

# Appunti di Statistica per la Tecnologia e l'Industria

Daniele Zago

3 luglio 2021

# INDICE

<b>I</b>	<b>Introduzione</b>	<b>1</b>
	<b>Lezione 1: Introduzione al corso</b>	<b>2</b>
1.1	Problema di base . . . . .	2
	<b>Lezione 2: Overview delle procedure di controllo</b>	<b>6</b>
2.1	Fase I . . . . .	7
2.2	Fase II . . . . .	10
<b>II</b>	<b>Fase I</b>	<b>12</b>
	<b>Lezione 3: Carte di controllo parametriche di fase I</b>	<b>13</b>
3.1	Carte Shewhart . . . . .	14
	<b>Lezione 4: Test di ipotesi multipli</b>	<b>17</b>
4.1	Confronti multipli . . . . .	17
4.2	Carte combinate . . . . .	18
	<b>Lezione 5: Statistica multivariata</b>	<b>19</b>
5.1	Statistiche di sintesi . . . . .	19
5.2	Matrici normali di dati . . . . .	20
5.3	Distribuzione Wishart . . . . .	21
	<b>Lezione 6: Carte Shewhart e Change Point di Fase I</b>	<b>23</b>
6.1	Carte di controllo Change-Point . . . . .	23
	<b>Lezione 7: Carte di Fase I non parametriche</b>	<b>26</b>
7.1	Metodi di permutazione . . . . .	26
7.2	Verifica $X_1 \sim Q_1(\vartheta_1)$ vs. $X_2 \sim Q_2(\vartheta_2)$ . . . . .	26
7.3	Permutazioni multivariate . . . . .	28
	<b>Lezione 8: Carte di Fase I non parametriche (ii)</b>	<b>30</b>
8.1	Calcolo dei limiti di controllo . . . . .	30
	<b>Lezione 9: Trasformazione per ranghi</b>	<b>33</b>
9.1	Distribuzione dei ranghi . . . . .	33
	<b>Lezione 10: Trasformazione per ranghi (ii)</b>	<b>36</b>
<b>III</b>	<b>Fase II</b>	<b>38</b>
	<b>Lezione 11: Introduzione alla Fase II</b>	<b>40</b>
11.1	Generalità sulla fase II . . . . .	40
11.2	Average Run Length (ARL) . . . . .	42
	<b>Lezione 12: Carte di controllo di Fase II</b>	<b>43</b>

12.1	Carta Shewhart . . . . .	43
12.1.1	Disegno della Shewhart . . . . .	44
12.2	Carta EWMA . . . . .	45
12.2.1	Estensioni dell'EWMA . . . . .	46
12.3	Carta CUSUM . . . . .	47
<b>Lezione 13: Disegno di EWMA e CUSUM</b>		<b>50</b>
<b>Lezione 14: Cambiamento <math>\delta</math> incognito</b>		<b>52</b>
14.1	Carte combinate . . . . .	52
14.2	EWMA adattiva . . . . .	53
<b>Lezione 15: Cambiamento <math>\delta</math> incognito (ii)</b>		<b>57</b>
15.1	Carte self-starting . . . . .	57
15.2	Generalized Likelihood Ratio (GLR) . . . . .	59
15.3	Legame tra CUSUM e verosimiglianza . . . . .	60
<b>Lezione 16: Disegno di carte complesse</b>		<b>62</b>
16.1	Disegno dell'AEWMA . . . . .	62
16.2	Approssimazioni stocastiche . . . . .	63
16.2.1	Algoritmo di Robbins e Monro . . . . .	63
16.2.2	Accelerazione del metodo RM . . . . .	64
<b>Lezione 17: Capizzi e Masarotto (2017)</b>		
«Phase I Distribution-Free Analysis of Multivariate Data»		<b>66</b>
17.1	Introduzione . . . . .	66
17.2	Stage I: Preprocessing . . . . .	67
17.3	Stage II: Numero di shift . . . . .	68
<b>Lezione 18: Capizzi e Masarotto (2017) (ii)</b>		<b>69</b>
18.1	Stage III: Aggregazione e $p$ -value . . . . .	69
18.2	Stage IV: Pruning post-segnale . . . . .	69
<b>Lezione 19: Steiner et al. (2000)</b>		
«Monitoring Surgical Performance Using Risk-Adjusted Cumulative Sum Charts»		<b>71</b>
<b>Lezione 20: Shen et al. (2016)</b>		
«Self-Starting Monitoring Scheme for Poisson Count Data With Varying Population Sizes»		<b>74</b>
20.1	Introduzione . . . . .	74
20.2	Calcolo dei limiti di controllo . . . . .	75
<b>Lezione 21: Shen et al. (2016) (ii)</b>		<b>76</b>
21.1	Confronto con altri competitor . . . . .	76
21.2	Studio dell'ARL . . . . .	76
21.3	Conclusioni . . . . .	77
<b>Lezione 22: Altre carte di Fase II</b>		<b>78</b>
22.1	Carte univariate non parametriche . . . . .	78
22.1.1	“Magnifiche tre” su Wilcoxon . . . . .	78

22.1.2	Change point su Wilcoxon . . . . .	79
22.2	Carte multivariate parametriche . . . . .	79
22.2.1	Carte combinate . . . . .	80
22.2.2	$T^2$ di Hotelling . . . . .	80
<b>Lezione 23: Carte di Fase II multivariate</b>		<b>81</b>
23.1	Carte multivariate parametriche (ii) . . . . .	81
23.1.1	CUSUM di Healy . . . . .	81
23.1.2	CUSUM di Crosier . . . . .	81
23.1.3	EWMA multivariata . . . . .	82
23.2	Maledizione della dimensionalità . . . . .	83
23.2.1	Articolo LAR Fase II . . . . .	84
23.2.2	Richiami sulle componenti principali . . . . .	84
<b>Lezione 24: Carte di Fase II multivariate (ii)</b>		<b>86</b>
24.1	Monitoring delle componenti principali . . . . .	86
24.2	Statistiche non invarianti per direzione . . . . .	87
24.3	Carta regression-adjusted . . . . .	88
<b>IV Journal Club</b>		<b>90</b>
<b>Lezione 25: Capizzi e Masarotto (2011)</b>		
«A Least Angle Regression Control Chart for Multidimensional Data»		<b>91</b>
25.1	Introduzione . . . . .	91
25.2	Caso non strutturato . . . . .	92
<b>Lezione 26: Capizzi e Masarotto (2011) (ii)</b>		<b>93</b>
26.1	Dati funzionali . . . . .	93
26.2	Processi Multistage . . . . .	94
26.3	Applicazione . . . . .	95
<b>Lezione 27: Capizzi e Masarotto (2011) (iii)</b>		<b>97</b>
27.1	Selezione delle variabili . . . . .	97
<b>Lezione 28: C. Zou et al. (2008)</b>		
«Monitoring Profiles Based on Nonparametric Regression Methods»		<b>100</b>
28.1	Test di non linearità di Azzalini e Bowman (1993) . . . . .	101
<b>Lezione 29: C. Zou et al. (2008) (ii)</b>		<b>102</b>
29.1	Sorveglianza di $g(\cdot)$ . . . . .	102
<b>Lezione 30: Abdella et al. (2019)</b>		
«An Adaptive Thresholding-Based Process Variability Monitoring»		<b>104</b>
<b>Lezione 31: Zhang e Woodall (2015)</b>		
«Dynamic Probability Control Limits for Risk-Adjusted CUSUM Charts Based on Multiresponses»		<b>107</b>

<b>Lezione 32: Hawkins e Maboudou-Tchao (2008)</b>	
«Multivariate Exponentially Weighted Moving Covariance Matrix»	<b>109</b>
<b>A Approfondimento sulle distribuzioni multivariate</b>	<b>112</b>
A.1 Distribuzioni ellittiche . . . . .	112
A.2 Distribuzioni stabili . . . . .	113
A.3 Data Depth . . . . .	114
A.4 Spatial signs . . . . .	115
<b>Riferimenti bibliografici</b>	<b>116</b>

Parte I

# Introduzione

## LEZIONE 1: INTRODUZIONE AL CORSO

2020-09-30

In questo corso si studierà la *stabilità di un processo*, del quale generalmente sono incognite le forme delle distribuzioni e gli istanti di cambiamento del processo. In generale,

1. Si modella un processo a regime.
2. Si costruisce un metodo per stabilire quando un processo è fuori controllo.

Queste metodologie sono nate per trattare argomentazioni di temi tecnologici, chimici, fisici, ..., ma sono progressivamente state estese ad altri ambiti, quali il biomedico, finanziario, etc. ...

Dal momento che le variabili da tenere sotto controllo sono numerose, nel tempo sono state sviluppate delle metodologie per coniugare la sorveglianza del processo con la selezione delle variabili (lasso, adaptive lasso, ...), in modo da individuare le variabili che con maggiore probabilità hanno portato all'allarme.

### 1.1 Problema di base

Nel *Statistical Process Control* si raccolgono dati da un processo, che può essere produttivo, fisico, ambientale, etc ...

Questi dati sono in generale autocorrelati secondo il tempo o lo spazio, e si vuole sviluppare una metodologia per rispondere alla seguente domanda:

*Il meccanismo che ha generato le osservazioni è sempre stato lo stesso, oppure a un certo punto è cambiato nel tempo/spazio?*

Nel caso in cui la risposta sia la seconda, si è anche interessati a individuare il prima possibile *quando* e *in che modo* è cambiato. La rapidità con cui si riesce ad individuare un andamento fuori controllo è una delle principali metriche con cui si valuta la performance una carta di controllo. Banalmente, se qualcosa va storto vogliamo rendercene conto il prima possibile per non sprecare tempo/risorse/...

**Def. (Processo in controllo)**

Un processo si dice è *in controllo* se tutte le osservazioni sono state generate dallo stesso meccanismo.

**Def. (Processo fuori controllo)**

Un processo è *in controllo* se **non** tutte le osservazioni sono state generate dallo stesso meccanismo.

**Def. (Carta di controllo)**

Si dice *carta di controllo* uno strumento di analisi dati che combina procedure analitiche e grafiche per determinare se il processo è in controllo o meno. In generale, è una statistica, cioè una funzione dei dati campionari, che si riporta sul grafico in funzione del tempo.

Per quasi tutto il corso ci occuperemo di determinare la carta di controllo più appropriata per ciascuno specifico problema. Esempi di applicazioni recenti dello *Statistical Process Control* (SPC) sono:

- › *Analisi testuale*: individuare comportamenti fraudolenti nelle mail tra dirigenti di un'azienda
- › *Social network*: analisi dei flussi di comunicazione in sotto-reti sociali (terroristi, ...).
- › *Riconoscimento facciale*: individuare contorni basilari per individuare la stessa persona con occhiali ecc. . .

Il problema reale non è il *big data* in quanto dimensione, ma in quanto *complessità*. A volte i dati sono strutturati in modo complesso, come ad esempio funzioni Figure 1, e vanno trattati in modo specifico secondo la loro natura.

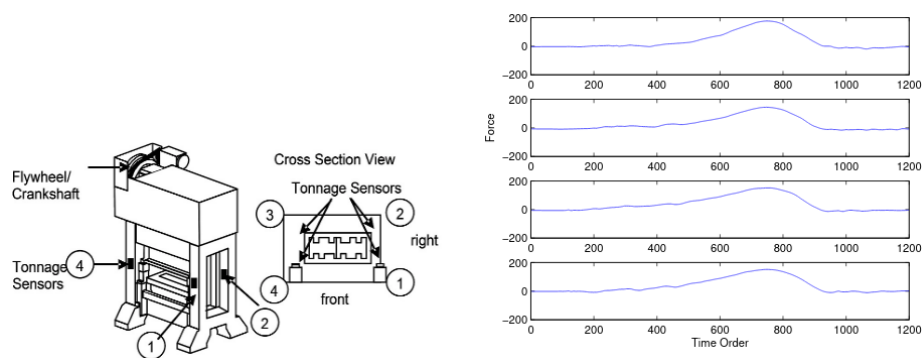


Figura 1: Curve che descrivono la forza misurata da ciascun sensore.

Per rispondere a una domanda sperimentale, si può procedere ad esempio:

1. Caratterizzando la curva in condizioni di stabilità (la *popolazione*).
2. Sorvegliare la forma della curva sia *nel tempo* sia *tra i sensori* rispetto alla caratterizzazione sotto l'ipotesi di stabilità.

Un altro problema di sorveglianza funzionale si ha in Figure 2, nel quale si vuole sorvegliare che la forma incisa dall'acido in un semiconduttore sia sufficientemente simile al profilo standard.



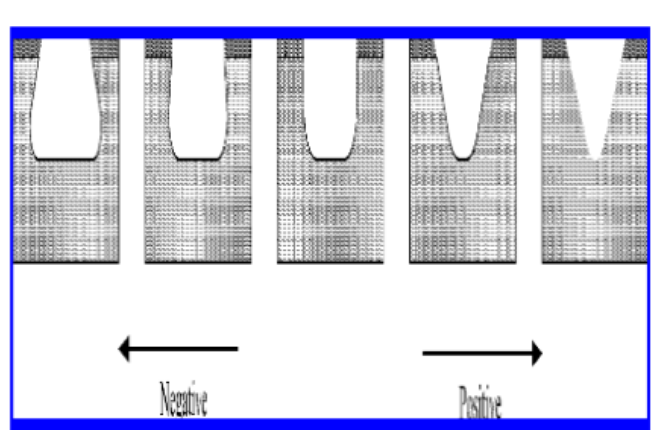


Figura 2: Incisione di un semiconduttore, con deviazioni positive e negative dalla situazione di stabilità.

Si vuole sorvegliare che la forma sia corretta, ad esempio studiandone l'angolo tra la parte discendente e la parte centrale piatta.

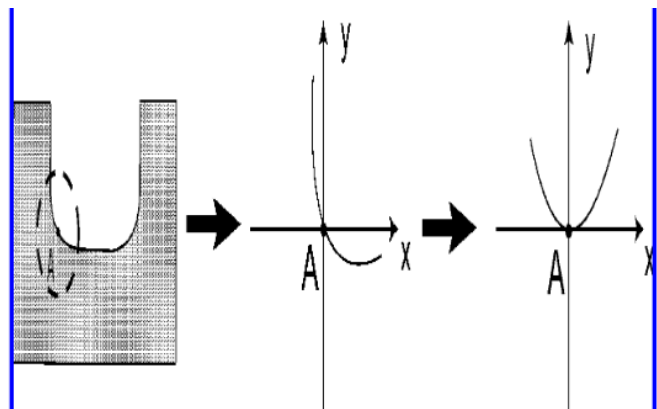


Figura 3: Normalizzazione dell'angolo di incisione rispetto al sistema di coordinate cartesiane standard.

Anche in questo caso i dati sono curve, che vengono rappresentate in modo globale in Figure 4.

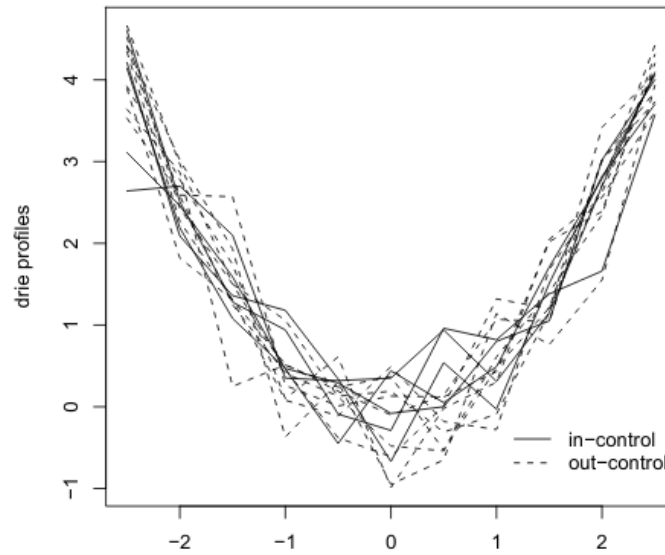


Figura 4: Curve osservate in controllo e fuori controllo per il processo di incisione.

I profili delle curve sono stati generati usando un modello lineare quadratico, con coefficienti diversi tra processo in controllo e fuori controllo

$$\text{In controllo } y_{i,j} = 1.55 + 0.62(x_j^2 - 2.5) + \varepsilon_{i,j}$$

$$\text{Fuori controllo } y_{i,j} = 1.55 + 0.67(x_j^2 - 2.5) + \varepsilon_{i,j}$$

## LEZIONE 2: OVERVIEW DELLE PROCEDURE DI CONTROLLO

2020-10-01

Le procedure di SPC si possono suddividere in due grandi famiglie:

- › Procedura *di Fase 1*: quando si dispone di poche conoscenze sul processo, si vuole raccogliere un insieme di dati in modo *offline* per stimare quando il processo è in controllo. Si estraggono campioni in base allo schema di campionamento previsto e si studia la distribuzione.
- › Procedura *di Fase 2*: il processo viene sorvegliato *sequenzialmente* per identificare il più rapidamente possibile quando il processo è fuori controllo. Fase di campionamento prospettico, si eseguono test di ipotesi ad ogni istante temporale per verificare la distribuzione  $F_0$  vs.  $F_1$ .
- › Procedura *di Fase 3*: in genere utilizziamo le osservazioni di fase 1 e 2 per rivalutare la nostra conoscenza delle caratteristiche del processo.

### Problematiche di fase I

La fase I è importante quando non si hanno informazioni sufficienti per la fase 2:

1. Non si conosce il processo in condizioni di stabilità: è gaussiana?  $t$  Student? ...
2. Se si conosce la forma, se ne conoscono i parametri?

In tal caso si effettua un'analisi *retrospettiva* offline, estraendo un numero più grande possibile di osservazioni per verificare se il processo è sotto controllo.

È un po' un circolo vizioso, perché

Distrib. ignota  $\xrightarrow{\text{quindi}}$  Studio stabilità  $\xrightarrow{\text{però}}$  Non conosco la distrib.  $\xrightarrow{\text{quindi}}$  Studio stabilità?  $\rightarrow \dots$

In letteratura è stato stabilito che, per non incorrere in grosse problematiche, bisogna usare un approccio non parametrico per lo studio della stabilità del processo in Fase I.

Infatti, i valori dei parametri stimati in Fase I si propagano in Fase II facendo finta che siano i veri valori dei parametri del processo. Sbagliando le stime dei parametri, si produrrebbero gravi errori nel controllo successivo.

### Problematiche di fase II

Nel caso in cui non si rifiuti l'ipotesi, si effettua una serie di test (*carta di controllo*) per un campione sequenziale e potenzialmente infinito di dati.

#### Def. (Run length)

Si definisce *run length* il numero di campioni osservati prima di lanciare l'allarme. In questo contesto, la *run length* è una variabile casuale e deve essere il più vicino possibile al vero momento in cui il processo è andato fuori controllo.

### Osservazione

Se l'allarme suona al 12° campione, significa che sono state necessarie  $n \cdot 12$  osservazioni per lanciare l'allarme.

Di questa run length, si studia in media quanto ci mette la carta di controllo per dichiarare il processo fuori controllo. In tal caso si parla di *average run length* della carta, che può essere di due tipologie:

- › Average run length IC ( $ARL_0$ ): media del numero di campioni IC prima del falso allarme, legata alla probabilità di I tipo  $\alpha$ ; vorremmo sia più grande possibile, perché tanti falsi allarmi fanno *calare l'attenzione* dell'operatore, il che è gravissimo in contesti come la biostatistica.
- › Average run length FC ( $ARL_1$ ): media del numero di campioni FC prima di un vero allarme, vorremmo sia piccolissimo per poter chiamare l'allarme il prima possibile. Se l' $ARL_0$  è legata ad  $\alpha$ , questo errore sarà legato alla potenza del test  $1 - \beta$  del test.

Come nella statistica classica, si fissa un valore di average run length IC ( $\alpha$ ) e si modifica la procedura in modo da avere la minor average run length FC ( $1 - \beta$ ).

### Problematiche di fase III

In processi caratterizzati short runs, cioè nei quali si producono pochi campioni e in genere non si dispone di storici, è impossibile avere due fasi separate.

In tal caso si hanno procedure *self-starting*, che contemporaneamente aggiornano le stime e controllano il processo, tramite una finestra mobile.

Se per sfortuna il processo è già fuori controllo, si ha la problematica della *window of opportunity*: inefficienza se si hanno cambiamenti vicini all'origine del monitoraggio oppure se sono cambiamenti di piccola entità.

## 2.1 Fase I

Si considerino  $m$  campioni di ampiezza  $n$   $\mathbf{x}_{i,j} \in \mathbb{R}^p$ , da distribuzioni  $p$ -variate:

$$\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n} \sim F_1(\cdot)$$

$$\mathbf{x}_{2,1}, \dots, \mathbf{x}_{2,n} \sim F_2(\cdot)$$

$$\vdots$$

$$\mathbf{x}_{m,1}, \dots, \mathbf{x}_{m,n} \sim F_m(\cdot)$$

Lo studio delle osservazioni può essere per dati individuali ( $n = 1$ ) o per “sottogruppi razionali” ( $n > 1$ ). Ad esempio, si possono valutare diverse caratteristiche (e.g. altezza, spessore, densità se  $p = 3$ ) e se ne possono osservare  $n = 5$  per ciascuna rilevazione.

Nella fase I, l'analisi di questi campioni valuta le seguenti domande:

- › Il processo è stabile? Il che corrisponde al sistema di ipotesi

$$\begin{cases} H_0 : F_1 = F_2 = \dots = F_m = F_0 \\ H_1 : \text{Non tutte le } F_i \text{ sono uguali} \end{cases}$$

- › Condizionato alla stabilità, come è fatta  $F_0$ ? (Forma/parametri)
- › Condizionato alla instabilità, che cambiamenti sono avvenuti e quando?
- › Condizionato alla instabilità, si può selezionare un sottocampione in controllo per studiarlo in stabilità?

### Esempio (Fase I non parametrica)

Campionamento per singole osservazioni ( $n = 1$ ) per una sola caratteristica ( $p = 1$ ), in cui si effettua un'analisi di fase I non parametrica (RS/P).

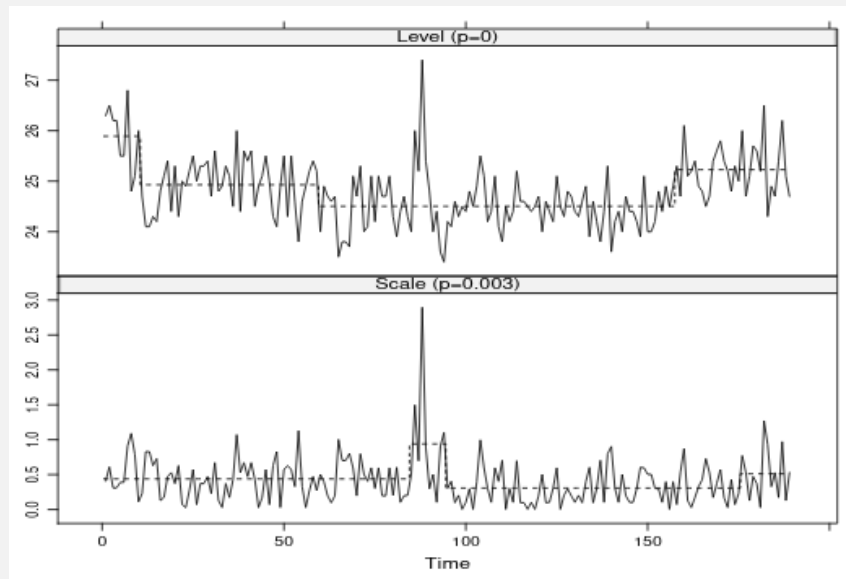


Figura 5: controlloFonderiaFase1

Si produce un  $p$ -value per indicare se è avvenuto un cambiamento nel parametro di posizione o di scala. Dall'analisi post-segnale, si osservano i momenti in cui è avvenuto il cambiamento dei parametri di posizione e scala.

### Esempio (Fase I multivariata)

In questo caso si possono avere più modi di rifiutare l'ipotesi nulla, in particolare possono variare i parametri di una componente, due, tutte assieme, ...

Inoltre, ci possono anche essere *segnali spuri* sulla stessa variabile, ovvero quando il parametro della distribuzione è variato, ma suona la carta di controllo di un altro parametro.

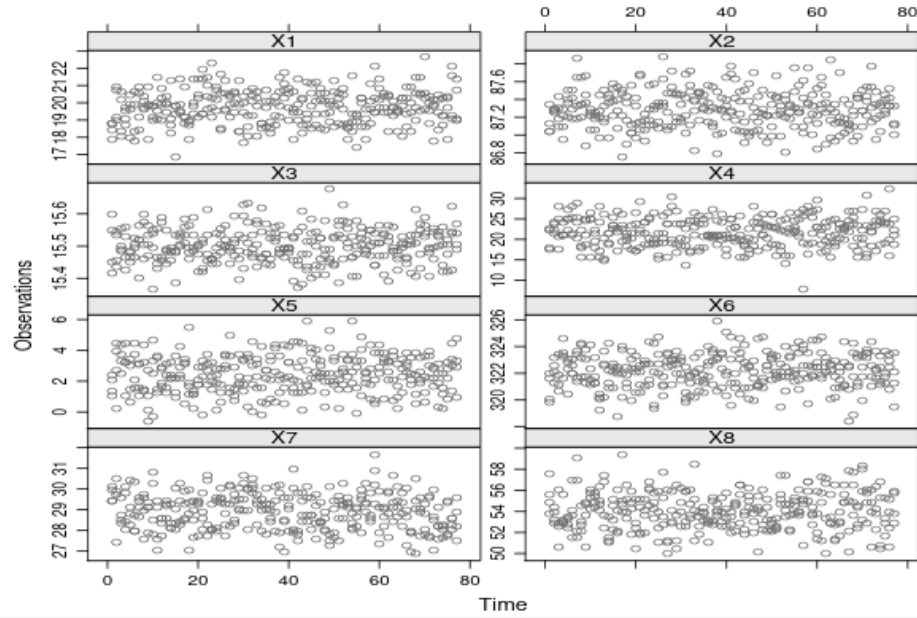


Figura 6: faseImultivariata

Applicando la stessa procedura RS/P di prima sui sottogruppi razionali ( $n = 4$ ) e assumendo che le  $p$  variabili siano indipendenti, si ottiene:

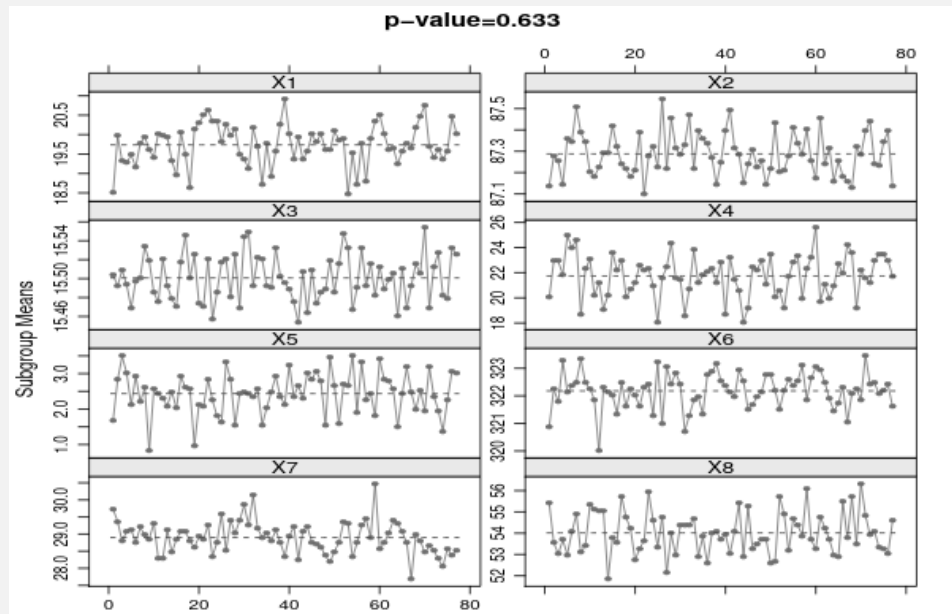


Figura 7: faseImultivariataNonParametrica

Il limite di questa procedura è assumere che le variabili siano indipendenti, cosa che non sempre può essere ragionevole.

## 2.2 Fase II

Nella situazione di fase II, ad un istante ignoto di tempo  $\tau$ , il processo va fuori controllo. Per semplicità, si assume che la distribuzione successivamente al tempo di cambiamento  $\tau$  rimanga costante,

$$F_\tau = F_{\tau+1} = \dots$$

Il sistema da controllare diventa allora

$$\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n} \sim F_1(\cdot) \quad \text{IC}$$

$$\mathbf{x}_{2,1}, \dots, \mathbf{x}_{2,n} \sim F_2(\cdot) \quad \text{IC}$$

$$\vdots$$

$$\mathbf{x}_{\tau-1,1}, \dots, \mathbf{x}_{\tau-1,n} \sim F_{\tau-1}(\cdot) \quad \text{IC}$$

$$\mathbf{x}_{\tau,1}, \dots, \mathbf{x}_{\tau,n} \sim F_\tau(\cdot) \quad \text{FC}$$

$$\mathbf{x}_{\tau+1,1}, \dots, \mathbf{x}_{\tau+1,n} \sim F_{\tau+1}(\cdot) \quad \text{FC}$$

$$\vdots$$

In questa situazione, si vuole *segnalare al più presto* che il cambiamento sia avvenuto, per cui i dati sono analizzati online. Le metriche di performance sono diverse da quelle in fase I, perché lì si vuole controllare la probabilità di falso allarme  $\implies$  corretta classificazione.

In fase II, i dati sono prospettici e ci si concentra sulla velocità di falso allarme. Al tempo  $i$ -esimo, si usano tutti i dati disponibili  $\mathbf{x}_{1,*}, \mathbf{x}_{2,*}, \dots, \mathbf{x}_{i,*}$  per verificare il sistema di ipotesi

$$\begin{cases} H_0 : \tau > i & \text{cambiamento non ancora avvenuto} \\ H_1 : \tau \leq i & \text{cambiamento già avvenuto} \end{cases}$$

Nella letteratura probabilistica, si indica  $H_0 : \tau = \infty$ , per indicare che non sia mai avvenuto.

Le carte di controllo si dividono in:

- › Carte *senza memoria*: usano solo il campione al tempo  $i$ . La carta è sensibile a i cambiamenti molto grandi, perché non tiene conto del passato.
- › Carte *con memoria*: cumulano l'informazione dal tempo 1 al tempo  $i$ .

Dopo aver segnalato un allarme, si vuole inoltre individuare quando il cambiamento è avvenuto, ovvero stimare  $\tau$ . Alcune carte permettono di stimare  $\tau$  tramite MLE.

Infine, si può anche provare a stimare l'ampiezza del cambiamento, quando possibile.

### Esempio (Carta di controllo LAR-EWMA)

I dati sono generati IC per  $i = 1, \dots, 5$  e FC per  $i = 6, \dots, 14$ . In questo caso, si ha un

punto (modello lineare) per ogni profilo funzionale.

Il limite di controllo è basato sulla soglia del test che lancia l'allarme una volta che viene superato. Essendo la EWMA una carta con memoria, lancio l'allarme al 12° profilo, ma prima si osserva un trend crescente: già dal 6° in poi si osserva una deviazione dal controllo.

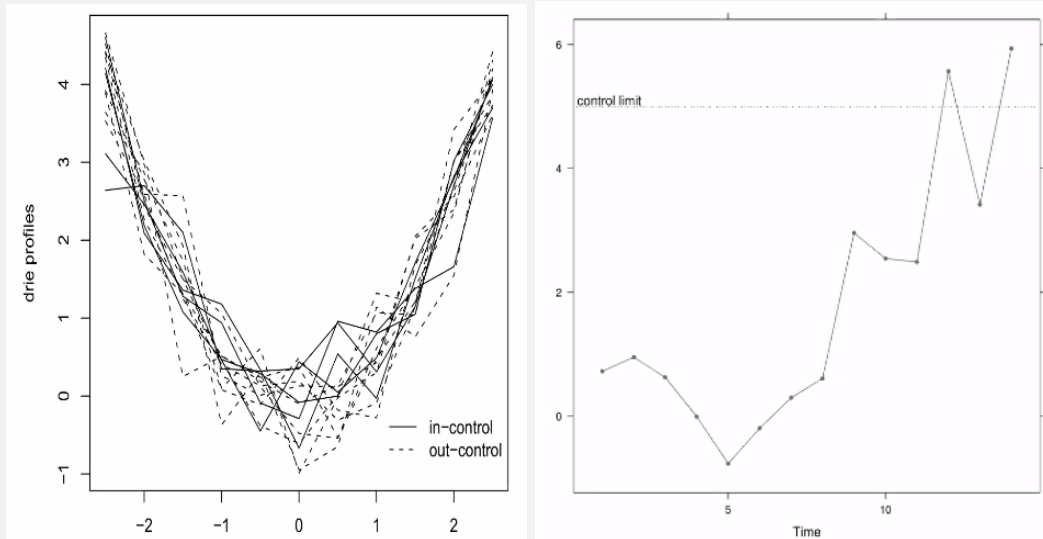


Figura 8: Carta di controllo che utilizza l'algoritmo LAR per identificare quali variabili abbiano portato alla chiamata dell'allarme.



---

Parte II

Fase I

## LEZIONE 3: CARTE DI CONTROLLO PARAMETRICHE DI FASE I

2020-10-02

In ragione degli esempi sopra, in fase I c'è il rischio di osservare campioni fuori controllo. Per poter eliminare tali campioni, bisogna essere in grado di risalire alle cause che hanno portato alla segnalazione dell'allarme. Dopodiché, l'operatore deve valutare se il valore anomalo sia stato causato da variazioni fuori dall'ordinario oppure casuali (falso allarme).

### Esempio (Andamenti fuori controllo)

Specialmente per il parametro di scala, ci sono comportamenti tipici di parametri fuori controllo.

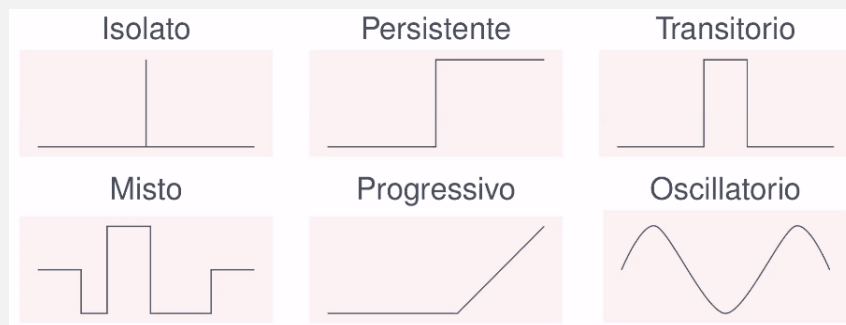


Figura 9: Tipici comportamenti fuori controllo di un parametro del processo.

- › *Isolato*: analogo degli outlier, si potrebbero studiare dal punto di vista del machine learning con metodi di *anomaly detection*.
- › *Persistente* o *step* o *sustained*: il salto viene mantenuto senza mai ritornare indietro.
- › *Transitorio* e *misto*: i più difficili da identificare in breve tempo.
- › *Progressivo*: il parametro cresce in modo monotono, ad esempio nei fenomeni di usura; se la pendenza è lieve, la carta può far fatica a riconoscere questo andamento, così come nel caso di uno step persistente molto piccolo.

“Pendenza : Progressivo = Step : Persistente”

In generale, gli schemi di controllo sono efficienti soprattutto sull'isolato e persistente, mentre per gli altri casi è di interesse la robustezza.

### Tassonomia delle carte di controllo

Si assume di conoscere la forma funzionale della distribuzione; la *tassonomia* delle carte di controllo è la seguente:

- › *Tipo di variabili*:
  - binarie, conteggio, continue, miste, complessi, ...
  - univariate, multivariate, alta dimensionalità

- › *Tipo di carta:*
  - Shewhart: senza memoria, decisioni basate solo sul campione corrente
  - Change-point: analisi sequenziale
  - Shewhart + change-point
  - Derivanti da carte II (CUSUM, EWMA): con memoria, CUMulative SUMs oppure [Exponentially Weighted Moving Average](#).
- › *Struttura:* individuali vs sottogruppi vs multiscale
- › *Assunzioni:* parametriche, non parametriche, dipendenza temporale
- › *Caratteristiche sorvegliate:* posizione + scala, intera distribuzione, ...

### 3.1 Carte Shewhart

*Riferimenti* Qiu (2014, §3)

Generalmente è accettato l'utilizzo di carte senza memoria, perché per la stabilità del processo si utilizzano dati di tipo retrospettivo, dove non serve monitorare in real time.

Siano  $i = 1, \dots, m$  campioni retrospettivi di dimensione  $n$ , si supponga di calcolare delle statistiche di sintesi (media, varianza, mediana, ...). Queste statistiche vengono graficate contro il tempo assieme a dei *limiti di controllo*, che permettono di discriminare dati in controllo vs. fuori controllo.

Eventualmente, i campioni fuori controllo vengono eliminati (o solo il più estremo) e la procedura viene ripetuta fino a che tutti i campioni risultano in controllo.

Una volta terminato il processo, le stime dei parametri vengono propagate alla fase successiva come se fossero i “veri” valori dei parametri del processo in controllo.

#### Esempio (Carta $\bar{X}$ )

Si consideri la carta sotto l'ipotesi che  $p = 1$  e assumendo che in controllo  $x_{i,j} \sim \mathcal{N}(\mu_0, \sigma_0^2)$  i.i.d. Per ogni sottogruppo si calcola la media

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{i,j},$$

e poi si tracciano i limiti di controllo in modo simile ad un test di ipotesi classico, una volta scelta la soglia  $L$  di cui si parlerà più avanti:

$$\text{LCL} = \hat{\mu}_0 - L \frac{\hat{\sigma}_0}{\sqrt{n}},$$

$$\text{UCL} = \hat{\mu}_0 + L \frac{\hat{\sigma}_0}{\sqrt{n}},$$

dove  $\hat{\mu}_0$  e  $\hat{\sigma}_0$  sono stime di  $\mu_0$  e  $\sigma_0$ . Come stimatori dei parametri, si può usare le medie e le

varianze complessive

$$\hat{\mu}_0 = \frac{1}{m} \sum_{i=1}^m \bar{x}_i,$$

$$\hat{\sigma}_0 = \frac{1}{m} \sum_{i=1}^m \frac{s_i}{d_n},$$

con  $d_n$  determinato in modo da garantire che  $\mathbb{E}[s_i/d_n] = \sigma_i$  se  $x_{i,j} \sim \mathcal{N}(\mu_0, \sigma_0^2)$ . La procedura è iterativa, perché se si elimina un campione vanno ricalcolate le soglie di controllo.

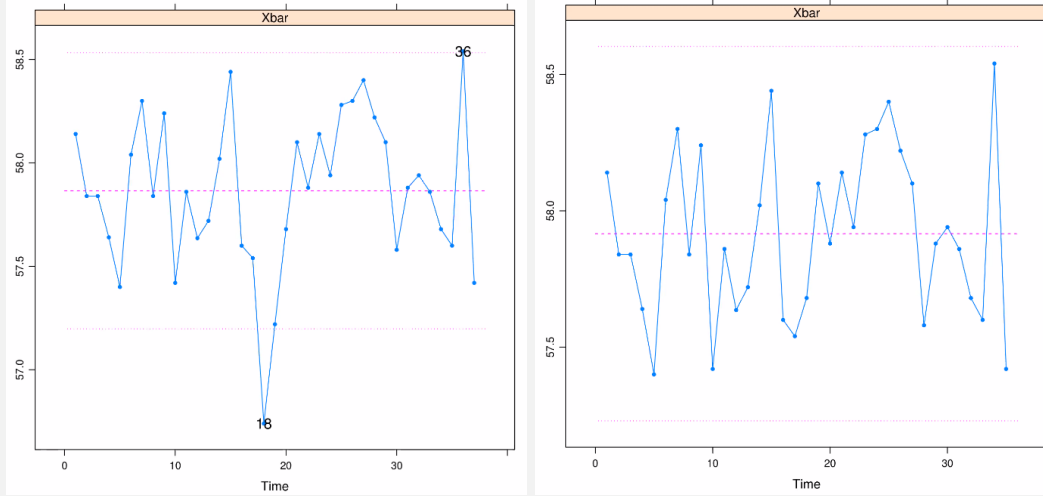


Figura 10: *Sinistra*: esempio di carta Shewhart (o carta  $\bar{X}$ ). *Destra*: la stessa carta  $\bar{X}$ , una volta terminato il processo ed eliminati i campioni identificati come fuori controllo durante la procedura (18, 19 e 36)..

Una volta effettuata questa procedura, si fa finta che le osservazioni rimanenti provengano dalla vera distribuzione in controllo e si stimano i parametri da portare in fase II.

### Osservazioni

1. Poiché si usa la deviazione standard nella carta  $\bar{X}$ , bisogna sorvegliare che anche  $\sigma_0^2$  sia stabile.
2. Con questo processo si troncano i valori anomali, fino a quando il corpo centrale dei dati corrisponde alla distribuzione assunta. Si sta testando un'ipotesi distributiva, con il problema che ci potrebbero essere campioni fuori controllo.

### Esempio (Carta S)

I limiti di controllo vengono calcolati quando il processo è in controllo, per cui va monitorata anche la varianza della distribuzione. Si definisce la carta  $S$  usando

$$s_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{i,j} - \bar{x}_i)^2},$$

con limiti di controllo

$$LCL = L_1 \hat{\sigma}_0, \quad UCL = L_2 \hat{\sigma}_0.$$

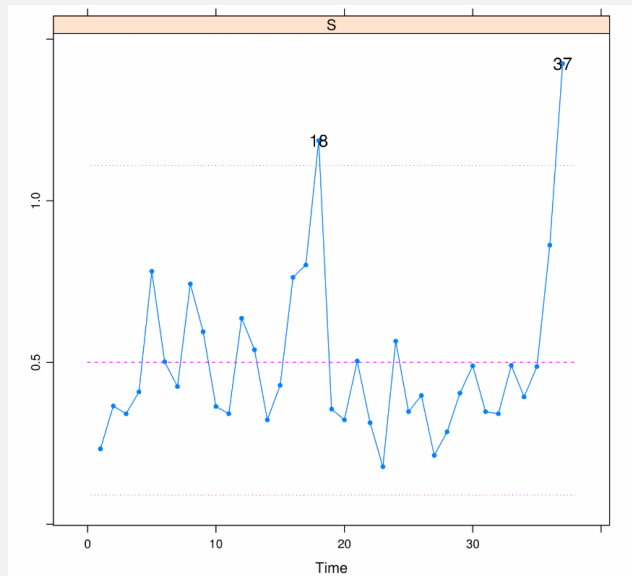


Figura 11: Carta  $S$

Nelle carte precedenti, si può scegliere di usare anche stimatori robusti come la mediana, in modo da tutelarsi rispetto a variazioni da parte di singoli sottogruppi.

Bisogna ora studiare come calcolare i limiti di controllo, che erano state definite per la carta  $\bar{X}$  come

$$LCL = \hat{\mu}_0 - L \frac{\hat{\sigma}_0}{\sqrt{n}}$$

$$UCL = \hat{\mu}_0 + L \frac{\hat{\sigma}_0}{\sqrt{n}}$$

e per la carta  $S$  come

$$LCL = L_1 \hat{\sigma}_0, \quad UCL = L_2 \hat{\sigma}_0.$$

I parametri delle carte vengono scelti affinché le carte abbiano delle performance desiderabili.

## LEZIONE 4: TEST DI IPOTESI MULTIPLI

2020-10-07

Gli estremi  $L_1, L_2, L$  delle carte corrispondono ai valori critici di un test, che sbaglia quando dichiara instabile un processo che non lo è. In generale, nel contesto dei *test di ipotesi multipli*, si è interessati a due quantità

1. *False Discovery Probability*  $\alpha$  : prob. che almeno una osservazione lanci un falso allarme
2. *False Discovery Rate*  $\alpha_i$  : prob. di falso allarme per singola osservazione.

Nei confronti multipli si vuole controllare la FAP, che non deve essere inflazionata. Per la carta  $\bar{X}$ , la condizione diventa

$$P(\text{false alarm}) = P_{IC} \left( \max_{i=1, \dots, m} \frac{|\mathbf{x}_i - \hat{\mu}_0|}{\hat{\sigma}_0 / \sqrt{n}} > L \right) = \alpha,$$

mentre per la carta  $S$  sono necessarie due condizioni:

$$\begin{cases} P(\text{false alarm}) = P_{IC} \left( \min_i \frac{s_i}{\hat{\sigma}_0} < L_1 \text{ o } \max_i \frac{s_i}{\hat{\sigma}_0} > L_2 \right) = \alpha & \text{controllo FAP} \\ P_{IC} \left( \min_i \frac{s_i}{\hat{\sigma}_0} < L_1 \right) = P_{IC} \left( \max_i \frac{s_i}{\hat{\sigma}_0} > L_2 \right) & \text{equi-tailed} \end{cases}$$

### 4.1 Confronti multipli

Si consider un insieme di ipotesi saggate simultaneamente

$$\begin{cases} H_0 : \vartheta_1 = \vartheta_0, \vartheta_2 = \vartheta_0, \dots \\ H_1 : \bar{H}_0 \end{cases}$$

dove la statistica  $t_j$  sagga l'ipotesi  $H_{0j}$ . Si definiscono le quantità

$$\text{FAR}_i = \alpha_i = P_{IC}(t_i \in R_i);$$

$$\text{FAP} = \alpha = 1 - P_{IC} \left( \bigcap_{i \in I} \{t_i \in A_i\} \right).$$

#### Osservazioni

- ›  $P_{IC}$  dipende dalla distribuzione *congiunta* delle statistiche test, per cui complicata da calcolare.
- › Le  $T_j$  in generale sono dipendenti dagli stessi limiti di controllo.
- › Per  $m \rightarrow \infty$ , si ha che  $\text{FAP} \xrightarrow{\text{a.s.}} 1$ . Tuttavia, la dipendenza tende a 0, per cui si può utilizzare la *correzione di Sidak* assumendo indipendenza delle statistiche test

$$\alpha = 1 - \prod_{i=1}^m (1 - \alpha_i) = 1 - (1 - \alpha_i)^m.$$

- › Al crescere del numero di campioni di fase I, per mantenere costante la FAP è necessario correggere molto.

**Metodo dell'unione e intersezione**

Data una famiglia di Hp. nulle, allora

- ›  $H_0$  si *accetta* se si accetta ogni  $H_{0i} \implies H_0 = \bigcap_{i \in I} H_{0i} \implies A = \bigcap_{i \in I} A_i$
- ›  $H_0$  si *rifiuta* se si rifiuta almeno una  $H_{0i} \implies H_1 = \bigcup_{i \in I} H_{0i} \implies R = \bigcup_{i \in I} R_i$ .

Poiché  $R = \bigcup_{i \in I} R_i$ , allora

$$R = \bigcup_{i \in I} \{t : |t_i| > L_\alpha\}$$

$$\iff \max_i |t_i| > L.$$

Usando la correzione di Bonferroni, si ha una regola semplice per controllare la probabilità

$$\text{FAR}_i = \frac{\text{FAP}}{m} = \frac{\alpha}{m}$$

**4.2 Carte combinate**

Invece di avere due carte separate, si sorvegliano entrambe e si lancia un allarme per  $\bar{X}$  oppure per  $S$ :

$$P_{IC} \left( \max_i \frac{|x_i - \hat{\mu}_0|}{\hat{\sigma}_0 / \sqrt{n}} > L \text{ o } \min_i \frac{s_i}{\hat{\sigma}_0} < L_1 \text{ o } \max_i \frac{s_i}{\hat{\sigma}_0} > L_2 \right) = \alpha \quad (\text{carta congiunta})$$

$$P_{IC} \left( \max_{i=1, \dots, m} \frac{|x_i - \hat{\mu}_0|}{\hat{\sigma}_0 / \sqrt{n}} > L \right) = P_{IC} \left( \min_i \frac{s_i}{\hat{\sigma}_0} < L_1 \text{ o } \max_i \frac{s_i}{\hat{\sigma}_0} > L_2 \right) \quad (\text{stessa prob.})$$

$$P_{IC} \left( \min_i \frac{s_i}{\hat{\sigma}_0} < L_1 \right) = P_{IC} \left( \max_i \frac{s_i}{\hat{\sigma}_0} > L_2 \right) \quad (\text{equi-tailed})$$

In generale si trattano in modo simmetrico, ma in condizioni di allarme si guarda prima la carta  $S$ , visto che i limiti per  $\bar{X}$  dipendono da  $\hat{\sigma}_0$ .

Le osservazioni vanno sempre cancellate una per volta, dopo che ci si è assicurati che fossero effettivamente dati fuori controllo (outliers). Altrimenti, potrebbe essere un problema relativo a

- a) Campione contaminato (vedi punti leva, ...)
- b) Assunzioni errate

## LEZIONE 5: STATISTICA MULTIVARIATA

2020-10-08

Il punto di partenza è la *matrice dei dati*  $X$ , che nel caso in questione si assumerà proveniente da una distribuzione normale multivariata.

In particolare,  $X \sim \mathcal{M}_{m \cdot n \times p}(\mathbb{R})$ , in quanto si hanno  $n$  sottogruppi razionali di numerosità  $m$ , da una distribuzione  $p$ -variata.

### 5.1 Statistiche di sintesi

Dalla matrice  $X$ , si possono ottenere le statistiche di sintesi

- › *Vettore medio*:  $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_n)^\top = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} X^\top \mathbb{1}_n$
- › *Matrice di varianze/covarianze*:  $S = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top = \frac{1}{n} X^\top H X$ , dove  $H$  è la *matrice di centramento*

$$H = \frac{1}{n} (I - \mathbb{1}_n \mathbb{1}_n^\top).$$

Si ricorda che  $H$  è simmetrica e idempotente.

- › *Matrice di correlazione*:  $R = D^{-1} S D^{-1}$ , con  $D = \text{diag}(s_{jj})$

In generale, tutta l'informazione campionaria è contenuta nella matrice  $X^\top X$ , mentre standardizzando le osservazioni si perde l'informazione sulla media e varianza.

L'informazione sulla variabilità si può sintetizzare con le statistiche

- › *Varianza generalizzata*:  $|S| = \prod_{j=1}^p \lambda_j$ . In generale, si ha che è proporzionale al volume generato dai vettori di deviazione  $\mathbf{e}^{(j)} = \mathbf{x}^{(j)} - \bar{\mathbf{x}}_j \cdot \mathbb{1}$

$$|S| = \frac{\text{volume}}{(n-1)^p}.$$

Inoltre, il luogo dei punti  $W = \{(\mathbf{x} - \bar{\mathbf{x}})^\top S^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \leq c^2\}$  definisce un iperellissoide, il cui volume è dato da

$$\text{vol}(W) = k_p |S|^{1/2} c^p.$$

- › *Varianza totale (inerzia)*:  $\text{tr}(S) = \sum_{j=1}^p \lambda_j$

Pro	Contro
1. La varianza generalizzata offre un'indicazione della variabilità in termini di volume	1. Diverse $S$ con lo stesso $ S $
2. Interpretabilità in termini di dispersione	2. Influenzata dalle singole variabili
	3. La matrice $R$ offre indicazioni indipendenti da trasformazioni di scala



## 5.2 Matrici normali di dati

Le trasformazioni lineari sono utili per ridurre la dimensione del problema da  $\mathbb{R}^n$  o  $\mathbb{R}^p$  a loro sottospazi, oppure per dare rappresentazioni più convenienti.

Verranno spesso considerate trasformazioni del tipo

$$\mathbf{y} = A\mathbf{x} + \mathbf{b} \quad \text{e} \quad \mathbf{y} = XA^\top + \mathbb{1}\mathbf{b}^\top.$$

In tal caso, il vettore medio è  $A\boldsymbol{\mu}$  e la var/cov è  $ASA^\top$ .

### Trasformazioni notevoli

1. *Media nulla*:  $Y = HXA^\top$
2. *Centramento*:  $Y^c = HX \implies$  media nulla e varianza invariata.
3. *Standardizzazione*:  $Y^s = HXD^{-1} \implies$  media nulla, varianze unitarie e correlazioni.
4. *Ortogonalizzazione*:  $Y^o = HXS^{-1/2} \implies$  media nulla, matrice di var/cov pari a  $I_p$ .

Assumendo di osservare  $x \sim \mathcal{N}(\mu, \Sigma)$  e  $y \sim \mathcal{N}(\nu, \Sigma)$ , si possono costruire le seguenti quantità basate sulla distanza

$$\Delta(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{y}) \quad \text{var. casuale}$$

$$\Delta(\mathbf{x}, \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad \text{var. casuale}$$

$$\Delta(\boldsymbol{\mu}, \boldsymbol{\nu}) = (\boldsymbol{\mu} - \boldsymbol{\nu})^\top \Sigma^{-1} (\boldsymbol{\mu} - \boldsymbol{\nu}) \quad \text{numero}$$

#### Def. (Matrice normale di dati)

Date  $\mathbf{x}_1, \dots, \mathbf{x}_n$  tali che  $\mathbf{x}_i \perp \mathbf{x}_j$  e  $\mathbf{x}_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ , si definisce *matrice normale di dati* la matrice

$$X = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix},$$

ovvero la matrice le cui righe sono generate da normali indipendenti.

#### Teo. (Caratterizzazione delle m.n.d.)

Se  $X_{n \times p}$  è matrice normale di dati da  $\mathcal{N}_p(\mu, \Sigma)$  e  $Y_{m \times q} = AXB$ , con  $A_{m \times n}$  e  $B_{p \times q}$  matrici fissate. Allora,  $Y$  è una m.n.d. se e solo se sono soddisfatte le seguenti condizioni:

- a)  $A\mathbb{1}_n = \alpha\mathbb{1}_m$  per qualche scalare  $\alpha$ , oppure  $B^\top \boldsymbol{\mu} = 0$ .
- b)  $AA^\top = \beta I_m$  per qualche scalare  $\beta$ , oppure  $B^\top \Sigma B = \mathbf{0}_{q \times q}$ .

In tal caso,  $Y$  è una m.n.d. da  $\mathcal{N}_q(\alpha B^\top \boldsymbol{\mu}, \beta B^\top \Sigma B)$ .

**Corollario**

Se  $X_{n \times p}$  è una matrice normale di dati da  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ , allora  $Z = XB$  è sempre matrice normale di dati da  $\mathcal{N}_q(B^\top \boldsymbol{\mu}, B^\top \Sigma B)$ .

**Conseguenze**

1. Il vettore medio campionario è  $\bar{\mathbf{x}} = \frac{1}{n} X^\top \mathbf{1}_n \sim \mathcal{N}_p(\boldsymbol{\mu}, \frac{1}{n} \Sigma)$ .
2. Se  $X_{n \times p}$  m.n.d. da  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$  e  $Y = AXB$ ,  $Z = CXD$  con matrici  $A, B, C, D$  fissate, allora  $Y$  è indipendente da  $Z$  se e solo se
  - $B^\top \Sigma D = \mathbf{0}$ , oppure
  - $AC^\top = \mathbf{0}$
3. Indipendenza del vettore medio campionario  $\bar{\mathbf{x}}$  da  $S$ .

**5.3 Distribuzione Wishart****Def. (Distribuzione di Wishart)**

Siano  $X$  m.n.d. con righe  $\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_i, \Sigma)$ ,  $\mathcal{M} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_m)$  matrice dei vettori medi e  $M = X^\top X$ . Allora, la matrice  $M$  ha distribuzione di *Wishart non centrale*,  $M \sim W_p(\Sigma, \mathcal{M}, m)$ , con matrice di scala  $\Sigma$ , matrice di non centralità  $\mathcal{M}$  e gradi di libertà  $m$ .

**Osservazioni**

- › Se  $\mathcal{M} = \mathbf{0}_{m \times p}$ , allora  $X$  è m.n.d. da  $\mathcal{N}_p(\mathbf{0}, \Sigma)$  e la densità di  $M \sim W_p(\Sigma, m)$  è

$$f_M(M) = \frac{|M|^{(m-p-1)/2} \exp\{-(\text{tr } \Sigma^{-1} M)/2\}}{2^{\frac{mp}{2}} |\Sigma|^{n/2} \Gamma_p(\frac{n}{2})}.$$

- › Si ha che  $\mathbb{E}[M] = \mathbb{E}[\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top] = \sum_{i=1}^n \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top] = m\Sigma$ .

**Teo. (Trasformazione della Wishart)**

Se  $M \sim W_p(\Sigma, m)$  è una *Wishart centrale*,  $B_{p \times q}$  matrice qualsiasi, allora

$$B^\top M B \sim W_q(B^\top \Sigma B, m),$$

in particolare se  $B = a \mathbf{1}_p$ ,  $a M a \sim W_q(a^2 \Sigma, m)$ .

**Osservazioni**

- › La *standardizzazione* è tale che  $B = \Sigma^{-1/2}$ , dunque

$$\Sigma^{-1/2} M \Sigma^{-1/2} \sim W_p(I_p, m)$$

- › Nel caso particolare in cui  $p = 1$ , si ha che  $W_1(\sigma^2, m) = \sigma^2 \chi_m^2$ .
- › Gli elementi sulla diagonale sono marginalmente

$$\frac{m_{ii}}{\sigma_{ii}} \sim \chi_m^2.$$

Si definisce *distanza di Mahalanobis campionaria* tra  $\bar{\mathbf{x}}$  e  $\boldsymbol{\mu}$  la distanza di Mahalanobis ottenuta sostituendo a  $\Sigma^{-1}$  un suo stimatore  $(n-1)S^{-1}$ :

$$D(\bar{\mathbf{x}}, \boldsymbol{\mu})^2 = (n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top S^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}).$$

Si ha che  $D(\bar{\mathbf{x}}, \boldsymbol{\mu})^2 \sim T^2(p, n-1)$ , con

$$T^2(p, m) = \frac{mp}{m-p+1} F_{p, m-p+1},$$

per cui il confronto nella carta di controllo verrà effettuato con i quantili di una  $F$  riscalata.

### Applicazione

Date le quantità  $\bar{\mathbf{x}}$  e  $S$ , si ha che

$$\frac{n-p}{p}(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top S^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim F_{p, n-p}.$$

Quindi, usando il test del rapporto di verosimiglianza per saggiare le ipotesi

$$\begin{cases} H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \end{cases},$$

si hanno due casi:

1. *Varianza nota*:  $LRT = (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \stackrel{H_0}{\sim} \chi_p^2$
2. *Varianza ignota*:  $LRT = (n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top S^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \stackrel{H_0}{\sim} T^2(p, n-1)$ .

## LEZIONE 6: CARTE SHEWHART E CHANGE POINT DI FASE I

2020-10-10

Nella fase I bisogna studiare la *stabilità* delle osservazioni, oltre ad applicare test di *bontà di adattamento*. A questo proposito si può usare una famiglia di *distribuzioni ellittiche*, anche se ci sono delle difficoltà nell'identificare quale membro della famiglia sia il più adatto.

Si può utilizzare una funzione della distanza di  $M$  che guarda l'atipicità della media campionaria rispetto al baricentro, una volta stimate  $\mu$  e  $\Sigma$  in condizioni di stabilità.

Per ogni sottogruppo  $i$  si ha uno scalare  $T^2 = n(\bar{\mathbf{x}}_i - \hat{\boldsymbol{\mu}}_0)^\top \hat{\Sigma}_0^{-1}(\bar{\mathbf{x}}_i - \hat{\boldsymbol{\mu}}_0)$ , con l'allarme che suona se per un  $i$  si ha  $T_i^2 > L$ .

Con una carta Shewhart basata sulla  $T^2$  è necessario valutare a quale delle variabili  $X_j$  imputare il cambiamento fuori controllo (*analisi post segnale*).

### Stimatori multivariati

Stimatori classici  $\hat{\boldsymbol{\mu}}_0 = \frac{1}{m} \sum_{i=1}^m \bar{\mathbf{x}}_i$  e  $\hat{\Sigma}_0 = \frac{1}{m} \sum_{i=1}^m S_i$ , che valgono se si ipotizza la stessa distribuzione delle osservazioni.

Alternativamente, si hanno stimatori *robusti*, che schiacciano le stime sulle osservazioni più coerenti alla nuvola di punti:

- › MCD ([Minimum Covariance Determinant](#))
- › MVE ([Minimum Volume Ellipsoid](#))

Carte Shewart	
Pro	Contro
<ul style="list-style-type: none"> <li>– Facili da interpretare, relative all'osservazione corrente</li> <li>– Molto efficienti se nel campione di fase I si hanno cambiamenti isolati di grande dimensione</li> </ul>	<ul style="list-style-type: none"> <li>– Inefficiente se si hanno cambiamenti medio-piccoli, in quanto non sfrutta la memoria del processo (informazione cumulata)</li> </ul>

### 6.1 Carte di controllo Change-Point

Si ipotizza che esista un  $\tau$  incognito dopo il quale la distribuzione del processo cambia. Sono carte *retrospettive* parametriche basate sul TRV, in cui degli  $m$  campioni ce ne sono  $\tau - 1$  in controllo e

$m - \tau$  fuori controllo:

$$\left. \begin{array}{l} \mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n} \sim p_0 \\ \mathbf{x}_{2,1}, \dots, \mathbf{x}_{2,n} \sim p_0 \\ \vdots \\ \mathbf{x}_{\tau-1,1}, \dots, \mathbf{x}_{\tau-1,n} \sim p_0 \end{array} \right\} \text{IC}$$

$$\left. \begin{array}{l} \mathbf{x}_{\tau,1}, \dots, \mathbf{x}_{\tau,n} \sim p_1 \\ \mathbf{x}_{\tau+1,1}, \dots, \mathbf{x}_{\tau+1,n} \sim p_1 \\ \vdots \\ \mathbf{x}_{m,1}, \dots, \mathbf{x}_{m,n} \sim p_1 \end{array} \right\} \text{OC}$$

Si assume che tutte le osservazioni siano i.i.d e la transizione  $p_0 \rightarrow p_1$  avvenga tramite il cambiamento del parametro  $(\vartheta_0, \psi) \rightarrow (\vartheta_1, \psi)$ . Il test rapporto di verosimiglianza si calcola per  $\tau = 2, \dots, m$  come

$$\begin{aligned} W_\tau &= 2 \log \frac{L \text{ sotto } H_1}{L \text{ sotto } H_0} \\ &= 2 \log \frac{\sup_{\vartheta_0, \vartheta_1, \psi} \prod_{i=1}^{\tau-1} L_i(\vartheta_0, \psi) \prod_{i=\tau}^m L_i(\vartheta_1, \psi)}{\sup_{\vartheta, \psi} \prod_{i=1}^m L(\vartheta, \psi)}. \end{aligned}$$

La statistica di controllo è  $W_\tau$  per  $\tau = 2, \dots, m$ , che verifica l'ipotesi che i parametri siano cambiati dopo un certo punto  $\tau$ .

Il valore di  $\tau$  che rende massimo  $W_\tau$  è la SMV  $\hat{\tau}$ , per cui è necessario conoscere la distribuzione del massimo per trovare la soglia  $L$ :

$$P_{IC} \left( \max_{\tau=2, \dots, m} W_\tau > L \right) = \alpha.$$

In generale, questo si ottiene via simulazione.

#### Esempio ( $p = 1$ , modello normale)

Se si assume che in controllo  $x_{ij} \sim N(\mu_1, \sigma_1)$ , mentre fuori controllo  $x_{ij} \sim N(\mu_2, \sigma_2^2)$  e  $\tau$  è noto, la carta diventa un test  $t$ -Student a due campioni.

Sotto  $H_0$ , la verosimiglianza è

$$\ell_0(\mu, \sigma^2) = -\frac{m}{2} \log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2} - \frac{m}{2\sigma^2} (\bar{x} - \mu)^2,$$

che calcolata nella SMV  $\hat{\mu} = \bar{x}, \hat{\sigma}^2$  diventa

$$\ell_0(\hat{\mu}, \hat{\sigma}^2) = -\frac{m}{2} \log(2\pi\hat{\sigma}^2) - \frac{m}{2}.$$

Sotto  $H_1$ , si può fattorizzare in  $\ell = \ell_1 + \ell_2$ :

$$\ell(\mu_1, \mu_2, \sigma_1, \sigma_2) = -\frac{m_1}{2} \log(2\pi\hat{\sigma}_1^2) - \frac{m_1}{2} - \frac{m_2}{2} \log(2\pi\hat{\sigma}_2^2) - \frac{m_2}{2}.$$

La statistica test è allora

$$\begin{aligned} W_{[m_1, m_2]} &= -2(\ell_0 - \ell) \\ &= m \log \left[ \hat{\sigma}^2 (\hat{\sigma}_1^2)^{-m_1/m} (\hat{\sigma}_2^2)^{-(m-m_1)/m} \right]. \end{aligned}$$

Dato  $T = m\hat{\sigma}^2$ , la *devianza totale* è

$$T = B + W = [m_1(\bar{x}_1 - \bar{x})^2 + m_2(\bar{x}_2 - \bar{x})^2] + [m_1\hat{\sigma}_1^2 + m_2\hat{\sigma}_2^2].$$

Sapendo che  $\bar{x} = \frac{m_1\bar{x}_1 + m_2\bar{x}_2}{m}$ , si può scrivere

$$\begin{aligned} \bar{x}_1 - \bar{x} &= \frac{m_2}{m}(\bar{x}_1 - \bar{x}_2) \\ \bar{x}_2 - \bar{x} &= -\frac{m_1}{m}(\bar{x}_1 - \bar{x}_2) \end{aligned}$$

da cui si ha che

$$B = \frac{m_1 m_2}{m^2} (\bar{x}_1 - \bar{x}_2)^2,$$

dunque

$$T = \frac{m_1 m_2}{m^2} (\bar{x}_1 - \bar{x}_2)^2 + \frac{m_1 \hat{\sigma}_1^2 + m_2 \hat{\sigma}_2^2}{m}.$$

Sostituendo  $\hat{\sigma}^2$ , la statistica può essere fattorizzata in  $W_{m_1, m_2}^*$  e  $W_{m_1, m_2}^{**}$ :

$$W_{[m, m_2]} = m \log \left[ \frac{1 + c(r^2 - 1)}{r^{2c}} \right] + m \log \left[ 1 + \frac{c(1 - c)}{1 + c(r^2 - 1)} d^2 \right],$$

dove  $c = \frac{m_1}{m}$ ,  $d = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}^2}$  e  $r = \frac{\hat{\sigma}_1}{\hat{\sigma}_2}$ .

In questo modo, si ha un pezzo che dipende solo da  $\sigma_1/\sigma_2$  e uno che dipende solo dallo scostamento delle medie campionarie.

In generale,  $W_{[m, m_2]}$  è asintoticamente  $\sim \chi_2^2$ , per cui valori grandi portano a grandi cambiamenti di media o varianza. Con la scomposizione sopra, è possibile risalire a quale dei due parametri è cambiato.

### Osservazione

La distribuzione della statistica varia al variare del valore di  $m_1$ , in particolare  $\mathbb{E}_{IC} [W_{m_1}]$  è più grande se  $m_1$  o  $m_2$  sono piccoli. Una modifica che si è rivelata utile in pratica utilizza

$$W_\tau^* = \frac{W_\tau}{\mathbb{E}_{IC} [W_\tau]}.$$

## LEZIONE 7: CARTE DI FASE I NON PARAMETRICHE

2020-10-16

Ci sono due possibilità per utilizzare metodi non parametrici:

1. Uso carte standard con limiti di controllo *aggiustati* con metodi non parametrici, come metodi di permutazione o bootstrap non parametrico.
2. Usare *statistiche distribution-free*, in generale metodi basati sui *ranghi*

### 7.1 Metodi di permutazione

*Riferimenti* Ernst (2004)

I metodi di permutazione sono tecniche di ricampionamento che permettono di effettuare test di ipotesi non parametrici, usando deboli ipotesi distributive.

**Ipotesi** Sotto  $H_0$ , si assume che

1. Le osservazioni siano *scambiabili*, cioè per ogni permutazione  $\sigma$

$$p(x_1, x_2, \dots, x_n) = p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}).$$

2. La distribuzione della statistica  $T$  non cambia permutando le osservazioni.

**Metodo** Dato un campione  $\mathbf{y}$  si calcola  $t = T(\mathbf{y})$  e la si confronta con la distribuzione dei  $T$  ottenuta permutando i dati.

I test di permutazione sono *distribution-free*, nel senso che le ipotesi distributive si limitano a scambiabilità e identica distribuzione. Non è necessario assumere indipendenza tra le osservazioni, perché si può utilizzare il *bootstrap a blocchi* (Kreiss e Paparoditis, 2011) nel caso ci sia autocorrelazione.

I test sono esatti, cioè garantiscono  $\alpha$  nominale sia

- › *condizionatamente* al dataset osservato
- › *incondizionatamente*, in quanto sono applicabili a qualunque campione estratto da una distribuzione di permutazioni.

### 7.2 Verifica $X_1 \sim Q_1(\vartheta_1)$ vs. $X_2 \sim Q_2(\vartheta_2)$

Si vuole verificare l'ipotesi nulla  $Q_1 = Q_2 = Q$ , ovvero

$$\begin{cases} H_0 : X_1 \stackrel{d}{=} X_2 \\ H_1 : X_1 \stackrel{d}{>} X_2 \end{cases}$$

inoltre si può anche verificare

$$\begin{cases} H_0 : \vartheta_1 = \vartheta_2 = \vartheta \\ H_1 : \vartheta_1 > \vartheta_2 \end{cases}$$

Sotto  $H_0$  i dati osservati da  $X_1$  e  $X_2$  sono *scambiabili* da  $Q$  ignota, dunque

$$X_{1,1}, \dots, X_{1,m} \sim Q, \quad X_{2,1}, \dots, X_{2,n} \sim Q.$$

**Idea** Ci si condiziona a un insieme di stat. sufficienti per  $Q$ , che in questo caso è l'insieme di dati osservati.

Se le osservazioni sono scambiabili, si possono concatenare i due gruppi in un singolo vettore (*pooled vector*)

$$\mathbf{Y} = (Y_1, \dots, Y_N) = (X_{1,1}, \dots, X_{1,m}, X_{2,1}, \dots, X_{2,n}).$$

### Osservazioni

1. Sotto  $H_0$ , anche tutte le componenti di  $\mathbf{Y}$  sono scambiabili.
2. Lo spazio di permutazione campionario  $\mathcal{Y}$  è dato da

$$|\mathcal{Y}| = \{(Y_{\sigma(1)}, \dots, Y_{\sigma(N)}) : \sigma \text{ permutazione di } 1, \dots, N\},$$

per cui  $|\mathcal{Y}| = N!$ .

3. Ci si condiziona al campione  $\mathbf{Y}$  osservato, per cui

$$P_{H_0}(Y_1 = y_1, \dots, Y_N = y_N | \mathbf{Y} \in \mathcal{Y}) = \frac{1}{N!}.$$

### Procedura per un test di permutazione

1. Si definiscono una ipotesi nulla  $H_0$  e una statistica  $T$ , ad esempio  $T = T_1 - T_2$ .
2. Dato  $\mathbf{Y}$  si genera  $\mathcal{Y}$ , dato da tutte le permutazioni degli elementi, e  $T_j$  è la statistica calcolata sulla  $j$ -esima permutazione.
3. Si trova la *distribuzione di permutazione di  $T$*  condizionata a  $\mathbf{Y}$  e si ottiene il  $p$ -value condizionato

$$p = P_{H_0}(T > T_d | \mathbf{Y}) = \frac{\sum_{j=1}^{N!} \mathbb{1}_{T_j > T_d}}{N!}.$$

### Osservazioni

- › I test di permutazione così ottenuti sono condizionati alla prima  $n$ -upla, distribution-free e condizionatamente esatti.
- › Siccome  $\alpha$  è controllato per tutti i possibili dataset iniziali  $\mathbf{Y}$ , i test sono anche *incondizionatamente esatti*.
- › Condizionatamente al dataset iniziale, le possibili combinazioni in due gruppi di dimensione  $n$  e  $m$  sono  $\binom{n}{m}$ , per cui si può semplificare i calcoli in

$$p = P_{H_0}(T > T_d | \mathbf{Y}) = \frac{\sum_{j=1}^{\binom{n}{m}} \mathbb{1}_{T_j > T_d}}{\binom{n}{m}}.$$

**Problema** Calcolare tutte le possibili suddivisioni è proibitivo.



**Soluzione** Simulazione Monte Carlo usando permutazioni casuali per ottenere un  $p$ -value approssimato.

Si generano  $B$  permutazioni casuali di  $\mathbf{Y}$ , ottenendo un campione casuale dalla distribuzione di permutazione di  $T$  :

$$(T_1^*, T_2^*, \dots, T_B^*)$$

In questo modo si ottiene una stima non distorta e consistente della distribuzione di  $T$ , usando la F.d.R. empirica

$$\hat{F}(z|\mathbf{Y}) = \hat{F}_B^*(z) = \frac{\sum_{j=1}^B \mathbb{1}_{T_j^* \leq z}}{B},$$

e del  $p$ -value condizionato, calcolando

$$\hat{p} = P(T^* \geq T_d|\mathbf{Y}) = \frac{\sum_{j=1}^B \mathbb{1}_{T_j^* \geq T_d}}{B}.$$

Perché usare permutazioni e non il bootstrap non parametrico?

#### Bootstrap non parametrico

- › Risultati approssimati, ma comunque ok.
- › Richiede  $n \cdot m$  grande.

#### Permutazioni

- › Risultati (quasi) esatti per ogni  $m$  e  $n$ .
- › Richiede assenza di autocorrelazione, altrimenti per forza bootstrap.

### 7.3 Permutazioni multivariate

In questo caso si concatenano  $m$  sottogruppi razionali di dimensione  $n$ , per cui si ottiene un array di dimensione  $p \times mn$ . Sia  $\mathbf{Y} \in \mathcal{S}$  con densità  $q(\cdot)$ , allora

$$P(\mathbf{Y} = \mathbf{a} | \mathbf{Y} \in \mathcal{Y}) = \begin{cases} \frac{q(\mathbf{a})}{\sum_{i=1}^{N_{\mathcal{Y}}} q(\mathbf{Y}_i)} & \text{se } \mathbf{a} \in \mathcal{Y} \\ 0 & \text{altrimenti} \end{cases},$$

in particolare se  $q(\mathbf{Y}_1) = \dots = q(\mathbf{Y}_{N_{\mathcal{Y}}})$ , si ha

$$P(\mathbf{Y} = \mathbf{a} | \mathbf{Y} \in \mathcal{Y}) = \begin{cases} \frac{1}{N_{\mathcal{Y}}} & \text{se } \mathbf{a} \in \mathcal{Y} \\ 0 & \text{altrimenti} \end{cases},$$

Siano ora  $\mathbf{x}_{i,j} \in \mathbb{R}^p$ , con  $i = 1, \dots, m$  e  $j = 1, \dots, n$ . Il pooled array diventa la matrice  $\mathbf{Y}_{p \times mn}$

$$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_{mn}) = \left( \underbrace{\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n}}_{1^\circ \text{ sottogruppo}}, \underbrace{\mathbf{x}_{2,1}, \dots, \mathbf{x}_{2,n}, \dots, \mathbf{x}_{m,1}, \dots, \mathbf{x}_{m,n}}_{2^\circ \text{ sottogruppo}} \right),$$

$$\mathbf{y}_1 = \begin{pmatrix} \mathbf{x}_{1,1} & x_{112} & \dots & x_{11m} \\ x_{121} & x_{122} & \dots & x_{12m} \\ \vdots & \vdots & \dots & \vdots \\ x_{1p1} & x_{1p2} & \dots & x_{1pm} \end{pmatrix},$$

in cui ogni sottogruppo  $\mathbf{x}_{i,j}$  è un'osservazione  $p$ -variata. La funzione di densità congiunta di  $\mathbf{Y}$  è, se i.i.d.,

$$q(\mathbf{Y}) = \prod_{i=1}^{mn} p(\mathbf{y}_i) = \prod_{i=1}^n \prod_{j=1}^m p(\mathbf{x}_{i,j}).$$

Sotto ipotesi di scambiabilità delle osservazioni, ci sono  $(mn)!$  modi di riorganizzare le colonne del pooled array.

### Osservazione

- › L'ipotesi di scambiabilità vale per le colonne della matrice, ma non per le righe. Ogni osservazione è multivariata e non si può rimescolare.

## LEZIONE 8: CARTE DI FASE I NON PARAMETRICHE (II)

2020-10-21

### 8.1 Calcolo dei limiti di controllo

I risultati suggeriscono di calcolare i limiti di controllo per una statistica in modo che

$$P_{\text{perm}}(\text{allarme}) = \alpha,$$

dove

$$P_{\text{perm}}(\mathbf{Y} = \mathbf{a}) = \begin{cases} \frac{1}{(nm)!} & \text{se } \mathbf{a} \in \mathcal{Y} \\ 0 & \text{altrimenti} \end{cases} \quad (1)$$

Se il processo è in controllo, la (1) garantisce che

$$P_{\text{IC}}(\text{allarme}) = \alpha.$$

Infatti,

$$P_{\text{IC}}(\text{allarme}) = \mathbb{E}_{\mathcal{Y}} [P_{\text{IC}}(\mathbf{Y} | \mathbf{Y} \in \mathcal{Y})].$$

**Complicazione 1**  $P_{\text{perm}}(\cdot)$  è una distribuzione discreta, quindi trovare esattamente  $\alpha$  non è facile. I “gradini” però sono piccoli, quindi non si va lontani dall’ $\alpha$  cercato.

**Complicazione 2** Di nuovo,  $(nm)!$  è molto grande, per cui si procede via simulazione Monte Carlo.

#### Esempio (Carta basata sulla $T^2$ di Hotelling)

Si basa, per  $i = 1, \dots, m$ , su

$$T_i^2 = n(\bar{\mathbf{x}}_i - \hat{\boldsymbol{\mu}}_0)^\top \hat{\Sigma}_0^{-1}(\bar{\mathbf{x}}_i - \hat{\boldsymbol{\mu}}_0),$$

con un limite di controllo  $L$ . Siccome interessa rifiutare quando *almeno una* è fuori controllo, bisogna calcolare via simulazione  $L$  tale che

$$P_{\text{perm}}\left(\max_{i=1, \dots, m} T_i > L\right) \approx \alpha.$$

In tal caso, la procedura è

1. Si genera  $\mathbf{Y}^{(h)} \sim P_{\text{perm}}$ .
2. Si calcolano  $\hat{\boldsymbol{\mu}}_0^{(h)}, \hat{\Sigma}_0^{(h)}$  nello stesso modo in cui si sono calcolati  $\hat{\boldsymbol{\mu}}_0$  e  $\hat{\Sigma}_0$  dai dati originali.
3. Si calcola la statistica

$$z^{(h)} = \max_{i=1, \dots, m} n(\bar{\mathbf{x}}_i^{(h)} - \hat{\boldsymbol{\mu}}_0^{(h)})^\top (\hat{\Sigma}_0^{(h)})^{-1}(\bar{\mathbf{x}}_i^{(h)} - \hat{\boldsymbol{\mu}}_0^{(h)})$$

4. Si calcola il quantile empirico di livello  $1 - \alpha$  dalla distribuzione di  $z^{(1)}, \dots, z^{(H)}$ .

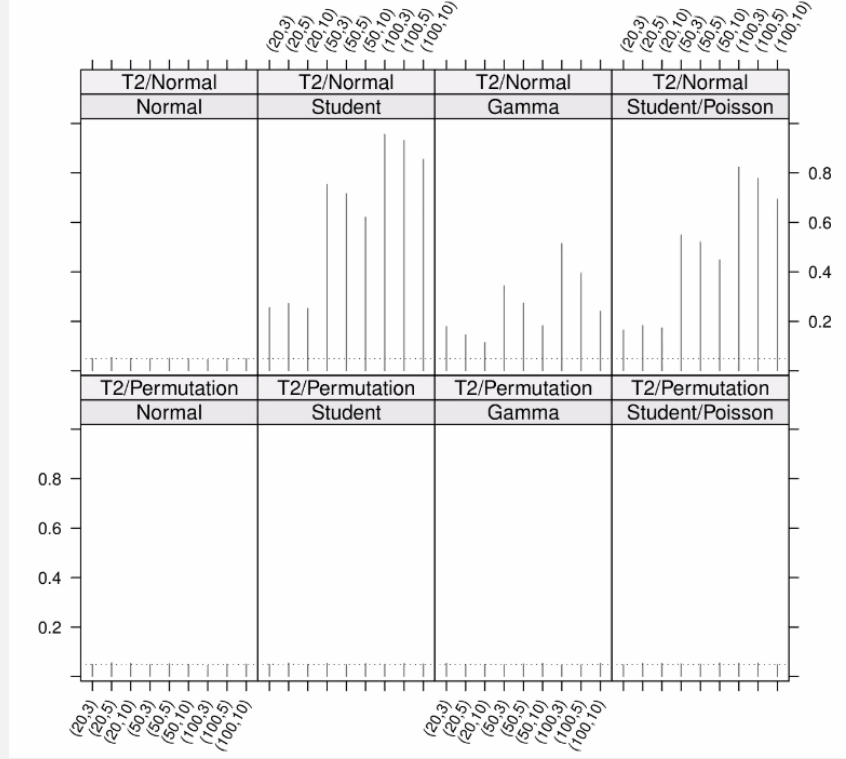


Figura 12: Probabilità di falso allarme nominale e reale nel caso parametrico e non parametrico, con le varie deviazioni dalla normalità fuori controllo.

$$\mathbf{x}_{i,j} \stackrel{\text{IC}}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

$$\mathbf{x}_{i,j} \stackrel{\text{FC}}{\sim} \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\delta}, \Sigma)$$

Probabilità di lanciare un vero allarme in funzione di  $\lambda$ , che rappresenta una misura standardizzata di scostamento dal baricentro.

$$\lambda = \sqrt{\boldsymbol{\delta}^\top \Sigma^{-1} \boldsymbol{\delta}}.$$

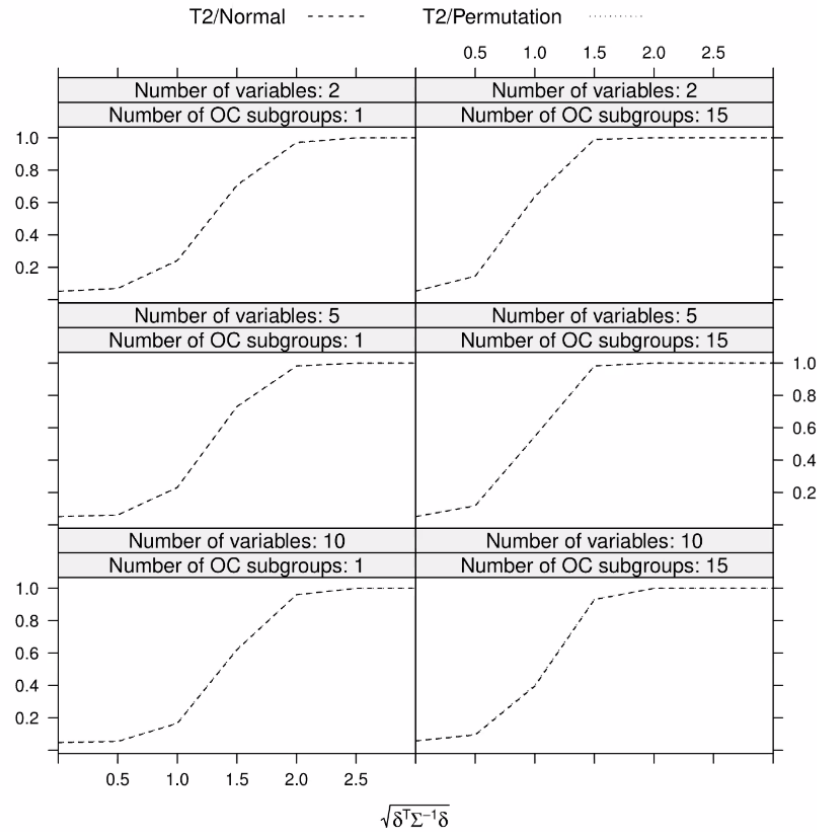


Figura 13: Performance (potenza) della  $T^2$  di Hotelling per un processo gaussiano, caso parametrico e non parametrico. La potenza ottenuta in modo non parametrico non è influenzata né dalla dimensionalità  $p$  né dal numero di sottogruppi  $m$

Nelle carte non parametriche c'è spesso un problema di *segnale spurio*, nel senso che l'allarme viene suonato dalla media quando in realtà è cambiata la varianza.

## LEZIONE 9: TRASFORMAZIONE PER RANGHI

2020-10-22

**Def. (Trasformazione rango)**

Siano  $y_1, y_2, \dots, y_N \sim F(\cdot)$ , la trasformazione

$$r_i = \# \{y \leq y_i\}$$

si definisce *trasformazione rango* di  $y$  e si indica  $r_i = \text{rank } y_i$ .

**Osservazione**

- › Il vettore dei ranghi  $\mathbf{r} = (r_1, r_2, \dots, r_N)$  è una permutazione di  $1, 2, \dots, N$ .
- › Il vettore  $\mathbf{R}$  è uniformemente distribuito su  $\mathcal{S}_r = \{\mathbf{r} \text{ permutazione di } 1, \dots, N\}$

$$P(\mathbf{R} = \mathbf{r}) = \frac{1}{N!}$$

- › I momenti di base derivano dalla distribuzione uniforme discreta:

$$\mathbb{E}[R_j] = \frac{N+1}{2}$$

$$\mathbb{V}[R_j] = \frac{(N+1)(N-1)}{12}$$

$$\text{Cov}(R_i, R_j) = -\frac{N+1}{12}$$

$$\sum_{i=1}^N R_i = \frac{N(N+1)}{2}$$

Sotto  $H_0$ , i ranghi sono *scambiabili* ma *dependenti*, in quanto la somma è vincolata. Stare attenti a correggere la varianza della media se si sorveglia  $\bar{R}$ .

- › Permutando  $y_i$ , i ranghi  $r_i$  permutano nello stesso modo, in quanto dipendono dai dati ordinati in modo crescente:

$$\text{rank } y_{\sigma(i)} = r_{\sigma(i)}.$$

**9.1 Distribuzione dei ranghi**

Siano  $Y_{(1)} < \dots < Y_{(N)}$  le statistiche di ordine di  $(Y_1, Y_2, \dots, Y_N) \stackrel{i.i.d}{\sim} F(\cdot)$  continua. Allora,  $F(Y_{(1)}) < \dots < F(Y_{(N)})$  sono distribuite come le statistiche d'ordine di  $U \sim \text{Unif}(0, 1)$ .

Inoltre, poiché  $F(\cdot)$  è monotona i ranghi di  $y_1, \dots, y_N$  sono uguali ai ranghi di  $u_1, \dots, u_N$ .

$$\text{rank } y_i \stackrel{\text{monot.}}{=} \text{rank}(F(y_i)) = \text{rank } u_i.$$

Inoltre, se  $u_1 = F(y_1), \dots, u_n = F(y_N)$ , si ha che  $U_i \stackrel{i.i.d}{\sim} \text{Unif}(0, 1)$  e di conseguenza

$$P(U_i \leq a) = a.$$

Considerato un insieme di dati, sotto  $H_0$  vale la scambiabilità e si può scrivere il pooled vector

$$\mathbf{Y} = (x_{1,1}, \dots, x_{m,n}) \xrightarrow{\text{rank}} \mathbf{R} = (r_{1,1}, \dots, r_{m,n}).$$

Si può calcolare una statistica di controllo  $T(\mathbf{R})$ , che è funzione del campione  $y_1, \dots, y_N$  solo attraverso i suoi ranghi.

**Importante** La statistica  $T(\mathbf{R})$  ha la stessa distribuzione sotto  $H_0$  qualsiasi sia la distribuzione dei dati.

### Esempio (Test non parametrici)

I test di Wilcoxon e Kruskal-Wallis sono versioni non parametriche dei test  $t$  / ANOVA.

Considero  $X_1 \sim F_1$ ,  $X_2 \sim F_2$ , si vuole verificare

$$H_0 : F_1 = F_2 \quad \text{vs.} \quad H_1 : F_2 = F_1(x - \Delta x).$$

Dati due campioni di dimensione  $m, n$  con  $m < n$ , si calcola il pooled vector unendo i ranghi dei due campioni

$$(\underbrace{R_1, \dots, R_m}_{\text{rank } X_1}, \underbrace{Q_1, \dots, Q_n}_{\text{rank } X_2}),$$

e si considera la statistica  $W = \sum_{i=1}^n R_i - \frac{m(m+1)}{2}$ .

Il valore  $\frac{m(m+1)}{2}$  è il valore che assumerebbe la somma dei ranghi se al primo campione corrispondesse la prima parte del vettore dei ranghi, cioè se  $F_2 = F_1(x - \Delta x)$ .

### Esempio (Carta di Fase I basata sui ranghi)

*Riferimenti* Jones-Farmer, Jordan et al. (2009)

Sorveglia la “posizione” del processo, con una costruzione analoga a  $\bar{X}$ . Si calcola per ogni sottogruppo la media dei ranghi

$$\bar{r}_i = \frac{1}{n} \sum_{j=1}^n r_{i,j},$$

media dei ranghi del sottogruppo. Dal momento che

$$\mathbb{E}[\bar{r}_i] = \frac{nm+1}{2}, \quad \mathbb{V}[\bar{r}_i] = \frac{(nm-n)(nm+1)}{12n},$$

i limiti usati sono:

$$\frac{nm+1}{2} \pm L \sqrt{\frac{(nm-n)(nm+1)}{12n}}.$$

Nell'articolo si standardizzano le variabili e si usa l'approssimazione normale, ma conviene

utilizzare un approccio non parametrico esatto.



## LEZIONE 10: TRASFORMAZIONE PER RANGHI (II)

### Esempio (Carta di Fase I basata sui ranghi (ii))

*Riferimenti* Jones-Farmer e Champ (2010)

Nel secondo articolo, gli autori costruiscono una carta analoga alla carta  $S$  tramite

$$\bar{Q}_i = \frac{1}{n} \sum_{j=1}^n q_{i,j}^2,$$

dove  $q_{i,j}$  sono i ranghi degli scostamenti dalla mediana del pooled vector:

$$q_{i,j} = \text{rank} \{ |x_{i,j} - m| : i = 1, \dots, m; j = 1, \dots, n \}.$$

I limiti sono

$$\bar{Q}_i \pm L,$$

con  $L$  calcolato nel solito modo via simulazione.

### Osservazione

Siccome si utilizza la mediana  $m$  del vettore originale, non vale esattamente l'indipendenza della distribuzione dei  $Q_{i,j}$  da quella del processo generatore dei dati. In realtà, empiricamente si osserva che l'indipendenza è soddisfatta.

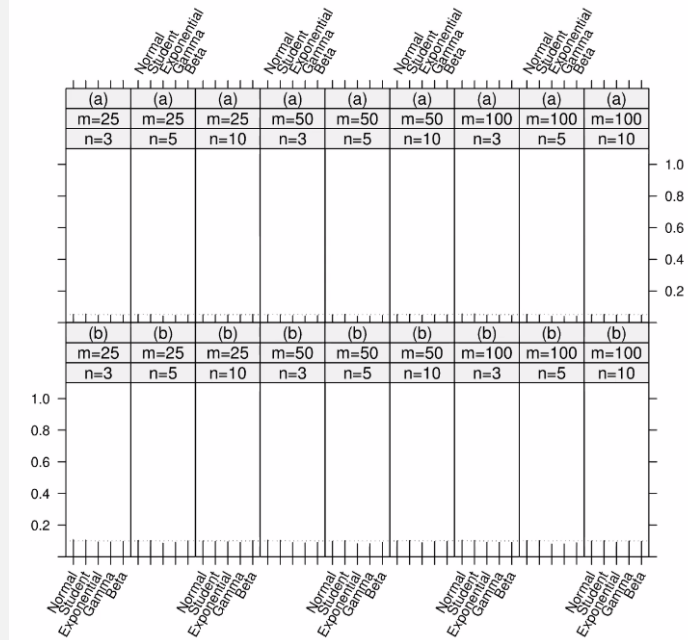


Figura 14: Probabilità di falso allarme per le due carte proposte da Jones-Farmer, Jordan et al. (2009) e Jones-Farmer e Champ (2010).

A livello di potenza, sotto normalità la carta  $\bar{X} - S$  applicata ai ranghi ha una performan-

ce pessima rispetto alla carta di permutazione (*solida*) e alla carta normale (*tratteggiata*). Globalmente, il metodo basato sulle permutazioni sembra migliore.

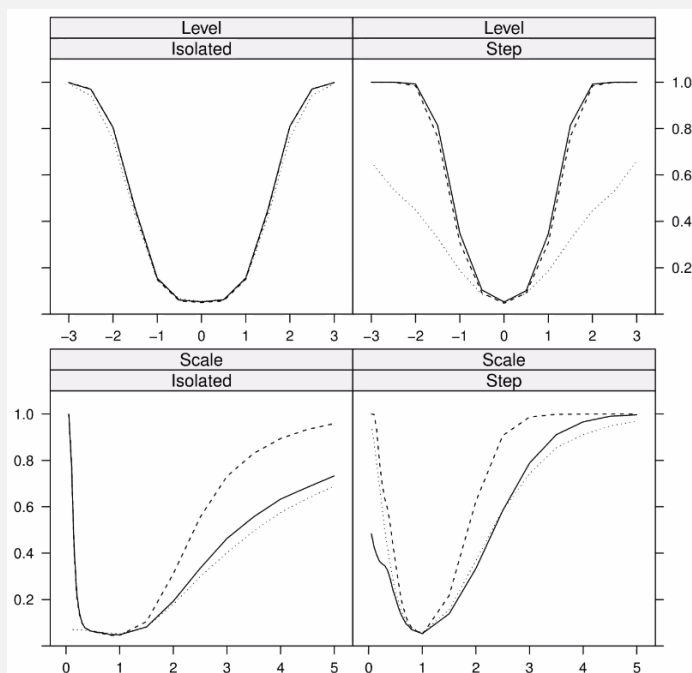


Figura 15: Probabilità di allarme corretto con  $m = 25$  sottogruppi di dimensione 5, distribuzione normale. L'“isolated” è un cambiamento in media (level) o varianza (scale) di un solo sottogruppo. Nel caso “step” solo gli ultimi 10 sottogruppi.

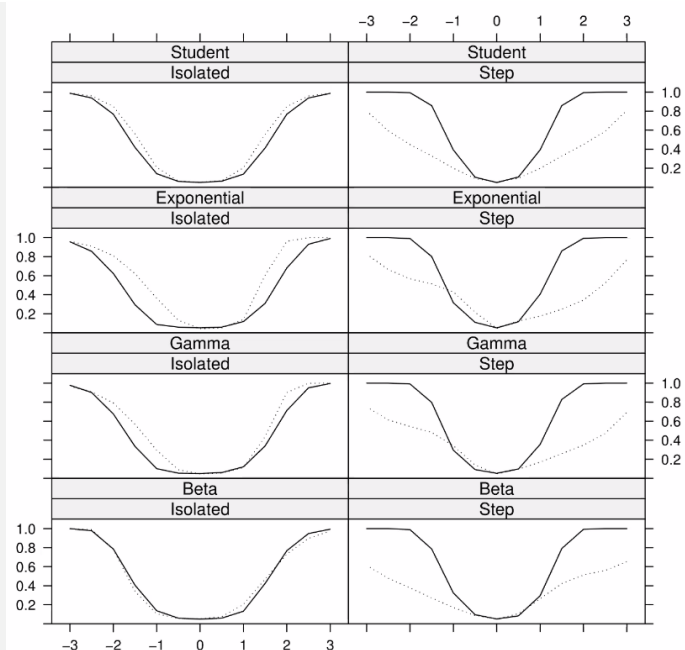


Figura 16: Quando il processo non proviene da una distribuzione normale, le differenze tra metodo di permutazione (*solida*) e ranghi (*puntini*) è molto diversa nel caso di cambiamenti isolati o persistenti.

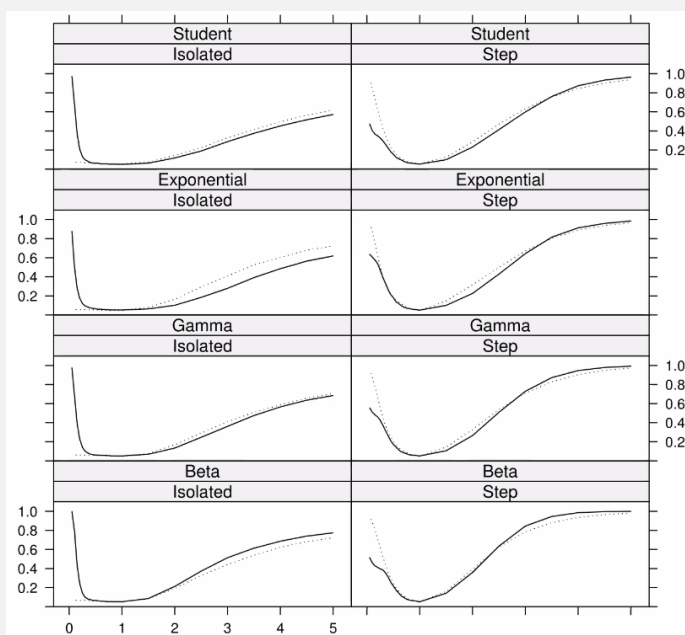


Figura 17: Per quanto riguarda la varianza, i due metodi si comportano sostanzialmente in modo uguale.

---

Parte III

Fase II

## LEZIONE 11: INTRODUZIONE ALLA FASE II

2020-10-28

## 11.1 Generalità sulla fase II

Si pensa ad un processo che comincia in controllo, ma che ad un *istante ignoto* di tempo  $\tau$  diventi fuori controllo:

$$\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n} \sim F_1(\cdot) \quad \text{IC}$$

$$\mathbf{x}_{2,1}, \dots, \mathbf{x}_{2,n} \sim F_2(\cdot) \quad \text{IC}$$

$$\vdots$$

$$\mathbf{x}_{\tau-1,1}, \dots, \mathbf{x}_{\tau-1,n} \sim F_{\tau-1}(\cdot) \quad \text{IC}$$

$$\mathbf{x}_{\tau,1}, \dots, \mathbf{x}_{\tau,n} \sim F_{\tau}(\cdot) \quad \text{FC}$$

$$\mathbf{x}_{\tau+1,1}, \dots, \mathbf{x}_{\tau+1,n} \sim F_{\tau+1}(\cdot) \quad \text{FC}$$

$$\vdots$$

In generale si assume un cambiamento costante

$$F_{\tau} = F_{\tau+1} = F_{\tau+2} = \dots$$

Ad ogni istante si usano tutti i dati raccolti per verificare il sistema di ipotesi

$$\begin{cases} H_0 : \tau > i \\ H_1 : \tau \leq i \end{cases}$$

In pratica, si disegna una statistica di controllo  $S_i$  variabile nel tempo, basata su tutto il passato della serie:

$$S_i = s(\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,n}, \dots, \mathbf{x}_{i-1,1}, \dots, \mathbf{x}_{i-1,n}, \dots, \mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n}).$$

## Osservazione: Differenze tra Fase 1 e Fase 2

## Fase I

- › Si cerca un campione rappresentativo di un processo in controllo
- › Metriche di interesse sono FAP e FAR.

## Fase II

- › Assumendo di conoscere il processo IC, si vuole verificare che il processo non vada fuori controllo.
- › Metriche di interesse sono la velocità di allarme e il controllo dei falsi allarmi.

In Fase II, un approccio parametrico significa che si conosce *esattamente*  $F_0(\cdot; \hat{\vartheta})$ , perché i parametri sono stati stimati accuratamente in Fase I. Inoltre, si possono effettuare test di bontà di adattamento per verificare se vi è conformità alle assunzioni parametriche.

Le carte di controllo di Fase II possono avere *memoria* di diversi tipi:

Durata

- › *Infinita*: al tempo  $i$  sono usati tutti i campioni  $1, \dots, i$ . Problematica se non ci sono scritture ricorsive per il test (e.g. *Generalized Likelihood Ratio*).
- › *Finita*: al tempo  $i$  si usano i campioni  $i - k, \dots, i$ .
- › *Nulla*: al tempo  $i$  si usa solo l' $i$ -esimo campione.

Peso

- › *Uniforme*: campioni tutti con lo stesso peso (CUSUM)
- › *Non uniforme*: campioni con pesi differenti (EMWA).

**Def. (Statistica di controllo)**

Si indichi con  $g(\mathbf{x}_j) = g(x_{1j}, \dots, x_{nj})$  una funzione del  $j$ -esimo sottogruppo razionale. Per controllare il processo si utilizza una *statistica di controllo* al tempo  $t$ , che in generale si può esprimere come

$$w_t = a_t + \sum_{j=0}^t b_j g(\mathbf{x}_{t-j}), \quad j = 1, \dots, t.$$

Le carte di Fase II più importanti sono le “magnifiche tre”:

1. *Shewhart*: senza memoria, ci si basa solo sul campione attuale:

$$w_t = g(\mathbf{x}_t), \quad a_t = 0, \quad b_j = \begin{cases} 1 & \text{se } j = t \\ 0 & \text{altrimenti} \end{cases}.$$

2. *CUSUM*: tutte le statistiche pesano in modo uguale.
3. *EWMA*: osservazioni pesate in modo decrescente da una *costante di liscio esponenziale*  $\lambda \in (0, 1)$ .

$$\begin{aligned} w_t &= \lambda g(\mathbf{x}_t) + \overbrace{(1 - \lambda)w_{t-1}}^{\text{memoria}} \\ &\vdots \\ &= \sum_{j=0}^{t-1} \lambda(1 - \lambda)^j \mathbf{x}_{t-j} + (1 - \lambda)^t w_0 \\ a_t &= (1 - \lambda)^t w_0; \\ b_j &= \lambda(1 - \lambda)^j. \end{aligned}$$

Per  $\lambda \rightarrow 0$  è a memoria uniforme, buona per cambiamenti persistenti, mentre per  $\lambda \rightarrow 1$  diventa una Shewhart, buona per cambiamenti repentini.

## 11.2 Average Run Length (ARL)

Per la carta Shewhart, che si basa sui test di ipotesi classici, la probabilità di errore di I tipo  $\alpha$  e la potenza  $1 - \beta$  sono legate alla ARL.

**Def. (Run length)**

Si definisce *run length* il tempo al primo allarme,

$$RL = \inf \{t : w_t \neq (LCL, UCL)\}$$

In Fase II, poiché non si sa quanti test di ipotesi verranno eseguiti, si preferisce controllare il suo valore atteso:

**Def. (Zero-state Average Run Length)**

Si definisce la *zero-state average run length*  $ARL_\tau$  come l'average run length con il processo che va fuori controllo al tempo  $\tau$

$$ARL = \begin{cases} \mathbb{E}[RL|\tau = \infty] & \text{tempo medio al falso allarme} \\ \mathbb{E}[RL|\tau = 1] & \text{tempo medio all'allarme se FC a } t = \tau \end{cases}.$$

A volte se  $\tau = \infty$ , si indica con  $ARL_0$ .

**Def. (Steady state Average Run Length)**

Si definisce la *steady-state average run length* come l'ARL del processo osservato a partire da un tempo  $\tau$

$$ARL = \mathbb{E}[RL - \tau + 1 | RL \geq \tau].$$

Poiché per  $\tau \rightarrow \infty$  raggiunge rapidamente un asintoto, a volte si considera il processo che “sta in controllo per un po’”

$$ARL = \lim_{\tau \rightarrow \infty} \mathbb{E}[RL - \tau + 1 | RL \geq \tau].$$

Per disegnare una carta di controllo, è di interesse tutelarsi rispetto ai falsi positivi, quindi che  $\mathbb{E}_\infty[RL]$  sia molto grande (analogo di  $\alpha$  fissato).

Condizionatamente a questo, si deve trovare i parametri della carta che danno  $\mathbb{E}_\tau[RL - \tau + 1 | RL \geq \tau]$  più piccolo possibile (analogo di  $\min \beta$ ).

Queste condizioni si traducono nella stima dei parametri della carta, attraverso la soluzione di un minimo vincolato

## LEZIONE 12: CARTE DI CONTROLLO DI FASE II

2020-10-29

### 12.1 Carta Shewhart

La carta Shewhart saggia l'ipotesi  $H_0 : \mu = \mu_0$  contro  $H_1 : \mu \neq \mu_0$ , tramite il test

$$\frac{\bar{x}_i - \mu_0}{\sqrt{\sigma_0^2/n}} \stackrel{H_0}{\sim} \mathcal{N}(0, 1),$$

considerando ogni sottogruppo razionale come indipendente l'uno dall'altro.

La carta è dunque

#### Limiti di controllo per una carta Shewhart sulla media

La statistica è  $\bar{x}_i$ , con i limiti di controllo

$$\mu_0 \pm L\sqrt{\frac{\sigma_0^2}{n}}.$$

#### Osservazione

› Il tempo al primo allarme è

$$RL = \inf \left\{ i = 1, 2, \dots : \frac{|x_i - \mu_0|}{\sqrt{\sigma_0^2/n}} > L \right\}$$

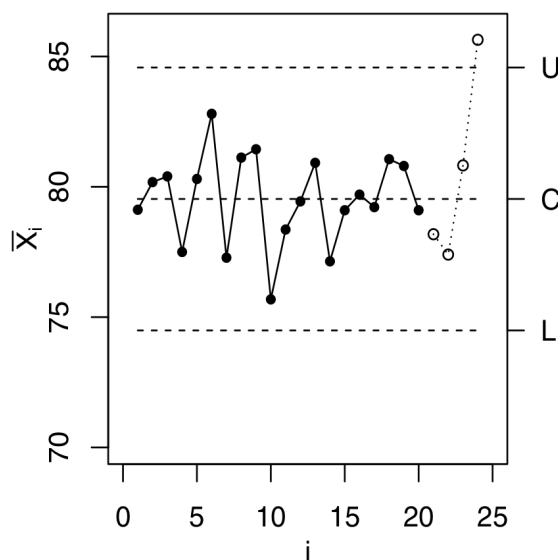


Figura 18: Carta Shewhart con processo che viene chiamato fuori controllo a  $t = 25$ .



### 12.1.1 Disegno della Shewhart

Per il disegno della carta Shewhart, si effettua la seguente minimizzazione vincolata

$$\begin{aligned} \min_L \mathbb{E}_\tau [\text{RL}] \\ \text{s.t. } \mathbb{E}_\infty [\text{RL}] = B. \end{aligned}$$

Per il processo in controllo, si ha che  $\text{RL} \sim \text{Geom}(\alpha)$ , in quanto

$$P(\text{RL} = r | \tau = \infty) = \alpha(1 - \alpha)^{r-1},$$

e dunque si fissa

$$\mathbb{E}_\infty [\text{RL}] = \frac{1}{\alpha} = B.$$

Fuori controllo, RL è ancora una geometrica, questa volta con probabilità  $1 - \beta$

$$P(\text{RL} = r | \tau = 1) = (1 - \beta)\beta^{r-1}, \quad r \geq 1.$$

$L$  si trova fissando  $\alpha$ , mentre la potenza si calcola

$$\text{ARL}_1 = \mathbb{E}_1 [\text{RL} | \tau = 1] = \frac{1}{1 - \beta}.$$

$L$  in modo da soddisfare il vincolo su  $\alpha$ , ovvero

$$P(w \notin (\text{LCL}, \text{UCL})) = \alpha \implies L = z_{1 - \frac{1}{2B}}$$

Supponendo di avere uno shift fuori controllo

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

la probabilità di errore di tipo II per  $\alpha$  fissato è

$$\begin{aligned} \beta(\delta) &= P(\mu_0 - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} | H_1) \\ &= P(-\delta\sqrt{n} - z_{1-\alpha/2} \leq \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} \leq -\delta\sqrt{n} + z_{1-\alpha/2} | H_1) \\ &= \Phi(-\delta\sqrt{n} + z_{1-\alpha/2}) - \Phi(-\delta\sqrt{n} - z_{1-\alpha/2}). \end{aligned}$$

Da qui, si può valutare la curva di average run length fuori controllo in funzione di  $\delta$ , per diversi valori di  $n$ :

$$\text{ARL}_1(\delta) = \frac{1}{1 - \beta(\delta)}.$$

#### Osservazioni

- › ARL decresce all'aumentare di  $\delta$ .
- › Decrescita dipende dalla dimensione campionaria  $n$ .

## 12.2 Carta EWMA

Osservando la formula, i pesi  $b_j$  sono una sequenza geometrica decrescente con rapidità di caduta esponenziale determinata da  $\lambda$ :

$$\sum_{j=0}^{t-1} \lambda(1-\lambda)^j = \lambda \frac{1 - (1-\lambda)^t}{1 - (1-\lambda)} = 1 - (1-\lambda)^t,$$

si ha una decrescita esponenziale dei pesi.

La carta EWMA sorveglia la media, in quanto se  $g(\mathbf{x}_t) = \mathbf{x}_t$  e  $\omega_0 = \mathbb{E}_{\text{IC}} [\mathbf{x}_t]$ , si ha che

$$\begin{aligned} \mathbb{E}_{\text{IC}} [w_t] &= \sum_{j=0}^{t-1} \lambda(1-\lambda)^j \mathbb{E}_{\text{IC}} [\mathbf{x}_{t-j}] + (1-\lambda)^t \mathbb{E}_{\text{IC}} [\mathbf{x}_t] \\ &= (1 - (1-\lambda)^t) \mathbb{E}_{\text{IC}} [\mathbf{x}_t] + (1-\lambda)^t \mathbb{E}_{\text{IC}} [\mathbf{x}_t] \\ &= \mathbb{E}_{\text{IC}} [\mathbf{x}_t]. \end{aligned}$$

Per la varianza, si ha un andamento che raggiunge un asintoto all'aumentare di  $t$ :

$$\begin{aligned} \mathbb{V}_{\text{IC}} [w_t] &= \sum_{j=0}^{t-1} \lambda^2 (1-\lambda)^{2j} \mathbb{V} [X_{t-j}] + (1-\lambda)^{2t} \mathbb{V} [w_0] \\ &= \sigma_0^2 \left( \frac{\lambda}{2-\lambda} \right) (1 - (1-\lambda)^{2t}) \\ &\xrightarrow{t \rightarrow \infty} \sigma_0^2 \left( \frac{\lambda}{2-\lambda} \right). \end{aligned} \tag{2}$$

### Limiti di controllo per una carta EWMA sulla media

I limiti di controllo sono

$$\mu_0 \pm L \sqrt{\sigma_0^2 \left( \frac{\lambda}{2-\lambda} \right) (1 - (1-\lambda)^{2t})}$$

### Osservazione

L'allarme viene suonato al tempo

$$\text{RL} = \inf \left\{ i = 1, 2, \dots : |w_t - \mu_0| / \sqrt{\mathbb{V} [w_t]} > L \right\}.$$

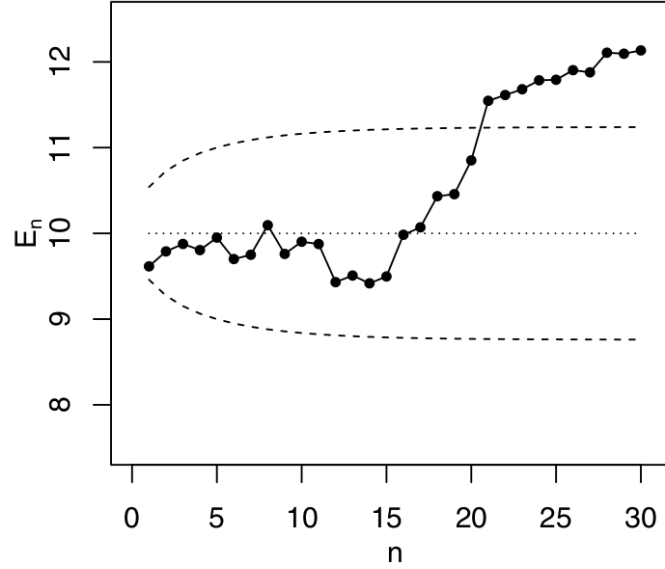


Figura 19: Esempio di carta EWMA con  $\lambda = 0.1$  per un processo di media  $\mu_0 = 10$  e varianza  $\sigma_0^2 = 1$ .

### Osservazioni

- › Con più sottogruppi razionali, si ha

$$w_t = \lambda \bar{x}_t + (1 - \lambda)w_{t-1},$$

con le immediate modifiche

$$\mathbb{E}_{IC} [w_t] = \mu_0,$$

$$\mathbb{V}_{IC} [w_t] = \frac{\sigma_0^2}{N} \left( \frac{\lambda}{2 - \lambda} \right) (1 - (1 - \lambda)^{2t}).$$

- › Siccome la convergenza è di ordine  $(1 - \lambda)^{2t}$ , c'è chi usa direttamente il limite asintotico (2). In generale è meglio usare quelli esatti.

In generale, la statistica di controllo fluttua attorno a  $\mu_0$  se il processo è in controllo, altrimenti si ha una deriva crescente o decrescente. Se c'è un cambiamento nel parametro, *non torna più* attorno a  $\mu_0$  perché vengono ponderate variabili con media  $\neq \mu_0$ .

In Figura 19 si vede una deriva crescente a partire dall'osservazione 22, per cui anche qui si può avere una stima grezza di  $\tau$ .

#### 12.2.1 Estensioni dell'EWMA

Si possono scegliere i parametri  $\lambda, L$  dell'EWMA in modo che siano ottimali a seconda della dimensione  $\delta$  della variazione di parametro. Nella realtà non si conosce  $\delta$ , per cui è importante che la carta sia il più possibile robusta rispetto a deviazioni di ogni dimensione.

### Carte multiple

Non conoscendo il parametro, la strada delle *carte multiple* usa tante carte EWMA con  $\lambda$  diversi che corrono contemporaneamente, e si lancia l'allarme quando la prima delle carte suona.

### Carte combinate

Alternativamente, una *carta combinata* EWMA, con  $\lambda$  piccolo per deviazioni persistenti, unita a una Shewhart per cambiamenti repentini, e si lancia quando la prima delle due suona.

### EWMA adattiva

Una soluzione molto buona è l'uso dell'*adaptive* EWMA introdotta in Capizzi e Masarotto (2003), dove si impone  $\lambda_t$  variabile in base a un errore di previsione.

L'EWMA adattiva risolve anche il problema di *inerzia*, cioè la lentezza dell'EWMA standard ad adattarsi a cambiamenti repentini in direzione opposta.

## 12.3 Carta CUSUM

Carta con memoria uniforme, data dalle somme cumulate delle deviazioni dal valore target  $\mu_0$ ,

$$w_t = w_{t-1} + (x_t - \mu_0) = \sum_{j=1}^t (x_j - \mu_0).$$

È immediato verificare che

$$\begin{cases} \mathbb{E}_{\text{IC}} [w_t] = 0, \\ \mathbb{E}_{\text{FC}} [w_\tau] = \underbrace{(\mu_1 - \mu_0)}_{=\delta\sigma} \cdot (t - \tau + 1), \end{cases}$$

$$\mathbb{V} [w_t] = t\sigma_0^2.$$

Anche in questo caso si osserva un random walk con drift nel caso in cui il processo vada fuori controllo, in quanto

$$\text{IC} : w_t \sim \mathcal{N}(0, t\sigma_0^2)$$

$$\text{FC} : w_t \sim \mathcal{N}(\delta \cdot (t - \tau + 1), t\sigma_0^2)$$

Usando l'ultimo momento prima del drift come stima di  $\hat{\tau}$ , si può stimare lo shift con

$$\delta = \frac{\mathbb{E} [w_t] - \mathbb{E} [w_\tau]}{t - \tau} \implies \hat{\delta} = \frac{w_{\text{RL}} - w_{\hat{\tau}}}{t - \hat{\tau}}.$$

**Problema** Calcolare i limiti di controllo non è immediato.

**Soluzione** Carta CUSUM tabulare, i cui limiti sono tabulati in funzione di  $\delta$ .

Si può utilizzare il *reference value*  $k = \delta/2$  per costruire una statistica CUSUM tabulare. Introdotta da Page (1954), la statistica di controllo per shift positivi è

$$P_n^+ = P_{n-1}^+ + (X_n - \mu_0) - k,$$

dove la carta suona un allarme se

$$P_n^+ - \min_{0 \leq i < n} P_i^+ > L.$$

Si può verificare che la carta è equivalente a definire la seguente carta di controllo per shift positivi:

$$S_0 = 0,$$

$$S_t^+ = \max \{0, S_{t-1}^+ + (x_t - \mu_0) - k\},$$

Con questa scrittura, se la deviazione  $(x_t - \mu_0)$  è  $> k$ , allora si cumula alla statistica, altrimenti non c'è evidenza di uno shift fuori controllo e si resetta.

Analogamente, si può disegnare la carta per shift negativi

$$S_0^- = 0,$$

$$S_t^- = \min \{0, S_{t-1}^- + (x_t - \mu_0) - k\}.$$

Mettendo insieme le due statistiche di controllo, si ottiene:

#### **Carta CUSUM tabulare bidirezionale**

La statistica di controllo  $S_t$  è data, per  $S_0^+ = S_0^- = 0$ , da

$$S_t^+ = \max \{0, S_{t-1}^+ + (x_t - \mu_0) - k\},$$

$$S_t^- = \min \{0, S_{t-1}^- + (x_t - \mu_0) - k\},$$

con limiti di controllo

$$S_t^+ > L, \quad S_t^- < -L.$$

#### **Osservazioni**

- › L'allarme viene suonato al tempo

$$RL = \inf \{i = 1, 2, \dots : S_i^+ > L, S_i^- < -L\}.$$

- › Il reference value  $k$  è l'analogo per la CUSUM del  $\lambda$  della EWMA.
- › Come stima grezza di  $\tau$  si può prendere l'ultima volta che si è visitato lo 0 prima di uscire dal controllo, per cui dalla Figura 20 si avrebbe  $\hat{\tau} = 9$ .

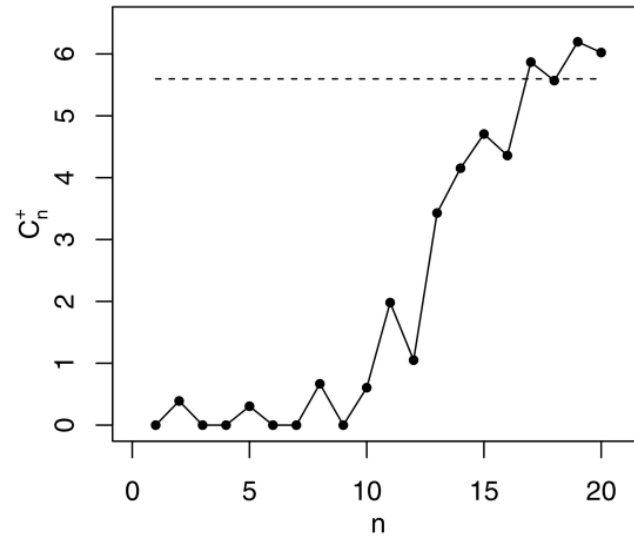


Figura 20: Carta CUSUM tabulare.

## LEZIONE 13: DISEGNO DI EWMA E CUSUM

2020-10-30

Per disegnare le carte EWMA e CUSUM bisogna trovare i parametri di interesse sulla base di  $\delta$  e  $\tau$  prefissati. Indicando con  $\vartheta$  i parametri della carta, è necessario risolvere

$$\begin{aligned} \min_{\vartheta_1, \vartheta_2} \mathbb{E}_{\tau, \delta} [\text{RL}(\vartheta) - \tau + 1 | \text{RL} > \tau] \\ \text{s.t. } \mathbb{E}_{\infty} [\text{RL}(\vartheta)] = B, \end{aligned} \quad (3)$$

oppure l'analogo che tutela rispetto a processi che rimangono per molto tempo in controllo

$$\begin{aligned} \min_{\vartheta_1, \vartheta_2} \lim_{\tau \rightarrow \infty} \mathbb{E}_{\tau, \delta} [\text{RL}(\vartheta) - \tau + 1 | \text{RL} > \tau] \\ \text{s.t. } \mathbb{E}_{\infty} [\text{RL}(\vartheta)] = B, \end{aligned} \quad (4)$$

### Osservazioni

- › L'ottimo per la CUSUM è  $k = \delta/2$ , per cui si massimizza solo in  $L$  il vincolo.
- › (3) e (4) sono due equazioni integrali, in quanto si calcolano dei valori attesi. Generalmente si risolvono con metodi numerici.
- ›  $B$  è l'average run length prima del falso allarme che l'operatore è disposto a tollerare.
- › Soluzioni si trovano con le funzioni `xewma.crit/xcusum.crit`, `xewma.arl/xcusum.arl`, `xewma.ad/xcusum.ad` della libreria `spc`

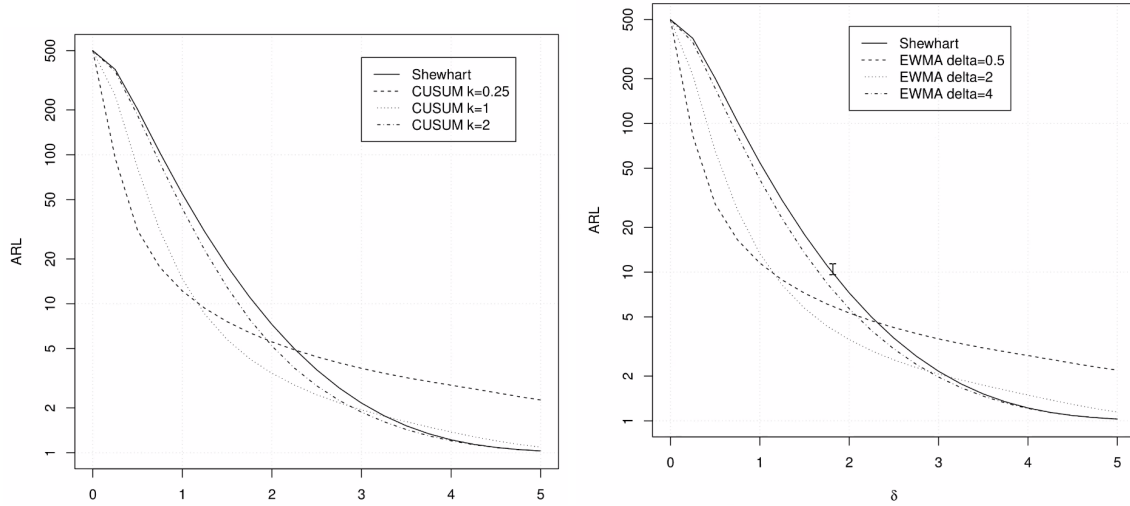


Figura 21: *Sinistra*: Confronto tra Shewhart e CUSUM con diversi valori di  $k = \frac{\delta}{2}$ . *Destra*: confronto tra Shewhart e CUSUM con diversi valori di  $\delta$ .

### Conclusioni

1. Shewhart offre protezione per cambiamenti grandi e isolati di parametro, ma CUSUM/EWMA sono comunque migliori della Shewhart: si possono scegliere  $k$  tali che si comportino quasi sempre meglio.

2. CUSUM/EWMA per un certo  $\delta$  hanno performance povere se cambiamenti sono lontani da  $\delta$ .
3. Tra CUSUM ed EWMA non c'è particolarmente differenza se ottimizzo per  $\delta$  uguale.

*Domanda: cosa se ne fa uno della Shewhart in Fase II?*

Nulla, perché andrebbe usata solo in Fase I :-).



## LEZIONE 14: CAMBIAMENTO $\delta$ INCOGNITO

2020-11-04

Quando il cambiamento è tale che sia  $\tau$  che  $\delta$  sono incogniti, è necessario usare tecniche più robuste di controllo di processo: Si colma la mancanza di conoscenza di  $\delta$  usando ad esempio:

1. *Carte combinate*, ottimizzate su diversi  $\delta$ . Ad esempio, è molto popolare la combinazione

$$\underbrace{\text{EWMA/CUSUM}}_{\text{cambiamenti piccoli}} + \underbrace{\text{Shewhart}}_{\text{cambiamenti grandi}}$$

2. Versioni *adattive* delle carte, AEWMA/ACUSUM, in cui i dati aggiornano la stima di  $\lambda_t$  e  $k_t$ . La migliore al momento è la AEWMA.
3. *Self-starting*: né Fase I né Fase II, congiuntamente si aggiorna la stima dei parametri e si controlla il processo.
4. *Generalized Likelihood Ratio*: parametri incogniti prima e dopo l'istante di cambiamento  $\tau$ .

### 14.1 Carte combinate

Si considerino  $H$  carte CUSUM, ognuna con un reference value  $k_1, \dots, k_H$  che siano ragionevolmente spaziate (di solito  $H = 3$ ). Si segnala l'allarme la prima volta che una delle carte esce dai propri limiti

$$S_i^+(k) > L \quad \text{o} \quad S_i^-(k) < -L, \quad i = 1, \dots, H.$$

La complessità sta nell'ottimizzazione di (4), in quanto i vincoli diventano

$$\left\{ \begin{array}{l} \text{Carta combinata: } \mathbb{E}_{\text{IC}} [\min(\text{RL}_1, \dots, \text{RL}_H)] = B \\ \text{No preferenza: } \mathbb{E}_{\text{IC}} [\text{RL}_1] = \mathbb{E}_{\text{IC}} [\text{RL}_2] = \dots = \mathbb{E}_{\text{IC}} [\text{RL}_H] \end{array} \right.$$

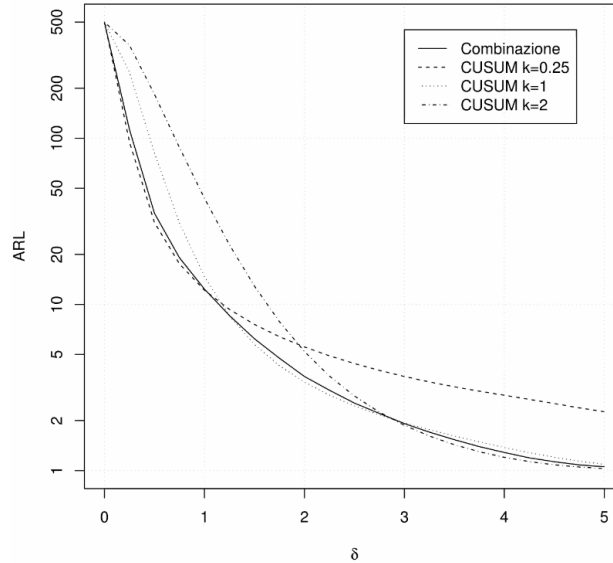


Figura 22: ARL per diversi  $\delta$ , le carte combinate offrono una protezione comparabile con la migliore delle tre carte.

## 14.2 EWMA adattiva

*Riferimenti* Capizzi e Masarotto (2003)

La carta EWMA standard si può scrivere in termini di un errore di previsione al tempo  $t$

$$\begin{aligned} S_t &= (1 - \lambda)S_{t-1} + \lambda x_t \\ &= S_{t-1} + \lambda \cdot \underbrace{(x_t - S_{t-1})}_{e_t = \text{errore di prev.}} \end{aligned}$$

L'idea è di utilizzare una carta con  $\lambda = \lambda_t$ , che modifichi il parametro sulla base dell'ampiezza dell'errore di previsione. Infatti,

- ›  $\lambda$  piccolo  $\implies$  cambiamenti piccoli.
- ›  $\lambda$  grande  $\implies$  cambiamenti grandi.

Si utilizza una scrittura del tipo

$$S_t = S_{t-1} + \varphi(x_t - S_{t-1}).$$

dove  $\varphi(e) = \lambda \cdot e$  per la EWMA. Per generalizzarla con  $\lambda_t$  variabile, si sceglie  $\varphi(e)$  in modo che

1.  $\varphi(\cdot)$  crescente e  $\varphi(e) = \text{sgn}(e)\varphi(|e|)$ , ovvero aggiustamento grande per errori grandi e in direzione dell'errore.
2. Data una soglia  $k$ , per  $|e| < k$ ,  $\varphi'(e) \approx \lambda$ , cioè si comporta come una EWMA per errori piccoli.
3. Per errori grandi,  $|e| > k$ ,  $\varphi'(e)$  cresce e si avvicina a 1 in modo che assomigli a una Shewhart.

Una delle funzioni che soddisfa questi requisiti è la *funzione di Huber*, definita come

$$\varphi(e) = \begin{cases} e + (1 - \lambda)k & \text{se } e < -k \\ \lambda e & \text{se } -k \leq e \leq k \\ e - (1 - \lambda)k & \text{se } e > k \end{cases}$$

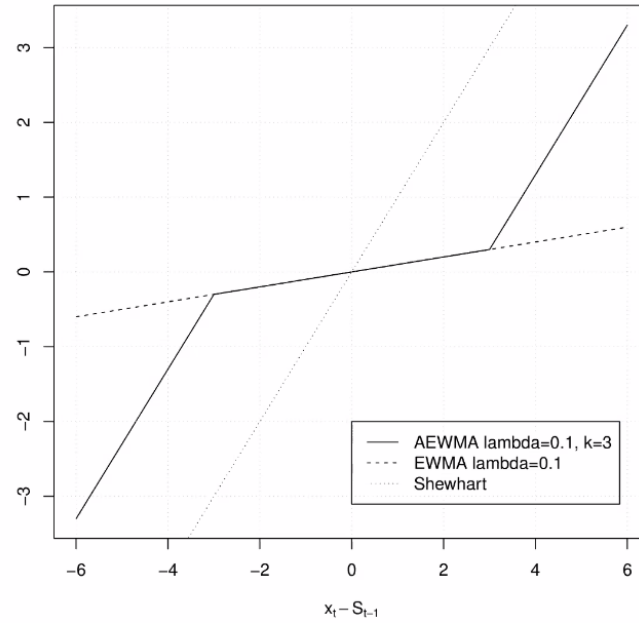


Figura 23: Funzione di Huber, una versione intermedia tra i due andamenti. È necessario introdurre un parametro aggiuntivo  $k$  per scegliere l'ampiezza del cambiamento

### EWMA adattiva

Si costruisce la statistica di controllo al tempo  $i$ -esimo come

$$\begin{aligned} S_i &= S_{i-1} + \lambda_i(x_i - S_{i-1}) \\ &= S_{i-1} + w_i(x_i - S_{i-1}) \end{aligned}$$

dove

$$w_i = \begin{cases} 0 & \text{se } x_i = S_{i-1} \\ \frac{\varphi(|x_i - S_{i-1}|)}{|x_i - S_{i-1}|} & \text{se } x_i \neq S_{i-1} \end{cases}$$

La statistica diventa allora

$$S_i = w_i x_i + (1 - w_i) S_{i-1}.$$

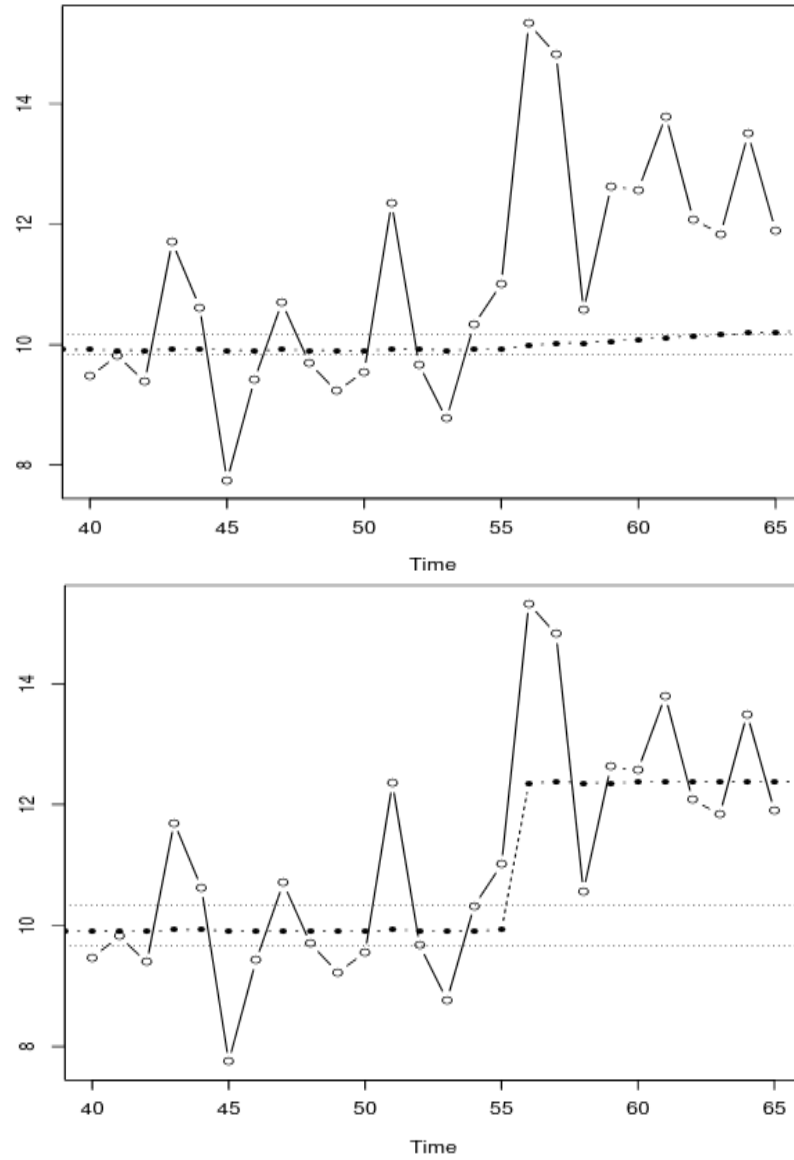


Figura 24: EWMA standard (*sopra*) e AEWMA (*sotto*) per lo stesso processo. Si nota il fenomeno di drift per la EWMA standard, mentre invece la AEWMA risponde immediatamente.

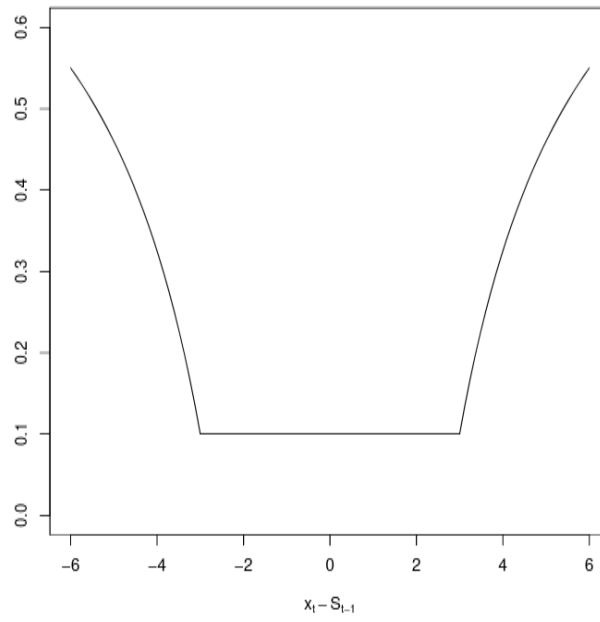


Figura 25: Andamento dei pesi usati dalla AEWMA al variare dell'errore  $e$ .

LEZIONE 15: CAMBIAMENTO  $\delta$  INCOGNITO (II)

2020-11-05

## 15.1 Carte self-starting

Si usano in tutte le situazioni nelle quali si devono fare cicli brevi di produzioni (*short runs*) e non si può effettuare la Fase I. Ad ogni tempo  $t$ , la carta aggiorna i parametri  $\vartheta_t$  in controllo e verifica l'instabilità nei parametri.

## Assunzioni

1. In controllo,  $(X_t)_{t \in \mathbb{N}} \stackrel{i.i.d}{\sim} \mathcal{N}(\mu_{IC}, \sigma_{IC}^2)$
2. Si ha un cambiamento *persistente* dei parametri (altrimenti va tutto in vacca)

$$X_i \sim \mathcal{N}(\mu_{IC}, \sigma_{IC}^2) \quad \text{per } i = 1, 2, \dots, \tau$$

$$X_i \sim \mathcal{N}(\mu_{IC} + \delta\sigma_{IC}, \sigma_{IC}^2) \quad \text{per } i > \tau$$

Man mano che le osservazioni risultano disponibili, se  $a_t = x_t - \bar{x}_{t-1}$ , allora l'aggiornamento degli stimatori dei parametri è

$$\bar{x}_t = \bar{x}_{t-1} + \frac{1}{t} a_t,$$

$$s_t^2 = s_{t-1}^2 + \frac{1}{t} \left( a_t^2 - \frac{t}{t-1} s_{t-1}^2 \right),$$

a partire dai valori iniziali stimati  $\bar{x}_2$  e  $s_2^2$ .

**Carta di controllo self-starting gaussiana**

In ogni istante di tempo si fa un test  $t$ -student per confrontare i campioni

$$\{x_t\} \quad \text{con} \quad \{x_{t-1}, \dots, x_1\},$$

tramite la statistica test assumendo uguale varianza (pooled)

$$T_t = \frac{(x_t - \bar{x}_{t-1})}{\sqrt{s_{t-1}^2 \left( \frac{1}{t-1} + 1 \right)}} \sim t_{t-2}, \quad t \geq 3.$$

## Osservazioni

› Si ha che

$$\mathbb{E}_{IC} [T_t] = 0$$

$$\mathbb{E}_{FC} [T_t] = \frac{\delta(\tau - 1)}{s_{t-1} \sqrt{t(t-1)}}.$$

In particolare, fuori controllo  $\mathbb{E}_{FC} [T_t] \xrightarrow{t \rightarrow \infty} 0$ , il che è non buono (Figura 26).

› Le statistiche  $T_t$  sono indipendenti, ma non identicamente distribuite al variare di  $t$ . Si utilizza allora la trasformazione  $Q$  di Quesenberry, data da

$$Q_t = \Phi^{-1}(G_{t-2}(T_t)) \stackrel{i.i.d}{\sim} \mathcal{N}(0, 1) \quad (5)$$

dove  $G_\nu$  è la cdf di una  $t_\nu$ . In questo modo, i limiti di controllo non dipendono da  $t$  e si può sorvegliare la media di  $Q_t \sim \mathcal{N}(0, 1)$ .

- › Nel caso in cui  $X_t$  abbia un'altra distribuzione (e.g. Poisson come si vedrà in Shen et al. (2016)), la media in controllo non è nulla e si ha una distorsione nell' $ARL_1$ .

### Problema delle self-starting

Le self-starting sono molto influenzate da  $\delta$  e  $\tau$ , in particolare vanno bene se si individua il cambiamento in una finestra vicina al punto di vero cambiamento. Se invece la carta non chiama subito e  $\delta$  è piccolo, le osservazioni fuori controllo vengono incluse nell'aggiornamento dei parametri.

*Le self-starting si comportano bene, a meno che ci sia un cambiamento  $\delta$  piccolo in prossimità ( $\tau$  piccolo) dell'origine.*

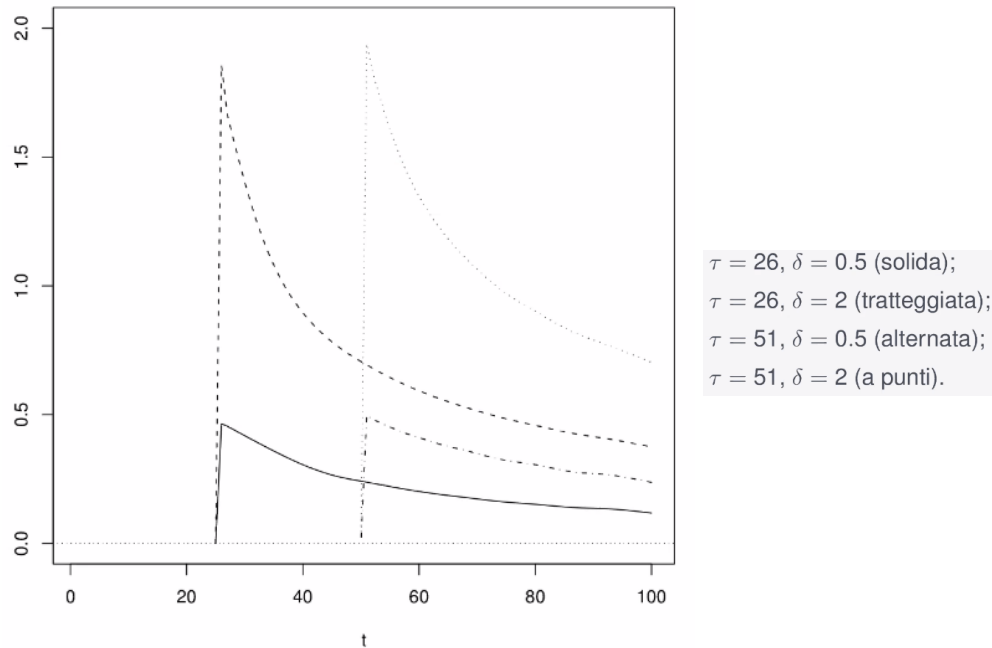


Figura 26: Valore atteso di  $Q_t$  per vari valori di ampiezza  $\delta$  e istante di cambiamento  $\tau$ .

### Osservazione

La ricerca corrente è sull'uso di tecniche di machine learning per provare a risolvere questi problemi.

Una volta calcolata la statistica  $Q_t$ , si può utilizzare per costruire la versione self-starting delle carte di controllo classiche:

#### Self-starting Shewhart

$$RL = \inf \{t \geq 3 : Q_t < LCL \text{ o } Q_t > UCL\}.$$

**Self-starting CUSUM**

$$C_t^L = \min \{0, C_{t-1}^L + Q_t + k\}$$

$$C_t^U = \max \{0, C_{t-1}^U + Q_t - k\}$$

$$\text{RL} = \inf \{t \geq 3 : C_t^L < \text{LCL} \text{ o } C_t^U > \text{UCL}\}.$$

**Self-starting EWMA**

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

$$\text{RL} = \inf \left\{ t \geq 3 : |Z_t| > h\sqrt{\lambda/(2-\lambda)} \right\}.$$

**15.2 Generalized Likelihood Ratio (GLR)**

Le carte basate sul GLR effettuano un test di ipotesi sequenziale, basato sul rapporto di verosimiglianza, assumendo  $\tau$  ignoto.

La verosimiglianza per un processo di cui si osserva  $x_1, x_2, \dots, x_i$  che va fuori controllo al tempo  $\tau$  incognito è

$$L(\mu_0, \sigma_0) = \prod_{j=1}^{\tau-1} f(x_j; \mu_0, \sigma_0) \prod_{j=\tau}^i f(x_j; \mu_0 + \delta\sigma_0, \sigma_0)$$

Per ogni  $\tau = 1, \dots, i$  bisogna saggiare il test

$$H_0 : \delta = 0 \quad \text{vs.} \quad H_1 : \delta \neq 0$$

e la statistica test, semplificando i fattori comuni per  $j = 1, \dots, \tau - 1$ , è data da

$$\begin{aligned} S_i &= \max_{\tau=1, \dots, i} \text{LRT}_i(\tau) \\ &= \max_{\tau=1, \dots, i} \frac{\sup_{\delta} \prod_{j=1}^{\tau-1} f(x_j; \mu_0, \sigma_0) \prod_{j=\tau}^i f(x_j; \mu_0 + \delta\sigma_0, \sigma_0)}{\prod_{j=1}^{\tau-1} f(x_j; \mu_0, \sigma_0) \prod_{j=\tau}^i f(x_j; \mu_0, \sigma_0)} \\ &= \max_{\tau=1, \dots, i} \frac{\sup_{\delta} \prod_{j=\tau}^i f(x_j; \mu_0 + \delta\sigma_0, \sigma_0)}{\prod_{j=\tau}^i f(x_j; \mu_0, \sigma_0)}. \end{aligned}$$



**Schema di controllo GLR**

In uno schema GLR si segnala un allarme quando

$$\max_{\tau=1,\dots,i} \left| \sum_{j=\tau}^i \frac{x_j - \mu_0}{\sigma_0 \sqrt{i - \tau + 1}} \right| > L,$$

per una soglia  $L$  che viene scelta in modo che

$$\mathbb{E}_{IC} [\text{RL}] = B.$$

**Osservazione**

Il costo computazionale della doppia massimizzazione in  $\delta, \tau$  è enorme, per cui di solito si sceglie di utilizzare una finestra mobile di ampiezza  $M$  e considerare

$$S_i = \max_{\tau=\max\{1, i-M\}, \dots, i} R_i(\tau).$$

Per  $X_i \stackrel{i.i.d}{\sim} \mathcal{N}(0, 1)$ , Jr e Lou (2010) consigliano  $M = 400$ .

**15.3 Legame tra CUSUM e verosimiglianza**

La statistica CUSUM deriva da un test di log-verosimiglianza, in quanto

$$\log \text{LRT} = \sum_{i=1}^n \log \frac{f_1(x_i, \mu_1)}{f_0(x_i, \mu_0)} = \sum_{i=1}^n \psi_i,$$

dove se si definisce  $K = \frac{\delta}{2} \sigma_0$ , si ha

$$\psi_i = \frac{\mu_1 - \mu_0}{\sigma_0^2} \cdot \left( x_i - \mu_0 - \underbrace{\frac{\mu_1 - \mu_0}{2}}_{\frac{\delta}{2} \sigma_0} \right) = \frac{\mu_1 - \mu_0}{\sigma_0^2} ((x_i - \mu_0) - K),$$

con  $k = \delta \sigma_0^2 / 2$ . Si può allora ottenere la carta CUSUM scrivendo  $S_n$  come

$$\begin{aligned} S_n &= \frac{\sigma_0^2}{\mu_1 - \mu_0} \log \text{LRT} \\ &= \sum_{i=1}^{n-1} (x_i - \mu_0 - k) + (x_n - \mu_0 - k) \\ &= S_{n-1} + ((x_i - \mu_0) - k) \end{aligned}$$

Sfruttando il legame tra CUSUM e test di log-verosimiglianza, si può allora scrivere una versione della CUSUM più generale per distribuzioni diverse dalla gaussiana, ad esempio quando i processi IC ed OC sono

$$X_i \stackrel{IC}{\sim} \text{Ber}(\pi_0) \quad i = 1, \dots, \tau - 1$$

$$X_i \stackrel{FC}{\sim} \text{Ber}(\pi_1) \quad i \geq \tau$$

**Domanda** È obbligatorio avere un punteggio di log-verosimiglianza che sia statico rispetto ad  $i$ , ovvero  $\pi_0$  e  $\pi_1$  fissati? In realtà no, sarebbe meglio avere parametri che variano in funzione di covariate, ad esempio come si vedrà più avanti in Steiner et al. (2000)

$$X_i \sim \text{Ber}(g(\mathbf{z}_i^\top \vartheta)),$$

$$X_i \sim \text{Pois}(h(\mathbf{z}_i^\top \vartheta)).$$

Processo di Poisson con media in funzione di un parametro  $\vartheta$ , allora diventa un **processo non omogeneo**.

## LEZIONE 16: DISEGNO DI CARTE COMPLESSE

2020-11-06

*Riferimenti* Capizzi e Masarotto (2003)

Qiu, 2014, §5.4.2

### 16.1 Disegno dell'AEWMA

Nell'AEWMA si ha vuole una carta che funzioni bene sia per piccoli sia per grandi cambiamenti. Il disegno è più complesso e procede in questo modo:

1. Si sceglie un  $ARL(0, \vartheta) = B$  in controllo e si scelgono due deviazioni, una “piccola”  $\mu_1$  e una “grande”  $\mu_2$ .
2. Si trovano i parametri  $\vartheta^*$  che danno una buona ARL rispetto a  $\mu_2$ ,

$$\vartheta^* = \underset{\vartheta}{\operatorname{argmin}} ARL(\mu_2, \vartheta)$$

$$\text{s.t. } ARL(0, \vartheta) = B.$$

3. Si sceglie una costante piccola  $\alpha \approx 0.05$  e si trovano i parametri  $\hat{\vartheta}$  che danno una buona ARL rispetto a  $\mu_1$ , ma che si scostino di poco rispetto alla ARL di prima,

$$\hat{\vartheta} = \underset{\vartheta}{\operatorname{argmin}} ARL(\mu_1, \vartheta)$$

$$\text{s.t. } ARL(0, \vartheta) = B,$$

$$ARL(\mu_2, \hat{\vartheta}) \leq (1 + \alpha)ARL(\mu_2, \vartheta^*) \quad (\text{“vicina” a } \vartheta^* \text{ di prima})$$

Questo problema si può risolvere usando ad esempio le catene di Markov, scrivendo tutti i possibili stati della matrice di transizione e calcolandola.

**Domanda esame** Quali sono i metodi studiati per il calcolo dei limiti di controllo?

1. Analitici
2. Numerici
  - Integrazione numerica (libreria `spc`)
  - Catene di Markov
  - Approssimazioni stocastiche (libreria `saControlLimits`)
  - Bootstrap e metodi di permutazione

Distinzione tra i metodi che trovano i parametri della carta (*disegno ottimale*) e quelli che trovano i limiti di controllo (*calibrazione*) a partire dai parametri. La libreria `saControlLimits` permette di calcolare i limiti di controllo sulla base di  $\lambda$  e  $k$ .

## 16.2 Approssimazioni stocastiche

*Riferimenti* Capizzi e Masarotto (2016)

Soluzioni di un'equazione stocastica non lineare tramite metodi iterativi. Bisogna trovare un criterio per stoppare questo metodo iterativo, che si chiama regola di arresto di Robbins e Monro. Ci si ferma la prima volta che si ottiene un margine di precisione prefissato.

Si assume di avere una statistica di controllo

$$z_t = z_t(x_t, x_{t-1}, \dots, x_1),$$

con la performance stabilita come il seguente tempo di arresto

$$\text{RL} = \inf \{t : Z_t > h\},$$

dove  $h$  è scelto per garantire l'ARL desiderata, oppure altre proprietà come FAP, quantili, ... Se gli schemi sono più complessi, ad esempio con carte multiple, si utilizzano metodi diversi.

In generale, si definisce  $h$  come il limite di controllo che fornisce

$$h \implies \mathbb{E}[\text{RL}] = \text{ARL}_0.$$

**Assunzioni** È necessario essere in grado di simulare la run length standardizzata

$$S = \frac{\text{RL} - \text{ARL}_0}{\text{ARL}_0} \sim \mathcal{P}_h,$$

per un valore fissato di  $\text{ARL}_0 = B$ . Si vuole effettuare una serie di simulazioni Monte Carlo per trovare in modo approssimativo la soluzione dell'equazione

$$g(h) = \mathbb{E}_h[S] = 0,$$

che consente di ricavare il limite di controllo  $h$  cercato.

### 16.2.1 Algoritmo di Robbins e Monro

1. Si effettuano simulazioni da

$$S_r = \frac{\text{RL}_r - \text{ARL}_0}{\text{ARL}_0}, \quad r = 0, 1, \dots$$

2. Si trova  $h^*$  radice di  $g(h)$  tramite l'iterazione ricorsiva

$$h_{r+1} = h_r - \frac{1}{r+1} A S_r, \quad r = 0, 1, \dots,$$

con  $A$  matrice di *guadagno* dello schema.

**Proprietà** Convergenza asintotica

$$h_r \xrightarrow{r \rightarrow \infty} h^*.$$

Una scelta efficiente di  $A$  è

$$A = \left( \frac{\partial g(h)}{\partial h^\top} \Big|_{h=h^*} \right)^{-1}.$$

### Problemi

1. Matrice jacobiana è incognita
2. Come si modifica il metodo affinché la convergenza sia garantita per  $r$  finito?

#### 16.2.2 Accelerazione del metodo RM

Metodo di due stadi con una inizializzazione

1. Primo stadio fissa un valore iniziale e la matrice di guadagno si muove in modo adattivo.

$$\tilde{h}_{r+1} = \tilde{h}_r - A_{\text{fix}} s_r, \quad r = 0, 1, \dots, N_{\text{fix}} - 1$$

con  $\tilde{h}_0$  scelto, ...

2. Il secondo stadio usa una media delle ultime  $N$  iterazioni per trovare la soluzione  $h$ , fermandosi quando l'accuratezza desiderata è trovata. Dato  $h_0 = \tilde{h}_{N_{\text{fix}}}$  e  $\bar{h}_0$ ,

$$h_{r+1} = h_r - \frac{1}{(r+q)^q} A s_r$$

$$\bar{h}_{r+1} = \bar{h}_r + \frac{1}{r+1} (h_r - \bar{h}_r)$$

e ci si ferma quando

$$|g(\bar{h}_r)| \leq \gamma$$

Quando  $r$  è grande, approssimativamente si ha

$$g(\bar{h}_r) \sim \mathcal{N}\left(0, \frac{1}{r} \mathbb{E}_{h^*} [S^2]\right),$$

dunque ci si ferma a

$$N_{\text{PR}} = \inf \left\{ N > N_{\text{min}} : N \geq \left( \frac{z}{\gamma} \right)^2 \frac{1}{N} \sum_{r=1}^N s_r^2 \right\}$$

e si prende  $\bar{h}_{N_{\text{PR}}}$  come stima finale di  $h$ .

Metodo complicato, ma dà delle buone soluzioni.

In particolare si possono applicare ai sistemi complessi, ad esempio nel caso delle carte multiple.

#### Esempio (Approssimazioni stocastiche per carte multiple)

Dati  $p$  schemi di controllo, ciascuno con la propria run length

$$\text{RL}_i = \inf \{t > 0 : z_{i,t} > h_i\}, \quad i = 1, \dots, p,$$

e si cerca la run length combinata

$$RL = \min\{RL_1, RL_2, \dots, RL_n\}.$$

Nel caso degli schemi multipli si risolve  $\mathbb{E}_{\mathbf{h}}[\mathbf{S}] = 0$  con lo score

$$\mathbf{s} = \left( \frac{\min\{RL_1, RL_2, \dots, RL_n\} - ARL_0}{ARL_0} + \frac{RL_i - \overline{RL}}{ARL_0} \right)^{\top}$$

**Osservazione**

Le approssimazioni stocastiche possono richiedere un po' di tempo per ottenere le soluzioni.

## LEZIONE 17: CAPIZZI E MASAROTTO (2017)

### «PHASE I DISTRIBUTION-FREE ANALYSIS OF MULTIVARIATE DATA»

2020-11-12

Si utilizza un metodo ibrido di Fase I, una procedura ricorsiva che implementa *adaptive lasso* (H. Zou e Hastie (2005)) e *forward search*.

#### 17.1 Introduzione

Gli unici articoli che in letteratura al tempo implementavano metodi non parametrici di Fase I multivariati erano:

Bell et al. (2014)    Dati raggruppati con distribuzione ellittica (Appendice A.1)  
Shewhart sui ranghi ottenuti dalla *Mahalanobis depth* (Appendice A.3)

Cheng e Shiau (2014)    Dati raggruppati con distribuzione ellittica  
Shewhart sugli *spatial signs* (Appendice A.4)

#### Problematiche

- › **Solo per dati raggruppati**, non si possono applicare a osservazioni singole.
- › **Non sono realmente distribution-free**, perché assumono distribuzione ellittica.
- › **Usano depth e spatial signs**, ma si potrebbero utilizzare entrambe in modo combinato, per avere informazioni sia sulla dimensione sia sulla direzione del cambiamento.
- › **Si usano solo carte Shewhart**, per cui non sfruttano l'ordinamento temporale dei dati e non permettono di individuare cambiamenti change-point o altri tipi di shift.

I limiti di controllo si possono applicare a tutte le distribuzioni ellittiche, una volta che li calcolo assumendo  $MVN(\mu, \Sigma)$ .

#### Proposta

Si generalizza il metodo di Recursive Segmentation and Permutation introdotto in Capizzi e Masarotto (2013), che effettua le seguenti procedure:

1. Si controlla che il parametro di posizione sia stabile.
2. Se risulta non stabile, si applica una procedura post-segnale per identificare gli istanti temporali più plausibili di cambiamento.

Si assume **sparsità** dei cambiamenti, poiché se *tutto* si modifica significa che il processo è completamente andato in vacca (cit. Nik).

La procedura è composta da quattro stadi:

1. Preprocessing, con trasformazione dei dati in *multivariate rank signs* (Appendice A.4).
2. Si usa una procedura *stepwise forward* per calcolare  $K < m$  *plausibili* cambiamenti di shift con corrispondenti statistiche test,  $K$  fissato dall'operatore.

3. Si aggregano le  $K$  statistiche elementari e si calcola un  $p$ -value via permutazioni per verificare la stabilità globale.
4. Se il processo è fuori controllo (step 3), si utilizza un *adaptive LASSO* ottimizzato rispetto ad un appropriato criterio di informazione per selezionare un sottoinsieme di variabili coinvolte nel cambiamento.

## 17.2 Stage I: Preprocessing

*Lecture consigliate* Oja (2010) (Statistica non parametrica multivariata)

### Standardizzazione

I dati vengono standardizzati in modo non parametrico

$$\mathbf{z}_{i,j} = \mathbf{S}^{-1/2}(\mathbf{x}_{i,j} - \ell),$$

con  $\mathbf{S}$  e  $\ell$  stimatori robusti di location e scatter matrix:

$$\mathbf{S} = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^\top$$

$$\ell = \mathbf{S}^{1/2} \text{spatial median} \left\{ \mathbf{S}^{-1/2} \bar{\mathbf{x}}_1, \dots, \mathbf{S}^{-1/2} \bar{\mathbf{x}}_m \right\},$$

dove la *spatial median* è definita come il minimizzatore delle distanze

$$\min_{m \in \mathbb{R}^g} \frac{1}{N} \sum_{i=1}^N \|\mathbf{x}_i - m\|_1 \quad (6)$$

**Problemi** La mediana spaziale (6) non è soddisfacente quando ci si allontana dall'ipotesi di simmetria. Si utilizza una trasformazione “affine equivariante” (invariante per rotazioni) data da

$$\mathbf{Y} \rightsquigarrow \mathbf{Y} \mathbf{S}^{-1/2} \rightsquigarrow \hat{\ell} = \text{spatial median}(\mathbf{Y} \mathbf{S}^{-1/2})$$

$$\rightsquigarrow \ell = \mathbf{S}^{1/2} \hat{\ell} = \mathbf{S}^{1/2} \text{spatial median}(\mathbf{Y} \mathbf{S}^{1/2}).$$

Successivamente, si calcolano gli *spatial signs* per rimuovere l'effetto della dimensione

$$U(\mathbf{z}_{i,j}) = \|\mathbf{z}_{i,j}\|^{-1} \mathbf{z}_{i,j}.$$

Data la Mahalanobis depth  $\|\mathbf{z}_{i,j}\|$ , che dà informazione sulla distanza dall'origine, si calcolano i ranghi

$$\mathbf{r} = \text{rank} \{ \|\mathbf{z}_{1,1}\|, \dots, \|\mathbf{z}_{m,n}\| \}$$

e si calcolano i ranghi con segno multivariati, indicando con  $F_{\chi_g^2}(\cdot)$  la funzione di ripartizione di una  $\chi_g^2$ .

$$\mathbf{u}_{i,j} = \sqrt{F_{\chi_g^2}^{-1} \left( \frac{r_{ij}}{mn+1} \right)} \frac{\mathbf{z}_{i,j}}{\|\mathbf{z}_{i,j}\|}.$$

Questi ranghi hanno la stessa direzione di  $\mathbf{z}_{i,j}$ , ma sono riscalandi in modo da avere la dimensione attesa da un processo gaussiano in controllo.



### Osservazioni

Questa trasformazione ha una serie di proprietà interessanti:

1. Questa trasformazione permette di avere un comportamento migliore rispetto alle carte ottimizzate sulle distribuzioni.
2. I dati sono ruotati in modo da tenere conto delle correlazioni (SURE regression).
3. Migliora la performance per dati con code pesanti.
4. È la combinazione ottimale tra ranghi spaziali e depth values (dimostrato da altra gente).

### 17.3 Stage II: Numero di shift

Quando il cambiamento non è isolato, lo approssimo aggiungendo parametri che rimangono permanentemente. Indicando con  $\mu_0$  il livello di stabilità,  $\delta_i$  un cambiamento a step permanente e  $\gamma_i$  un cambiamento isolato, si utilizza:

$$\begin{aligned}
 \text{LOCATION} &= \text{PERSISTENTE} + \text{ISOLATO} \\
 \mu_1 &= \mu_0 + \gamma_1 \\
 \mu_2 &= \mu_0 + \delta_2 + \gamma_2 \\
 \mu_3 &= \mu_0 + \delta_2 + \delta_3 + \gamma_3 \\
 &\vdots \quad \quad \quad \vdots \\
 \mu_m &= \mu_0 + \sum_{i=2}^{m-1} \delta_i + \gamma_m
 \end{aligned} \tag{7}$$

### Osservazioni

- › Questo modello permette di catturare pattern anche più generali, cumulando step isolati e permanenti.
- › Ci sono  $m - 2$  parametri  $\delta \in \mathbb{R}^g$  e  $m$  vettori  $\gamma \in \mathbb{R}^g$ , dunque  $2g(m - 1)$  parametri totali.

Fissato  $K$  possibili cambiamenti nei dati, si deve testare l'ipotesi di stabilità

$$\begin{cases} H_0 : & \text{tutti i } \delta \text{ e } \gamma \text{ sono nulli} \\ H_1 : & \text{almeno uno dei } \delta \text{ o } \gamma \text{ non sono nulli} \end{cases}$$

È importante assumere **sparsità** degli shift, ovvero gli shift non avvengono in tutti gli istanti ( $\delta_i = \mathbf{0}$ ) e coinvolgono solo alcune delle variabili ( $\delta_{ij} = 0$ ).

Le  $K$  statistiche elementari corrispondono a un numero plausibile di  $\delta, \gamma$  che risultano  $\neq 0$ . Sono le **varianze spiegate** da modelli di regressione elementari, perché (7) è la scrittura di un modello lineare.

## LEZIONE 18: CAPIZZI E MASAROTTO (2017) (II)

2020-11-13

Il metodo adatta un modello con **forward search** a partire da  $\hat{\mathbf{u}}_i^{(0)} = \bar{\mathbf{u}} = (nm)^{-1} \sum_{i,j} \mathbf{u}_{ij}$ :

$$\hat{\mathbf{u}}_i^{(k)} = \hat{\beta}_0^{(k)} + \hat{\beta}_1^{(k)} \xi_i^{(1)} + \dots + \hat{\beta}_i^{(k)} \xi_i^{(k)}, \quad i = 1, \dots, m,$$

con parametri

$$\xi_i^{(k)} = \begin{cases} I(i = \tau^{(k)}) & \text{shift isolato} \\ I(i \geq \tau^{(k)}) & \text{shift step} \end{cases}$$

L'update avviene tramite

$$\hat{\mathbf{u}}_i^{(k)} = \hat{\mathbf{u}}_i^{(k-1)} + \beta_i^{(k)} \xi_i^{(k)},$$

scegliendo quello che minimizza la somma del quadrato dei residui al passo precedente.

**Perché forward search e non LASSO, ...?**

- › Forward search è ben conosciuta in generale.
- › Si usa come screening iniziale (Stage 2) e ha una buona performance di screening. In generale, si suggerisce  $K = \sqrt{m}$  e poi successivamente si farà pruning.
- › I test basati sulla forward search hanno una buona performance.

**18.1 Stage III: Aggregazione e  $p$ -value**

Le statistiche elementari sono

$$W_{\text{obs}} = \max_{k=1, \dots, K} \frac{T_k - \mathbb{E}_{\text{stable}}[T_k]}{\sqrt{\mathbb{V}_{\text{stable}}[T_k]}},$$

con  $p$ -value definito come

$$p = P_{\text{stable}}(W \geq W_{\text{obs}}).$$

Siccome non si conosce la distribuzione in controllo, i momenti e il  $p$ -value vengono calcolati via **permutazioni** dei ranghi con segno  $\mathbf{u}_i$ .

**18.2 Stage IV: Pruning post-segnale**

Se  $p < \alpha$  si lancia l'allarme e si cerca di identificare la presenza di shift con un modello **adaptive LASSO** (H. Zou e Hastie (2005)), condizionandosi agli shift identificati nella forward search:

$$\begin{aligned} \mathbf{u}_{i,j} &= \beta_0 + \beta_1 \xi_i^{(1)} + \dots + \beta_k \xi_i^{(K)} + \hat{e}_{i,j} \\ &= \mathbf{S}^{-1/2} \boldsymbol{\delta}_0 + \mathbf{S}^{-1/2} \boldsymbol{\delta}_1 \xi_i^{(1)} + \dots + \mathbf{S}^{-1/2} \boldsymbol{\delta}_K \xi_i^{(K)} + \hat{e} \end{aligned}$$

Si stima utilizzando

$$\min D(\boldsymbol{\delta}_0, \dots, \boldsymbol{\delta}_K) + \lambda \sum_{k=1}^K \sum_{h=1}^g \frac{\|\boldsymbol{\delta}_i\|_1}{\|\hat{\boldsymbol{\delta}}_i^{\text{ols}}\|_1},$$

scegliendo il valore di  $\lambda$  che produce il migliore valore di **Extended BIC**, che ha un parametro  $\gamma$  la cui scelta dà il trade-off tra protezione da shift ampi e shift medio-piccoli.

Nell'articolo si confronta la procedura con altre procedure di Fase I, con distribuzioni appartenenti alla famiglia delle distribuzioni ellittiche e non. Le performance vengono valutate con la potenza in base alla dimensione standardizzata dello shift.

Dalle simulazioni, si osserva che

1. Si comporta in modo migliore delle carte ottimizzate per le distribuzioni ellittiche, anche quando la distribuzione è ellittica.
2. Con che precisione si identifica lo step? Si studia se viene identificato con una finestra temporale ampia al massimo pari a 5.

Si utilizza la trasformazione normale per avere proprietà di ottimalità e rapidità di calcolo della selezione delle variabili.

## LEZIONE 19: STEINER ET AL. (2000)

### «MONITORING SURGICAL PERFORMANCE USING RISK-ADJUSTED CUMULATIVE SUM CHARTS»

2020-11-18

**Intro** Si tratta di un articolo che porta dei contributi di Fase I e Fase II.

**Obiettivo** Si vuole sorvegliare l'esito di un intervento e segnalare se vi è un cambiamento sistematico del tasso di insuccesso degli interventi chirurgici, in particolare tenendo conto della gravità del paziente:

1. Individuare rapidamente decrescite (miglioramenti) o aumenti (peggioramenti) dell'odds ratio di morte.
2. Valutare la performance chirurgica in funzione di esperienza del chirurgo e gravità del paziente.

Si considera una carta CUSUM tradizionale di tipo

$$S_t = \max \{0, S_{t-1} + W_t\}, \quad t = 1, 2, 3, \dots$$

Si ha un modello  $Y_t \sim \text{Ber}(p_t)$ , dove  $y_t$  è 1 se il paziente muore, per cui  $W_t$  è il peso basato sul rapporto di log-verosimiglianza per il paziente  $t$ -esimo

$$\begin{aligned} W_t &= \log \frac{f(y_t|p_1)}{f(y_t|p_0)} \\ &= \log \frac{p_1^{y_t}(1-p_1)^{1-y_t}}{p_0^{y_t}(1-p_0)^{1-y_t}} = \begin{cases} \log \left( \frac{1-p_1}{1-p_0} \right) & \text{se } y = 0 \\ \log \left( \frac{p_1}{p_0} \right) & \text{se } y = 1 \end{cases} \end{aligned} \quad (8)$$

Si verifica l'ipotesi  $H_0 : p = p_0$  contro  $H_1 : p_1 > p_0$  con il log-rapporto di verosimiglianza

$$y_t - \underbrace{\frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{1-p_0}{1-p_1} - \log \frac{p_0}{p_1}}}_{K \text{ reference value}}$$

dunque

$$S_t = \max \{0, S_{t-1} + W_t\} = \max \{0, S_{t-1} + y_t - K\},$$

con  $K$  che rende minima l'ARL<sub>1</sub> per  $p_1$  prefissato.

#### Idea e conclusione dell'articolo

Si combina una carta con un modello di regressione logistico per aggiustare il valore di  $W_t$  sulla base della stima del rischio **pre-operatorio** di ogni paziente. Se non si correggesse in base alla gravità del paziente, con una CUSUM standard si potrebbe concludere che chirurghi più giovani sono più bravi di quelli con esperienza, visto che tendenzialmente ricevono pazienti meno gravi.

Si dispone di predittori sull'individuo, per cui  $Y_t \sim \text{Ber}(g(\mathbf{x}_t))$ , che viene stimato su un campione di training. La misura da controllare è l'odds ratio, in particolare si ha come *odds di morte a priori*

$$\text{ODDS}_2 = \frac{p_t}{1-p_t} = \frac{g(\mathbf{x}_t)}{1-g(\mathbf{x}_t)},$$

mentre bisogna controllare gli odds di sopravvivenza in controllo e fuori controllo

$$R_0 = \frac{\text{ODDS}_1}{\text{ODDS}_2} = \frac{\frac{u_t}{1-u_t}}{\frac{p_t}{1-p_t}}$$

$$R_A = \frac{\text{ODDS}_3}{\text{ODDS}_2} = \frac{\frac{q_t}{1-q_t}}{\frac{p_t}{1-p_t}}$$

In generale, nell'articolo si assume  $R_0 = 1$  come accettabile per la performance chirurgica, mentre  $R_A > R_0$  indica un peggioramento della performance chirurgica. Sotto  $H_0$ , la probabilità di morte è

$$R_0 = \frac{\frac{u_t}{1-u_t}}{\frac{p_t}{1-p_t}} \implies u_t = \frac{R_0 p_t}{1-p_t + R_0 p_t}.$$

Analogamente, sotto  $H_1$  si ha che

$$R_A = \frac{\frac{q_t}{1-q_t}}{\frac{p_t}{1-p_t}} \implies q_t = \frac{R_A p_t}{1-p_t + R_A p_t}.$$

Tornando al punteggio di log-verosimiglianza (8), sostituendo diventa

$$W_t = \begin{cases} \log \frac{1-p_1}{1-p_0} & \text{se } y = 0 \\ \log \frac{p_1}{p_0} & \text{se } y = 1 \end{cases}$$

$$= \begin{cases} \log \frac{1-q_t}{1-u_t} & \text{se } y = 0 \\ \log \frac{q_t}{u_t} & \text{se } y = 1 \end{cases}$$

$$= \begin{cases} \log \frac{1-p_t+R_0 p_t}{1-p_t+R_A p_t} & \text{se } y = 0 \\ \log \frac{R_A(1-p_t+R_0 p_t)}{R_0(1-p_t+R_A p_t)} & \text{se } y = 1 \end{cases}$$

### Osservazioni

- › Una volta aggiustato per il rischio del paziente, l'ARL delle carte si comportano in modo coerente anche stratificando rispetto all'abilità del chirurgo (Figura 3).
- › La performance della CUSUM si può stratificare a posteriori secondo il livello di competenza dei chirurghi.
- › Grafico in Figura 4 dà la robustezza della procedura rispetto alle stime dei parametri del modello di regressione via bootstrap.

**Problemi**

Per migliorare questa analisi bisognerebbe utilizzare un modello logistico più ampio, per tenere conto di altre covariate e dell'abilità del chirurgo nell'odds a priori, perché la variabilità bootstrap che si vede per i chirurghi con maggiore esperienza andrebbe spiegata meglio.

## LEZIONE 20: SHEN ET AL. (2016)

### «SELF-STARTING MONITORING SCHEME FOR POISSON COUNT DATA WITH VARYING POPULATION SIZES»

2020-11-19

#### 20.1 Introduzione

Si introduce una carta self-starting, quindi combinazione di Fase I e Fase II, per sorvegliare conteggi di Poisson nel caso in cui il processo sia non omogeneo (popolazioni di ampiezza  $n_i$  variabile).

$$\text{In controllo : } X_i | n_i \sim \text{Pois}(\vartheta_0 n_i)$$

$$\text{Fuori controllo : } X_i | n_i \sim \text{Pois}(\vartheta_1 n_i)$$

In questo caso si applica alla sorveglianza dell'incidenza di melanomi, cercando di individuare il prima possibile se aumenta l'occurrence rate.

**Procedura** Per disegnare la carta di controllo si possono usare le trasformazioni  $Q$  di Quesenberry (5), che nel caso gaussiano (test  $t$ ) portano a una distribuzione  $\mathcal{N}(0, 1)$ . Tuttavia, per una Poisson non si ottengono quantità pivotali e sono necessari limiti di controllo dinamici, che in questo caso sono ottenuti con un bootstrap parametrico condizionato.

Si assume un modello Poisson con numerosità variabile e shift di tipo step fuori controllo:

$$X_i | n_i \sim \begin{cases} \text{Pois}(\vartheta_0 n_i) & i = -m_0 + 1, \dots, 0, 1, 2, \dots, \tau - 1 \\ \text{Pois}(\vartheta_1 n_i) & i = \tau, \tau + 1, \dots \end{cases}$$

Si possono costruire le carte standard

$$\text{Shewhart : } \vartheta_0 \pm L\sqrt{\vartheta_0}$$

$$\text{CUSUM : } W_i = \log \frac{f(x_i | \vartheta_1)}{f(x_i | \vartheta_0)} = x_i - \frac{\vartheta_1 - \vartheta_0}{\log \vartheta_1 - \log \vartheta_0}$$

In questo paper, si propone un metodo che procede nel modo seguente:

- › Si stimano i parametri con le osservazioni disponibili  $(X_{-m_0+1}, n_{-m_0+1}), \dots, (X_0, n_0)$ :

$$\hat{\vartheta}_t = \frac{\sum_{j=-m_0+1}^{t-1} X_j}{\sum_{j=-m_0+1}^{t-1} n_j},$$

- › Li si standardizza con

$$S_t = \frac{X_t - n_t \hat{\vartheta}_t}{\sqrt{n_t \hat{\vartheta}_t}}$$

e si calcola una statistica simile alla EWMA ma con barriera riflettente per proteggersi rispetto all'inerzia (si potrebbe usare l'AEWMA?):

$$\begin{cases} R_0 = 0 \\ R_t = \max \{0, (1 - \lambda)R_{t-1} + \lambda S_t\} \end{cases}$$

con run length  $RL = \inf \{t : Z_t > h_t\}$  e limiti di controllo dinamici  $h_t$ .

- › Bisogna calcolare i limiti di controllo per un  $\alpha$  specificato  $h_1(\alpha), h_2(\alpha), \dots, h_i(\alpha), \dots$  in modo da controllare il tasso di falso allarme condizionato:

$$\begin{cases} P(Z_1 > h_1(\alpha) | n_1) = \alpha \\ P(Z_1 > h_1(\alpha) | n_t, Z_i \leq h_i(\alpha), 1 \leq i < t) = \alpha \quad \text{per } t = 2, 3, \dots, \end{cases} \quad (9)$$

**Problema** Siccome la distribuzione di  $Z_t | n_t$  non è pivotale e dipende da  $\vartheta_0$ , è difficile approssimare la sequenza  $h_i(\alpha)$ .

**Soluzione** Metodo bootstrap parametrico *condizionato* per il calcolo di  $h_i(\alpha)$ .

## 20.2 Calcolo dei limiti di controllo

Il calcolo avviene tramite bootstrap parametrico, per cui ad ogni istante di tempo si stima il parametro e si simulano delle osservazioni per calcolare i limiti.

Attenzione, perché siccome si mantiene costante la probabilità condizionata (9), è necessario *condizionarsi* ai limiti precedenti:

1. Si aggiorna la stima di  $\vartheta_t$  sulla base dei dati  $i = 1, \dots, t - 1$ :

$$\hat{\vartheta}_t = \frac{\sum_{i=-m_0+1}^{t-1} y_i}{\sum_{i=-m_0+1}^{t-1} n_i},$$

2. Con la stima attuale  $\hat{\vartheta}_t$ , si generano  $i = 1, \dots, N$  osservazioni  $Y_{i,t} \sim \text{Pois}(n_t \hat{\vartheta}_t)$ , da cui si ottiene un campione di statistiche di controllo:

$$\begin{aligned} S_{i,t} = \frac{Y_{i,t} - \hat{\vartheta}_t n_t}{\sqrt{n_t \hat{\vartheta}_t}} &\rightsquigarrow \mathbf{R}(t) = \{R_1(t), \dots, R_N(t)\} \\ &= \{(1 - \lambda)R_1(t-1) + \lambda S_{1,t}, \dots, (1 - \lambda)R_N(t-1) + \lambda S_{N,t}\}. \end{aligned}$$

Una volta ordinato il vettore  $\mathbf{R}(t)$ , il limite di controllo simulato al tempo  $t$  è  $R_{[H]}(t)$ , con  $[H] = [(1 - \alpha) \cdot N] = \text{round}((1 - \alpha) \cdot N)$ .

3. Si confronta la statistica  $Z_t$  con  $R_{[H]}(t)$  e, se non si interrompe la procedura, si aggiorna  $\mathbf{R}(t)$  ricampionando  $N$  tra i valori di  $\mathbf{R}(t)$  più piccoli del limite di controllo (*bootstrap condizionato*), ovvero  $\{R_{[1]}(1), \dots, R_{[H]}(1)\}$ .



## LEZIONE 21: SHEN ET AL. (2016) (II)

2020-11-20

### 21.1 Confronto con altri competitor

La carta proposta viene confrontata con diversi competitor:

Competitor	Descrizione
TR-1	Stima $\hat{\vartheta}_0$ e lo tiene fisso, fa abbastanza schifo.
TR-2	Assume $\vartheta_0$ noto e si fa bootstrap condizionato da $\text{Pois}(n_j \vartheta_0)$
TR-3	Approssima con $\text{Bin}(n_t/N_t)$ e usa $Q$ di Quesenberry.
TR-4	Complicatissima, ma si assume che $n_t$ sia costante, con qualche modifica.

Si mostra che TR-1, TR-2 e TR-3 in generale sono poco soddisfacenti, mentre TR-4 è competitiva (tradotto: praticamente equivalente) a quella proposta nell'articolo.

### 21.2 Studio dell'ARL

La variabilità della procedura, e quindi dell'ARL, dipende da due sorgenti:

$$\text{Variabilità} = \text{Variabilità campionaria} + \underbrace{\text{Variabilità bootstrap} \mid \text{campione storico}}_{\text{CONDIZIONATA}}$$

#### Paragrafo 3.2: Average ARL

Si ottiene un'unica ARL per ogni set di valori storici di dimensione  $m_0$  simulati.

Si valuta l'average run length in funzione di  $m_0$ ,  $\text{ARL}(m_0)$ , che diventa una variabile casuale dal momento che il campione è esso stesso casuale: se ne studia allora il valore medio  $\text{AARL}(m_0)$ .

Per un particolare  $\vartheta_0$  si generano  $m_0$  valori preliminari e poi si calcola l'ARL in controllo. Si effettuano  $M$  generazioni e si calcola l'average run length in controllo per lo specifico  $m_0$ .

$$\left. \begin{array}{l} y_{-m_0+1}^{(1)}, \dots, y_0^{(1)} \longrightarrow \text{RL}^{(1)}(m_0) \\ y_{-m_0+1}^{(2)}, \dots, y_0^{(2)} \longrightarrow \text{RL}^{(2)}(m_0) \\ \vdots \\ y_{-m_0+1}^{(K)}, \dots, y_0^{(K)} \longrightarrow \text{RL}^{(K)}(m_0) \end{array} \right\} \longrightarrow \text{AARL}(m_0)$$

#### Paragrafo 3.3: ARL condizionata

Si verifica anche che l'ARL condizionata non sia troppo variabile: si simulano 100 insiemi di valori storici e per ognuno di esso si simulano 400 RL dallo stesso insieme di valori storici e si fa la media

campionaria.

$$\begin{array}{c}
 y_{-m_0+1}^{(1)}, \dots, y_0^{(1)} \longrightarrow \left\{ \begin{array}{c} RL^{(1,1)} \\ RL^{(1,2)} \\ \vdots \\ RL^{(1,400)} \end{array} \right\} \\
 y_{-m_0+1}^{(2)}, \dots, y_0^{(2)} \longrightarrow \left\{ \begin{array}{c} RL^{(2,1)} \\ RL^{(2,2)} \\ \vdots \\ RL^{(2,400)} \end{array} \right\} \\
 \vdots \\
 y_{-m_0+1}^{(K)}, \dots, y_0^{(K)} \longrightarrow \left\{ \begin{array}{c} RL^{(K,1)} \\ RL^{(K,2)} \\ \vdots \\ RL^{(K,400)} \end{array} \right\}
 \end{array} \bigg\} \longrightarrow \text{AARL}(m_0)$$

Mostrano che l'ARL condizionata ha poche fluttuazioni rispetto a quanto dichiarato.

### 21.3 Conclusioni

Questa procedura non porta a chissà che grande vantaggio di performance IC e OC rispetto alla procedura TR-4 (paragrafo 3.2). Il valore aggiunto che porta l'articolo è la tecnica di calcolo dei limiti di controllo dinamici via bootstrap.

**Da saper raccontare all'esame:**

1. Bootstrap parametrico per limiti dinamici.
2. Sapere cos'è una carta self-starting e quali sono le problematiche (info storiche, distribuzioni, dimensione di  $\tau$  e  $\delta$ , window of opportunity, ...).
3. Trasformazione  $Q$  di Quesenberry per distribuzioni non normali.
4. Differenze tra  $\text{ARL}_0$  condizionata e non condizionata.
5. Le altre procedure con cui viene confrontata, specialmente TR-4.

## LEZIONE 22: ALTRE CARTE DI FASE II

2020-11-25

Riferimenti Qiu, 2014, pp. 327–328, 332–333 (BASE)

Qiu, 2014, pp. 338–341 (CHANGE-POINT)

## 22.1 Carte univariate non parametriche

Alcuni test non parametrici non hanno nulla da invidiare a quelli basati su distribuzioni parametriche. La logica è sempre la stessa, con  $\mathbf{x}_j \sim F_0(\mathbf{x})$  per  $j = 1, \dots, \tau-1$  e  $\mathbf{x}_j \sim F_1(\mathbf{x})$  per  $j = \tau, \dots$

Si assume che  $F_0$  non sia nota o appartenente a un modello parametrico. Si assume però che dalla Fase I sia stato possibile ottenere un campione in controllo, a partire dal quale effettuare i confronti.

L'idea è calcolare la somma dei ranghi  $W_i$  del campione disponibile al tempo  $i$ , una volta combinato col campione della Fase I:

$$\underbrace{\mathbf{x}_{1-m}, \dots, \mathbf{x}_0}_{\text{IN CONTROLLO}} \quad \underbrace{\mathbf{x}_i}_{?}$$

## Osservazione

- ›  $W_i$  è la statistica del test [Wilcoxon-Mann-Whitney](#) per verificare che i due campioni abbiano la stessa “posizione”.

## 22.1.1 “Magnifiche tre” su Wilcoxon

Si può dimostrare che la distribuzione di  $W_1, W_2, \dots$  non dipende da  $F_0(\cdot)$ , per cui si può controllare il valore atteso che, in controllo, è pari a

$$\mu_W = \frac{n(n+m+1)}{2}, \quad \sigma_W^2 = \frac{nm(n+m+1)}{12}.$$

Fuori controllo, ci si aspetta che  $\mathbb{E}[W_i] \neq \mu_W$ , per cui si può applicare una carta di controllo Shewhart/CUSUM/EWMA alla statistica standardizzata

$$\tilde{W}_i = \frac{W_i - \mu_W}{\sigma_W},$$

poiché è tale che

$$\mathbb{E}_{\text{IC}}[\tilde{W}_i] = 0$$

$$\mathbb{E}_{\text{FC}}[\tilde{W}_i] \neq 0$$

Carta	Statistica	Limiti
Shewhart	$\tilde{W}_i = (W_i - \mu_W)/\sigma_W$	$\pm L$
CUSUM +	$S_i^+ = \max\{0, S_i^+ + \tilde{W}_i - k\}$	$L$
EWMA	$S_i = (1 - \lambda)S_{i-1} + \lambda\tilde{W}_i$	$\pm L$

I limiti di controllo, poiché la distribuzione dei  $W_i$  non dipende dalla distribuzione di  $\mathbf{x}_i$ , si possono calcolare simulando da una qualunque distribuzione continua e utilizzando il bootstrap condizionato

come in Shen et al. (2016), per imporre che

$$P(S_1 > L_1) = \frac{1}{B}$$

$$P(S_i > L_i | S_1 \leq L_1, \dots, S_{i-1} \leq L_{i-1}) = \frac{1}{B}$$

### 22.1.2 Change point su Wilcoxon

Alternativamente, si può usare la versione change-point delle carte basate sui  $\tilde{W}_i$ , che ora si calcolano per verificare l'ipotesi che non sia avvenuto un cambiamento prima di  $i$ :

$$\tilde{W}_i = \max_{\tau=1, \dots, i-1} \left| \frac{W(i, \tau) - \mu_W(i, \tau)}{\sigma_W(i, \tau)} \right|,$$

dove  $W(i, \tau)$  è la statistica di Wilcoxon che verifica l'ipotesi nulla che i campioni

$$\underbrace{\{\mathbf{x}_{1-m}, \dots, \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\tau-1}\}}_{\text{PRIMA DI } \tau} \quad \text{e} \quad \underbrace{\{\mathbf{x}_\tau, \dots, \mathbf{x}_i\}}_{\text{DOPO } \tau}$$

provengano dalla stessa distribuzione.

In questo caso, il calcolo dei limiti di controllo è più oneroso ma la carta è anche in generale più efficiente.

## 22.2 Carte multivariate parametriche

Nel caso multivariato, studieremo solo le carte che proteggono rispetto a un cambiamento nella *media* della distribuzione e non rispetto a un cambiamento di scala.

Assumendo  $\boldsymbol{\mu}_1$  e  $\Sigma_0$  note (o comunque accuratamente stimate in Fase I), si ha il solito schema di controllo di processo:

$$\begin{aligned} \mathbf{x}_1 &\sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0) \\ &\vdots \\ \mathbf{x}_{\tau-1} &\sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0) \\ \mathbf{x}_\tau &\sim \mathcal{N}(\boldsymbol{\mu}_0 + \boldsymbol{\delta}, \Sigma_0) \\ \mathbf{x}_{\tau+1} &\sim \mathcal{N}(\boldsymbol{\mu}_0 + \boldsymbol{\delta}, \Sigma_0) \end{aligned}$$

dove  $\boldsymbol{\delta}$  è il cambiamento non standardizzato che contiene le seguenti informazioni:

1.  $\{i : \delta_i \neq 0\} \rightarrow$  Quali sono le variabili coinvolte nel cambiamento.
2.  $\boldsymbol{\delta}/\|\boldsymbol{\delta}\| \rightarrow$  Direzione del cambiamento
3.  $\|\boldsymbol{\delta}\| \rightarrow$  Dimensione del cambiamento

### 22.2.1 Carte combinate

Poiché  $\mathbf{x}_i \in \mathbb{R}^p$ , si potrebbero utilizzare  $p$  carte univariate, una per ciascuna variabile, e calcolare l'ARL in controllo in modo che siano bilanciate:

$$P(\text{carta 1 chiama l'allarme}) = P(\text{carta 2 chiama l'allarme}) = \dots = P(\text{carta } p \text{ chiama l'allarme})$$

Tuttavia, questo schema non tiene conto della correlazione tra le variabili, che è un fattore molto rilevante nel caso multidimensionale (Figura 27).

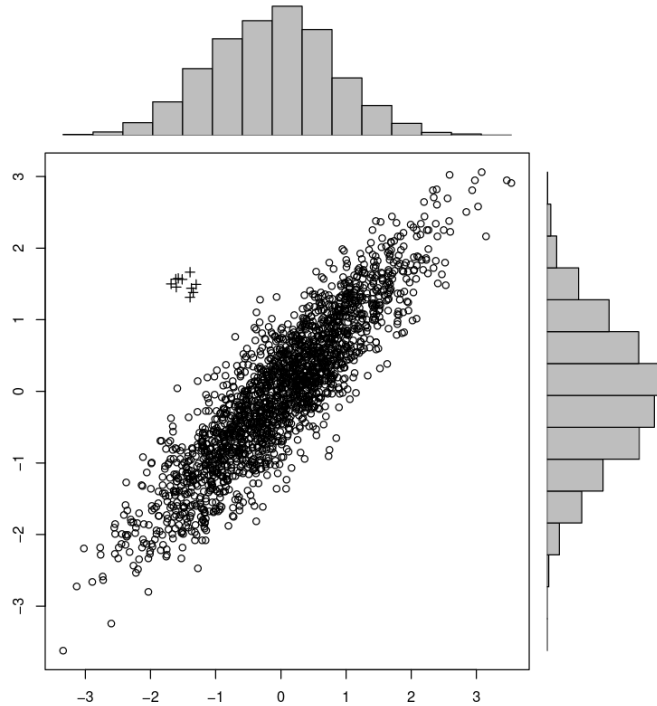


Figura 27: Le osservazioni indicate con il + sono “accettabili” per ciascuna distribuzione marginalmente, ma chiaramente outlier rispetto al resto delle osservazioni.

### 22.2.2 $T^2$ di Hotelling

Abbiamo già visto questa carta, la differenza è che ora si basa sulla distribuzione  $\chi^2$  invece della  $F$  riscalata, poiché si conoscono i parametri in seguito alla Fase I:

$$T_i^2 = (\mathbf{x}_i - \boldsymbol{\mu}_0)^\top \Sigma_0^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0) \stackrel{H_0}{\sim} \chi_p^2.$$

## LEZIONE 23: CARTE DI FASE II MULTIVARIATE

2020-11-26

### 23.1 Carte multivariate parametriche (ii)

*Riferimenti* Qiu, 2014, pp. 275–289

Healy (1987)	(CUSUM di Healy)
Crosier (1988)	(CUSUM di Crosier)
Lowry et al. (1992)	(EWMA multivariata)

#### 23.1.1 CUSUM di Healy

Lo schema proposto da Healy (1987) consiste nell'applicare un CUSUM univariato su  $\mathbf{x}$ : si calcola il rapporto di verosimiglianza

$$\log \frac{f_1(\mathbf{x}_i)}{f_0(\mathbf{x}_i)} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\top \Sigma_0^{-1} \mathbf{x}_i - \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\top \Sigma_0^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

$$S_i = \max \{0, S_{i-1} + \mathbf{a}^\top (\mathbf{x}_i - \boldsymbol{\mu}_0) - 0.5d\},$$

per una particolare direzione  $\mathbf{a}$  ottenuta per verificare il test di ipotesi

$$H_0 : \mathbf{a}^\top \boldsymbol{\mu} = \mathbf{a}^\top \boldsymbol{\mu}_0 \quad \text{per ogni } \mathbf{a} \in \mathbb{R}^p.$$

Utilizzando il metodo di unione ed intersezione, che abbiamo già visto nella Lezione 4 quando si parlava di test multipli, si ottiene la seguente statistica test:

$$\max_{\mathbf{a} \in \mathbb{R}^p} \frac{\mathbf{a}^\top (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \mathbf{a}}{\mathbf{a}^\top S \mathbf{a}} \quad (10)$$

con soluzione massimizzante

$$\mathbf{a} = \frac{S^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)}{\sqrt{(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top S^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)}}$$

con  $d = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\top \Sigma_0^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}$  parametro di non centralità. Questa statistica considera solo cambiamenti da  $\boldsymbol{\mu}_0$  a  $\boldsymbol{\mu}_0 + \boldsymbol{\delta}$ , perché  $d$  è la lunghezza di Mahalanobis di  $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$ .

#### 23.1.2 CUSUM di Crosier

Crosier (1988) costruisce una versione multivariata della CUSUM che sia simile al caso univariato, in cui si ha

$$\begin{array}{ccc} \text{Somma cumulata delle} & + & \text{“compressione” verso} \\ \text{deviazioni dalla media} & & \text{lo zero} \end{array},$$

e successivamente la trasforma in uno scalare per poterla monitorare nel tempo.

La sua carta di controllo è un semplice plug-in multivariato della CUSUM standard, ovvero

$$S_i = \max \{0, S_{i-1} + (\mathbf{x}_i - \boldsymbol{\mu}_0) - \mathbf{k}\}$$

**1) Come si interpreta  $\max\{\mathbf{0}, \mathbf{b}\}$  se  $\mathbf{0}, \mathbf{b} \in \mathbb{R}^p$ ?**

Sia  $\mathbf{d}_i = \mathbf{S}_{i-1} + (\mathbf{x}_i - \boldsymbol{\mu}_0)$ , in questa carta si confrontano i vettori in termini di lunghezza rispetto alla norma di Mahalanobis data da

$$\|\mathbf{v}\|_{\Sigma_0^{-1}} = \sqrt{\mathbf{v}^\top \Sigma_0^{-1} \mathbf{v}},$$

ponendo  $\mathbf{S}_i = \mathbf{0}$  se

$$\|\mathbf{d}_i\|_{\Sigma_0^{-1}} \leq \|\mathbf{k}\|_{\Sigma_0^{-1}}.$$

Altrimenti, siccome nel caso univariato la statistica viene contratta di un fattore  $k$  verso lo zero, qui si utilizza un fattore di compressione che dipende dalla norma dei vettori:

$$\mathbf{S}_i = \mathbf{d}_i \cdot \left(1 - \frac{\|\mathbf{k}\|_{\Sigma_0^{-1}}}{\|\mathbf{d}_i\|_{\Sigma_0^{-1}}}\right).$$

Infatti, se si vuole effettuare uno shrinkage, il vettore  $\mathbf{k}$  deve essere tale che  $\mathbf{k}^\top \Sigma_0^{-1} \mathbf{k} = k^2$  e che abbia la stessa direzione del vettore  $\mathbf{S}_{i-1} + (\mathbf{x}_i - \boldsymbol{\mu}_0)$ , per cui

$$\mathbf{k} = \underbrace{\|\mathbf{k}\|_{\Sigma_0^{-1}}}_{\text{LUNGHEZZA}} \cdot \underbrace{\frac{\mathbf{d}_i}{\|\mathbf{d}_i\|_{\Sigma_0^{-1}}}}_{\text{DIREZIONE}}.$$

Per rappresentarlo come grafico, si controlla

$$Y_i = \sqrt{\mathbf{S}_i^\top \Sigma_0^{-1} \mathbf{S}_i},$$

con lo schema che chiama un allarme se

$$\mathbf{S}_i^\top \Sigma_0^{-1} \mathbf{S}_i > L.$$

Il limite di controllo che si trova in modo standard imponendo  $\text{ARL}_0 = B$ , anche se in questo caso la massimizzazione è abbastanza più complessa.

**Osservazione**

Si mostra che il comportamento fuori controllo dipende

**23.1.3 EWMA multivariata**

Lowry et al. (1992) introduce la versione multivariata dell'EWMA, che inizializza  $\mathbf{S} = \boldsymbol{\mu}_0$  e aggiorna la statistica di controllo con

$$\mathbf{S}_i = (I_{p \times p} - \lambda)\mathbf{S}_{i-1} + \lambda\mathbf{x}_i,$$

con lo schema che segnala un allarme la prima volta che

$$\frac{2-\lambda}{\lambda}(\mathbf{S}_i - \boldsymbol{\mu}_0)^\top \Sigma_0^{-1}(\mathbf{S}_i - \boldsymbol{\mu}_0) > L.$$

Alternativamente, se si dispone di informazioni a priori riguardo il peso delle variabili si può usare

una variante con diverso liscio per ogni componente

$$\mathbf{S}_i = (I_{p \times p} - \Lambda) \mathbf{S}_{i-1} + \Lambda \mathbf{x}_i,$$

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p),$$

ma l'ottimizzazione si complica notevolmente e bisogna conoscere in anticipo quali variabili siano importanti. Nella pratica, nessuno usa questa variante.

Come sempre, si possono usare CUSUM/EWMA multipli, carte GLR multivariate **wang2013**, carte EWMA adattive multivariate **mahmoud2010**, CUSUM adattive **dai2010**, ...

Anche nel multivariato ci sono schemi self-starting **hawkins2007**, **hawkins2010**, **capizzi2010**, che però hanno dei grossi problemi per quanto riguarda la stima della matrice di dispersione.

## 23.2 Maledizione della dimensionalità

In letteratura è uno dei problemi principali, ovvero quello che accade muovendosi in dimensioni più elevate.

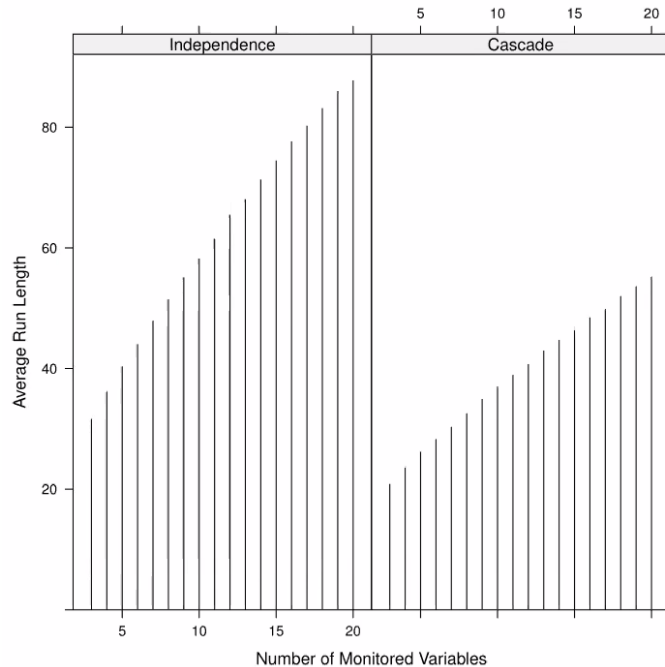


Figura 28:  $ARL_1$  per una MEWMA al crescere del numero delle variabili indipendenti (*sinistra*) o con correlazione  $\rho^{|i-j|}$  (*destra*); tutte le carte sono ottimizzate per  $ARL_0 = 500$  e lo shift è nelle prime due variabili,  $\boldsymbol{\delta} = (0.5, 0.5, 0, \dots)^\top$ . Si osserva che  $d = \sqrt{\boldsymbol{\delta}^\top \Sigma_0^{-1} \boldsymbol{\delta}}$  non dipende da  $p$ .

### Osservazioni sulla Figura 28

- › Impatto negativo dell'aumento di  $p$
- › L'impatto è diverso se è presente una struttura di dipendenza



- › Capire come è fatto  $\delta$  e come la sparsità influenza l'ARL.

### Possibili soluzioni

1. Utilizzare *componenti principali* per ridurre la dimensionalità.
2. Assumere che il cambiamento coinvolga *una sola variabile*.
3. Cercare in modo *adattivo* le direzioni (LASSO, LAR, ...).

#### 23.2.1 Articolo LAR Fase II

In **MPhase1** si effettua una selezione degli istanti temporali con l'adaptive LASSO, mentre nell'articolo di Fase II si usa il LAR per selezionare le variabili.

Le finalità dei due articoli sono differenti:

1. **MPhase1**: sparsità sul numero e tipologia di *shift temporali*.
2. **??**: sparsità sul numero di *variabili che shiftano*.

Articolo di Fase II parametrico, che ipotizza un cambiamento nel parametro di posizione e un aumento di variabilità:

$$\begin{cases} \mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0) & i < \tau \text{ in controllo} \\ \mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_0 + \boldsymbol{\delta}, \Sigma_0) & i \geq \tau \text{ fuori controllo} \end{cases}$$

#### 23.2.2 Richiami sulle componenti principali

Si decompone una matrice di varianza e covarianza come

$$\Sigma_0 = V\Omega V^\top,$$

dove

$$\Omega = \text{diag}(\omega_1, \omega_2, \dots, \omega_p), \quad V = (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_p),$$

con autovalori  $\omega_1 \geq \omega_2 \geq \dots \geq \omega_p \geq 0$  e autovettori associati  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ .

Per  $u < p$ , si mantengono solo le prime  $u$  componenti

$$\Omega_0 = \text{diag}(\omega_1, \omega_2, \dots, \omega_u), \quad V_0 = (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_u).$$

Le prime  $u$  componenti di  $\mathbf{x}_i$  sono calcolate come se il processo sia in controllo:

$$\mathbf{y}_{u_i} = V_0^\top (\mathbf{x}_i - \boldsymbol{\mu}_0)$$

e la rappresentazione di  $\mathbf{x}_i$  rispetto alle prime  $u$  componenti principali è, come se fosse un modello lineare,

$$\mathbf{x}_i = \underset{p \times 1}{\boldsymbol{\mu}_0} + \underset{p \times u}{V_0} \underset{u \times 1}{\mathbf{y}_{u_i}} + \underset{p \times 1}{\mathbf{r}_i}$$

mentre la componente residua non spiegata dalle prime  $u$  componenti principali è

$$\mathbf{r}_i = \mathbf{x}_i - \boldsymbol{\mu}_0 - V_0 \mathbf{y}_{u_i} = \mathbf{x}_i - \boldsymbol{\mu}_0 - V_0 (V_0^\top (\mathbf{x}_i - \boldsymbol{\mu}_0)) = (I - V_0 V_0^\top) (\mathbf{x}_i - \boldsymbol{\mu}_0).$$

Con questa scrittura,

$$\mathbb{E} [\mathbf{r}_i] = (I - V_0 V_0^\top) \mathbb{E} [\mathbf{x}_i - \boldsymbol{\mu}_0] = \mathbf{0}, \quad \mathbb{V} [\mathbf{r}_i] = \Sigma_0 - V_0 \Omega_0 V_0^\top.$$

Si vuole allora introdurre una carta per sorvegliare le due componenti, spiegata e residua, in modo parallelo.

## LEZIONE 24: CARTE DI FASE II MULTIVARIATE (II)

2020-11-29

### 24.1 Monitoring delle componenti principali

Si assume di aver calcolato le prime  $u$  componenti principali  $\mathbf{y}_{u_i}$  che, poiché sono combinazioni lineari di v.c. Gaussiane, sono ancora normali

$$\begin{aligned} \text{IC : } \mathbf{y}_i &\sim \mathcal{N}(\mathbf{0}_u, \Omega_0) \\ \mathbf{r}_i &\sim \mathcal{N}(\mathbf{0}_p, \Sigma_0 - V_0 \Omega_0 V_0^\top), \\ \text{FC : } \mathcal{N}(V_0^\top \delta, \Omega_0) \\ &\mathcal{N}((I - V_0 V_0^\top) \delta, \Sigma_0 - V_0 \Omega_0 V_0^\top). \end{aligned} \tag{11}$$

#### Problemi

- › Il  $\delta$  viene moltiplicato per  $I - V_0 V_0^\top$ , per cui l'interpretabilità dello shift va a farsi benedire.
- › Siccome gli shift sulle  $x_j$  non si trasmettono immediatamente alle  $y_i$  a causa dei pesi nella combinazione lineare, non si ha la stessa reattività di allarme che si avrebbe sulle variabili originali.
- › La quantità di shift che viene trasferito sulla trasformazione non è detto che sia completo, lo stesso che accade anche studiando i residui di un modello per osservazioni autocorrelate.

Una soluzione consiste nel sorvegliare congiuntamente le componenti principali e i residui, ovvero

$$\begin{cases} \mathbf{y}_1, \dots, \mathbf{y}_u \\ q_i = \mathbf{r}_i^\top \mathbf{r}_i \end{cases}$$

Dal momento che

$$\mathbb{E}_{\text{IC}} [q_i] = \mathbb{E}_{\text{IC}} [\mathbf{r}_i^\top \mathbf{r}_i] = \mathbb{E}_{\text{IC}} [\text{tr}(\mathbf{r}_i \mathbf{r}_i^\top)] \stackrel{(*)}{=} \text{tr}(\mathbb{E}[\mathbf{r}_i \mathbf{r}_i^\top]) = \text{tr}(\Sigma_0 - V_0 \Omega_0 V_0^\top) \stackrel{(**)}{=} \sum_{j=u+1}^p \lambda_j,$$

dove al passaggio (\*) è stata usata la proprietà di [linearità](#) della traccia e varianza in controllo di  $\mathbf{r}_i$  data da (11), mentre al passaggio (\*\*) si è usato il fatto che la [traccia di una matrice è uguale alla somma degli autovalori](#).

#### Osservazioni

- ›  $q_i$  riflette la “dimensione” delle restanti  $p - u$  componenti principali.
- › Sotto normalità, le PC sono indipendenti sia tra di loro sia con i residui.

Di solito quello che si fa è applicare una carta multivariata alle prime  $u$  componenti principali per sorvegliarne la media (media IC =  $\mathbf{0}_u$ ) assieme ad uno schema per sorvegliare la somma  $q_i$  dei quadrati dei residui:

$$E_i = \max \{ \mathbb{E}_{\text{IC}} [q_i], (1 - \lambda) E_{i-1} + \lambda q_i \},$$

con l'allarme che viene chiamato se  $E_i > L_q$ .

## 24.2 Statistiche non invarianti per direzione

Le statistiche basate sulle forme quadratiche ( $T^2$ , MEWMA, MCUSUM) sono *invarianti*, ovvero la performance non dipende della direzione del vettore OC:

$$T(W\mathbf{x}_i + \mathbf{b}, W\boldsymbol{\mu}_0 + \mathbf{b}) = T(\mathbf{x}_i, \boldsymbol{\mu}_0), \quad \text{se } W \text{ ortogonale.}$$

### Singola direzione

Supposto di conoscere che il cambiamento  $\boldsymbol{\delta}$  si trovi lungo una retta individuata da un vettore noto  $\mathbf{f}$ , ovvero

$$\boldsymbol{\delta} = \beta \cdot \mathbf{f},$$

il modello diventa allora

$$\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_0 + \underbrace{\beta \cdot \mathbb{1}_{i \geq \tau}}_{b_i} \cdot \mathbf{f}, \Sigma_0).$$

In tal caso, si calcola la stima ai minimi quadrati (generalizzati) di  $b_i = \beta \cdot \mathbb{1}_{i \geq \tau}$

$$\hat{b}_i = (\mathbf{f}^\top \Sigma_0^{-1} \mathbf{f})^{-1} \mathbf{f}^\top \Sigma_0^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0), \quad \mathbb{V}[\hat{b}_i] = (\mathbf{f}^\top \Sigma_0^{-1} \mathbf{f})^{-1}.$$

A questo punto, ad ogni istante di tempo  $i$  si testa il sistema di ipotesi

$$\begin{cases} H_0 : \hat{b}_i \sim \mathcal{N}(0, \mathbf{f}^\top \Sigma_0^{-1} \mathbf{f}) \\ H_1 : \hat{b}_i \sim \mathcal{N}(\beta, \mathbf{f}^\top \Sigma_0^{-1} \mathbf{f}) \end{cases}$$

per cui, poiché valore atteso e varianza IC li conosciamo, si possono utilizzare gli schemi univariati per controllare le stime  $\hat{b}_i$  in funzione del tempo.

### Multiple direzioni

Supponendo di voler verificare che la direzione del cambiamento sia all'interno di un *insieme finito* di  $u$  possibili direzioni, ovvero

$$\boldsymbol{\delta} \in \{\beta_1 \mathbf{f}_1, \beta_2 \mathbf{f}_2, \dots, \beta_u \mathbf{f}_u\},$$

il modello in questione diventa allora

$$\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_0 + b_{i,1} \mathbf{f}_1 + \dots + b_{i,u} \mathbf{f}_u, \Sigma_0),$$

e analogamente il sistema di ipotesi è

$$\begin{cases} b_{i,1} = \dots = b_{i,u} = 0 \\ \text{UNO SOLO dei } b_{i,j} \neq 0 \end{cases} \quad \left( \begin{array}{l} \text{non può shiftare in più} \\ \text{direzioni assieme} \end{array} \right)$$

Si dispone ad ogni istante di tempo di  $u$  stime ai minimi quadrati, una per ogni direzione nota.

Per ciascuna di queste si costruisce una carta EWMA inizializzando  $S_{0,1} = \dots = S_{0,u} = 0$  e

$$z_{i,j} = \frac{\frac{\mathbf{f}_j^\top \Sigma_0^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_0)}{\mathbf{f}_j^\top \Sigma_0^{-1} \mathbf{f}_j}}{\sqrt{\mathbf{f}_j^\top \Sigma_0^{-1} \mathbf{f}_j}} = \frac{\mathbf{f}_j^\top \Sigma_0^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_0)}{\sqrt{\mathbf{f}_j^\top \Sigma_0^{-1} \mathbf{f}_j}} =$$

da cui si aggiornano

$$S_{i,j} = (1 - \lambda)S_{i-1,j} + \lambda z_{i,j}$$

e l'allarme viene segnalato se

$$\max \{|S_{i,1}|, |S_{i,2}|, \dots, |S_{i,u}|\} > L \sqrt{\frac{\lambda}{2 - \lambda}}.$$

### Osservazioni

- › Poiché sono standardizzate, si può usare un solo limite di controllo e le statistiche sono *confrontabili tra loro*.
- › Si assume che tutte le  $p$  variabili possano subire un cambiamento nel parametro di posizione, ma ciascuna in una sola delle  $u$  direzioni.

## 24.3 Carta regression-adjusted

Regression adjusted, si ha un numero di direzioni pari al numero di variabili osservate e si assume che una sola di queste possa subire un cambiamento.

### Assunzione

Il cambiamento coinvolge *una sola delle variabili* sorvegliate. Si assume allora che le direzioni siano quelle delle singole variabili:

$$\mathbf{f}_1 = \mathbf{e}_1, \quad \mathbf{f}_2 = \mathbf{e}_2, \quad \dots, \quad \mathbf{f}_p = \mathbf{e}_p.$$

Si sorvegliano  $z_{i,j}$  residui standardizzati e aggiustati rispetto alla regressione.

Sotto l'ipotesi di processo in controllo, date  $X_1, \dots, X_p$  le variabili sorvegliate, al tempo  $i$  si regredisce  $X_{i,j}$  sulle  $p - 1$  rimanenti

$$\operatorname{argmin}_{\beta} \left( \mathbb{E}_{\text{IC}} \left[ X_j - \beta_{j,0} - \sum_{h \neq j} \beta_{j,h} X_h \right] \right)^2,$$

per cui sia  $v_j^2$  il minimo della varianza dei residui.

In controllo,  $\mathbf{x}_i$  e  $\mathbf{z}_i$  sono marginalmente  $\mathcal{N}(0, 1)$

$$z_{i,j} = \frac{x_{i,j} - \beta_{j,0} - \sum_{h \neq j} \beta_{j,h} x_{i,h}}{v_j}.$$

Fuori controllo, se solo  $X_j$  shifta,

$$X_j \sim \mathcal{N}(\delta, 1)$$

$$\mathbb{E}[z_{i,j}] = \delta(1 - R_j^2).$$

**Osservazione**

$z_{i,j}$  contiene informazione su quello che succede al tempo  $i$  alla  $j$ -esima variabile e che non è spiegabile in funzione della sua relazione con le altre variabili.

---

Parte IV

## Journal Club

## LEZIONE 25: CAPIZZI E MASAROTTO (2011) «A LEAST ANGLE REGRESSION CONTROL CHART FOR MULTIDIMENSIONAL DATA»

2020-12-02

L'obiettivo di questo paper è introdurre una metodologia per sorvegliare processi multivariati, con particolare attenzione a

1. Individuare *sottoinsiemi* di variabili potenzialmente fuori controllo.
2. Monitorare sia cambiamenti in media sia aumenti di variabilità.
3. Adattare la metodologia sia a dati *non strutturati* che a dati *strutturati*, ad esempio nella sorveglianza di dati funzionali o processi multistage.

### 25.1 Introduzione

Si assume che il processo sia di tipo change-point

$$\mathbf{y}_t \sim \begin{cases} \mathcal{N}_n(\boldsymbol{\mu}, \Sigma) & t < \tau \quad \text{in controllo} \\ \mathcal{N}_n(\boldsymbol{\mu} + \boldsymbol{\delta}, \Omega) & t \geq \tau \quad \text{fuori controllo} \end{cases} \quad (12)$$

con  $\Omega - \Sigma \succ 0$  (definita positiva) e si assume di poter approssimare il cambiamento come

$$\boldsymbol{\delta} \approx F\boldsymbol{\beta},$$

con  $F$  matrice del disegno delle possibili direzioni note fuori controllo e  $\boldsymbol{\beta}$  parametri incogniti.

La matrice  $F$  può essere utilizzata come matrice di “scenari” fuori controllo, in quanto si possono ipotizzare diverse tipologie di cambiamento nei valori attesi come vedremo nei casi strutturati.

$\boldsymbol{\beta}$  ha un'interpretazione diversa a seconda di che tipologia di dati si osservano, ma in ogni caso riassume l'ampiezza del cambiamento.

In particolare, la metodologia proposta vuole cercare una soluzione *sparsa*, dal momento che ad elevata dimensionalità

- › È inverosimile che *tutte* subiscano un cambiamento.
- › C'è molto rumore e le metodologie standard di test di ipotesi perdono di efficacia.

Le soluzioni che sfruttano i metodi di proiezione (PCA, PLS) hanno degli svantaggi, in quanto funzionano se le variabili sono correlate ma non è chiaro

1. Quanto cambiamento si trasferisce sulle proiezioni?
2. Come si interpretano gli shift in termini di variabili originali?

Ci si vuole concentrare sul *sottoinsieme* di variabili che più plausibilmente sono responsabili dell'allarme, usando algoritmi di selezione delle variabili.

La selezione delle variabili avviene tramite il LAR, ma altre procedure come LASSO, Forward Search ecc. non impattano particolarmente sulla performance (capizzi2014).



## 25.2 Caso non strutturato

I dati non strutturati sono il caso più semplice, in cui può imporre  $F = I$  nel modello (12), cioè le colonne di  $F$  indicano la variabile la cui media è cambiata.

Il processo fuori controllo ha valore atteso

$$\mathbb{E} [\mathbf{y}_t - \boldsymbol{\mu}] = \mathbf{b}_t = \boldsymbol{\beta} \cdot \mathbb{1}_{t \geq \tau},$$

per cui l'andamento fuori controllo è direttamente interpretabile.

Ad ogni istante di tempo, si ha la stima ai minimi quadrati

$$\hat{\mathbf{b}}_i = (F^\top \Sigma_{0i}^{-1} F)^{-1} F^\top \Sigma_0^{-1} \mathbf{x}_i,$$

### Differenze rispetto al regression-adjusted

- › Regression-adjusted sono  $u$  carte univariate e si controllano in contemporanea.
- › Si può applicare una carta di controllo multivariata a  $\hat{\mathbf{b}}_i$  invece che alle  $\mathbf{x}_i$ . In questo modo, si possono individuare contemporaneamente intensità e direzione del cambiamento, oltre a ridurre la dimensionalità del problema.

## LEZIONE 26: CAPIZZI E MASAROTTO (2011) (II)

2020-12-03

## 26.1 Dati funzionali

Sorveglianza di dati funzionali, in questo caso si deve “sorvegliare” una relazione, stimata in Fase I a partire da coppie  $(x_i, y_i)$ , data da una combinazione lineare di funzioni di base

$$y_i = \sum_{j=1}^p \beta_j f_j(x_i),$$

e l’allarme corrisponde a segnalare quando una delle componenti  $f_j(\cdot)$  va fuori controllo.

Nella scrittura del modello (12), questo corrisponde ad assumere che

$$\begin{aligned} \mathbb{E}_{\text{OC}} [\mathbf{y}_t - \boldsymbol{\mu}] = F\boldsymbol{\beta} &= \begin{pmatrix} f_1(x_1) & f_2(x_1) & \dots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_p(x_2) \\ \vdots & \vdots & \dots & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_p(x_n) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \\ &= \begin{pmatrix} f_1(x_1) \\ f_1(x_2) \\ \vdots \\ f_1(x_n) \end{pmatrix} \beta_1 + \begin{pmatrix} f_2(x_1) \\ f_2(x_2) \\ \vdots \\ f_2(x_n) \end{pmatrix} \beta_2 + \dots + \begin{pmatrix} f_p(x_1) \\ f_p(x_2) \\ \vdots \\ f_p(x_n) \end{pmatrix} \beta_p, \end{aligned}$$

per cui il vettore  $\beta_j$  indica l’intensità dello shift della componente  $j$  della base di funzioni che descrive la relazione.

L’ipotesi sui parametri diventa dunque

$$\begin{cases} H_0 : \beta_j = \beta_0, \sigma_j = \sigma_0 \\ H_1 : \beta_j \neq \hat{\beta}_0 \text{ oppure } \sigma_j > \sigma_0 \end{cases}$$

Indicando con  $F_{\mathcal{N}(0,1)}^{-1}$  e  $F_{\chi_{n-k}^2}$  rispettivamente la funzione quantile della normale e la funzione di ripartizione della distribuzione  $\chi_{n-k}^2$ , si può scrivere il vettore  $(k+1)$ -dimensionale da controllare

$$\mathbf{z}_j = \begin{pmatrix} \hat{\beta}_j \\ F_{\mathcal{N}(0,1)}^{-1} \left( F_{\chi_{n-k}^2} \left( \frac{(n-k)\hat{\sigma}_j}{\sigma_0} \right) \right) \end{pmatrix},$$

che ha distribuzione in controllo

$$\mathbf{z}_j \sim \mathcal{N}(\boldsymbol{\mu}_z, \Sigma_z),$$

$$\boldsymbol{\mu}_z = \begin{pmatrix} \beta_0 \\ 0 \end{pmatrix},$$

$$\Sigma_z = \begin{pmatrix} \sigma_0(F^\top F)^{-1} & \mathbf{0}_k^\top \\ \mathbf{0}_k & 1 \end{pmatrix}.$$

Nell'articolo si pone pari a 0 la correlazioni tra  $\hat{\mathbf{b}}$  e  $\hat{\sigma}$ , che è valido sotto le ipotesi del modello lineare (indipendenza  $\hat{\beta}$  e  $\hat{\sigma}$  per il [teorema di Basu](#)), ma non nel caso più generale che viene considerato.

### Alternative

In generale, gli approcci più popolari per la sorveglianza di dati funzionali sono:

1. Assumere che valga un modello di regressione lineare con matrice del disegno data dalle  $f_j(\cdot)$  e sorvegliare la stabilità delle stime dei parametri  $(\hat{\beta}_t, \hat{\sigma}_t^2)$ .
2. Regressione non parametrica con lisciamiento esponenziale, si sorveglia la forma della funzione ottenuta.

Alcune metodologie di confronto tra funzioni comprendono la nested ANOVA, in cui ci sono osservazioni gerarchiche e si vuole valutare se c'è una variazione significativa tra i fattori:

$$\text{Variabilità} \left\{ \begin{array}{l} \text{linee produzione} \\ \text{lotti nelle linee} \\ \text{dentro lotti delle linee} \end{array} \right.$$

## 26.2 Processi Multistage

I processi di assemblaggio tipicamente avvengono su più stadi (processi *multistage*), in quanto si possono ottenere diverse misurazioni della stessa quantità a diversi istanti del processo (Figura 29).

La caratteristica principale dei processi multistage è la *propagazione dell'errore*, ovvero se il cambiamento avviene allo stadio  $\tau$  si propaga a tutti gli stadi successivi  $\tau+1, \tau+2, \dots$  fino all'output finale.

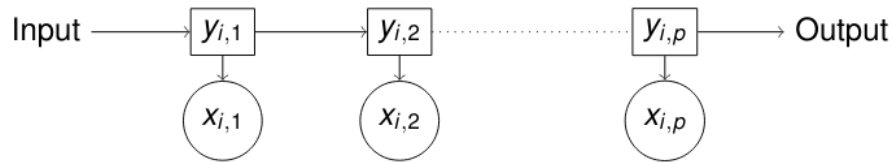


Figura 29: Esempio di processo multistage, in cui si osservano le misurazioni  $x_{i,j}$  che dipendono da quantità latenti non misurabili  $y_{i,j}$ .

Questa tipologia di processi ha una naturale modellazione mediante modelli state-space, che vengono utilizzati per descrivere contemporaneamente l'evoluzione di stati latenti ed osservabili.

Siano  $x_{i,j}$  l' $i$ -esima caratteristica osservabile al  $j$ -esimo stadio e  $y_{i,j}$  l' $i$ -esima caratteristica latente al  $j$ -esimo stadio, allora un modello state-space ha forma

$$\begin{cases} x_{i,j} = m_j + c_j y_{i,j} + v_{i,j} & v_{i,j} \sim \mathcal{N}(0, \sigma_{v,j}^2) \\ y_{i,j} = d_j y_{i,j-1} + w_{i,j} & w_{i,j} \sim \mathcal{N}(0, \sigma_{w,j}^2) \end{cases}$$

In particolare,  $v_{i,j}$  è l'errore di misura e  $w_{i,j}$  è il rumore latente allo stadio  $j$ . Si assumono inoltre la condizione iniziale  $y_{i,0} \sim \mathcal{N}(m_0, \sigma_y^2)$  e che  $m_i = 0$  per ogni  $i$ .

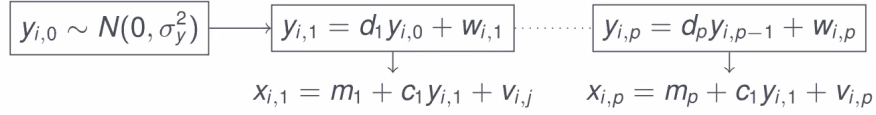


Figura 30: Esempio di modello state-space applicato a un processo multistage.

Ci interessa studiare la possibilità che ad un certo stadio  $\tau$  avvenga un cambiamento, ovvero

$$\begin{cases} x_{i,j} = m_j + c_j y_{i,j} + v_{i,j} \\ y_{i,j} = d_j y_{i,j-1} + \beta \mathbb{1}_{j \geq \tau} + w_{i,j} \end{cases}$$

Il cambiamento si propaga a tutti gli stadi successivi attraverso le costanti moltiplicative  $d_\tau, d_{\tau+1}, \dots, d_p$  per gli stati latenti e  $c_\tau, c_{\tau+1}, \dots, c_p$  per quelli osservabili.

### 26.3 Applicazione

Per poter stimare uno di questi modelli, bisogna essere in grado di calcolare il valore atteso di un'osservazione ad un qualunque stadio  $j$ . In particolare, si ha il seguente modello (cambio di notazione per seguire quella del paper,  $y_{t,i}$  osservabili e  $x_{t,i}$  latenti):

$$\begin{cases} y_{t,i} = \mu_i + c_i x_{t,i} + v_{t,i} \\ x_{t,i} = d_i x_{t,i-1} + \beta_i \mathbb{1}_{i \geq \tau} + w_{t,i} \end{cases}$$

Al tempo  $t$ , il I stage ha equazione:

$$\begin{cases} y_1 - \mu_1 = c_1 x_1 + v_1 \\ x_1 = d_1 x_0 + \beta_1 + w_1 \end{cases} \implies \begin{cases} y_1 - \mu_1 = c_1 d_1 x_0 + c_1 \beta_1 + v_1 \\ x_1 = d_1 x_0 + \beta_1 + w_1 \end{cases}$$

Al tempo  $t$ , il II stage è

$$\begin{cases} y_2 - \mu_2 = c_2 x_2 + v_2 \\ x_2 = d_2 x_1 + \beta_2 + w_2 \end{cases} \implies \begin{cases} y_2 - \mu_2 = c_1 d_1 d_2 x_0 + c_2 d_2 \beta_1 + c_2 d_2 w_1 + \beta_2 c_2 + c_2 w_2 + v_2 \\ x_1 = d_1 d_2 x_0 + d_2 \beta_1 + d_2 w_1 + \beta_2 + w_2 \end{cases}$$

In particolare, si individuano tre componenti  $A, B, C$  (vedi slides) che possono essere manipolati.  $A$  contiene le informazioni sulle costanti di trasferimento e componente di inizializzazione,  $B$  sulle componenti di trasferimento e  $\beta$ ,  $C$  sugli errori.

›  $A$  : si ha un vettore  $\mathbf{f}_1$  che coinvolge lo shift solo nel primo stadio

$$\begin{pmatrix} c_1 d_1 \\ c_2 d_1 d_2 \end{pmatrix} x_0 = d_1 \begin{pmatrix} c_1 \\ c_2 d_2 \end{pmatrix} x_0 = d_1 \mathbf{f}_1 x_0$$

›  $B$  : matrice  $F = (\mathbf{f}_1 \ \mathbf{f}_2)$  di cambiamento solo nel secondo stadio

$$\begin{pmatrix} c_1 \beta_1 \\ c_2 d_2 \beta_1 + c_2 \beta_2 \end{pmatrix} x_0 = d_1 \begin{pmatrix} c_1 & 0 \\ c_2 d_2 & c_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = F \beta$$

›  $C$  : matrice  $F = (\mathbf{f}_1 \ \mathbf{f}_2)$ , cambiamento solo nel secondo stadio

$$\begin{pmatrix} c_1 w_1 \\ c_2 d_2 w_1 + c_2 w_2 \end{pmatrix} x_0 = d_1 \begin{pmatrix} c_1 & 0 \\ c_2 d_2 & c_2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = F \mathbf{w}$$

Dunque, si ha un modello di tipo

$$\mathbf{y}_t - \boldsymbol{\mu} = d_1 \mathbf{f}_1 x_0 + F \boldsymbol{\beta} + F \mathbf{w}_t + \mathbf{v}_t,$$

per cui per un modello generico al tempo  $t$  si ha

$$\mathbf{y}_t - \boldsymbol{\mu} = F \boldsymbol{\beta} + G \mathbf{u}_t + \mathbf{v}_t,$$

da cui

$$\mathbb{E} [\mathbf{y}_t - \boldsymbol{\mu}] = d_1 \mathbf{f}_1 m_0 + F \boldsymbol{\beta},$$

$$\mathbb{V} [\mathbf{y}_t] = G^\top D_{\mathbf{w}} G + D_{\mathbf{v}},$$

dove  $D_{\mathbf{w}} = \text{diag}(\mathbf{w}_t)$  e  $D_{\mathbf{v}} = \text{diag}(\mathbf{v}_t)$ .

Nel caso di un processo ad  $n$  stadi si ha una matrice dei cambiamenti (nel paper è sbagliata)

$$F \boldsymbol{\beta} = \begin{pmatrix} c_1 & 0 & 0 & \dots & 0 \\ c_2 d_2 & c_2 & 0 & \dots & 0 \\ c_3 d_2 d_3 & c_3 d_3 & c_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ c_n \prod_{k=2}^n d_k & c_n \prod_{k=3}^n d_k & c_n \prod_{k=4}^n d_k & \dots & c_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$$

Si sorveglia la stabilità di  $\mathbb{E} [\mathbf{z}_t]$  con una statistica di tipo MEWMA (C. Zou et al., 2007)

## LEZIONE 27: CAPIZZI E MASAROTTO (2011) (III)

2020-12-04

Una volta scritto il modello sotto forma (12), le osservazioni  $\mathbf{y}_t - \boldsymbol{\mu}$  vengono accumulate con la statistica MEWMA

$$\mathbf{z}_t = (1 - \lambda)\mathbf{z}_{t-1} + \lambda(\mathbf{y}_t - \boldsymbol{\mu}).$$

In controllo, si ha che la distribuzione di  $\mathbf{z}_t$  è nota,

$$\begin{aligned}\mathbf{z}_t &= F\boldsymbol{\beta} + \mathbf{a}_t, \\ \mathbf{a}_t &\sim \mathcal{N}_n\left(\mathbf{0}_n, \frac{\lambda}{2-\lambda}\Sigma\right),\end{aligned}$$

e si potrebbe sorvegliare il processo tramite la verifica di ipotesi

$$\begin{cases} H_0 : \boldsymbol{\beta} = \mathbf{0}_p \\ H_1 : \boldsymbol{\beta} \neq \mathbf{0}_p \end{cases}.$$

Tuttavia, i test standard basati su questo approccio hanno una bassissima sensitività quando un piccolo sottoinsieme di variabili è fuori controllo (ipotesi di *sparsità*) e non si hanno indicazioni a priori sulla dimensione del cambiamento.

## 27.1 Selezione delle variabili

Si considera allora un problema diverso, ovvero identificare le colonne di  $F$  che corrispondano ai coefficienti diversi da zero:

1. Si inizializzano  $\beta_1 = \beta_2 = \dots = \beta_p = 0$ .
2. Si utilizza il LAR per aggiungere una variabile alla volta.
3. Si ottiene la soluzione ai minimi quadrati in  $p$  passi, perché ciascuna variabile entra nel modello una sola volta (leggere nell'articolo).

## Verifica di ipotesi tramite LAR

Sia  $j_i$  il predittore selezionato dal LAR all' $i$ -esimo step, allora ad ogni step  $k = 1, \dots, p$ , si considera l'insieme  $J_k = \{\text{"variabili selezionate allo step } k\}\}$ :

$$J_k = \{j_1, j_2, \dots, j_k\} \xrightarrow{\text{coefficienti}} \underbrace{\{\beta_{j_1}, \beta_{j_2}, \dots, \beta_{j_k}\}}_{\text{potenziali var. OC}}.$$

Al  $k$ -esimo step si verifica allora il sistema di ipotesi sulle variabili selezionate,

$$\begin{cases} H_{0,k} : \beta_{j_1} = \beta_{j_2} = \dots = \beta_{j_p} = 0 \\ H_{1,k} : \beta_{j_1} \neq 0, \dots, \beta_{j_k} \neq 0, \beta_{j_{k+1}} = 0, \dots, \beta_{j_p} = 0 \end{cases},$$

dove le componenti poste a 0 sotto  $H_{1,k}$  derivano dal fatto che le variabili non sono state selezionate dal LAR. Sotto il vincolo dato da  $H_{1,k}$ , la stima di  $\boldsymbol{\beta}_{t,k}$  è

$$\hat{\boldsymbol{\beta}}_{t,k} = (F^\top \Sigma^{-1} F)^{-1} F \Sigma^{-1} \mathbf{z}_{t,k},$$

e la statistica basata sul LRT è

$$\begin{aligned} S_{t,k} &= \hat{\mathbf{z}}_{t,k} \Sigma_{\mathbf{Z}}^{-1} \hat{\mathbf{z}}_{t,k} \\ &= \frac{2-\lambda}{\lambda} \hat{\mathbf{z}}_{t,k}^{\top} \Sigma^{-1} \hat{\mathbf{z}}_{t,k} \\ &= \frac{2-\lambda}{\lambda} \hat{\boldsymbol{\beta}}_{t,k}^{\top} F^{\top} \Sigma^{-1} F \hat{\boldsymbol{\beta}}_{t,k}. \end{aligned}$$

Date le  $k = 1, \dots, p$  statistiche test che verificano l'ipotesi che ciascun sottoinsieme di parametri di posizione sia nullo, si ha anche una statistica aggiuntiva per verificare che la matrice di scala al tempo successivo sia definita positiva

$$S_{t,p+1} = \max \left\{ 1, (1-\lambda)S_{t-1,p+1} + \lambda \frac{(\mathbf{y}_t - \boldsymbol{\mu})^{\top} \Sigma^{-1} (\mathbf{y}_t - \boldsymbol{\mu})}{n} \right\}.$$

$S_{t,1}$  verifica il miglior modello selezionando un parametro,  $S_{t,2}$  con due parametri, ...

Usando la solita procedura con Bonferroni, la statistica di controllo LAR-EWMA per verificare il test di ipotesi complessivo è

$$W_t = \max_{k=1, \dots, p+1} \frac{S_{t,k} - \mathbb{E}[S_{t,k}]}{\sqrt{\mathbb{V}[S_{t,k}]}} ,$$

dove valore atteso e varianza in controllo vengono approssimati uguali per ogni  $k$  per semplificare i calcoli. Il limite di controllo  $h$  in questo viene calcolato tramite approssimazioni stocastiche.

### Diagnosi post-segnale

A partire dal test di ipotesi descritto sopra, si possono calcolare le seguenti quantità di interesse:

$$\begin{aligned} \text{Statistica di controllo : } & W_t \\ \text{Tempo di allarme : } & \hat{\tau} = \min \{t : W_t > h\} \\ \text{Analisi post-segnale : } & \hat{k} = \min \left\{ k : \frac{S_{t,k} - \mathbb{E}[S_{t,k}]}{\sqrt{\mathbb{V}[S_{t,k}]}} > h \right\} \end{aligned}$$

### Cose richieste per l'esame

- › Leggere le comparazioni sotto diversi scenari, che vengono riassunti con la statistica

$$\text{RMI} = \frac{1}{N} \sum_{i=1}^N \text{RMI}_i = \frac{1}{N} \frac{\text{ARL}_{\delta_\ell} - \text{mARL}_{\delta_\ell}}{\text{mARL}_{\delta_\ell}},$$

dove  $\text{mARL} = \min \{ \text{ARL}_{\delta_\ell} \text{ di tutte le carte considerate} \}.$

- › Sapere quali sono gli altri competitors, vien fuori che praticamente vince quasi sempre.

MEWMA	Multivariate EWMA
REWMA	Regression-adjusted EWMA
LEWMA	Lasso-based EWMA

- › Vedere l'esempio funzionale delle 14 incisioni con modello quadratico.



## LEZIONE 28: C. ZOU ET AL. (2008)

### «MONITORING PROFILES BASED ON NONPARAMETRIC REGRESSION METHODS»

2020-12-09

#### Motivazione

In generale, la distribuzione esatta degli stimatori dei parametri di una regressione non lineare non è ottenibile e l'applicazione dei metodi parametrici in questo caso porta a risultati non soddisfacenti.

Inoltre, questo tipo di modelli in genere assumono che la forma della distribuzione sia la stessa, mentre i parametri sono le uniche quantità che possono variare.

#### Introduzione

In questo paper si considerano i parametri  $\beta$ , perché si assume di essere nel caso della regressione non parametrica

$$y = g(x) + \varepsilon.$$

Si sorvegliano congiuntamente la funzione di regressione  $g(\cdot)$  e la varianza del termine di errore  $\sigma^2$ .

In Fase I si stima  $g_0(\cdot)$  in modo talmente preciso da poterla utilizzare come benchmark con la quale confrontare i cambiamenti fuori controllo.

Sia  $j$  l'istante di tempo in cui è disponibile il campione e  $x_j^{(i)}$  la *singola* covariata (più covariate è un problema aperto) associata alla curva  $j$ -esima, allora si assume un modello per dati funzionali di tipo

$$y_{ij} = g(x_j^{(i)}, \beta) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_j^2), \quad (13)$$

dove la funzione  $g_0(\cdot)$  può essere qualunque cosa, purché sufficientemente liscia.

Si vuole verificare l'ipotesi di stabilità della funzione di regressione del processo tramite il sistema di ipotesi

$$\begin{cases} H_0 : g = g_0, \sigma = \sigma_0 \\ H_1 : g \neq g_0, \sigma = \sigma_0 \end{cases}$$

Fuori controllo si può stimare  $g(\cdot)$  tramite un lisciatore, ad esempio la regressione lineare locale

$$\hat{g}(x) = \sum_{i=1}^n w_i(x) y_i,$$

per ottenere una statistica test basata sul rapporto di verosimiglianza che diventa

$$\text{GLR} = \frac{1}{\sigma_0^2} [(Y - G_0)^\top (Y - G_0) - (Y - WY)^\top (Y - WY)]$$

che con un po' di pazienza si può riscrivere in una forma simile alla statistica ottenuta da Azzalini e Bowman (1993) per verificare la linearità relazione di regressione.

## 28.1 Test di non linearità di Azzalini e Bowman (1993)

Si vuole verificare la linearità di una relazione di regressione, ovvero dato

$$y_i = g(x_i) + \varepsilon_i,$$

verificare il sistema di ipotesi

$$\begin{cases} H_0 : g(\cdot) \text{ lineare} \\ H_1 : g(\cdot) \text{ non lineare} \end{cases}.$$

Dati i modelli di regressione

$$H_0 : y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \longrightarrow \text{stimatore OLS}$$

$$H_1 : y_i = g(x_i) + \varepsilon_i \longrightarrow \text{stimatore kernel}$$

Si usa una statistica analoga al test  $F$ , ricordando che la regressione lineare locale si può scrivere come  $\hat{\mathbf{y}} = W\mathbf{y}$ .

$$\begin{aligned} F &= \frac{\text{RSS}_0 - \text{RSS}_1}{\text{RSS}_1} \\ &= \frac{(\mathbf{y} - X\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - X\hat{\boldsymbol{\beta}}) - (\mathbf{y} - W\mathbf{y})^\top (\mathbf{y} - W\mathbf{y})}{(\mathbf{y} - W\mathbf{y})^\top (\mathbf{y} - W\mathbf{y})}. \end{aligned}$$

## LEZIONE 29: C. ZOU ET AL. (2008) (II)

2020-12-10

29.1 Sorveglianza di  $g(\cdot)$ 

Sia  $W$  la matrice di *smoothing* dello stimatore kernel, allora la curva stimata diventa

$$\hat{g}(x) = \sum_{i=1}^n W_i(x) y_i = \begin{pmatrix} W_1(x) & W_2(x) & \cdots & W_n(x) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

per cui  $\hat{y} = Wy$ , con

$$W = \begin{pmatrix} W_1(x_1) & W_2(x_1) & \cdots & W_n(x_1) \\ W_1(x_2) & W_2(x_2) & \cdots & W_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ W_1(x_n) & W_2(x_n) & \cdots & W_n(x_n) \end{pmatrix}.$$

La scelta della funzione di liscio non ha un grande impatto sulla procedura, per cui nel paper ne scelgono una (*local linear regression*) e usano quella.

La statistica GLR è

$$\text{GLR} = \frac{1}{\sigma_0^2} [(Y - G_0)^\top (Y - G_0) - (Y - WY)^\top (Y - WY)],$$

che usando i valori standardizzati

$$Z = \frac{1}{\sigma_0^2} (Y - G_0) = \frac{1}{\sigma_0^2} \begin{pmatrix} y_1 - g_0(x_1) \\ y_2 - g_0(x_2) \\ \vdots \\ y_n - g_0(x_n) \end{pmatrix} \stackrel{(13)}{\sim} \mathcal{N}(\mathbf{0}_n, I_n),$$

si può scrivere come

$$\text{LR} = Z^\top Z - (Z - WZ)^\top (Z - WZ).$$

Sotto  $H_0$ , la statistica converge in distribuzione a una distribuzione  $\mathcal{N}(\mu_z, \sigma_z^2)$ , per dei particolari valori di  $\mu_z$  e  $\sigma_z^2$  calcolati in appendice.

Questa statistica si può utilizzare in uno schema MEWMA per monitorare congiuntamente la stabilità di  $g(\cdot)$  e di  $\sigma^2$ .

Serve un test non parametrico anche per sorvegliare  $\sigma_j^2$ , in questo caso si utilizza la statistica

$$n\hat{\sigma}_j^2 = (Z - WZ)^\top (Z - WZ).$$

Come al solito, si può utilizzare la trasformazione di Quesenberry per ottenere una distribuzione normale

$$\tilde{\sigma}_j = \Phi^{-1}(\psi(n\hat{\sigma}_j^2)),$$

indicando con  $\Phi^{-1}(\cdot)$  la funzione quantile della normale standard e con  $\psi(\cdot)$  la funzione di

ripartizione di  $n\hat{\sigma}_j^2$  quando il processo è in controllo. Utilizzando l'assunzione di normalità, anche se  $n\hat{\sigma}_j^2$  è stimato in modo non parametrico, è una forma quadratica con distribuzione calcolabile come combinazione lineare di  $\chi_1^2$ .

Il calcolo esatto di questa distribuzione è dispendioso per un update online e si utilizza invece un metodo che fa il match dei primi tre momenti della distribuzione  $\chi^2$ .

Per il controllo, si utilizza una carta EWMA

$$E_j = \lambda U_j + (1 - \lambda)E_{j-1}$$

sulle quantità

$$U_j = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \\ \tilde{\sigma}_j \end{pmatrix},$$

che definendo  $\Sigma = \begin{pmatrix} V & 0 \\ 0 & 1 \end{pmatrix}$ , si segnala l'allarme se (attenzione che moltiplica per  $\Sigma$  e non  $\Sigma^{-1}$ )

$$Q_j = E_j^\top \Sigma E_j > L \frac{\lambda}{2 - \lambda}.$$

Procedura che attribuisce il cambiamento alla varianza oppure alla funzione  $g(\cdot)$  solo da leggere, perché è un po' complicato.

## LEZIONE 30: ABDELLA ET AL. (2019)

### «AN ADAPTIVE THRESHOLDING-BASED PROCESS VARIABILITY MONITORING»

2021-01-14

#### a) Obiettivi e introduzione

Si introduce una carta Shewhart di Fase II multivariata per sorvegliare la *stabilità* della matrice di varianze e covarianze sotto l'ipotesi di *sparsità* del cambiamento.

#### b) Stato dell'arte

In letteratura, la carta PLR (*Penalized Likelihood Ratio*) si basa sul rapporto di verosimiglianza penalizzato attraverso un parametro  $\lambda$ , che però ha una performance molto sensibile alla scelta di quest'ultimo.

Sotto  $H_0$  si ipotizza  $\mathbf{y}_t \sim \mathcal{N}(\boldsymbol{\mu}_0, I_p)$ , per cui le unità vengono premoltiplicate per  $\Sigma_0^{-1/2}$ .

#### c) Metodologia proposta

Assumendo un processo generatore dei dati

$$\mathbf{y}_t \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0) & \text{se } t = 1, \dots, \tau \\ \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_{OC}) & \text{se } t = \tau + 1, \tau + 2, \dots \end{cases},$$

si sorveglia la stabilità tramite un sistema di ipotesi

$$\begin{cases} H_0 : \Sigma - \Sigma_0 = 0 \\ H_1 : \Sigma - \Sigma_0 \neq 0 \end{cases}.$$

Si stima la matrice di varianza tramite  $S_t = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_{ti} - \bar{\mathbf{y}}_t)(\mathbf{y}_{ti} - \bar{\mathbf{y}}_t)^\top$  e si studia la versione *shrunked* della matrice delle differenze:

$$D_t = (d_{ijt})_{i,j} = S_t - \Sigma_0$$

$$\hat{D}_t = (\hat{d}_{ijt})_{i,j} = \left( d_{ijt} \cdot \max \{0, 1 - (\gamma_{ij}/d_{ijt})^\eta\} \right)_{i,j},$$

con  $\gamma_{ij}^* = \varphi \sqrt{\vartheta_{ij} n^{-1} \log p}$  e  $\vartheta_{ij} = \sigma_{ii}\sigma_{jj} + \sigma_{ij}^2$  calcolato in controllo.

Dopo aver standardizzato  $\hat{D}$  tramite

$$\hat{D}^* = (\hat{d}_{ijt}^*)_{i,j} = \left( \frac{d_{ijt}}{\sqrt{n\vartheta_{ij}}} \right)_{i,j},$$

si utilizza come statistica di controllo la [norma di Frobenius](#)

$$M_t = \|\hat{D}^*\|_F^2 = \sum_{i=1}^p \sum_{j=1}^n (\hat{d}_{ijt}^*)^2.$$

#### d) Confronto con metodi alternativi

Confronto con la carta PLR avendo fissato per entrambe  $ARL_0 = 200$  e valutando la performance tramite  $ARL_1$ .

Come scenari fuori controllo, si è valutato

1. Cambiamento diagonale  $\Sigma_{OC} = \begin{pmatrix} 1+\delta & 0 & \dots & 0 \\ 0 & 1 & \dots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$
2. Cambiamento fuori diagonale  $\Sigma_{OC} = \begin{pmatrix} 1 & \delta & \delta & \delta & \delta \\ \delta & 1 & \delta & \delta & \delta \\ \delta & \delta & 1 & \delta & \delta \\ \delta & \delta & \delta & 1 & \delta \\ \delta & \delta & \delta & \delta & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$
3. Cambiamento dentro e fuori diagonale  $\Sigma_{OC} = \begin{pmatrix} 1+\delta & \delta & & & 0 \\ \delta & 1+\delta & 1 & & \\ & & \ddots & & \\ 0 & & & 1 & \\ & & & & 1+\delta & \delta \\ & & & & \delta & 1+\delta \end{pmatrix}$

$\Sigma_{OC}$	$\delta$	PLR			ALT-norm $\varphi=1.00$
		$\lambda=0.05$	$\lambda=1.00$	$\lambda=1.50$	
$\Sigma_{OC_2}$	0.2	165.17	<b>123.25</b>	128.20	135.01
	0.4	114.54	70.04	77.98	<b>64.55</b>
	0.6	84.69	39.34	47.77	<b>25.01</b>
	0.8	59.64	21.50	29.26	<b>10.98</b>
	1.0	38.69	12.93	17.24	<b>5.87</b>
$\Sigma_{OC_3}$	0.2	89.06	179.10	183.48	<b>66.85</b>
	0.4	17.31	119.17	124.59	<b>6.36</b>
	0.6	3.78	49.39	86.18	<b>1.96</b>
	0.8	1.27	14.77	51.07	<b>1.25</b>
	1.0	1.00	5.22	28.38	<b>1.00</b>
$\Sigma_{OC_7}$	0.2	75.83	<b>38.39</b>	39.56	43.9
	0.4	23.39	8.64	10.29	<b>6.71</b>
	0.6	7.97	3.13	3.83	<b>2.20</b>
	0.8	3.68	1.76	2.05	<b>1.35</b>
	1.0	2.20	1.27	1.45	<b>1.10</b>

Figura 31: Confronto della performance per le carte PLR e ALT-norm. Le deviazioni standard (non riportate) sono dello stesso ordine di grandezza delle stime ottenute, il che è un problema.

In generale, la metodologia ALT-norm è migliore per quasi tutti i pattern, tuttavia la PLR non è stata disegnata per matrici in controllo diverse dall'identità.

#### e) Disegno della carta

La scelta dei parametri comporta la stima di  $\varphi$  e  $\eta$ , che sono legati alla sparsità di  $D_t$ . Entrambi i parametri contribuiscono a individuare shift di diverse dimensioni:

›  $\varphi$  grande  $\implies$  shift grandi.

- ›  $\eta$  grande  $\implies$  shift piccoli.

**f) Vantaggi, criticità, domande aperte**

**Pro**

- › È possibile effettuare un'analisi post-segnale stimando una soluzione sparsa che massimizzi la norma di Frobenius  $\|\hat{D}_k\|_q \leq s_0$ .
- › No limitazioni sulla struttura di  $\Sigma_0$ .
- › Bassa sensibilità rispetto agli iperparametri  $\eta$  e  $\varphi$  se considerati separatamente.
- › Semplice analisi post-segnale.

**Contro**

- › Calcolo dei limiti di controllo non descritti bene.
- › Alta variabilità dell'ARL<sub>1</sub>
- › Scenari OC considerati poco realistici.
- › Sensibilità rispetto a  $(\eta, \varphi)$  non studiata congiuntamente.
- › Sensibilità rispetto a  $s_0$  non studiata.

## LEZIONE 31: ZHANG E WOODALL (2015)

### «DYNAMIC PROBABILITY CONTROL LIMITS FOR RISK-ADJUSTED CUSUM CHARTS BASED ON MULTIRESPONSES»

2021-01-15

#### a) Obiettivi e introduzione

Carta di controllo di Fase II per sorvegliare la stabilità della probabilità di una distribuzione multinomiale.

Si vuole generalizzare il metodo di calcolo dei limiti di controllo dinamici di Steiner et al. (2000) (chirurghi), applicato alla sorveglianza di più condizioni post-operatorie dei pazienti.

#### b) Stato dell'arte

In letteratura Tang et al. (2015) utilizzano una regressione multilogit a odds proporzionali che considera più di due esiti operatori, tuttavia utilizza limiti di controllo *costanti*. Le problematiche di questo metodo stanno nel fatto che la performance risulta instabile se il rischio varia per ciascun paziente  $t = 1, 2, \dots$

Si parte dal paper di Tang et al. (2015), con un modello multinomiale

$$\text{logit} \{P(Y \leq k | S = s)\} = \alpha_k + \beta_s, \quad k = 0, \dots, J-1,$$

per un paziente con Parsonnet score  $S = s$ . Fuori controllo, la probabilità di un esito  $k \in \{0, 1, \dots, J\}$  è proporzionale rispetto ad un valore  $R$ , che si assume essere uguale per ogni categoria:

$$\pi_k^*(s) = R \frac{\sum_{i=0}^k \pi_i(s)}{1 - \sum_{i=0}^k \pi_i(s)}, \quad k = 0, 1, \dots, J.$$

Al tempo  $t$  si calcola il peso per aggiornare la carta CUSUM, ottenuto dall'esito osservato  $Y_t$  per il paziente  $t$ ,

$$W_t = \log \frac{\pi_{Y_t}^*(s)}{\pi_{Y_t}(s)},$$

per poi scrivere

$$C_t = \max \{0, C_{t-1} + W_t\}, \quad t = 1, 2, 3, \dots$$

#### c) Metodologia proposta

Gli autori propongono *limiti di controllo dinamici*  $h(\alpha)$  calcolati per soddisfare approssimativamente la probabilità di falso allarme condizionato:

$$h(\alpha) : \begin{cases} P(C_1 > h_1 | S_1 = s_1) = \alpha \\ P(C_t > h_t | S_t = s_t, \overbrace{C_1 < h_1, \dots, C_{t-1} < h_{t-1}}^{\text{non ancora segnalato}}) = \alpha \end{cases} \quad \text{per } t = 2, 3, \dots$$

Con i limiti di controllo dinamici, si ha  $RL_0 \sim \text{Geom}(\alpha) \implies ARL_0 \approx \frac{1}{\alpha}$ .

La procedura è analoga a quella di Steiner et al. (2000):



1. Genere  $N$  variabili categoriali

$$Y_{t,i} \sim \text{Categorical}(\pi_0(s_t), \pi_1(s_t), \dots, \pi_J(s_t)).$$

2. Calcolo  $N$  statistiche CUSUM

$$C_{t,i} = \max \{0, C_{t-1,l} + W_{t,i}\} \quad l = 1, 2, \dots, N,$$

dove  $C_{t-1,l}$  è stata ricampionata casualmente dall'insieme di statistiche CUSUM che non hanno segnalato l'allarme:

$$\{C_{t-1,i} : C_{t-1,i} \leq h_{t-1}(\alpha)\}.$$

3. Il limite di controllo è  $h_t(\alpha) = C_{t,(\lceil N(1-\alpha) \rceil)}$ .

#### d) Confronto con metodi alternativi

La procedura viene confrontata con la CUSUM proposta da Tang et al. (2015), che utilizza limiti costanti nel tempo.

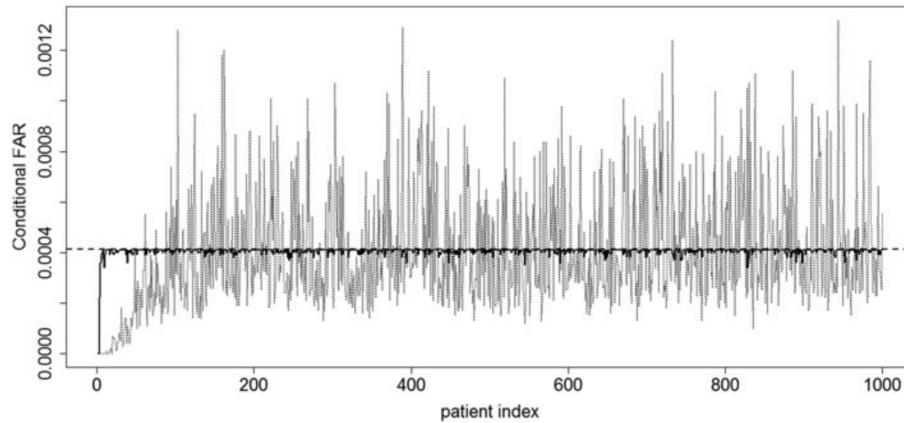


Figura 32: *False Alarm Rate* (FAR) condizionata per la carta con limiti dinamici (*nero*) e limiti costanti (*grigio*). Con limiti costanti, la FAR condizionata è molto più instabile al variare dei pazienti.

#### f) Vantaggi, criticità, domande aperte

Pro	Contro
<ul style="list-style-type: none"> <li>› Immediata generalizzazione di Steiner et al. (2000).</li> <li>› FAR condizionata raggiunge l'<math>\alpha</math> specificato molto velocemente e in modo stabile nel tempo.</li> </ul>	<ul style="list-style-type: none"> <li>› Computazionalmente costosa.</li> <li>› Non si discute la performance rispetto agli iperparametri <math>(\alpha, R, N)</math>.</li> <li>› Non viene studiata la performance fuori controllo, anche se è meno rilevante di una IC corretta.</li> <li>› Inerzia della CUSUM.</li> </ul>

## LEZIONE 32: HAWKINS E MABOUDOU-TCHAO (2008) «MULTIVARIATE EXPONENTIALLY WEIGHTED MOVING COVARIANCE MATRIX»

2021-01-15

### a) Obiettivi e introduzione

Introduzione di una carta di controllo EWMA multivariata di Fase II per sorvegliare la stabilità della matrice di varianze e covarianze. In particolare, questo metodo può essere applicato senza necessità di osservare sottogruppi razionali.

### b) Stato dell'arte

In letteratura sono presenti tre metodi di controllo della matrice di varianze e covarianze:

1. Montgomery e Wadsworth (1972), carta Shewhart sulla *varianza generalizzata*  $|S| = \prod_{j=1}^p \lambda_j$ .
  - Non può identificare aumenti e diminuzioni contemporaneamente.
2. Reynolds e Cho (2006), carta Shewhart sulla *varianza totale*  $\text{tr } S = \sum_{j=1}^p \lambda_j$ .
  - + Identifica aumenti di variabilità.
  - Non identifica diminuzioni di variabilità.
3. Alt (2004), statistica GLR

$$W_i = -(m-1) \left( p + \log \frac{|S_i|}{|\Sigma_0|} - \text{tr } \Sigma_0^{-1} S_i \right) \sim \chi_{p(p+1)/2}^2.$$

- + Identifica deviazioni in qualunque direzione.
- Inefficiente per  $m$  piccolo.

### c) Metodologia proposta

1. Si ortogonalizza usando l'inverso della decomposizione di Cholesky  $\Sigma_0 = A^{-1}(A^{-1})^\top$ :

$$U_i = A(X_i - \mu_0) \stackrel{\text{IC}}{\sim} \mathcal{N}(\mathbf{0}, I_p).$$

2. La stima al tempo  $t$  della matrice è

$$S_n = (1 - \lambda)S_{n-1} + \lambda U_n U_n^\top.$$

3. La statistica di controllo è la GLR di Alt (2004):

$$c_n = \text{tr } S_n - \log |S_n| - p,$$

che suona se si supera un limite di controllo costante  $h$  ottenuto tramite simulazione.

### d) Confronto con metodi alternativi

Si confronta questa carta con la Shewhart di Alt (2004), valutando l'ARL fuori controllo di entrambe, in tre scenari:

1. Cambiamenti nella varianza.
2. Cambiamenti nella covarianza.
3. Cambiamenti nella varianza e covarianza.

Poiché quella di Alt non si può utilizzare per  $m = 1$ , si sceglie  $m = 10$  e si stimano i parametri in modo da avere stessa ARL in base al numero di confronti.

Si osserva un comportamento generalmente migliore rispetto al competitor, anche se per piccoli decrementi della varianza si ha una ARL *distorta*, ovvero

$$ARL_1 > ARL_0.$$

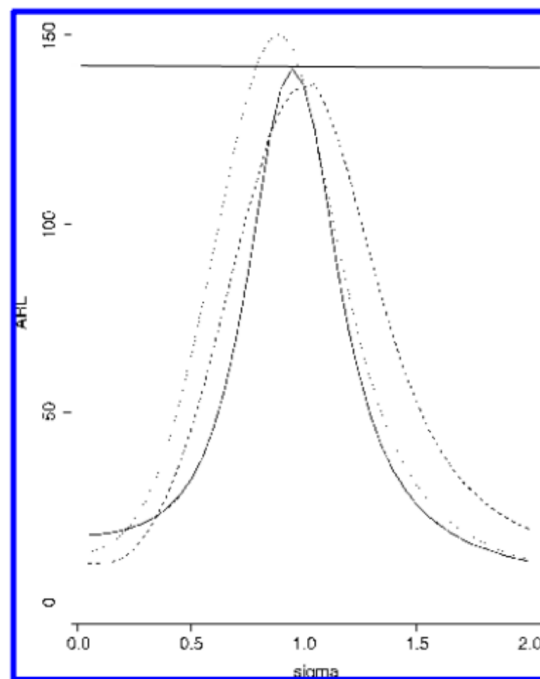


Figura 33: La carta si comporta in generale meglio di quella proposta da Alt (2004). Per piccoli decrementi di varianza e  $\lambda$  grande si può osservare un'ARL distorta.

#### e) Ulteriori aspetti positivi della metodologia

È possibile combinare la carta per monitorare anche cambiamenti nella media, in modo da avere una diagnostica post-segnale robusta rispetto a segnali spuri.

Si può utilizzare un approccio *regression-adjusted* per identificare il parametro fuori controllo nella procedura post-segnale, effettuando una regressione del  $j$ -esimo elemento di  $U$  sulle precedenti  $j - 1$  componenti:

$$U_j = \gamma_{0j} + \sum_{k=1}^{j-1} \gamma_{jk} U_k + e_j.$$

Quando il processo è IC, tutti i coefficienti sono nulli e  $\mathbb{V}[e_j] = 1$ , dunque ci sono tre test di verifica di ipotesi che permettono di studiare:

1.  $H_0 : \gamma_{0j} = 0$  stabilità della media.
2.  $H_0 : \gamma_{kj} = 0$  stabilità relazione tra regressori.
3.  $H_0 : \sigma_{e_j}^2 = 1$  stabilità della varianza.

**f) Vantaggi, criticità, domande aperte**

<b>Pro</b>	<b>Contro</b>
› Rilevazione sia aumenti sia diminuzioni di variabilità.	› Assunzione parametrica per la distribuzione.
› Carta combinata con analisi post-segnale per risolvere il problema dei segnali spuri.	› Non ci sono competitor naturali.
› Applicabile anche per $m = 1$ .	› ARL distorta per piccole diminuzioni di variabilità.
	› Metodologia che dipende dalla decomposizione di Cholesky e quindi dall'ordinamento delle variabili per l'analisi post-segnale.

## A APPROFONDIMENTO SULLE DISTRIBUZIONI MULTIVARIATE

### A.1 Distribuzioni ellittiche

**Def. (Distribuzione ellittica)**

Si dice che  $X$  ha distribuzione ellittica, e si indica con  $X \sim E_n(\mu, \Sigma, \psi)$ , se la sua funzione caratteristica  $\psi$  soddisfa l'equazione funzionale

$$\psi(t - \mu) = \psi(t^\top \Sigma t).$$

**Osservazioni**

- › Le distribuzioni ellittiche sono una generalizzazione della distribuzione normale multivariata.
- › Se  $X$  ha funzione di densità, ha forma

$$f_X(x) = \frac{c_n}{\sqrt{|\Sigma|}} g_n((x - \mu)^\top \Sigma^{-1} (x - \mu)),$$

con  $g(\cdot)$  chiamata *generatore di densità*. In tal caso, a volte si può indicare con  $X \sim E_n(\mu, \Sigma, g_n)$ .

- › Se  $X \sim E_n(\mu, \Sigma, \psi)$  e il suo valore atteso e matrice di covarianza esistono, questi saranno

$$\mathbb{E}[X] = \mu$$

$$\text{Cov}(X) = -\frac{\partial \psi(0)}{\partial t} \cdot \Sigma$$

- › Se  $X \sim E_n(\mu, \Sigma, \psi)$ , allora

$$AX + b \sim E_q(A\mu + b, A\Sigma A^\top, g_q)$$

**Corollario (Trasformazioni lineari)**

1. Ogni distribuzione marginale di una v.c. ellittica è una distribuzione ellittica.
2. Ogni combinazione lineare di distribuzioni ellittiche è una distribuzione ellittica.

- › Se  $X = (X_1, X_2, \dots, X_n)^\top \sim E_n(\mu, \Sigma, g_n)$ , allora

$$\sum_{i=1}^n X_i \sim E_1(\mathbb{1}_n^\top \mu, \mathbb{1}_n^\top \Sigma \mathbb{1}_n, g_1).$$

- › Ci sono distribuzioni ellittiche *skewed*, ad esempio la *normale multivariata asimmetrica*

$$f(x) = 2\varphi(x) \cdot \Phi(\lambda \cdot x),$$

con  $\lambda$  parametro di simmetria. È una distribuzione asimmetrica positivamente per  $\lambda > 0$  e negativamente per  $\lambda < 0$ .

Tabella 1: Esempi di generatori di densità

Distribuzione	$g(u)$
Cauchy	$(1 - u)^{-(n+1)/2}$
Exponential power	$\exp\{-r \cdot u^s\}$
Laplace	$\exp\{- u \}$
Logistica	$\frac{\exp\{- u \}}{(1 + \exp\{-u\})^2}$
Normale	$\exp\{-u/2\}$
$t$ Student	$(1 + \frac{u}{m})^{-(n+m)/2}, m \in \mathbb{N}$

## A.2 Distribuzioni stabili

### Def. (Distribuzione $\alpha$ -stabile)

Se  $X_1, \dots, X_n$  e  $X$  sono variabili casuali i.i.d, allora si dice che  $X$  segue una distribuzione  $\alpha$ -stabile se esistono due costanti  $d_n$  e  $c_n > 0$  tali che

$$X_1 + X_2 + \dots + X_n \stackrel{d}{=} c_n X + d_n.$$

### Osservazioni

› Le distribuzioni  $\alpha$ -stabili sono definite dalla funzione caratteristica

$$\psi_{\text{stable}}(t; \alpha, \beta, \sigma, \mu) = \begin{cases} \exp\{i\mu t - |\sigma t|^\alpha (1 - i\beta \operatorname{sgn}(t) \tan \frac{\pi\alpha}{2})\} & \text{se } \alpha \neq 1 \\ \exp\{i\mu t - \sigma|t|(1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \log |t|)\} & \text{se } \alpha = 1 \end{cases}$$

dove i parametri sono

$\alpha \in (0, 2]$	indice di stabilità
$\beta \in [-1, 1]$	indice di asimmetria
$\sigma \in (0, +\infty)$	parametro di scala
$\mu \in (-\infty, +\infty)$	parametro di posizione

› In generale hanno densità non in forma chiusa, eccetto per tre casi particolari

1.  $\alpha = 2$ , distribuzione gaussiana

$$f(x) = \frac{1}{2\sigma\sqrt{\pi}} e^{-\frac{(x-\mu)^2}{4\sigma^2}}.$$

2.  $\alpha = 1, \beta = 0$ , distribuzione di Cauchy

$$f(x) = \frac{\sigma}{\pi((x-\mu)^2 + \sigma^2)}.$$

3.  $\alpha = 1/2, \beta = \pm 1$ , distribuzione di Lévy

$$f(x) = \frac{\sqrt{\sigma}}{\sqrt{2\pi}(x-\mu)^{3/2}} e^{-\frac{\sigma}{2(x-\mu)}}.$$

› I momenti sono

$$\mathbb{E}|X|^p = \begin{cases} k_p < \infty & \text{se } 0 < p < \alpha \\ \infty & \text{se } p \geq \alpha \end{cases}; \quad \mathbb{E}[X] = \begin{cases} \mu & \text{se } \alpha > 1 \\ \infty & \text{se } \alpha \leq 1 \end{cases}$$

› Le code di una distribuzione  $\alpha$ -stabile decadono come una funzione di potenza

$$P(|X| > x) \propto C \cdot x^{-\alpha}, \quad x \rightarrow \infty.$$

### A.3 Data Depth

*Riferimenti* Bell et al., 2014

Il *data depth* misura la centralità di un'osservazione  $\mathbf{x}$  rispetto a una nube di punti osservata  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^p$ .

Estensione naturale delle statistiche d'ordine a dimensioni più elevate, ordinando i dati rispetto al centro della distribuzione.

#### Def. (Mahalanobis Depth)

La *Mahalanobis Depth* (MD) di  $\mathbf{x}$  rispetto a una funzione di distribuzione  $p$ -dimensionale  $F$  è definita come

$$\begin{aligned} \text{MD}(\mathbf{x}; F) &= (1 + d_{\Sigma}^2(\mathbf{x} - \boldsymbol{\mu}))^{-1} \\ &= (1 + (\mathbf{x} - \boldsymbol{\mu})^{\top} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}))^{-1}. \end{aligned}$$

#### Osservazioni

- › Valori piccoli di MD corrispondono a osservazioni anomale (elevata distanza).
- › MD non è una metrica robusta ed è appropriata quando  $F$  è ellittica. Contorni ellittici corrispondono a depth uguali.

#### Esempio (Carta $T^2$ basata sui ranghi del data depth)

Si costruisce il pooled array dato dalle osservazioni  $p$ -dimensionali

$$\mathbf{Y} = (\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n}, \dots, \mathbf{x}_{m,1}, \dots, \mathbf{x}_{m,n})$$

e si ottiene per ciascun vettore il data depth  $\text{MD}(\mathbf{x}_{i,j}; \hat{F}_N)$  rispetto alla funzione di ripartizione empirica  $\hat{F}_N$ .

Successivamente, si ordinano i valori in ordine decrescente di MD e si assegnano i ranghi ai valori così ordinati. Di questi ranghi, si calcola la media per sottogruppo razionale e si riporta su un grafico il loro valore standardizzato.

## A.4 Spatial signs

*Riferimenti* Cheng e Shiau, 2014

Dal momento che  $\mathbb{R}^p$  non può essere **totalmente ordinato**, per  $p > 1$ , è necessario introdurre un'estensione della funzione segno univariata

$$U(x) = \begin{cases} |x|^{-1}x & \text{se } x \neq 0 \\ 0 & \text{se } x = 0 \end{cases},$$

che indichi la *direzione* di un'osservazione rispetto all'origine:

$$U(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|^{-1}\mathbf{x} & \text{se } \mathbf{x} \neq \mathbf{0} \\ 0 & \text{se } \mathbf{x} = \mathbf{0} \end{cases},$$

con  $\|\cdot\|$  norma euclidea. Per  $A$  matrice di trasformazione, il vettore degli *spatial signs* è

$$\mathbf{u} = (u_i)_{i=1,\dots,m} = (U(A\mathbf{x}_i))_{i=1,\dots,m}.$$

Scelta una particolare matrice  $A^*$  e posto per questa

$$u_{i,j} = U(A^*(\mathbf{x}_i - \hat{\vartheta})), \quad i = 1, \dots, m, j = 1, \dots, n,$$

allora la statistica  $Q_i = np\mathbf{u}_i^\top \mathbf{u}_i$  è distribution-free rispetto alla famiglia di distribuzioni ellittiche.

### Esempio (Statistica di Hotelling basata sugli spatial signs)

Si utilizza la statistica

$$Q = n\mathbf{u}^\top \left( \frac{1}{n} \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i^\top \right)^{-1} \mathbf{u}.$$



## RIFERIMENTI BIBLIOGRAFICI

- Abdella, G. M. et al. (2019). «An Adaptive Thresholding-Based Process Variability Monitoring». In: *Journal of Quality Technology* 51.3, pp. 242–256.
- Alt, F. B. (2004). «Multivariate Quality Control». In: *Encyclopedia of Statistical Sciences*. American Cancer Society.
- Azzalini, A. e Bowman, A. (1993). «On the Use of Nonparametric Regression for Checking Linear Relationships». In: *Journal of the Royal Statistical Society: Series B (Methodological)* 55.2, pp. 549–557.
- Bell, R. C. et al. (2014). «A Distribution-Free Multivariate Phase I Location Control Chart for Subgrouped Data from Elliptical Distributions». In: *Technometrics* 56.4, pp. 528–538.
- Capizzi, G. e Masarotto, G. (2003). «An Adaptive Exponentially Weighted Moving Average Control Chart». In: *Technometrics* 45.3, pp. 199–207.
- Capizzi, G. e Masarotto, G. (2011). «A Least Angle Regression Control Chart for Multidimensional Data». In: *Technometrics* 53.3, pp. 285–296.
- Capizzi, G. e Masarotto, G. (2013). «Phase I Distribution-Free Analysis of Univariate Data». In: *Journal of Quality Technology* 45, pp. 273–284.
- Capizzi, G. e Masarotto, G. (2016). «Efficient Control Chart Calibration by Simulated Stochastic Approximation». In: *IIE Transactions* 48.1, pp. 57–65.
- Capizzi, G. e Masarotto, G. (2017). «Phase I Distribution-Free Analysis of Multivariate Data». In: *Technometrics* 59.4, pp. 484–495.
- Cheng, C.-R. e Shiau, J.-J. H. (2014). «A Distribution-Free Multivariate Control Chart for Phase I Applications». In: *Quality and Reliability Engineering International* 31.1, pp. 97–111.
- Crosier, R. B. (1988). «Multivariate Generalizations of Cumulative Sum Quality-Control Schemes». In: *Technometrics* 30.3, pp. 291–303.
- Ernst, M. D. (2004). «Permutation Methods: A Basis for Exact Inference». In: *Statistical Science* 19.4, pp. 676–685.
- Hawkins, D. M. e Maboudou-Tchao, E. M. (2008). «Multivariate Exponentially Weighted Moving Covariance Matrix». In: *Technometrics* 50.2, pp. 155–166.
- Healy, J. D. (1987). «A Note on Multivariate CUSUM Procedures». In: *Technometrics* 29.4, pp. 409–412.
- Jones-Farmer, L. A. e Champ, C. W. (2010). «A Distribution-Free Phase I Control Chart for Subgroup Scale». In: *Journal of Quality Technology* 42.4, pp. 373–387.
- Jones-Farmer, L. A., Jordan, V. et al. (2009). «Distribution-Free Phase I Control Charts for Subgroup Location». In: *Journal of Quality Technology* 41.3, pp. 304–316.

- Jr, M. R. R. e Lou, J. (2010). «An Evaluation of a GLR Control Chart for Monitoring the Process Mean». In: *Journal of Quality Technology* 42.3, pp. 287–310.
- Kreiss, J.-P. e Paparoditis, E. (2011). «Bootstrap methods for dependent data: A review». In: *Journal of the Korean Statistical Society* 40.4, pp. 357–378.
- Lowry, C. A. et al. (1992). «A Multivariate Exponentially Weighted Moving Average Control Chart». In: *Technometrics* 34.1, pp. 46–53.
- Montgomery, D. e Wadsworth, H. (1972). «Some Techniques for Multivariate Quality Control Applications». In: Transactions of the ASQC. Washington, DC.
- Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Lecture Notes in Statistics. New York: Springer-Verlag.
- Page, E. S. (1954). «Continuous Inspection Schemes». In: *Biometrika* 41.1/2, p. 100.
- Qiu, P. (2014). *Introduction to Statistical Process Control*. CRC Press.
- Reynolds, M. R. J. e Cho, G.-Y. (2006). «Multivariate Control Charts for Monitoring the Mean Vector and Covariance Matrix with Variable Sampling Intervals». In: *Sequential Analysis* 30.1, pp. 1–40.
- Shen, X. et al. (2016). «Self-Starting Monitoring Scheme for Poisson Count Data With Varying Population Sizes». In: *Technometrics* 58.4, pp. 460–471.
- Steiner, S. H. et al. (2000). «Monitoring surgical performance using risk-adjusted cumulative sum charts». In: *Biostatistics* 1.4, pp. 441–452.
- Tang, X. et al. (2015). «Risk-Adjusted Cumulative Sum Charting Procedure Based on Multiresponses». In: *Journal of the American Statistical Association* 110.509, pp. 16–26.
- Zhang, X. e Woodall, W. H. (2015). «Dynamic probability control limits for risk-adjusted CUSUM charts based on multiresponses». In: *Statistics in Medicine* 34.25, pp. 3336–3348.
- Zou, C. et al. (2007). «Monitoring General Linear Profiles Using Multivariate Exponentially Weighted Moving Average Schemes». In: *Technometrics* 49.4, pp. 395–408.
- Zou, C. et al. (2008). «Monitoring Profiles Based on Nonparametric Regression Methods». In: *Technometrics* 50.4, pp. 512–526.
- Zou, H. e Hastie, T. (2005). «Regularization and variable selection via the elastic net». In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.