostenere che sia tanto più di interesse rilevare l'i-esima variabile al tempo t quanto più il valore assoluto dell'i-esimo elemento di W_{t-1} è grande, cioè quanto più si ha evidenza che l'i-esima variabile stesse andando fuori controllo nei tempi precedenti. Per questo motivo, è conveniente selezionare con probabilità proporzionale a $|w_{t-1,i}|$ la i-esima variabile tra le q sottoposte a monitoraggio. Inoltre, va tenuto in considerazione che, per effetto del caso, una variabile possa non essere estratta per molto tempo. Pertanto, si definisce R_t vettore di p elementi interi $r_{t,i}$ che rappresentano il tempo passato (in termini di cicli di monitoraggio) rispetto a t dall'ultima volta che ognuna delle p variabili è stata osservata: $r_{ti} = k$ se l'i-esima variabile è stata inclusa tra le q sotto osservazione per l'ultima volta al tempo $t-k$. In generale, è tanto più di interesse rilevare l'i-esima variabile al tempo t quanto più tempo è passato dall'ultima volta che è stata osservata. Si definisce $r_{ti} = \begin{cases} r_{ti}^* \; se \; x_i \notin J_t \\ 0 \; se \; x_i \in J_t \end{cases}$ con J_t insieme delle q variabili oggetto di rilevamento al tempo t e $r_{t,i}^* = r_{t-1,i} +$

1 , $r_{0,i} = 0 \ \ \forall i \in \{1, \dots, p\}$. In questo modo, $r_{t,i}$ rappresenta l'informazione riguardo il monitoraggio o meno della variabile i al tempo t; $r^*_{t,i}$ da quanto non veniva monitorata, indipendentemente dal fatto se lo sia stata al tempo t o meno.

Per riassumere, la probabilità di selezionare l'i-esima variabile tra le q da monitorare al tempo t è posta proporzionale a:

- il modulo di $w_{t-1,i}$,
- il tempo intercorso dall'ultima osservazione della stessa.

È necessario tenere conto del fatto che la presente carta di controllo ha lo scopo di "difendere" da attacchi potenzialmente disegnati in maniera intelligente per aggirare il monitoraggio. Pertanto, le probabilità stesse di estrazione delle q variabili da valutare al passo t sono scelte in modo casuale, rispettando le assunzioni sulla proporzialità di cui sopra. Per ottenere questo, il vettore di probabilità z_t, $z_{t,i} = P(la\ variabile\ x_i\ venga\ monitorata\ al\ passo\ t)$, è assunto come realizzazione di una variabile casuale di Dirichlet di densità

$f\left(z_t = z_{t,1}, \dots, z_{t,p} \mid \alpha_{t,1}, \dots, \alpha_{t,p}\right) = \frac{\Gamma(\alpha_t)}{\Gamma(\alpha_{t,1}) * \dots * \Gamma(\alpha_{t,p})} {z_{t,1}}^{\alpha_{t,1}-1} * \dots * {z_{t,p}}^{\alpha_{t,p}-1}$,

$E\left(z_{t,i}\right) = \frac{\alpha_{t,i}}{\alpha_t}$, $\alpha_t = \sum_{i=1}^{p} \alpha_{t,i}$ e $\sum_{i=1}^{p} z_{t,i} = 1$.

In questo modo z_t è un vettore di probabilità i cui elementi sono proporzionali a $\left(\alpha_{t,1}, \dots, \alpha_{t,p}\right)$. Bisogna dunque porre $\alpha_{ti} \propto |w_{t-1,i}| * r^*_{t,i}$. Tale condizione è garantita dalla scelta degli $\alpha_{t,i} = |w_{t-1,i}| * r^*_{t,i} + k$, dove k è una costante arbitrariamente posta pari a 0.1.

COSTRUZIONE DELLA CARTA DI CONTROLLO

W_t statistica di base

Una volta determinato quali variabili includere nel monitoraggio, si costruisce la statistica $W_t = \{w_{ti}\}_{i=1,\dots,p}$ che riassume l'informazione di interesse. Per definirla in maniera esaustiva, si

consideri W_t^* con i-esimo elemento $w_{t,i}^*$ per le q variabili selezionate al tempo t ($i\ t.c.\ x_{t,i} \in J_t$): $w_{t,i}^* = (1-\lambda)^{r_{t,i}^*} w_{t-1,i}^* + \left(1-(1-\lambda)^{r_{t,i}^*}\right) x_{t,i}$ con $w_{ti}^* = 0\ \forall i$.

Tale statistica non tiene conto del fatto che $Var(w_{t,i})$ sia generalmente diversa da $Var(w_{t,j})$ per $i \neq j$. Per quanto riguarda la stima della varianza, si veda la parte successiva successiva: qui non interessa tanto la sua stima, quanto piuttosto definire W_t in maniera tale che i suoi elementi siano confrontabili tra loro. Per fare ciò, questi ultimi possono essere riscalati sulla loro varianza, per far sì che un valore particolarmente elevato di $w_{t,i}^*$ affetto però da grande variabilità sia valutato in maniera coerente con un valore inferiore ma affetto da variabilità minore. Infatti, è chiaro che solo la combinazione della media e della sua varianza fornisca tutta l'informazione disponibile necessaria a valutare se un $w_{t,i}$ sia "grande" rispetto agli altri.
Quanto detto finora deve tenere conto che soltanto q tra le p variabili vengono aggiornate al tempo t e perciò solo queste devono essere riscalate, per evitare di dividere per la propria varianza valori di $w_{t,i}^*$ già standardizzati. Pertanto, si definisce

$$w_{t,i} = \begin{cases} w_{t-1,i}\ se\ x_i \notin J_t \\ \frac{w_{t,i}^*}{Var(w_{t,i}^*)}\ se\ x_i \in J_t \end{cases} \text{ con } w_{0,i} = 0\ \forall i.$$

Si noti che per quanto riguarda le variabili inserite nel monitoraggio al tempo t vengono utilizzati gli $r_{t,i}^*$, cioè i tempi intercorsi dall'ultima volta che le variabili in J_t erano state inserite nella valutazione di W_t. Tale definizione consente di pesare meno le osservazioni più "vecchie", facendo decrementare esponenzialmente il parametro che moltiplica l'informazione ottenuta nel passato (W_{t-1}) con il tempo trascorso dall'ultimo monitoraggio.

È possibile definire in forma matriciale quanto illustrato.
Definita Λ_t come matrice diagonale $q \times q$ formata dai valori

$$l_{t,i} = \begin{cases} 1-(1-\lambda)^{r_{t,i}^*}\ se\ i \in J_t, \\ 0\ altrimenti \end{cases}$$

è possibile calcolare la forma vettoriale di $W_t^* = \left(I_p - \Lambda_t\right) W_{t-1}^* + \Lambda_t x_t$, con $W_0^* = 0_{p \times p}$.

Ω_t varianza di W_t^*

Sia definita $\Omega_t = \mathrm{Var}(\mathrm{W}_\mathrm{t}^*) = Var\big(\ (I_p - \Lambda_t)W_{t-1}^* + \Lambda_t x_t\big)$. Per calcolare la varianza della statistica test dunque necessario calcolare $Cov\big[\ (I_p - \Lambda_t)W_{t-1}^*, \Lambda_t x_t^*\ \big]$, che risulta particolarmente complesso. Tale difficoltà può essere aggirata considerando che i valori $\omega_{t,i\,i}$ sulla diagonale di Ω_t sono compresi tra σ_i vera varianza di x_i e $\frac{\lambda\sigma_i^2}{2-\lambda} * [1-(1-\lambda)^{2t}]$ varianza di una $EWMA$ costruita su x_i. Questi due estremi sono definiti dai due casi limite: il primo si verifica se la variabile x_i non viene *mai* inserita tra quelle da monitorare, il secondo se la variabile x_i viene *sempre* inserita tra quelle da monitorare. Sulla base di questa considerazione, è possibile stimare la varianza di W_t^* attraverso una sua approssimazione condizionata alle x_i campionarie, evitando dunque di considerare la covarianza tra le due quantità: $\widehat{\Omega}_t = \big(I_p - \Lambda_t\big)\Omega_{t-1}\big(I_p - \Lambda_t\big) + \ \Lambda_t\Sigma^*\Lambda_t$ con $\Sigma^* = Var(x_t^*)$.
Poiché si tratta di una procedura iterativa, è necessario definire un valore di partenza Ω_0. Ragionevolmente, essa è posto pari a una matrice di zeri, dimodoché inizialmente la varianza sia definita solamente da $\Lambda_t\Sigma^*\Lambda_t$, per poi essere aggiornata sulla base dei suoi valori precedenti. Tale scelta comporta che la prima volta in cui una variabile x_i viene selezionata il calcolo di $w_{t,i}^*$ non sia possibile: infatti, la sua varianza risulta pari a 0. Per tale ragione, se ciò avviene si pone $w_{t,i}^* = k$ arbitrario. Nelle simulazioni che seguono, k è sempre posto pari a 5.

Q_t statistica di controllo

Una volta determinato W_t^* e la sua matrice di varianza e covarianza si può procedere a calcolare la statistica di controllo da confrontare con i limiti ad ogni ciclo di monitoraggio. Per determinarla, sono scelti due tra i diversi approcci possibili (di seguito "tipo 1" e "tipo 2"):

1) la statistica di controllo è calcolata sulle q variabili selezionate al tempo t;

2) la statistica di controllo è calcolata sulle q variabili che danno luogo ai più grandi $w_{t,i}$ calcolati, indipendentemente dal fatto che siano state selezionate al tempo t.
   Tuttavia, se a) al tempo $t-1$ si e monitorata l'i-esima variabile e b) al tempo t questa non è stata selezionata tra le q che determinano l'aggiornamento di W_t^*,

anche se il valore assoluto del corrispondente $w_{t,i}$ rimane tra i q più elevati (e dunque dovrebbe entrare nel monitoraggio), si può valutare di escluderla dal confronto con i limiti di controllo a favore di una maggior sensibilità della carta, costruita in questo modo su meno variabili. Pertanto, si confronta con il limite di controllo la statistica costruita sulle $x_i \in J_t$ che sono più grandi della maggiore delle $w_{t,i}^*$ relative alle $x_i \notin J_t$.

Posto Q_t come vettore dei w_{ti} relativi alle variabili da utilizzare nella carta selezionate in base a uno dei due metodi presentati, si costruisce la statistica di controllo $T_t = Q_t^T \Omega_{\mathrm{Q_t}}^{-1} Q_t$.

Limiti di controllo

I limiti di controllo, definiti come i valori tali per cui se $T_t > L$ viene emesso un allarme, devono tenere conto del fatto che ad ogni ciclo di monitoraggio Q_t è costruito su variabili diverse. Pertanto, è necessario che L sia ridefinito in maniera dinamica ad ogni ciclo come L_t. Questo può essere fatto imponendo la condizione $P(Run\ Length > \tau \mid RL \geq \tau) = 1 - \frac{1}{ARL_{IC}}\ \forall\, \tau$, cioè che la probabilità che la carta non segnali un allarme al tempo τ, sapendo che non l'ha segnalato fino al tempo τ, sia pari al complemento a uno della probabilità di segnalare un falso allarme in una situazione in controllo. Si noti che $P(RL > \tau \mid RL \geq \tau) = P(T_\tau \leq L_\tau \mid T_1 \leq L_1, \dots, T_{\tau-1} \leq L_{\tau-1})$.
Valutati in questo modo, i limiti L_t possono essere generati via simulazione, conoscendo la distribuzione in controllo, assunta gaussiana con parametri noti.

La costruzione via simulazione del limite di controllo al tempo t è composta in 3 fasi:

1) Per $Nsim$ volte:
   a) generazione casuale delle $x_i, i = 1, \dots, p$ da una distribuzione in controllo,
   b) scelta delle q variabili con cui aggiornare W_t,
   c) calcolo di W_t^*,
   d) calcolo di $\widehat{\Omega}_t$,
   e) standardizzazione di W_t^* sulla base di $\widehat{\Omega}_t$ per ottenere W_t,

e) calcolo di Q_t sulla base di uno dei due approcci proposti,
f) calcolo di T^2.

2) Calcolo di L_t come quantile $1-\alpha$ della distribuzione empirica degli $Nsim$ valori di T^2 ottenuti al passo 1).

3) Sostituzione dei k valori W_t fuori controllo con valori ricampionati tra gli $Nsim-k$ W_t in controllo. Questo è necessario poiché i W_t, che rappresentano la distribuzione in controllo, sono calcolati in maniera ricorsiva: è quindi essenziale elidere i valori che escono dalla distribuzione desiderata, in modo che non diano origine a valori distorti di W_{t+1}.

SIMULAZIONE DELL'ATTACCO

La simulazione della situazione di attacco va costruita tenendo conto 1) delle Q_1 variabili che subiscono shift reversibili ad ogni tempo t e 2) delle Q_2 variabili che subiscono shift non reversibili (dovuti all'attacco o a cause esterne) nell'intero periodo di osservazione.

Si è proceduto:

- selezionando casualmente Q_2 variabili tra le p presenti modificate in maniera permanente a partire dal tempo tau;
- selezionando casualmente, ad ogni ciclo di monitoraggio a partire dal tempo tau, Q_1 variabili che vengono modificate in maniera reversibile.

Per semplicità si assume che i dati siano generati da una normale multivariata con media $\mu_x = \tilde{0}_{px1}$ e matrice di covarianza "a cascata" con diagonale pari a $1_{p\times 1}$ e generico elemento i,j pari a $\rho^{|i-j|}$ per $i = 1,\dots,p\ e\ j = 1,\dots,p$. Di segutio, ρ è posto pari a 0.7. La media è debitamente modificata per le Q_1 e le Q_2 variabili.
Per i dettagli si veda quanto segue.

Variabili con shift non reversibile

Come detto, l'interesse di chi compie l'attacco oggetto di studio è che le variabili subiscano una variazione della media reversibile. Tuttavia, alcune variabili, una volta modificate malevolmente, potrebbero non poter essere più riportate ad una situazione di controllo. È poi possibile che alcune variabili escano di controllo per motivi non determinati dall'attacco. Pertanto, essendoci due diverse tipologie di variazioni non reversibili, è necessario effettuare due valutazioni distinte.

1) Q_{2MAL} variabili modificate permanentemente a causa dell'attacco: il ciclo di monitoraggio c_j in cui si verifica la j-esima variazione malevola non reversibile ($j = 1, \dots, Q_{2MAL}$) è assunto variabile aleatoria uniforme in $\{1, T\}$ con T numero di cicli sottoposti alll'attacco. Tale shift viene applicato casualmente ad una tra le $p - Q_1$ caratteristiche di interesse che non subiscono altre variazioni legate all'attacco
2) Q_{2PROC} variabili modificate permanentemente per cause esterne: una variazione di questo tipo può verificarsi anche in sovrapposizione ad una dovuta all'attacco malevolo.

Per semplicità, poiché l'interesse è valutare le caratteristiche della carta di controllo sopra illustrata, si è scelto di definire un istante di tempo τ dopo il quale il processo inizia ad andare fuori controllo. Ciò significa che le medie delle Q_2variabili sono modificate secondo quanto descritto a partire esattamente dal tempo τ, siano tali alterazioni legate all'attacco o meno. Tale scelta, non completamente aderente ad una situazione reale – che trova piuttosto il suo corrispettivo in quanto esplicitato al paragrafo precedente – consente di effettuare i confronti riportati al termine di questa seconda sezione per valutare la bontà della carta di controllo, senza tuttavia distorgerne i risultati. Questo ragionamento consente anche di poter evitare di considerare le differenze tra i due tipi di shift non reversibili.

In sintesi, si selezionano casualmente $x_{variata_j}$ per $j = 1, \dots, Q_2$, variabili la cui media viene modificata in modo stabile a partire da tau. La variazione complessiva δ è assunta distribuita equamente tra le Q_2 variabili, cioè determina, ove presente, $x_{t,i} \sim N(\mu_{x_i} + \delta^*, \Sigma_{ii})$ con $\delta^* = \frac{\delta}{\sqrt{Q_2}}$.

Variabili con shift reversibile

Al tempo t le Q_1 variabili casualmente attaccate che subiscono una variazione reversibile sono individuate con probabilità uniforme tra le p caratteristiche di interesse, ognuna estratta con probabilità $\frac{1}{p}$.

La variazione complessiva ε è assunta distribuita equamente tra le Q_1 variabili, cioè determina $x_{ti} \sim N(\mu_{x_i} + \varepsilon^*, \Sigma_{ii})$ con $\varepsilon^* = \frac{\varepsilon}{\sqrt{Q_1}}$.

Tabella 8: scenari definiti dalle combinazioni di Q_1, Q_2, p, q, λ, δ e ε

	Situazione 1	Situazione 2	Situazione 3
p	20	50	100
Q_1	3	5	10
Q_2	5	8	13
q	5	7	8
λ	0.2	0.2	0.2
δ	2	2.5	4
ε	2	2.5	4

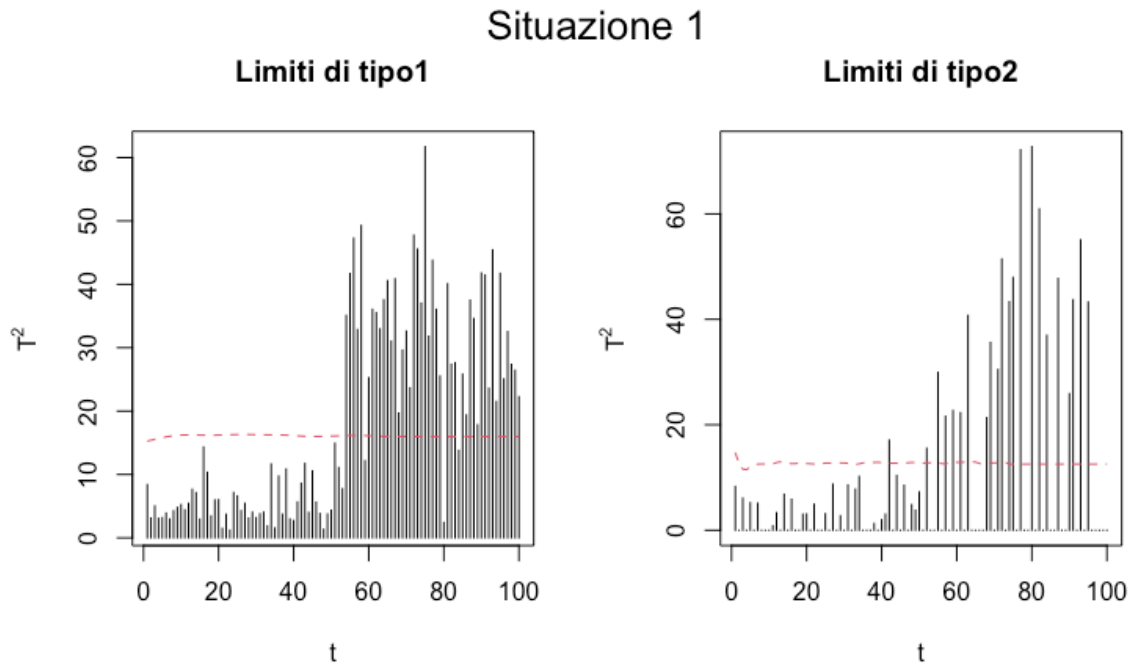
SIMULAZIONI

Per osservare i diversi comportamenti del sistema di QC a seconda dei vari scenari generati da queste due diverse tipologie di shift, sono state simulate, a titolo esemplificativo, diverse situazioni basandosi su combinazioni valori di Q_1, Q_2, p numero di variabili presenti, q numero di variabili che aggiornano W_t ad ogni ciclo, λ parametro della carta $MEWMA$ riportate nella Tabella 8.

Di seguito (Figura 2, 3 e 4) sono riportati i risultati di una singola applicazione per ognuna delle 3 situazioni individuate, ciascuna con i 2 limiti di controllo proposti. L'attacco inizia al tempo

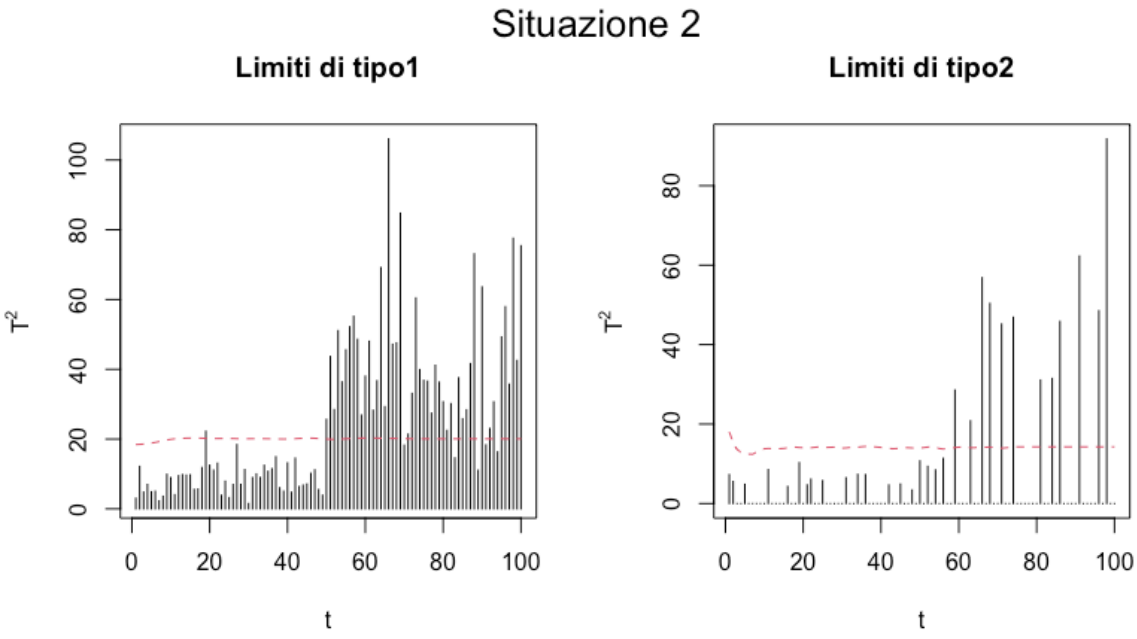
50 per ogni scenario. Sono riportati i tempi in cui la carta segnala il primo allarme dopo il tempo di uscita di controllo ed eventuali falsi allarmi.

Figura 2



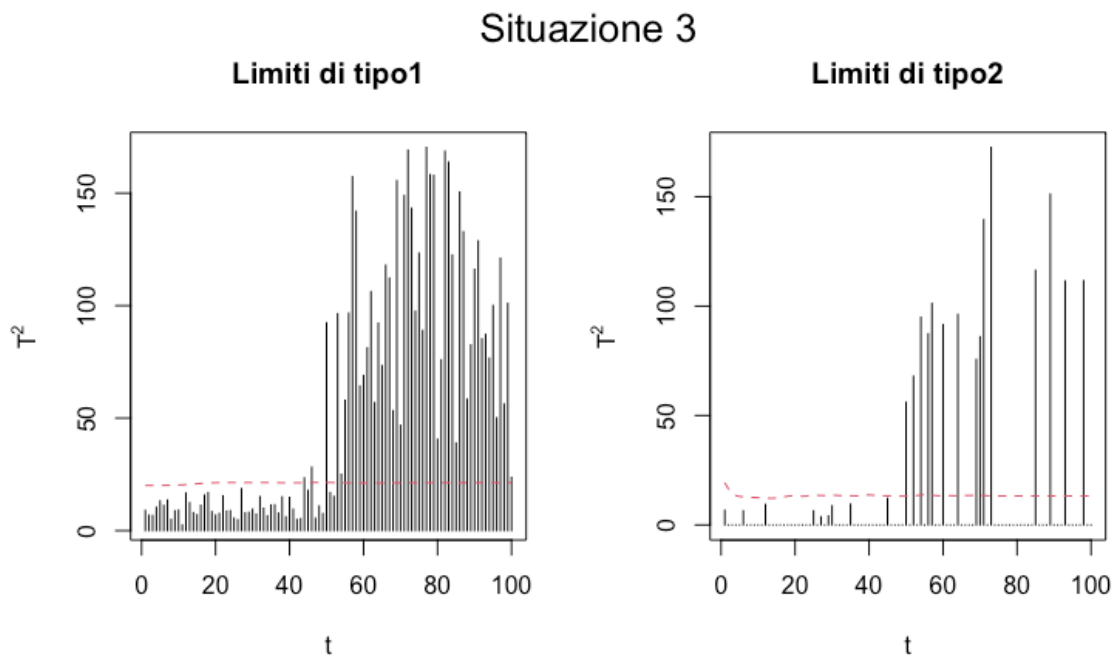
La prima carta segnala un allarme al tempo 54, la seconda un falso allarme al tempo 42 e un allarme al tempo 5.

Figura 3



La prima carta segnala un allarme al tempo 50, la seconda un allarme al tempo 52.

Figura 4



La prima carta segnala falsi allarmi ai tempi 44 e 46, un allarme al tempo 50; la seconda un allarme al tempo 50.

Si è proceduto effettuando 500 simulazioni per ogni combinazione di situazione e limiti di controllo, in modo da ottenere una stima degli ARL per ogni carta di controllo. I risultati sono riportati nella Tabella 9.

Tabella 9: valori della Run Length osservati per ogni situazione

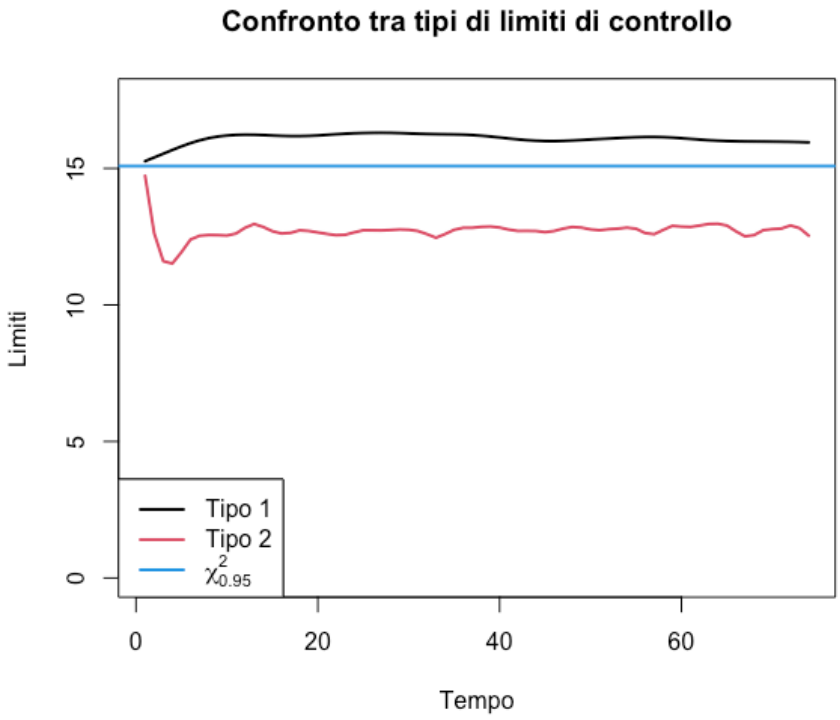
Situazione	Tipo di limiti	ARL stimato
1	1	1.462
	2	2.214
2	1	1.53
	2	2.31
3	1	0.33
	2	0.372

VALUTAZIONI CONCLUSIVE

Si può notare che la carta produca risultati in linea con quanto descritto nella fase di definizione per quanto riguarda il monitoraggio delle variabili maggiormente indicative circa possibili situazioni di fuori controllo: quando il processo è IC le variabili sono monitorate in modo uniforme, cioè ogni variabile è campionata all'interno della statistica test con frequenza pari alle altre. Tuttavia, tale frequenza aumenta - una volta FC - per le variabili che hanno subito shift reversibili. Per esempio, si considera una carta di controllo costruita su $p = 20$ variabili e monitorandone ad ogni passo $q = 5$, applicata a dei dati normali con matrice di varianza e covarianza "a cascata" di parametro $\rho = 0.8$, con $Q2 = 5$ variabili che subiscono shift costanti (nello specifico, le prime 5) e $Q1 = 3$ variabili che subiscono shift reversibili a partire dal tempo $\tau = 50$. La frazione di campionamento all'interno della statistica di controllo di almeno 1 tra le prime 5 variabili (quelle cioè che saranno poi soggette a shift non reversibile) è pari a 0.14. Una volta che il processo viene indotto ad essere fuori controllo (quando cioè le 5 variabili subiscono modifica in media), tale frazione di campionamento aumenta a 0.32. Questo accade perché la carta di controllo è costruita in modo tale da inserire più frequentemente, tra le variabili monitorate, quelle che presentano maggiore distanza dalla situazione di controllo. L'evidenza sperimentale che ciò avvenga indica che la carta rispetta nella pratica quanto ci si aspetta.

Per quanto riguarda i due diversi tipi di limiti di controllo, si noti che quelli di tipo 1 sono meno conservativi rispetto a quelli di tipo 2. Di seguito (Figura 5) si fornisce un esempio basandosi sulla medesima situazione illustrata per evidenziare le frequenze di campionamento: in nero i limiti lisciati di primo tipo, in rosso di secondo tipo, in blu il quantile 0.95 di una distribuzione Chi Quadro con $q = 5$ gradi di libertà. Quest'ultima è indicativa nell'assunzione che le statistiche T^2 si distribuscano come Chi Quadro e che pertanto i limiti di controllo debbano essere confrontabili con tale distribuzione.

Figura 5


Confronto tra tipi di limiti di controllo
Limiti
0
5
10
15
0
20
40
60
Tempo
Tipo 1
Tipo 2
$\chi^2_{0.95}$

CONCLUSIONI

Nella prima sezione è stata illustrata una carta di controllo alla Shewart che fosse utile nelle situazioni di Passive Joint Attacks. Si è poi valutato di proporre una seconda carta di controllo, basata su un'approccio EWMA, considerando il fatto che il processo possa subire delle variazioni non causate dall'attacco o che non rispettino gli obiettivi di chi lo compie. In particolare, alcune variabili possono subire modifiche non reversibili e, pertanto, inserire nella carta caratteristiche di "memoria" di quanto accaduto, conservando gli aspetti di randomizzazione evidenziati nella prima sezione. Nelle applicazioni effettuate, la carta proposta nella seconda sezione presenta le caratteristiche che ci si attendeva. In particolare, rileva variazioni nella media in maniera efficace e, per le tre situazioni proposte, rapidamente. Essa rispetta le indicazioni circa la proporzionalità tra la probabilità di monitorare le variabili e l'evidenza del loro shift ai tempi precedenti. I limiti di controllo, al netto della variabilità dovuta al fatto che sono calcolati via simulazione, rispettano le proprietà di stazionarietà proprie delle carte $MEWMA$. Si può concludere che la carta proposta sia ragionevolmente efficace nell'ambito di interesse.

Ulteriori approfondimenti possono essere compiuti effettuando una valutazione più rigorosa dell'efficienza della carta in termini di distribuzione della Run Length; applicazioni a scenari diversi da quelli proposti e a situazioni limite; modifiche alla procedura di scelta delle variabili inserendo una valutazione dell'importanza di ognuna di esse nella stima della qualità complessiva del processo; valutazione dei costi del secondo approccio rispetto al primo in situazioni analoghe. Inoltre, si può osservare che la carta proposta nella seconda sezione possa essere utilizzata efficacemente anche in situazioni di attacco dove le variabili variano solo in modo reversibile soltanto se lo shift - applicato alla media - è sempre nella stessa direzione, cioè se la media della i-esima variabile aumenta o diminuisce ogni volta che viene attaccata. In questa situazione, poiché le variazioni presentano sempre la medesima direzione, la $EWMA$ conserva l'informazione e la cumula in maniera progressiva. Tuttavia, se l'attacco determina variazioni positive e negative della media, la EWMA non è più efficace. Infatti, contenendo evidenza di cambiamenti ora in una direzione, ora in quella opposta, in media non raggiunge valori tali da emettere allarmi; pertanto, attacchi di questo tipo non sarebbero facilmente rilevati. D'altra parte, in tale situazione, rimanendo la media costante, la variabilità invece aumenterebbe: può essere utile dunque costruire una carte di controllo EWMA per il monitoraggio della varianza del processso.

BIBLIOGRAFIA

Albright, Brannan e Christina. 2020. Did Stuxnet take out 1,000 centrifuges at the Natanz enrichment plant? Institute for Science and International Security (ISIS)

Azzalini e Scarpa, Data Analysis and Data Mining. An Introduction, Oxford University Press, 2009

Capizzi e Masarotto. 2011. A least angle regression control chart for multidimensional data. *Technometrics*, 53:285–296

DeSmit, Elhabashy, Wells e Camelio. 2017. An approach to cyber-physical vulnerability assess- ment for intelligent manufacturing systems. *Journal of Manufacturing Systems* 43:339–51. doi:10.1016/j.jmsy. 2017.03.004.

Elhabashy, Dastoorian, Wells e Camelio. 2020. Random sampling strategies for multivariate statistical process control to detect cyber-physical manufacturing attacks. *Quality Engineering*. Doi: 10.1080/08982112.2020.1838541

Harrysson, Marcellin-Little e Horn. 2015. Applications of Metal Additive Manufacturing. *Veterinary Orthopedic Surgery*. JOM 67, 647–654. https://doi.org/10.1007/s11837-015-1295-x

Lorenzen e Vance. 1986. The economic design of control charts: A unified approach. *Technometrics* 28 (1):3–10. doi:10.1080/00401706.1986. 10488092

Montgomery, 2009, Introduction to Statistical Quality Control, *John Wiley & Sons, Inc.*

Robert e Assante & Tim Conway. 2014. German Steel Mill Cyber Attack. *Industrial Control Sistem*. SANS institute

Wang e Jiang. 2009. High-Dimensional Process Monitoring and Fault Isolation via Variable Selection, *Journal of Quality Technology*, 41:3, 247-258, DOI: 10.1080/00224065.2009.11917780

Wang e Jiang. 2009. High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology*, 41:247–258.

Zeltmann, Gupta, Tsoutsos et al. 2016. Manufacturing and Security Challenges in 3D Printing. JOM 68, 1872–1881 (2016). https://doi.org/10.1007/s11837-016-1937-7

Zou e Qiu. 2009. Multivariate statistical process control using LASSO. *Journal of American Statistical Association*, 104:1586–1596

APPENDICE CON CODICE R

```
# funzione che standardizza W. Se la varianza è vicina 0 (non ho
osservazioni
# precedenti) pongo pari a valore fissato Ritorna i pesi generati da una
# dirichelet
std.W <- function(W, omega, R, soglia, val.fissato) {
    idx <- which(diag(omega) <= soglia)
    W[-idx] <- W[-idx]/sqrt(diag(omega)[-idx])
    W[idx] <- val.fissato  # valore di default
    W <- rgamma(length(W), abs(W) * R + 0.01)
    W/sum(W)
}

## Radice quadrata di una matrice simmetrica positiva
## definita m e della sua pseudo-inversa
sqm <- function(m) {
    m <- eigen(m, symmetric = TRUE)
    m$vectors %*% (sqrt(pmax(0, m$values)) * t(m$vectors))
}

isqm <- function(m, eps = .Machine$double.eps/100) {
    m <- eigen(m, symmetric = TRUE)
    idx <- which(m$values > eps)
    m$vectors[, idx] %*% (sqrt(1/m$values[idx]) * t(m$vectors[,idx]))
}


R_MEWMA.limiti <- function(Sigma, q, lambda, limiti = c("tipo1", "tipo2"),
ARL0 = 100,
    Nsim = 100 * ARL0, K = 50 + round(log(0.001/lambda)/log(1 - lambda)),
    soglia = sqrt(.Machine$double.eps), val.fissato = 5, dimmi.dove.sei =
TRUE) {

    # tipo di limiti
    limiti <- match.arg(limiti)

    # spazio per le varie statistiche simulate, le pseudo matrici di
covarianza, ...
    p <- NROW(Sigma)
    W.sim <- matrix(0, nrow = p, ncol = Nsim)  # statistiche simulate
    Omega.sim <- array(0, dim = c(p, p, Nsim))  # lista delle matrici di
cov di W.sim parto dalla matrice di 0
    R <- matrix(0, nrow = p, ncol = Nsim)  # inizializzo a zero il vettore
R. ogni colonna si riferisce ad un W.sim
    Lt <- rep(0, K)  # limiti di controllo
    A <- sqm(Sigma)  ## radice di Sigma usata per generare i dati
    T2 <- numeric(Nsim)

    # ciclo principale
    for (j in 1:K) {
        x.sim <- A %*% matrix(rnorm(p * Nsim), p)  # dati simulati

        for (i in 1:Nsim) {
            # vettore di probabilità da cui estrarre le variabili da
monitorare e variabili
            # estratte
            prob <- std.W(W.sim[, i], Omega.sim[, , i], R[, i], soglia,
val.fissato)
            which.to.W <- sample(1:p, q, prob = prob)
```

```
            # definisco le grandezze utili al calcolo di W e T
            R[, i] <- R[, i] + 1  # aggiorno R (nella notazione del testo,
questo è R*)
            lr <- numeric(p)  # pesi per aggiornare le varie quantità
            lr[which.to.W] <- (1 - (1 - lambda)^R[which.to.W, i])
            R[which.to.W, i] <- 0  # finisco l'aggiornamento di R
assegnando a quelle appena inserite un ritardo di 0

            # aggiornamento di W.sim
            W.sim[, i] <- (1 - lr) * W.sim[, i] + lr * x.sim[, i]

            # aggiornamento di Omega
            LR1 <- diag(lr)  # matrice diagonale dipendente da R ritardi
            LR2 <- diag(1 - lr)  # appoggio per leggibilità
            Omega.sim[, , i] <- LR2 %*% Omega.sim[, , i] %*% LR2 +
                                LR1 %*% Sigma %*% LR1

            if (limiti == "tipo1") {
                qq <- which.to.W  # disegno la carta con i nuovi
inserimenti
            } else if (limiti == "tipo2") {
                # monitoro solo, tra quelle inserite in W al tempo t,
quelle con modulo maggiore
                # del modulo della più shiftata delle var in W 'non nuova'
                Ws <- numeric(p)
                idx <- diag(Omega.sim[,,i])>soglia
                Ws[idx] <- W.sim[idx,i]/sqrt(diag(Omega.sim[idx,idx,i]))
                qq <- intersect(which.to.W, which(abs(Ws) >= max(abs(Ws[R[,
i] != 0]))))
            }
            if (length(qq) == 0) {
                T2[i] <- 0
            } else {
                T2[i] <- sum((isqm(Omega.sim[qq, qq, i]) %*% W.sim[qq,
i])^2)
            }
        }
        Lt[j] <- quantile(T2, 1 - 1/ARL0)
        ## sostituzione delle traiettorie *fuori controllo*
        idx <- which(T2 > Lt[j])
        jdx <- sample(which(T2 <= Lt[j]), length(idx))
        W.sim[, idx] <- W.sim[, jdx]
        R[, idx] <- R[, jdx]
        Omega.sim[, , idx] <- Omega.sim[, , jdx]
        if (dimmi.dove.sei)
            cat(j, "/", K, "\n")
    }
    list(Sigma = Sigma, p = p, q = q, lambda = lambda, limiti = limiti,
ARL0 = ARL0,
        Nsim = Nsim, soglia = soglia, val.fissato = val.fissato, Lgrezzi =
Lt, L = predict(smooth.spline(Lt))$y)
}

rprocesso <- function(n, Sigma, tau = 1, Q2.index = NULL, delta = 3, Q1 =
0, epsilon = 3) {
    fisse <- Q2.index
    casuali <- Q1
    p <- NCOL(Sigma)
    x <- matrix(rnorm(p * n), n) %*% sqm(Sigma)
    OC <- seq(tau, n)
```

```
    if (!is.null(fisse)) {
        x[OC, fisse] <- x[OC, fisse] + delta
    }
    if (casuali > 0) {
        pp <- 1:p
        for (i in OC) {
            attaccate <- sample(pp, casuali)
            x[i, attaccate] <- x[i, attaccate] + epsilon
        }
    }
    x
}



R_MEWMA.applica <- function(obj, x, plot = TRUE) { # obj è un oggetto
output di R_MEWMA.limiti, x dati fase 2
    # estrazione di alcune caratteristiche dagli input
    p <- NCOL(x)
    n <- NROW(x)
    Sigma <- obj$Sigma
    soglia <- obj$soglia
    val.fissato <- obj$val.fissato
    q <- obj$q
    lambda <- obj$lambda
    limiti <- obj$limiti
    # MEWMA, ritardi e var(MEWMA) corrente
    Wt <- numeric(p)
    R <- integer(p)
    Omega <- matrix(0, p, p)
    # Oggetti da ritornare
    W <- matrix(0, n, p)
    Q <- matrix(0, n, q)
    T2 <- numeric(n)
    L <- obj$L
    if (length(L) > n) {
        L <- L[1:n]
    } else {
        L <- c(L, rep(L[length(L)], n - length(L)))
    }

    # ciclo principale
    for (i in 1:n) {
        prob <- std.W(Wt, Omega, R, soglia, val.fissato)
        Q[i, ] <- which.to.W <- sample(1:p, q, prob = prob)
        R <- R + 1
        lr <- numeric(p)
        lr[which.to.W] <- (1 - (1 - lambda)^R[which.to.W])
        R[which.to.W] <- 0
        Wt <- (1 - lr) * Wt + lr * x[i, ]
        LR1 <- diag(lr)
        LR2 <- diag(1 - lr)
        Omega <- LR2 %*% Omega %*% LR2 + LR1 %*% Sigma %*% LR1
        Ws <- numeric(p)
        idx <- diag(Omega) > soglia
        W[i, idx] <- Ws[idx] <- Wt[idx]/sqrt(diag(Omega)[idx])
        if (limiti == "tipo1") {
            qq <- which.to.W
        } else if (limiti == "tipo2") {
            qq <- intersect(which.to.W, which(abs(Ws) >= max(abs(Ws)[R !=
0])))
```

```
        }
        if (length(qq) == 0) {
            T2[i] <- 0
        } else {
            T2[i] <- sum((isqm(Omega[qq, qq]) %*% Wt[qq])^2)
        }
    }
    if (plot) {
        matplot(cbind(T2, L), type = c("h", "l"), lty = 1:2, col = 1:2,
xlab = "t",
            ylab = expression(T^2), main=paste("Limiti di", limiti))
    }
    invisible(list(T2 = T2, L = L, W = W, Q = Q))
}



# definisco le quantità per disegnare le 6 carte e stimare i 6 ARL

set.seed(10+06+2021)

p_1 <- 20
S_1 <- outer(1:p_1, 1:p_1, function(i, j) 0.8^abs(i - j))
Q1_1 <- 3
Q2_1 <- 5
q_1 <- 5
lambda_1 <- 0.2
delta_1 <- 2
epsilon_1 <- 2

p_2 <- 50
S_2 <- outer(1:p_2, 1:p_2, function(i, j) 0.8^abs(i - j))
Q1_2 <- 5
Q2_2 <- 8
q_2 <- 7
lambda_2 <- 0.2
delta_2 <- 2.5
epsilon_2 <- 2.5

p_3 <- 100
S_3 <- outer(1:p_3, 1:p_3, function(i, j) 0.8^abs(i - j))
Q1_3 <- 10
Q2_3 <- 13
q_3 <- 8
lambda_3 <- 0.2
delta_3 <- 4
epsilon_3 <- 4



# definisco i limiti di entrambi i tipi per ogni situazione
set.seed(10+06+2021)

limiti_1A <- R_MEWMA.limiti(S_1, q_1, lambda_1, limiti="tipo1")
limiti_1B <- R_MEWMA.limiti(S_1, q_1, lambda_1, limiti="tipo2")
limiti_2A <- R_MEWMA.limiti(S_2, q_2, lambda_2, limiti="tipo1")
limiti_2B <- R_MEWMA.limiti(S_2, q_2, lambda_2, limiti="tipo2")
limiti_3A <- R_MEWMA.limiti(S_3, q_3, lambda_3, limiti="tipo1")
limiti_3B <- R_MEWMA.limiti(S_3, q_3, lambda_3, limiti="tipo2")

plot(limiti_1A$L, ylim = c(0, 1.05 * max(limiti_1A$Lgrezzi)), type="l",
```

```
     main="Confronto tra tipi di limiti di controllo", ylab="Limiti",
xlab="Tempo",
     lwd=2)
abline(h = qchisq(0.99, q_1), col = 4, lwd=2)
points(limiti_1B$L, ylim = c(0, 1.05 * max(limiti_1B$Lgrezzi)), col=2,
type="l",
       lwd=2)
ex <- expression(chi[0.95]^2)
legend("bottomleft", legend=c("Tipo 1", "Tipo 2", ex),
       col=c(1,2,4), lty=rep(1,3), lwd=rep(2,3))
# plot(limiti_2A$Lgrezzi, ylim = c(0, 1.05 * max(limiti_2A$Lgrezzi)))
# lines(limiti_2A$L, col = 2)
# abline(h = qchisq(0.99, q), col = 4)
#
# plot(limiti_2B$Lgrezzi, ylim = c(0, 1.05 * max(limiti_2B$Lgrezzi)))
# lines(limiti_2B$L, col = 2)
# abline(h = qchisq(0.99, q), col = 4)


# genero i dati nei 3 scenari

set.seed(10+06+2021)
x_1 <- rprocesso(100, S_1, tau=50, Q2.index = c(1:5), delta=delta_1,
                 Q1 = Q1_1, epsilon=epsilon_1)
x_2 <- rprocesso(100, S_2, tau=50, Q2.index = c(1:8), delta=delta_2,
                 Q1 = Q1_2, epsilon=epsilon_2)
x_3 <- rprocesso(100, S_3, tau=50, Q2.index = c(1:13), delta=delta_3,
                 Q1 = Q1_3, epsilon=epsilon_3)

# applicazioni uniche per ottenere grafici
par(mfrow=c(1,2))

set.seed(10+06+2021)
a1 <- R_MEWMA.applica(limiti_1A, x_1)
b1 <- R_MEWMA.applica(limiti_1B, x_1)
mtext("Situazione 1", side = 3, line = -1.2, outer = TRUE, cex=1.6)

which(a1$T2>a1$L)
which(b1$T2>b1$L)

set.seed(10+06+2021)
a2 <- R_MEWMA.applica(limiti_2A, x_2)
b2 <- R_MEWMA.applica(limiti_2B, x_2)
mtext("Situazione 2", side = 3, line = -1.2, outer = TRUE, cex=1.6)

which(a2$T2>a2$L)
which(b2$T2>b2$L)

set.seed(10+06+2021)
a3 <- R_MEWMA.applica(limiti_3A, x_3)
b3 <- R_MEWMA.applica(limiti_3B, x_3)
mtext("Situazione 3", side = 3, line = -1.2, outer = TRUE, cex=1.6)

which(a3$T2>a3$L)
which(b3$T2>b3$L)

set.seed(10+06+2021)
a <- R_MEWMA.applica(limiti_1A, x_1)
which(a$T2 > a$L)
mean(apply(a$Q[1:50, ], 1, function(x) any( (x == 1) || (x == 2)||
```

```
                                                    (x == 3)|| (x == 4) || (x
== 5) ) ))
mean(apply(a$Q[51:100, ], 1, function(x) any( (x == 1) || (x == 2)||
                (x == 3)|| (x == 4) || (x == 5) ) ))
plot(ts(a$W)[,1:10],
     main="Andamento delle prime 10 statistiche di controllo",
     ylim=c(-2,7))

set.seed(10+06+2021)
a <- R_MEWMA.applica(limiti_1B, x_1)
which(a$T2 > a$L)
mean(apply(a$Q[1:50, ], 1, function(x) any((x == 1) || (x == 3))))
mean(apply(a$Q[51:100, ], 1, function(x) any((x == 1) || (x == 3))))
plot(ts(a$W))


### stime ARL

primi.veri.allarmi_1A <- NULL
primi.veri.allarmi_1B <- NULL
primi.veri.allarmi_2A <- NULL
primi.veri.allarmi_2B <- NULL
primi.veri.allarmi_3A <- NULL
primi.veri.allarmi_3B <- NULL

for (i in 1:500){
    arl.find_1A <- R_MEWMA.applica(limiti_1A, x_1, plot=F)
    arl.find_1B <- R_MEWMA.applica(limiti_1B, x_1, plot=F)
    arl.find_2A <- R_MEWMA.applica(limiti_2A, x_2, plot=F)
    arl.find_2B <- R_MEWMA.applica(limiti_2B, x_2, plot=F)
    arl.find_3A <- R_MEWMA.applica(limiti_3A, x_3, plot=F)
    arl.find_3B <- R_MEWMA.applica(limiti_3B, x_3, plot=F)
    allarmi_1A <- which(arl.find_1A$T2 > arl.find_1A$L)
    allarmi_1B <- which(arl.find_1B$T2 > arl.find_1B$L)
    allarmi_2A <- which(arl.find_2A$T2 > arl.find_2A$L)
    allarmi_2B <- which(arl.find_2B$T2 > arl.find_2B$L)
    allarmi_3A <- which(arl.find_3A$T2 > arl.find_3A$L)
    allarmi_3B <- which(arl.find_3B$T2 > arl.find_3B$L)

    primi.veri.allarmi_1A <- c(primi.veri.allarmi_1A, allarmi_1A[
which(allarmi_1A >= 50)[1]])
    primi.veri.allarmi_1B <- c(primi.veri.allarmi_1B, allarmi_1B[
which(allarmi_1B >= 50)[1]])
    primi.veri.allarmi_2A <- c(primi.veri.allarmi_2A, allarmi_2A[
which(allarmi_2A >= 50)[1]])
    primi.veri.allarmi_2B <- c(primi.veri.allarmi_2B, allarmi_2B[
which(allarmi_2B >= 50)[1]])
    primi.veri.allarmi_3A <- c(primi.veri.allarmi_3A, allarmi_3A[
which(allarmi_3A >= 50)[1]])
    primi.veri.allarmi_3B <- c(primi.veri.allarmi_3B, allarmi_3B[
which(allarmi_3B >= 50)[1]])

    print(i)
}
mean(primi.veri.allarmi_1A)-50
mean(primi.veri.allarmi_1B)-50
mean(primi.veri.allarmi_2A)-50
mean(primi.veri.allarmi_2B)-50
mean(primi.veri.allarmi_3A)-50
mean(primi.veri.allarmi_3B)-50
```