



**Appunti universitari**

**Tesi di laurea**

**Cartoleria e cancelleria**

**Stampa file e fotocopie**

**Print on demand**

**Rilegature**

**NUMERO: 2197A**

**ANNO: 2017**

# **A P P U N T I**

**STUDENTE: Bellone Marco**

**MATERIA: Simulazione dei sistemi gestionali - Prof. Alfieri**

Il presente lavoro nasce dall'impegno dell'autore ed è distribuito in accordo con il Centro Appunti.

Tutti i diritti sono riservati. È vietata qualsiasi riproduzione, copia totale o parziale, dei contenuti inseriti nel presente volume, ivi inclusa la memorizzazione, rielaborazione, diffusione o distribuzione dei contenuti stessi mediante qualunque supporto magnetico o cartaceo, piattaforma tecnologica o rete telematica, senza previa autorizzazione scritta dell'autore.

**ATTENZIONE: QUESTI APPUNTI SONO FATTI DA STUDENTIE NON SONO STATI VISIONATI DAL DOCENTE.  
IL NOME DEL PROFESSORE, SERVE SOLO PER IDENTIFICARE IL CORSO.**

# SIMULAZIONE DEI SISTEMI GESTIONALI

per noi e non può aiutarci se non abbiamo un modello di sistema.

I VANTAGGI nel suo uso sono

- sceglie correttamente
- tempo di compressione ed espansione
- capisce "Perché?"
- esplora possibili alternative ed alternative
- diagnostica problemi
- identifica i vincoli
- sviluppa la comprensione

Gli SVANTAGGI derivano dal fatto che

- non fornisce una risposta esatta, ma solo una stima / approssimazione
- la costruzione del modello richiede una formazione particolare
- i risultati della simulazione possono essere difficili da interpretare
- può richiedere molto tempo e denaro

Tipi di modelli di simulazione:

### STATICI

VS

### DINAMICI

Rappresentano un sistema ad un certo istante di tempo o un sistema che non dipende dal tempo

[Monte Carlo]

Rappresentano il sistema durante la sua evoluzione temporale

### DETERMINISTICI

VS

### STOCASTICI

Rappresentano sistemi caratterizzati dalla mancanza di casualità

- esperimenti con gli stessi dati in input danno gli stessi risultati
- non c'è variabilità

Rappresentano sistemi casuali

- Valori in input casuali danno risultati casuali
- c'è variabilità

### CONTINUI

VS

### DISCRETI

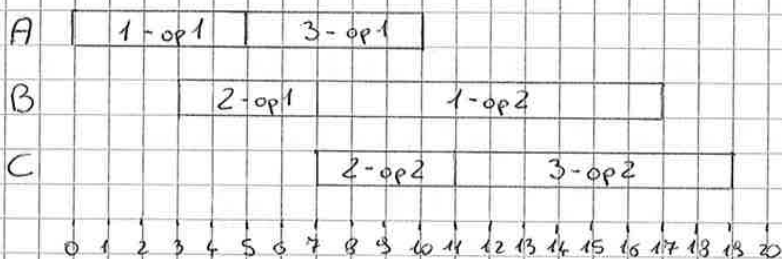
Rappresentano sistemi che "evolvono" in qualsiasi tempo

Rappresentano sistemi che evolvono a certi istanti di tempo



Bisogna tenere una LISTA DEGLI EVENTI ordinato

Evento	Tempo
Entra la parte 1 e viene processata	0
Entra la parte 2	3
Entra la parte 3	4
Fine del processamento della parte 1	5
Fine del processamento della parte 2	7
Fine del processamento della parte 3	10



Gli elementi di una simulazione sono:

- COMPONENTI DEL MODELLO che comprendono
  - Entità: qualcosa che fluisce nel sistema e ne cambia lo stato
  - Risorse: per processare o ricevere le entità
  - Code: dove le entità aspettano se la risorsa è occupata
- ATTIVITÀ
- EVENTI di cui si tiene traccia tramite una lista che annota le varie cose che si verificano durante la simulazione. Questo è controllato dall'avanzamento del "simulation clock" (= variabile che tiene conto del valore corrente del tempo)



verificarsi di un evento

- $Q(t)$ : numero di parti in coda
- $B(t)$ : utilizzo della macchina, può essere solo 0 o 1

### 1) Inizializzazione del sistema

L'inizializzazione può essere posta in  $t=0$  come momento in cui si verifica il primo evento oppure all'apertura, ad esempio, del negozio. Se si inizializza se e prescindere del verificarsi o meno di un evento, rischierci di introdurre un bias

↳ tendenza a deviare dal valore medio

### 2) Logica procedurale

Quando arriva un cliente, in particolare il primo, vengono a verificarsi due eventi:

- si mette in lista l'arrivo del cliente successivo
- se la macchina è libera inizio con la lavorazione

L'arrivo del secondo cliente è estratto da una variabile casuale che descrive gli arrivi del sistema

Qualora il secondo arrivi prima della fine della lavorazione del primo, si genera un nuovo interarrivo. Questo però mi implica la gestione di una coda.

### 3) Avanzamento

$t=0$        $Q(t)=0$  → non si sono generate code  
 $B(t)=1$  → macchine occupate



$t=1,73$  (arriva job 2)       $Q(t)=1$  → coda di 1 pezzo perché la macchina è occupata dal job 1  
 $B(t)=1$



Nella lista degli eventi non ho ancora il service level del job 2 perché non ha ancora iniziato la lavorazione



## SIMULATION PROJECTS

Nella simulazione di un progetto abbiamo varie fasi:

- Definizione del problema che porta il sistema reale ad un modello con assunzioni e semplificazioni; comprende la raccolta di dati che se non ci sono vanno dichiarati; si deve porre il livello di analisi.
- Analisi del problema che è l'obiettivo della simulazione; misura delle performance che vogliamo ottenere dalla simulazione; tempo di simulazione; scelta delle configurazioni del sistema che deve essere modellizzato;
- Data collection che porta all'identificazione dei componenti e delle relazioni funzionali; identificazione di variabili e parametri dei componenti; identificazione delle condizioni ambientali.
- Livello di dettaglio che dipende da come devono essere usati i risultati della simulazione; deve essere consistente con i dati disponibili; tempo e costi limitano il livello di dettaglio; si parte da un modello semplice e si migliora come necessita.
- Modello concettuale è il modello che verrà simulato, deve essere molto dettagliato e deve contenere le assunzioni fatte per rappresentare il sistema; può richiedere l'utilizzo di flowchart, UML, IDEF, ...
- Costruzione del modello bisogna scegliere il linguaggio di programmazione (possono da proposte generali come C++, C#, Java, ... o usando onde programmi orientati alla simulazione come Arena, Sigma, ...) e implementarlo il codice.



un'empirico e la CORRELAZIONE dove posso assumere l'indipendenza o escludere l'analisi della correlazione.

I risultati sono delle stime ed il confronto tra gli output è applicabile solo per gli stessi input.

Dato che questi sono statistiche, non avevo veri e propri confronti: lo stimatori puntuali e stimatori intervallari che vanno confrontati con i test d'ipotesi.

- DOE (Design of Experiments) sono scenari del sistema che possono essere comparati miscelandoli. Bisogna però decidere quanto simulare ed in che scenario.



clienti in attesa, dato un certo numero di operatori fissi. Quale coda mi aspetto?

Una variabile sarà la CODA IN INGRESSO ~~de cui conte~~ sarà aggiornato in base al numero di persone in coda in quel momento. Questo diminuisce se il cliente è servito o aumenta se un cliente viene lasciato in attesa.

Una seconda VARIABILE conterà il numero di operatori liberi.]

Le variabili di stato rappresentano lo STATO DEL SISTEMA

In primis bisogna capire PERCHÉ simulare il sistema, determinate le variabili ed il problema, cerco di ~~esporre~~ SEMPLIFICARLO il più possibile, una volta fatto ciò posso iniziare il modello concettuale.

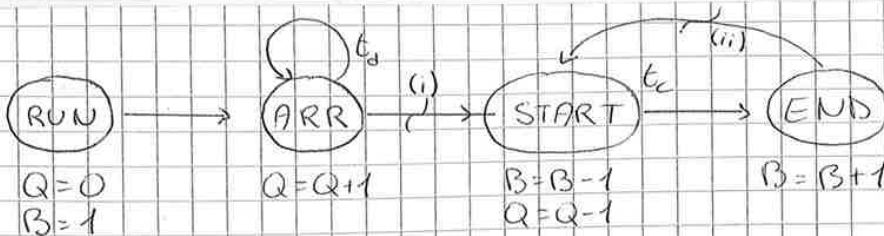
Nel sistema occorrono degli eventi, in base alla loro occorrenza <sup>può cambiare</sup> ~~può essere~~ 1 o più variabili di stato.

**NON** possiamo dire che un evento determini il cambiamento delle variabili di stato, ma è corretto dire che le variabili di stato non cambiano se non **occorrono degli eventi**. → il verificarsi di un evento è **CONDIZIONE NECESSARIA** per il cambiamento dello stato del sistema.

Gli eventi possono essere legati l'un l'altro.

Negli ERG i **NODI** sono gli **EVENTI**, gli **ARCHI** sono le **RELAZIONI** tra gli eventi. Le variabili di stato sono definite in opportune "maschere" tra cui e all'occorrenza dell'evento, sempre tramite maschere, vado a descrivere come questo mi fa cambiare il sistema.





(i)  $B \geq 1$

(ii)  $Q > 0$

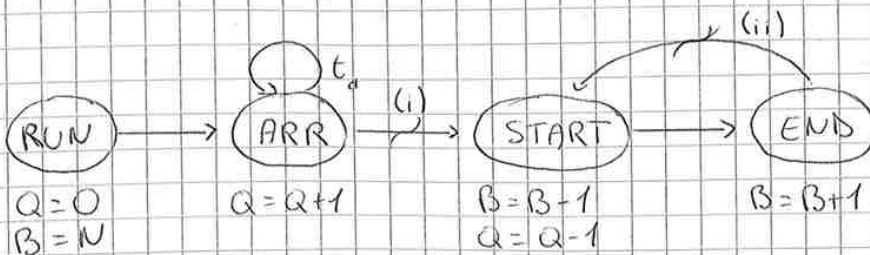
**RUN**: inizializzazione      **ARR**: arrivo al sistema

L'evento ARRIVO genera un altro evento ARRIVO dopo un tempo  $t_a$  ( $\rightarrow$  tempo di interarrivo)

Il primo arco condizionato non implica passaggio di tempo: se il server è libero, serve il cliente altrimenti questo ultimo aumenta la coda. La condizione col singolo server sarebbe  $B=1$ , quella ~~scritta~~ sotto è il caso di più server.

Dal START ad END lo il tempo  $t_c$  che è quello di processo del cliente. Una volta finita la lavorazione "torna indietro" per servire il prossimo cliente SE questo è presente (se c'è una coda)

**ESEMPIO 2** Considerando l'esempio precedente, si suppone ora la presenza di  $N$  server con coda condivisa (in parallelo).



(i)  $B \geq 1$

(ii)  $Q > 0$

La differenza rispetto al primo è che  $B$  è inizializzato ad  $N$



Per gestire la "percentuale", si estrae un valore casuale e se questo è inferiore al valore percentuale per cui risulta essere "buono" si accetta, altrimenti si rilavora.

Bisogna capire allo fine della lavorazione se "posso" prima il pezzo che deve essere rilavorato o quello che deve iniziare il processo.

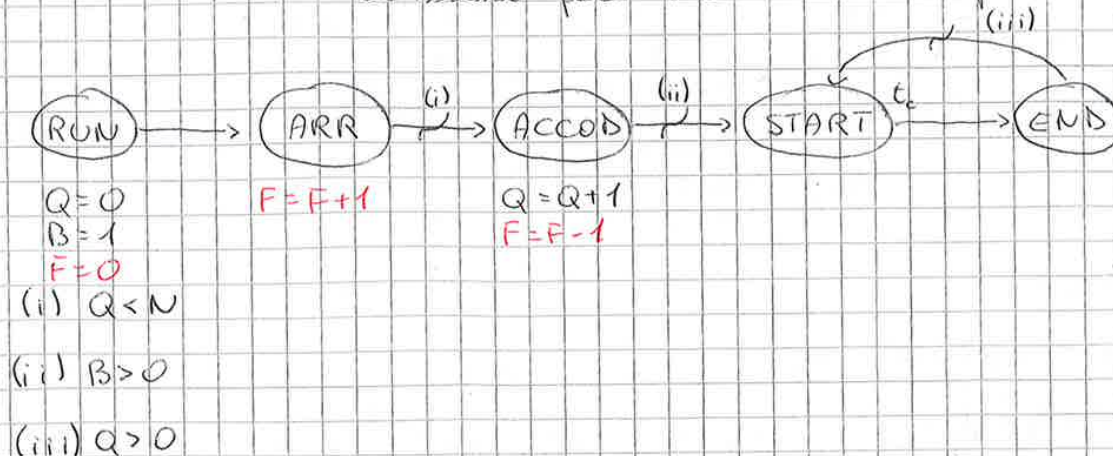
Posso trovarmi nella situazione in cui accadono due eventi: - uno sicuro di non entrare in conflitto - devo dare a questo una priorità di esecuzione.

Per questo ho il tempo  $t_s$ : dopo questo tempo, se la parte è venuta male, lo rilavoro altrimenti termino.

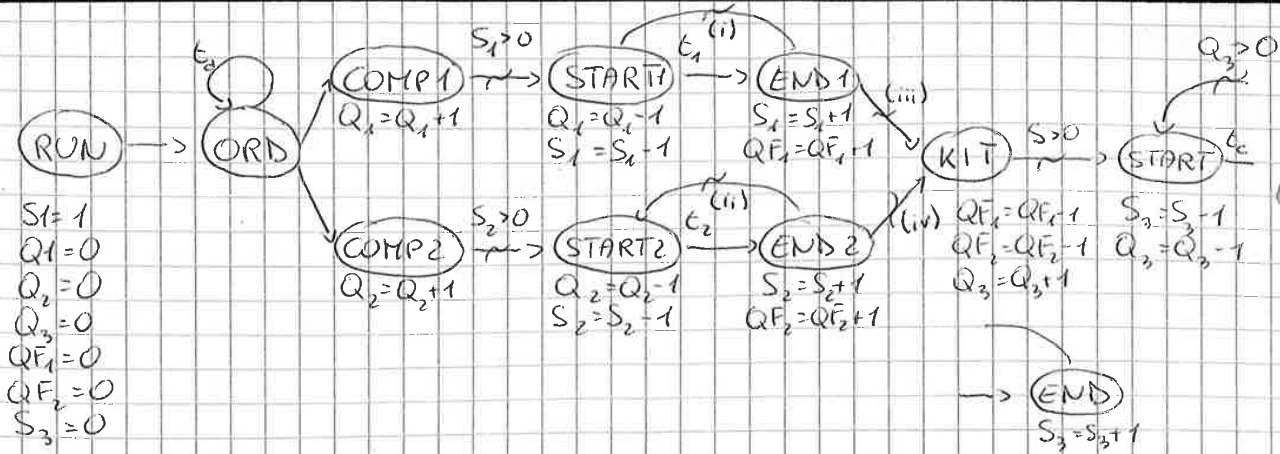
La macchina quindi, finito un pezzo, prende il successivo della coda, nel frattempo in  $t_s$  il pezzo viene controllato e, se non rispetto i parametri viene rilavorato.

Quando il tempo è rappresentato da una variazione ( $\Delta t$ ), la soluzione del sistema è una EQUAZIONE DIFFERENZIALE. Quando  $t$  rappresenta dei valori determinati, si ha il classico sistema di EVENTI DISCRETI.

**ESEMPIO 5** Servente singolo ma con spazio in coda limitato [ $N \rightarrow$  capienza coda]. (potizziamo che siano pezzi i clienti sono pazienti)

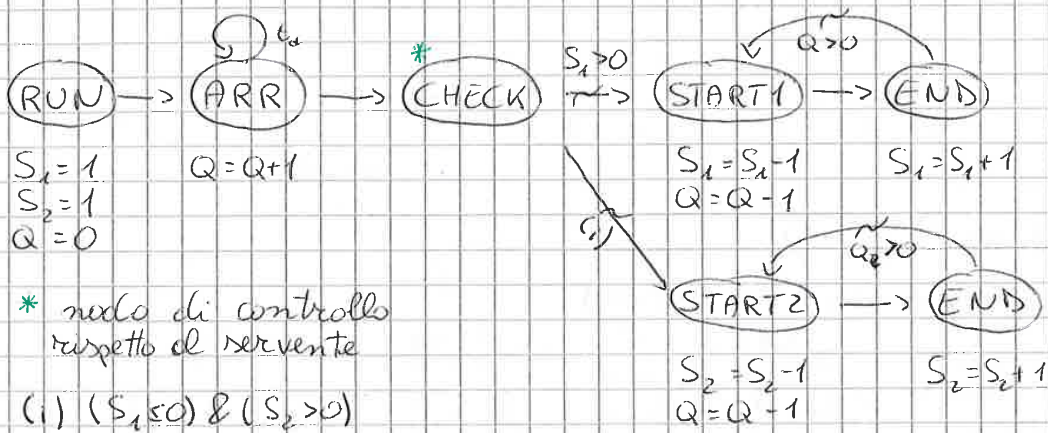






(i)  $Q_1 > 0$  (ii)  $Q_2 > 0$  (iii)  $Q_{F_2} > 0$  (iv)  $(Q_{F_2} > 0) \& (Q_{F_1} > 0)$

**ESEMPIO 7** Consideriamo il caso di tanti server in parallelo tutti diversi (di più lento, di più veloce) e, se possibile, uso il migliore



\* nodo di controllo rispetto al server

(i)  $(S_1 \leq 0) \& (S_2 > 0)$

Considerando l'esempio 1, come posso inserire una ~~simulazione a termine~~ SIMULAZIONE A TERMINE?

Ovvero come faccio a dire al mio sistema che la simulazione finisce dopo tot tempo?

Serve una tipologia di costrutto definita ARCHI DI

CANCELLAZIONE All'inizio della simulazione so quanto questo deve durare perciò, dopo un tempo  $t_s$ , a sera una chiusura da CLOSE ma partiammo degli archi tratteggiati che cancellano i nodi



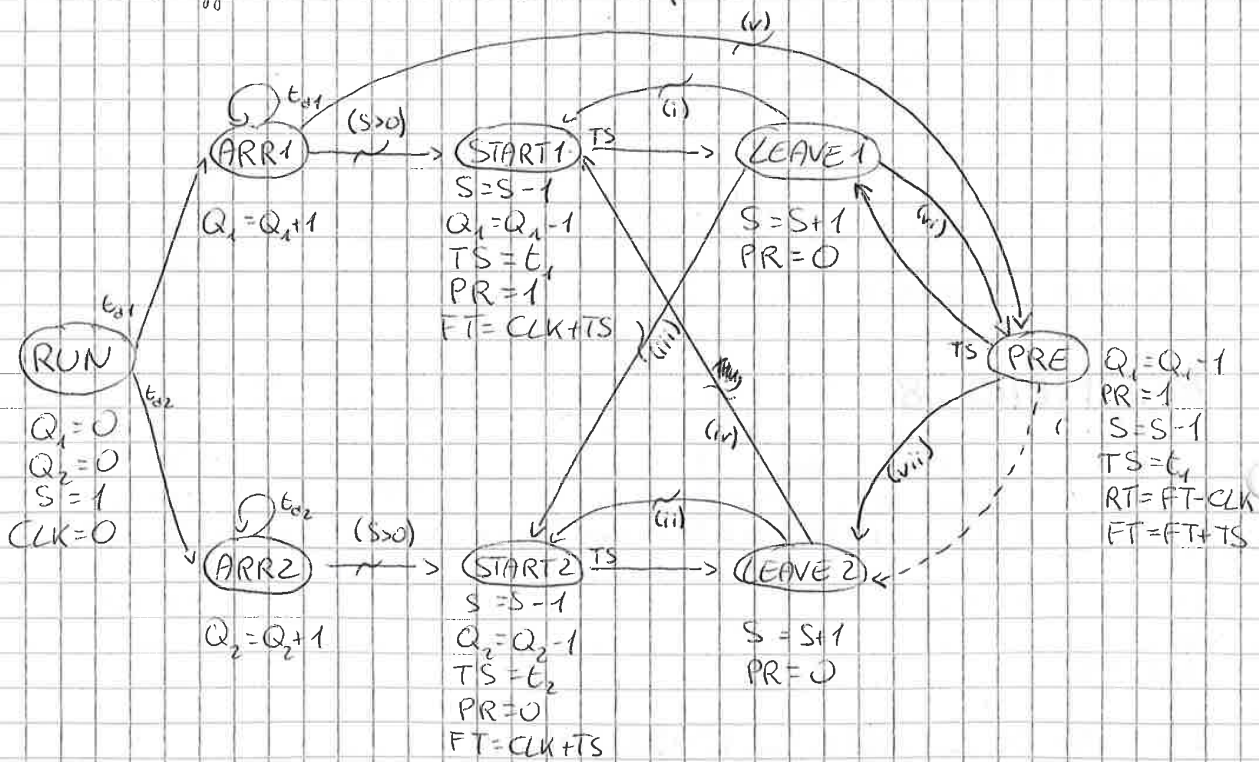
2013/14

ESERCIZIO 9 Situazione in cui lo un solo server ma due tipologie di clienti (urgente/non urgente)

**Con arco di cancellazione**

La prima cosa da decidere è come generare le due tipologie di cliente diverso. Abbiamo due alternative:

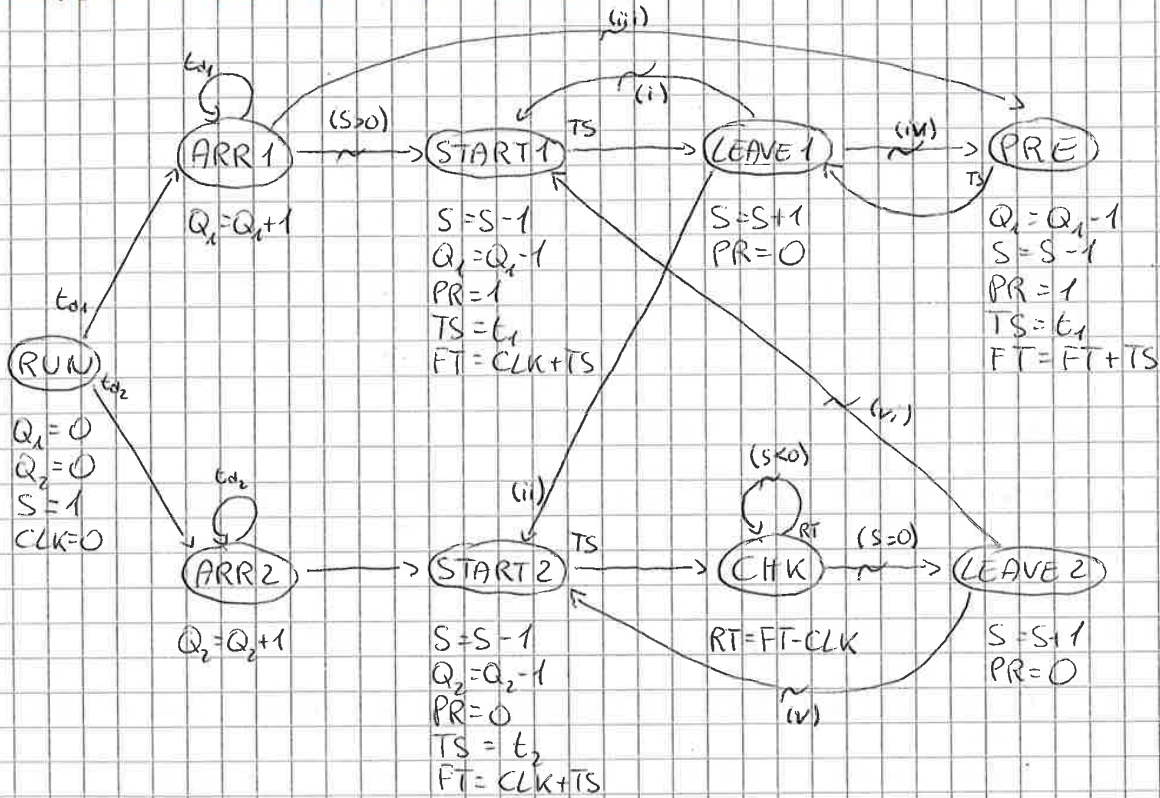
- avere due processi di generazione paralleli: che si fa quando i processi in arrivo stocastici sono diversi: ci sono due popolazioni diverse con tempi di interarrivo diversi e tutto si basa su due processi di arrivo paralleli.
- avere una popolazione uniforme in termini di arrivo e la differenziazione c'è dopo lo scelto.



- (i)  $(S>0) \& (Q_1>0)$       (ii)  $(S>0) \& (Q_2>0) \& (Q_1=0)$
- (iii)  $(S>0) \& (Q_1=0) \& (Q_2>0)$       (iv)  $(S>0) \& (Q_1=1)$
- (v)  $(S=0) \& (PR=0)$       (vi)  $(Q_1>0) \& (S=0)$
- (vii)  $(S>0) \& (Q_1=1)$

tipo 2, per dare precedenza ad uno di tipo 1. Il cliente non urgente dovrà essere finito sempre che non arrivino altri codici rossi, se fanno trascorre il finish time del codice verde

### Senza arco di cancellazione



- (i)  $(Q_1 > 0) \& (S > 0)$
- (ii)  $(S > 0) \& (Q_1 = 0) \& (Q_2 > 0)$
- (iii)  $(S = 0) \& (PR = 0)$
- (iv)  $(Q_1 > 0) \& (S > 0)$
- (v)  $(Q_1 = 0) \& (Q_2 > 0) \& (S = 0)$
- (vi)  $(Q_1 > 0) \& (S > 0)$

Non posso mettere più il LEAVE dopo lo START perché non posso più cancellare, ma dovrò avere un nodo di controllo ciclico che mi controllo quando posso finire e posso terminare solo quando non ci saranno più interruzioni

Posso un tempo  $TS=t_2$  io faccio un controllo. Nel CHECK calcolo il tempo rimanente e, con un ciclo, dopo un tempo  $RT$  torno sul nodo se  $S < 0$  ovvero mi indica esserci stato una interruzione. "Esco" quando  $S = 0$  ovvero è occupato da un job che lo aveva già occupato



Il numero non è per forza qualcosa di DISCRETO.

Come si può generare un numero casuale uniformemente distribuito? Es: dado: lancio e prendo il risultato diviso

per 6. Non esce mai 0, è uniforme, ma NON è un buon generatore. Posso migliorarlo lanciando sempre più dadi.

Non basta l'UNIFORMITÀ, ma cerchiamo anche la DENSITÀ dei numeri. Se poi il dado non è truccato, i numeri sono INDIPENDENTI tra loro.

Perché proprio l'uniforme per rappresentare la casualità?

Perché tutti i numeri hanno la stessa probabilità di uscire.

UNIFORMITÀ e INDIPENDENZA sono le caratteristiche che deve avere un generatore di numeri casuali.

Il problema può essere un problema di un generatore software di numeri casuali? Abbiamo un algoritmo che genera numeri, ma questo è una serie di istruzioni definite e priori (operazioni). Per evitare che ci sia in uscita lo stesso numero, si cercano di dare delle forme ricorsive. Il problema è che non c'è più nulla di casuale! I generatori di numeri casuali software sono generatori PSEUDO-CASUALI ovvero un esterno li vede come casuali, ma sono il risultato di una serie di step, sono una sequenza prevedibile e priori.

I numeri casuali sono una serie di numeri che devono avere due proprietà:

- 1) INDIPENDENZA ovvero quando viene generato un numero non deve avere indicazioni su quale sia il predecessore o il successore
- 2) UNIFORMITÀ perché ogni numero deve avere la stessa probabilità di essere estratto.



de farlo sembrare casuale. Condizione **necessaria e sufficiente** affinché un numero sia reputato casuale è de passi i test statistici.

Ci sono infiniti tipi di generatori, ma noi ci concentreremo su:

### - Generatori lineari Congruenti (LCG)

- congruenti misti
- congruenti moltiplicativi

### - Generatori Congruenti Generali (GCG)

## GENERATORI LINEARI CONGRUENTI

Generano dei numeri  $X$  che sono nell'intervallo  $[0, m-1]$  dove  $m$  è il **MODULO** e la generazione avviene:

$$X_{i+1} = (dX_i + c) \bmod m \quad i = 0, 1, 2, \dots$$

la generazione in  $i+1$  avviene prendendo il valore precedente moltiplicato per  $d$ , il **MOLTIPLICATORE**, aggiungendo un parametro  $c$  detto **INCREMENTO** e diviso per il modulo, prendendo il **RESTO** della divisione.

Per passare dalla  $X$  alla  $R$ , che è compreso in  $[0, 1]$ :

$$R_{i+1} = \frac{X_{i+1}}{m} \quad i = 0, 1, 2, \dots$$

La  $R$  siamo sicuri appartiene a  $[0, 1]$  perché la  $X$  appartiene all'intervallo  $[0, m-1]$  discreto. E sappiamo che  $X$  è in quell'intervallo perché è il resto di una divisione per interi.

Il ultimo parametro è la  $X_0$  detto **SEED**

## ESEMPIO

$$X_0 = 25 \quad d = 12 \quad c = 37 \quad m = 200$$

$$X_0 = 25 \quad \text{seed}$$



limitato, ma sono onde valori limitati: non tutti i punti dell'intervallo sono possibili.

I numeri  $R$  "controllano" onde de l'intervallo obbia

**MASSIMA DENSITA'** ovvero distanze non troppo larghe nell'intervallo e **MASSIMO PERIODO**, oltre all'uniformità e all'indipendenza.

Il valore dato ai vari parametri impatta fortemente sul numero in uscita: dovrebbero essere scelti in modo da **MASSIMIZZARE** uniformità, indipendenza e periodo.

Ci sono alcune condizioni base per far sì che il generatore possa almeno funzionare:

- il modulo deve essere strettamente maggiore di tutti gli altri valori;

È per avere un generatore lineare FULL PERIOD dobbiamo rispettare anche:

- l'incremento  $c$  deve essere primo relativo rispetto  $m$  (ovvero devono essere primi tra loro e solo 1 può essere un divisore comune);
- preso un numero  $q$  che divide  $m$ , questo deve anche dividere  $d-1$

Verificare le condizioni è molto difficile, soprattutto se considerando che un  $m$  "piccolo" vale quasi 2 miliardi. Valori del modulo usati sono  $2^{48}$  e  $2^{31}-1$ . Sono usate potenze di 2 per avere velocità ed efficienza nell'uso del generatore al computer.

Un esempio di GENERATORE LINEARE CONGRUENTE

**MOLTIPLICATIVO** è quando  $c=0$ . Questi non riescono mai a raggiungere il full period, ma l'unico periodo massimo che raggiungono è  $m-1$ .



e come si dispongono i pieni tra loro.

28/3/17

La prima famiglia di test che analizziamo è quella che si occupa dell'UNIFORMITÀ: vedremo il test di KOLMOGOROV-SMIRNOV e il CHI-QUADRO

## TEST DI KOLMOGOROV-SMIRNOV

Il test **KS** si basa sul confronto delle cumulate ovvero se come è fatto la  $F$  di una sequenza di osservazioni e ci si crea la cumulata empirica della stringa di numeri del campione generato e si confrontano.

Il confronto tra la cumulata teorica e la distribuzione cumulata empirica si baserà sul calcolo di una statistica che verrà confrontata con un livello di significatività con il valore tabulato della statistica.

### ALGORITMO

Abbiamo generato una serie di numeri  $(R_1, R_2, \dots, R_N)$  e dobbiamo testare se sono distribuiti secondo una uniforme tra 0 e 1.  $\{ F(x) = x \text{ se } 0 \leq x \leq 1 \}$

La distribuzione empirica  $S_N(x)$  lo trova contando tutti i numeri che sono più piccoli di  $x$  e li rapporta al totale del campione

$$S_N(x) = \frac{\text{numero di } R_1, R_2, \dots, R_N \leq x}{N}$$

Il confronto si basa su quello che è chiamato **DEVIAZIONE ASSOLUTA**, cerchiamo di calcolare il valore massimo

Questo si calcola come:  $D = |F(x) - S_N(x)|$

$D$  è una variabile casuale di cui si conosce la distribuzione del campione ed è tabulata come funzione di  $N$  (quantili  $D_{\alpha, N}$ ). Confrontiamo la  $D$  trovata con quella



## TEST CHI-QUADRO

Questo test confronta le  $f$ , ovvero la funzione di densità di probabilità.

La statistica è

$$\chi_0 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

↑ Effettivo
↑ Teorico

Si va a fare la sommatoria ~~tra~~ su  $n$  classi di dati.  $O_i$  ed  $E_i$  sono i valori che stanno nella classe. Ho bisogno di due variabili perché  $O_i$  sono i valori osservati,  $E_i$  sono i valori che mi aspetto nelle classi. [osservato - quello che mi aspetto rapportato a quello che mi aspetto].

La statistica che troviamo segue approssimativamente una chi-quadro con  $n-1$  gradi di libertà.

Calcolo la statistica, scelgo i gradi di libertà e vedo sulle tabelle: se  $\chi_0 > \chi_{\alpha, n-1}$  rifiuto l'ipotesi, se  $\chi_0 < \chi_{\alpha, n-1}$  accetto l'ipotesi.

La difficoltà qui è determinare il numero di classi: se sono troppo piccole o troppo grandi il test può risultare non significativo. Le classi devono essere costituite in modo tale che il valore medio della classe sia almeno 5. Dovrei avere almeno 5 classi.

Il KS si dimostra ~~MIGLIORE~~ MIGLIORE del  $\chi$  perché rappresenta meglio la realtà.

### ESEMPIO

$N = 100$  scelgo un  $n = 10$

Per quanto riguarda le  $O_i$ , conto in ogni intervallo quanti valori ci sono, per le  $E_i$ , il valore è dato da  $N/n = 10$ .

$L_{\min} = 0$ ,  $U_{\max} = 1$ ,  $L_i = 0.1(i-1)$ ,  $U_i = 0.1 \cdot i$   
 $L_i$  ed  $U_i$  sono gli estremi dei vari intervalli.



Da questi dati costruisco la statistica

$$z_0 = \frac{\hat{\rho}_{11}}{\hat{\sigma}_{11}}$$

che è una normale standardizzata.

$M$  è il valore più grande tale per cui il valore da cui partiamo più un numero di log più uno è minore uguale ai numeri totali.

$$i + (M+1)l \leq N \Rightarrow M \leq \frac{N-i}{l} - 1$$

**ESEMPIO** (vedi slides)

$$\hat{\rho}_{35} = \frac{12}{4+1} [(0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) + (0.27)(0.05) + (0.05)(0.36)] - 3$$

$$= -2.334$$

$$\hat{\sigma}_{35} = \frac{\sqrt{13.4 + 7}}{(4+1)} = 1.536$$

$$z_0 = -\frac{2.334}{1.536} = -1.52 \in (-1.96, 1.96)$$

accetto l'ipotesi

Perché i generatori di numeri casuali **FALLISCONO**?

Perché i test sono fatti su un campione.

Si basano sulla struttura e reticolo dei punti che da in output il generatore di numeri pseudo-casuali.

## TEST SPETTRALE

Sulla base delle formule ricorsive, calcolo le distanze tra tutti i possibili piani, tutte le possibili dimensioni che quel generatore può "provocare". Usarli e leggerne il risultato è molto difficile.

Quello che ricavo da un generatore di numeri casuali è una stringa di valori della quale posso ricavare sottostringhe diverse e spesso nelle simulazioni possono servire più



Esistono molti test teorici empirici ma nessuno di questi si è dimostrato migliore degli altri

Si usava un generatore di numeri casuali in base al caso se si doveva affrontare e i dati o disposizione.



l'output che è la variabile con l'input?  
 Ovvero, posso trovare un legame univoco?

Un generatore che abbia tutte queste caratteristiche NON ESISTE

## LA TRASFORMATTA INVERSA

"Prende" la  $F(x)$  e la inverte perché questo, qualunque sia la distribuzione, assume sempre valori tra 0 e 1.

È simile alle R; che ovvero gli stessi valori. Bisogna generare una distribuzione che abbia una certa  $F$  con parametri dipendenti dalla distribuzione.

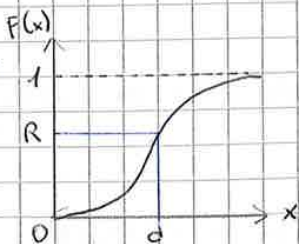
La trasformata inversa legge i numeri casuali come valori della  $F$  e li usa come elementi incogniti  $x$ .

La forma della  $F$  è nota e se è invertibile, per si può localmente trovare  $x$ .

Questo è il metodo **più facile** per generare numeri casuali.

È **robusto** perché non chiede parametri, è **efficiente** perché faccio un'unica operazione, permette una buona **sincronizzazione**, non richiede grandi spazi di memoria.

4/4/17



Se ho un numero casuale del generatore, lo vedo come  $F(x)$ . Vado sulla curva e ricevo il valore  $x$  della variabile casuale. Questo vuol dire "fare  $F^{-1}(R)$ " del ~~numero~~

valore iniziale. Una qualunque distribuzione di probabilità ha una funzione cumulata che ha valori tra 0 e 1: in teoria funziona per tutte, in pratica no!



Devo trovare quel valore che risolve  $P[X \leq x] = 0.65$ .

NON posso usare l'uguale perché nella funzione non ho nessun 0.65 Devo prendere il 2 o il 3? Prendo lo 0.9 perché contiene lo 0.65  $P[X \leq x] \geq 0.65$

Devo scegliere, tra i punti discreti, quello che mi contiene la R.

La trasformata inversa è molto flessibile ma non può essere espressa in "forma chiusa". Ha un'ottima sincronizzazione tra il valore della variabile e il numero consueto di uso per generare la variabile ed è il metodo da applicare se voglio ridurre rapidamente. Questo metodo induce forti correlazione nel generare variabili consuete, permette una loro facile sincronizzazione così da aumentare l'effettiva riduzione della varianza. ↓

## LA CONVOLUZIONE

Questo metodo è usato quando posso scrivere la variabile consueta  $X$  come somma di un certo numero di variabili consuete  $Y$  che sono indipendenti ed identicamente distribuite e più facili da generare della  $X$ .

$$X = Y_1 + Y_2 + Y_3 + \dots + Y_m$$

### ESEMPIO ERLANG ( $m, \lambda$ )

Con  $\lambda$  il rate ed  $m$  lo shape, il parametro di forma e il tasso della Erlang. Questo può essere visto come somma di variabili esponenziali con tasso  $\lambda$ .

Generiamo, ad esempio, 4 esponenziali con  $\lambda=2$ :  $R_1=0.5$ ,  $R_2=0.12$ ,  $R_3=0.41$ ,  $R_4=0.28$  e, usando la trasformata inversa, trovo:



della funzione di densità, non è una densità perché l'area sotto la curva è più grande di 1. Chiamiamo  $c$  l'area sottostante la funzione di maggioranza.

Se prendo la funzione  $t$  e la si divide per  $c$ , si trovano dei punti che sono minori o uguali ad 1.

La funzione  $r$ , ovviamente, dovrà essere più "facile" di quello di partenza e lo trovo dalla  $t$ . Poi genero un numero casuale  $R$  e confronto con il rapporto  $R \leq f/t$ . Genero un  $Y$  che segue la  $R$ , metti il valore di  $Y$  in  $f$  e  $t$ , genero  $R$  e fai il confronto.

### ESEMPIO BETA ( $\alpha_1, \alpha_2$ )

$$f(x) = \begin{cases} 60x^3(1-x)^2 & 0 \leq x \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

Poiché  $f$  ha un massimo in  $x = 0.6$ , la funzione massima è  $f(0.6) = 2.0736$

$$t(x) = \begin{cases} 2.0736 & 0 \leq x \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

Il  $c$  è pari a  $\int_0^1 2.0736 dx = 2.0736$ . lo trovo una maggiorante molto semplice.

La nostra  $r$  sarà una Uniforme tra 0 e 1.

Genero  $R_1 = Y = 0.46$  ed un numero casuale  $R_2 = 0.71$ .

Calcolo poi la  $f(Y) = 60 \cdot 0.46^3(1-0.46)^2$  e la  $t(Y) = 2.0736$

Devo avere che il numero casuale  $R_2$  sia minore o uguale

del rapporto  $f(Y)/t(Y) \Rightarrow R_2 = 0.71 \leq 0.82$

La condizione è rispettata perciò mi fermo. Se  $R_2$  fosse stato 0.81, avrei rigettato i valori e trovati di nuovi.

Il principio base di questo metodo è molto generale e può essere esteso alla generazione di punti in più dimensioni.

La funzione maggiorante imposta in modo efficiente



11/4/17

## PROCESSI STOCASTICI

Considerando che sto lavorando con dei processi, non genererò solo variabili, ma anche dei tempi di interarrivo: la maggiore difficoltà sta nel capire se le variabili sono indipendenti, sono generate in una sequenza.

Un **processo stocastico**  $P$  è una sequenza di osservazioni della stessa variabile casuale  $X$  come è osservata in

$$t = 1, 2, \dots, T \quad P: \{X_t\} = \{X_1, X_2, \dots, X_T\}$$

Se le osservazioni che compongono il processo sono indipendenti ed identicamente distribuite, si può semplicemente generare una sequenza di variabili dalle distribuzioni appropriate.

In molti casi le  $X_t$  sono identicamente distribuite, ma non indipendenti e l'**autocorrelazione** deve essere modellizzata in modo opportuno.

L'autocorrelazione tra due variabili può essere rappresentata come **Cor** ( $X_t, X_{t+l}$ ) con  $l = 1, \dots, p$  dove  $l$  è il **LAG** ovvero la distanza tra due variabili correlate.

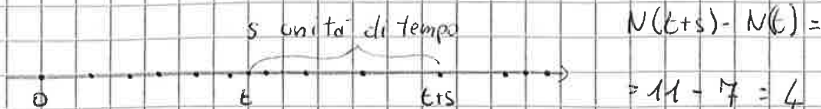
Solitamente, più vicine sono le variabili, più alta è la correlazione così che  $\text{Cor}(X_t, X_{t+l})$  diminuirà con  $l$ , fino a raggiungere lo 0 per valori di LAG più grandi di  $p$ .

Discuteremo alcuni processi:

- **Processi di arrivo**
  - processi di Poisson
  - arrivi batch
- **Catene di Markov**



2. le entità arrivano una alla volta;  
 3. il numero di arrivi in un certo intervallo di tempo è dato da  $N(t+s) - N(t)$



e questo non dipende da ciò che succede prima di  $t$ ;

4. questa differenza non dipende da  $t$ , perciò  $N(t+s) - N(t)$  non dipende da  $t$ . Questo vuol dire che il numero di entità che arrivano in un intervallo, dipende dalla sua grandezza e non dal punto di partenza.

Se valgono tutte queste condizioni,

$$P[N(t+s) - N(t) = k] = \frac{e^{-\lambda s} (\lambda s)^k}{k!} \quad k = 0, 1, 2, \dots \text{ e } s, t \geq 0$$

ovvero la probabilità che arrivi un certo numero  $k$  di entità in un intervallo lungo  $s$ , è esprimibile tramite una Poisson.  $A_i \sim \text{Exp}(\lambda)$  ovvero i tempi di interarrivo sono delle esponenziali con tasso  $\lambda$ , chiamato **tasso di processo** e segue che:

$$E[N(s)] = \lambda s \quad E[N(1)] = \lambda \quad E[A_i] = 1/\lambda$$

Come si genera un processo di Poisson stazionario?

Innanzitutto, un processo è **STAZIONARIO** quando il **tasso non cambia nel tempo**,  $\lambda$  rimane lo stesso.

Generarlo è facile perché arriva 1 item alla volta perciò non devo generare il numero di clienti e il tempo di interarrivo è esponenziale, perciò avrò delle variabili che seguono questa distribuzione.

**ALGORITMO**

Genero la variabile  $R \sim U(0,1)$  indipendente da qualsiasi variabile casuale generata in precedenza. Da qui ritorno



non stazionario se:

1. i clienti arrivano uno allo volta
2.  $N(t+s) - N(t)$  non dipende da  $\{N(u), 0 \leq u \leq t\}$  ovvero i clienti che arrivano in una unità di tempo, non dipendono da quello che è successo prima.

Si può trovare una formula per la probabilità che arrivino  $k$  entità in un certo intervallo di tempo

$$P[N(t+s) - N(t) = k] = \frac{e^{-b(t,s)} [b(t,s)]^k}{k!}$$

La formula è simile alla precedente, ma invece di avere  $\lambda \cdot s$ , è presente  $b(t,s)$  ovvero la differenza tra

$$\Lambda(t+s) - \Lambda(t) = \int_t^{t+s} \lambda(y) dy.$$

$\lambda(t)$  è ora detta **funzione del tasso** e  $\Lambda(t)$  è la **funzione del valore atteso** e sono legate dal fatto che  $\lambda$  è la derivata di  $\Lambda$  e  $\Lambda$  è il valore atteso degli arrivi  $N(t)$ . Bisogna comunque stimare  $\lambda$  e  $\Lambda$  da un set di dati.

Uno dei più semplici e comuni metodi per generare un processo di Poisson non stazionario è detto **THINNING** (dimagrimento).

### ALGORITMO

L'idea è quella di generare un processo con il tasso  $\lambda$  più alto possibile, ipotizzando che il processo sia stazionario e poi, nell'intervallo in cui mi trovo, tengo o meno dei valori in modo casuale.

In primis, quindi, si sceltono i veri indici  $\lambda^* = \max_{0 \leq t \leq T} \lambda(t)$ ,

$t=0, i=1$ . Poi generiamo un numero casuale

indipendente da quelli precedenti e imponiamo

$t = t - 1/\lambda^* \ln R_1$ , come se fosse un processo poissoniano stazionario.



della distribuzione.

Dato  $X(t)$  il numero totale di clienti individuali arrivati al tempo  $t$  e  $B_i$  il numero totale di clienti singoli nel batch  $i$ , si ha  $X(t) = \sum_{i=1}^{N(t)} B_i$  per  $t \geq 0$

La  $B$  viene generata secondo la variabile casuale de  $B$  segue.

Si può anche avere un processo composto non stazionario dove per generare i vari istanti di tempo in cui arriva il lotto successivo, si usa il processo di Poisson non stazionario con il "dimagrimento": generato un istante di tempo, quando si è sicuri di accettarlo, si estrae dalla forma della distribuzione un secondo  $P_{B_i}$  numero.

## CATENE DI MARKOV

Sono un processo stocastico semplice ma sono un po' di contesto nel processo Poissoniano: guardandolo la forma delle probabilità, la catena di Markov si trova al tempo  $t$  in un certo stato, la voglio trovare.

$$P[X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0] = P[X_{t+1} = j | X_t = i] = P_{i,j}$$

Voglio trovare la probabilità che in  $t+1$ , ad esempio, per il processo sistema, si trovi nello stato  $j$ .

Se questo evento è una catena di Markov, il condizionamento a questo evento rispetto al percorso fino a quel momento, si riduce al soltanto al condizionamento prima di questo.

Tutta la probabilità che in  $t+1$  ci si trovi allo stato  $j$  dipende dal fatto che in  $t$  ci si sia trovati nello stato  $i$  e tutta questa rappresenta la **PROBABILITÀ DI**



Ora si genera quello che sarà lo stato di arrivo, come si fa?

	1	2	3	4	5	6	7
1							
2							
3	0.3	0.2	0.1	0.1	0.05	0.2	0.05
4			x				
5							
6							
7							

Abbiamo la riga 4 che è una distribuzione di probabilità empirica. Usando, ad esempio, la trasformata inversa si determinano

lo stato di arrivo sia 3. La catena di Markov è a tempo discreto e  $t$ , di conseguenza, non ci interessa. In  $t=1$  sarò in 3. Da qui poi genero quello che sarà il prossimo punto di arrivo, ad esempio 5.

Sulle righe si trovano le probabilità con cui devono essere estratti i numeri corrispondenti. Da queste si trova la ~~curva~~ curva cumulata e poi, con il metodo

	$P(x)$	$F(x)$
1	0.3	0.3
2	0.2	0.5
3	0.1	0.6
4	0.1	0.7
5	0.05	0.75
6	0.2	0.95
7	0.05	1

(ad esempio) della trasformata inversa con variabili discrete empiriche, si genera un numero  $R$  e si vede dove si trova.

Il pedice della  $X$  è il tempo, i.e.;

sono degli istanti.

Con ~~la~~ All'inizio scelgo lo stato iniziale ponendo  $X_0 = i_0$  e setto  $t=1$ . Dopo genero  $Y_0$ , ovvero una variabile casuale che è distribuita secondo le ~~probabilità~~ probabilità delle righe  $i$ -esime. A questo punto  $X_1$  sarà uguale al

numero generato  $X_1 = Y_0$ . Se  $t < N$  setto  $i = X_t$ , genero  $Y_i$ , sposto in avanti  $t$  ( $t = t+1$ ) e imposto di nuovo  $X_t = Y_i$ . In caso contrario mi fermo.

In uno stato non rimango una unità di tempo, ma il tempo indicato dalla sua distribuzione: la matrice di



estraggo  $R$  e genero  $R_1 = -(1/Q_5) \cdot \ln R_e$  e se non sono arrivato all'istante finale mi sposto, generando un altro numero che indicherò il mio salto, ad esempio 7. Qui genero un altro tempo, lo confronto, siamo ancora prima di  $T$  e vedo l'altro salto da fare.

Portando dalle catene a tempo continuo, possiamo en

**PROCESSI "SEMI-MARKOV"** con l'unica differenza che l'holding time, il tempo di permanenza in uno stato non è più esponenziale, ma segue una distribuzione diversa.

### **ALGORITMO**

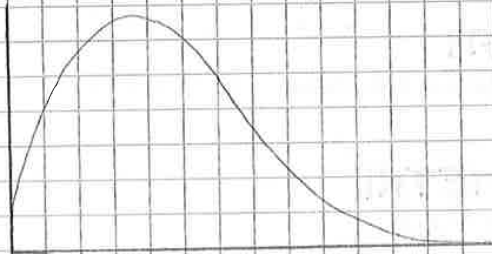
È esattamente uguale a quello precedente, ma il numero generato  $H_{i_0}$  segue una funzione  $F_{i_0}$



I dati sono assunti indipendenti ed identicamente distribuiti e l'appropriatezza della famiglia di distribuzione è basata solo sulla sua forma.

Dalle STATISTICHE sui dati possiamo coprire la forma della loro distribuzione. Ad esempio se la mediana è simile alla media, la distribuzione sarà quasi SIMMETRICA; per le distribuzioni continue a coefficiente di variazione vicino ad 1, è indicata l'ESPONENZIALE, ...

A volte un grafico può portare in a sbagliare



Questa funzione può essere una lognormale, una gamma o una Weibull.

Lo primo però ha un coefficiente

di variazione sempre positivo, mentre le altre due la forma corrisponde ad un coefficiente  $< 1$  e un parametro  $\alpha > 1$ .

Si possono anche avere istogrammi di frequenze e diagrammi quantile-quantile (Q-Q)

2/5/17

Usare il q-q plot senza l'istogramma delle frequenze può risultare errato perché questo ultimo mi dà l'idea di quali possano essere le frequenze, mentre il primo mi presuppone una sola distribuzione

## ISTOGRAMMI DI FREQUENZA

Tramite i dati determino degli intervalli in ognuno dei quali conto quanti dati finiscono e creo un rettangolo proporzionale al numero di dati. Il problema è che sono



ordine del più piccolo al più grande, l' $i$ -esimo valore di questo scalo è uno stima del  $\frac{i-\alpha}{n+1-2\alpha}$  quantile (a valore di noi dobbiamo scegliere tra 0 e 1). Al variare di  $\alpha$  sono tutte "approssimazioni" di  $1/n$ .

Se  $F$  è una funzione di distribuzione della distribuzione da abbiamo scelto, plottando  $x_i$  con  $F^{-1}((i-\alpha)/(n+1-2\alpha))$  avremo una linea approssimativamente dritta se la distribuzione fits bene i dati.

Nella figura abbiamo sulle ascisse i quantili della normale standard e sulle ascisse i quantili del campione in input: i pallini mi danno i quantili della normale contro il quantile di ricavo assumendo che  $F$  sia una normale. Lo scostamento tra retta e pallino mi dice se la distribuzione normale è esatta o meno.

Gli ISTOGRAMMI possono aiutare ad identificare la corretta famiglia di distribuzione. Capito però che la stessa forma può essere associata a differenti famiglie, soprattutto se la distribuzione cambia forma drasticamente al variare del valore dei parametri.

In questi casi si usa la **BASE FISICA** della distribuzione: e secondo del processo che si considera, dal punto di vista fisico si possono avere può immaginare la forma dei dati. Questo è fatto onde grazie alle **CARATTERISTICHE**

**FISICHE**. Ad esempio, volendo modellizzare il numero di parti difettose di un campione, ci si immagina ~~una~~ seguano una binomiale. Gli occorri sono poissoniani e, di conseguenza, i tempi di interarrivo sono esponenziali.

Alcuni esempi di famiglie di distribuzioni sono

**BINOMIALE** da modello il numero di successi in un



certa distribuzione, dico che sono una famiglia che segue quella distribuzione) il metodo calcolo i momenti teorici della distribuzione (~~NON~~ media, NON varianza ma una sua approssimazione, ...) come una funzione dei parametri uguali ai momenti trovati usando la collezione di dati.

### ESEMPIO ESPONENZIALE

6.5602	0.7284	3.6874	1.5149	0.1087	2.5146	10.926
0.8308	2.5135	1.2671	6.6285	0.2056	5.0226	3.9708
0.6776	2.1548	0.7882	2.7790	3.8955	9.8715	

$$E[X] = 1/\lambda$$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{6.5602 + 0.7284 + \dots + 9.8715}{20} = 2.8411$$

$$\frac{1}{\hat{\lambda}} = 2.8411 \Rightarrow \hat{\lambda} = 0.3520$$

### ESEMPIO UNIFORME

8.9468	6.8383	6.0301	5.4333	8.8597	6.0284
6.9414	7.7589	6.1448	8.2097	7.4224	5.7592
8.8097	5.5030	6.4703	6.1869	7.6544	5.4375
7.0266	5.5242				

$$E[X] = \frac{\alpha + \beta}{2} \quad \text{Var}[X] = \frac{(\beta - \alpha)^2}{12}$$

$$\bar{X} = \frac{\sum x_i}{n} = 6.8553 \quad S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = 1.4475$$

$$\begin{cases} \frac{\hat{\alpha} + \hat{\beta}}{2} = 6.8553 \\ \frac{(\hat{\beta} - \hat{\alpha})^2}{12} = 1.4475 \end{cases} \Rightarrow \begin{cases} \hat{\alpha} = 4.7714 \\ \hat{\beta} = 8.9392 \end{cases}$$



ne e parametri, dobbiamo solo valutare la conformità dei dati; ora invece bisogna stimare la conformità dei dati con una distribuzione scelta da noi.

Un'ulteriore difficoltà è data dalla lettura in output dei dati perché tendono a sotto o sopra stimare i valori se i parametri sono stimati con il METODO DEI MOMENTI.

Se nel corso della generazione di numeri si "battava" il generatore se non si passava un test, ora non è detto che debba essere escluso, ma che la stima dei dati sia stata fatta in modo errato.

## TEST CHI-QUADRO

Il test è valido per un campione grande e quando la distribuzione ~~dei~~ dei parametri è stata stimata con la massima verosimiglianza. Il problema può risultare la scelta del numero di CELLE in cui partizionare i dati. La PROCEDURA è identica a quella dei numeri casuali.

1. Scegliere il numero di celle e partizionare i dati.
2. Calcolare quante osservazioni cadono in ogni cella, questo numero è  $N$  ovvero la Frequenza empirica del numero di osservazioni nelle classi.

Si calcola poi la Frequenza teorica, ovvero il numero totale delle osservazioni <sup>moltiplicato per la</sup> ~~da~~ ~~calcolato~~ ~~una~~ probabilità di essere in quella cella.

Se la distribuzione è discreta:  $p_i = p(x_i) = p(X=x_i)$ ;

se la distribuzione è continua:  $p_i = \int_{l_i}^{r_i} f_X(x) dx = F_X(r_i) - F_X(l_i)$

3. Calcolare la statistica

$$\chi_0^2 = \sum_{i=1}^k \frac{(N_i - E_i)^2}{E_i}$$



Numero cella $j$	Intervallo cella $[l_j, r_j)$	Numero di osservazioni $N_j$	Frequenza relativa $\hat{p}_j$	Probabilità teorica $p_j$
1	[10, 12)	13	0.13	0.10
2	[12, 14)	9	0.09	0.10
3	[14, 16)	8	0.08	0.10
4	[16, 18)	9	0.09	0.10
5	[18, 20)	12	0.12	0.10
6	[20, 22)	8	0.08	0.10
7	[22, 24)	13	0.13	0.10
8	[24, 26)	10	0.10	0.10
9	[26, 28)	10	0.10	0.10
10	[28, 30)	8	0.08	0.10

$$\chi_0^2 = \sum_{j=1}^{10} \frac{(N_j - E_j)^2}{E_j} = \sum \frac{n^2 ((N_j - E_j)/n)^2}{E_j} = \sum \frac{n^2 (\hat{p}_j - p_j)^2}{np_j} = \frac{n}{p} \sum (\hat{p}_j - p_j)^2 = 4.4$$

Scelgo  $\alpha = 0.10$  con  $10 - 2 - 1 = 7$  gradi di libertà  $\Rightarrow \chi_{0.1, 7}^2 = 12$

Si accetta  $H_0$

## TEST DI KOLMOGOROV-SMIRNOV

Il test chi-quadro richiede la partizione in celle, che non è possibile nelle DISTRIBUZIONI CONTINUE.

Il test K S compara le funzioni di distribuzione ~~empirica~~ cumulate empiriche con la equivalente teorica della distribuzione ipotizzata.

Questo non richiede un grande campione e non richiede che i dati siano clusterizzati in celle.

Il problema è che KS presuppone parametri **NOTI** e non stimati.

Anche qui la ~~procedura~~ <sup>PROCEDURA</sup> è uguale

1. Ordina le  $n$  osservazioni in modo crescente
2. Costruisci la funzione di distribuzione empirica

$$\hat{F}_x(x) = \frac{\max \{j : x_j \leq x\}}{n}$$

3. Opera il test statistico

$$D_0 = \sup_x \{|\hat{F}_x(x) - F_x(x)|\}$$



Supponiamo di avere ~~una~~ la distribuzione in figura e un  $\alpha = 0.1$  e uno per  $\alpha = 0.01$ . Se il test statistico è 24,  $H_0$  è rifiutato per  $\alpha = 0.1$  ed accettato per  $\alpha = 0.01$ .

Il p-value è la coda destra rispetto alla statistica di riferimento, ovvero quale è la probabilità limite della zona di accettazione/rifiuto data da una certa statistica.

PIÙ GRANDE è il p-value, MIGLIORE è la simulazione.

Un grande p-value indica una grande possibilità di errore se rifiuto  $H_0$ , e quindi un buon fit.

Al contrario, un piccolo p-value implica un rischio trascurabile di rifiuto  $H_0$  e quindi un fit "povero".

Più stretto è l'intervallo di confidenza e meglio è.

Non sempre ci troviamo in situazioni a singola distribuzione, ma può capitare che quelle in uscita sia la somma di distribuzioni diverse e questo si può osservare quando si fanno più mode o quando il comportamento a destra o sinistra della mode non è sempre crescente o decrescente. Prendo "pezzi" della distribuzione in uscita e ad ognuno adotto una

distribuzione adeguata. Dico che la distribuzione somma delle varie distribuzioni è una combinazione lineare, ripetute per la quantità di dati usati per fitare la distribuzione. È una **DISTRIBUZIONE MULTI MODALE** di cui è facile fare il FITTING, non il TEST. Il secondo problema è quello dei dati non indipendenti.

Si può operare così:

1. Partizionare i ~~vari~~ dati del campione  $n$  in distribuzioni unimodali del sotto campione  $n_1, n_2, \dots, n_m$
2. Fitare le distribuzioni in modo separato per ogni sotto-



Le variabili casuali della simulazione rappresentano una delle traiettorie che il sistema può seguire. Una sola replica potrebbe non bastare per o non essere sufficiente per capire se la strada presa è quella giusta. L'idea è quella di fare un esperimento statistico cioè costruire un campione di misure di performance con de usero per fare una **media campionaria** dove il campione è il risultato delle singole ripetizioni.  $E[Y] = \mu$

Del punto di vista statistico abbiamo punti campionati da una distribuzione e se ne faccio la media campionaria, so della statistica che è uno stimatore del valore atteso della popolazione del throughput.

Con l'analisi degli input si vanno a definire le **distribuzioni di probabilità** degli input. L'output è anch'esso una variabile ~~casuale~~ casuale, funzione delle variabili casuali in ingresso di cui è difficile dare una distribuzione. Posso farlo avendo tante repliche, tanti valori in uscita e andando a fare istogrammi, Q-Q plot e cercando di definire la distribuzione.

NON SERVE molto perché il mio capo vorrà un numero in uscita e poi non mi soffermo sulla ~~distribuzione~~ distribuzione della misura in uscita, ma cerco di stimarlo nel modo più preciso possibile ovvero trovo un intervallo di confidenza.

La prima possibilità che ci dà ARENA è quella di fare la **media**  $\hat{\mu} = \bar{Y} = (1/n) \sum Y_i$

Per fare l'**intervallo di confidenza** dobbiamo sapere il livello di confidenza e bisogna stimare l'ampiezza in modo tale che sia il livello di confidenza la probabilità



$$\text{Var}[\bar{Y}] = \text{Var}\left[\frac{\sum_i Y_i}{n}\right] = \frac{1}{n^2} \sum_i \text{Var}[Y_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Sostituisco in quest'ultima formula lo stimo  $\hat{\sigma}^2$  che conosco, ed ottengo

$$\text{Var}[\bar{Y}] = \frac{\sum_i (Y_i - \bar{Y})^2}{n(n-1)}$$

## Distribuzione T

Assumiamo di avere  $n$  campioni da un esperimento di simulazione  $(Y_1, \dots, Y_n)$ ; se sono distribuiti normalmente la statistica  $t$

$$t = \frac{\hat{\mu} - \mu}{\hat{\sigma}(\hat{\mu})} = \frac{\bar{Y} - E[Y]}{S/\sqrt{n}}$$

è distribuita secondo la  $t$  con  $n-1$  gradi di libertà; se  $Y_1, \dots, Y_n$  non sono normalmente distribuiti, la statistica  $t$  di sopra è approssimativamente distribuita secondo una  $t$  grazie al **Teorema del Limite Centrale**, che richiede una grande numerosità campionaria.

$$\hat{\sigma}^2 = S^2 = \frac{\sum_i (Y_i - \bar{Y})^2}{n-1} \quad \frac{Y - \mu}{S/\sqrt{n}} \approx t_{n-1}$$

$$P\left[\bar{Y} - t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \leq \mu \leq \bar{Y} + t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}\right] = 1 - \alpha$$

$$c = t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

Cosa devo fare nella **SIMULAZIONE A TEMPO TERMINE**

- Fisso il livello di confidenza e scelgo il numero  $n$  di repliche da fare in modo da avere un certo valore di ampiezza  $c$ .

Bisogna stare attenti ad alcuni punti:



Fissiamo lo  $c$  che è quello che vogliamo ottenere dalla simulazione.

Scegliamo un livello di confidenza  $1-\alpha$

Calcoliamo  $n$  tramite:

$$c = t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \Rightarrow n = \left( \frac{t_{n-1, 1-\frac{\alpha}{2}}}{c} \right)^2 S^2$$

Nell'algoritmo **1 due passi** fisso un numero di repliche, che devono essere almeno pari a 5 su cui calcolare lo  $S$ . Nel passo 2 determino lo  $n$  usando lo  $S$  precedente e scegliendo, ~~ho~~ sulle tavole, la  $t$  che corrisponde all' $n_0$  (numero di repliche pilota fissate nel punto precedente).

Il problema è che il calcolo dello  $n$  include la statistica  $t$  delle tavole che ha un numero di gradi di libertà pari a quello iniziale.

[Il metodo a 2 passi prevede: **scegli  $n_0$ , fissa  $\alpha$ , cerca  $t$ , fissa  $c$ , fai il calcolo, confronta se  $n \geq n_0$** ]

Metodo **iterativo** il quale mi impone di eseguire almeno 2 repliche perché per stimare la varianza 1 solo non basta. Calcolo lo  $S$ . Calcolo la  $t$ , quindi l'ampiezza dell'intervallo di confidenza, ~~del~~ risultato.

Una volta calcolato il  $c$  con l' $n$ , numero di repliche, lo si confronta con il  $c$  desiderato. Se  $c(n) \leq c^*$  ci si ferma e si hanno il numero di repliche adeguate.

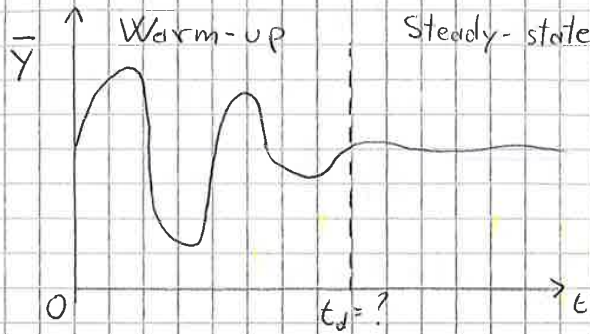
In questa situazione si ha coerenza totale in tutto il modello e si hanno  $n$  stime dello  $S$  ed  $n$  dello  $t$ .

Cosa succede se la simulazione mi serve per spiegare una ~~variabile~~, **PROBABILITÀ** (di questo, servono occupato...)?



transitorio iniziale perché se il sistema stabilizza il suo comportamento dopo un certo tempo, all'inizio si comporterà in modo diverso.

Rappresentandolo graficamente la nostra misura di performance, avremo



L'**OBIETTIVO** dell'output analysis, quando parliamo di simulazione all'equilibrio, è **IDENTIFICARE** quando ciò avviene. Questo perché non voglio tenere in considerazione della parte iniziale del transitorio: a me servono solo i valori all'**EQUILIBRIO**.

Il numero di repliche non attenua questo problema. Le soluzioni sono: - non considerare i dati facenti parte del transitorio - allungare la simulazione tantissimo.

Il transitorio comprende in sé tante distribuzioni diverse: dall'inizio fino all'equilibrio la distribuzione dei miei dati CAMBIA continuamente fino ad assestarsi. Le distribuzioni nel transitorio hanno una variabilità molto più elevata di quella che mi aspetto di trovare all'equilibrio.

Vediamo come calcolare lo **STIMA PUNTUALE** nella simulazione all'equilibrio.

Abbiamo le solite misure di una singola replica:

$$Y_{warm-up}, Y_d \quad , \quad Y_{steady-state}, Y_m$$

warm up                      steady state



**2 Euristiche** che sono utili se dobbiamo automatizzare il processo perché ci indicano in che punto finisce il transitorio. (esempio regola **MSER**)

**3 Metodi QUANTITATIVI** che sono rigorosi, quasi esatti, ma molto complessi

## METODO DI WELCH

Si fanno **ALMENO 5** repliche e per ognuna collezioniamo un certo numero di osservazioni  $Y_{ij}$  ( $i \rightarrow$  repliche  $j \rightarrow$  osservazioni)

Una volta raccolte queste informazioni faccio una media, come per dire "misuro il tempo di attesa di tutti i miei clienti" e poi ne faccio la media.

$$\textcircled{1} Y_{11}, Y_{12}, Y_{13}, \dots, Y_{1m}$$

$$\textcircled{2} Y_{21}, Y_{22}, Y_{23}, \dots, Y_{2m}$$

...

$$\textcircled{n} Y_{n1}, Y_{n2}, Y_{n3}, \dots, Y_{nm}$$

$$\hline \bar{Y}_1 \quad \bar{Y}_2 \quad \bar{Y}_3 \quad \dots \quad \bar{Y}_m$$

A questo punto calcolo una media mobile su una determinata finestra  $w$  di spostamento, ~~che~~ "scorrendo" (devo scegliere l'ampiezza, solitamente  $w \leq m/4$ ). Una volta scelta l'ampiezza, il numero di repliche e la media, la guardiamo e vediamo quando secondo noi converge ad un preciso valore ovvero le oscillazioni vengono smorzate.

Si andrà a plottare come si muove il tempo medio del cliente 1, 2, ..., andrà a vedere la media sui tempi medi dei primi  $x$  clienti all'interno della finestra e



$$Y_3(3) = \frac{1}{2 \cdot 3 - 1} \sum_{k=-2}^2 \bar{Y}_{3+k} = \frac{1}{5} (\bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3 + \bar{Y}_4 + \bar{Y}_5)$$

$$Y_4(3) = \frac{1}{2 \cdot 3 + 1} \sum_{k=-3}^3 \bar{Y}_{4+k} = \frac{1}{7} (\bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3 + \bar{Y}_4 + \bar{Y}_5 + \bar{Y}_6 + \bar{Y}_7)$$

$$Y_5(3) = \frac{1}{2 \cdot 3 + 1} \sum_{k=-3}^3 \bar{Y}_{5+k} = \frac{1}{7} (\bar{Y}_2 + \bar{Y}_3 + \bar{Y}_4 + \bar{Y}_5 + \bar{Y}_6 + \bar{Y}_7 + \bar{Y}_8)$$

$$Y_6(3) = \frac{1}{2 \cdot 3 + 1} \sum_{k=-3}^3 \bar{Y}_{6+k} = \frac{1}{7} (\bar{Y}_3 + \bar{Y}_4 + \bar{Y}_5 + \bar{Y}_6 + \bar{Y}_7 + \bar{Y}_8 + \bar{Y}_9)$$

$$Y_7(3) = \dots \quad Y_8(3) = \dots \quad Y_9(3) = \dots$$

Metto su un simulatore, faccio la media delle medie tramite Welch e poi osservo il grafico. Il **problema** è capire dove finisce il transitorio.

Quindi nel metodo di Welch:

- **m** deve essere grande abbastanza così che tutte le fonti di variabilità si sono manifestate
- bisogna plotare un gran numero di valori di  $w$  e sceglierne il valore più piccolo per il quale il plot corrispondente è "ragionevolmente smorzato"
- la scelta di **d** è soggettiva
- si deve aumentare **n** se non si riesce ad identificare **d**

## MSER (Marginal Standard Error Rule)

Questo metodo si applica solitamente alla SINGOLA REPLICA. Cosa fa?

Sceglie il transitorio iniziale risolvendo un problema di ottimizzazione ovvero: su un'unica replica in cui vengono raccolte **n** osservazioni, viene scelto un **d**



o più piccolo della varianza vera di un termine che contiene il numero di osservazioni sulle quali faccio il calcolo ed il coefficiente di autocorrelazione

$$E[S^2] = \sigma^2 \left[ 1 - 2 \frac{\sum_{j=1}^{m-1} (1 - \frac{j}{m}) \rho_j}{m-1} \right]$$

DA NON  
IMPARARE  
A MEMORIA

Se  $\rho > 0$ , la varianza stimata con quelle osservazioni è più piccola della varianza vera.

Il problema che sorge ora è: se c'è correlazione posso trovare lo stimatore puntuale ~~per~~ la media non la bias, non lo stimatore intervallato perché tra varianza e bias trovo già un intervallo di confidenza.

Ci sono vari modi per aggirare il problema, uno di questi è il **METODO BATCH DELLE MEDIE**

ovvero non metto tutto insieme, calcolo la media su sottoinsieme di dati in modo da ammassare la "questione" autocorrelazione: divido le mie osservazioni in un insieme ed ognuno lo rappresento con il valore medio (non affetto da correlazione). Quando vedo e stimare la varianza guardando soltanto i pochi punti considerati e la loro media, SE lo scatto bene gli intervalli, lo **eliminato** la correlazione.

La penalità che si paga a priori è in primis il non considerare le statistiche che mi dicevano essere correlazione; devo anche scegliere correttamente la dimensione del batch e il suo numero.

Come identifico questo intervallo?

**ALGORITMO DI LAW & CARSON** (euristico)

1. Sette dei parametri ( $j=1, n_0=800, n_1=800$ ) e raccogliere  $n_1$  osservazioni



## COMMON RANDOM NUMBERS (CRN)

Si usano quasi esclusivamente per confrontare alternative di sistema, quindi potrei volere il metodo che si sarebbe potuto scegliere per minimizzare la varianza del costo della mensa.

Quando confronto **configurazioni alternative** l'obiettivo è quello di dire che se ho due sistemi con lo stesso input, devo vedere quale funziona meglio.

Dare le stesse distribuzioni in ingresso **NON** garantisce di avere gli stessi numeri casuali utilizzati allo stesso scopo: il confronto che ne risulta è "zoppicante".

Con la riduzione della varianza, oltre che a mantenere sincronizzato le due simulazioni, devo ridurre la varianza delle misure in output. Perché voglio poter confrontare quello che esce dai due sistemi pensando che sia solo il frutto del comportamento del sistema e non di una elezione "messa sotto".

Questa tecnica tende ad involvere una **correlazione** tra le configurazioni, infatti viene anche chiamata **correlated sampling** per abbattere la varianza.

### Fondamento logico

Si fanno due configurazioni alternative  $X_1$  e  $X_2$ , in cui si sono fatte un certo numero di repliche;  $X_{1j}$  e la performance ottenute nella  $j$ -esima replica della configurazione 1 e  $X_{2j}$  lo stesso per la seconda configurazione.

Confrontando vogliamo stimare ~~una~~ la differenza, in valore atteso, della performance medio di quelle dei due sistemi:

Se la differenza è sostanzialmente nulla, vuol dire che anche con il cambiamento non ottengo alcun impatto sulla



introdurre la correlazione e diminuire la varianza.

Il problema che sorge è come fare ad avere correlazione positiva.

**NON È POSSIBILE** dire a priori, in modo indipendente dal sistema che stiamo simulando come si fa ad avere correlazione positiva, se è possibile farlo e, nel caso, se i "trucchi" sono efficaci ovvero se veramente la varianza si riduce e di quanto.

## Applicabilità

- L'aumento dell'efficienza dipende molto dai sistemi di voto e confronto e soprattutto dal fatto che rispondono allo stesso modo al valore delle variabili casuali di genere. (Se genero un tempo di processo elevato, l'utilizzo aumenta, se genero un tempo di processo più basso, l'utilizzo diminuisce e questo deve accadere in entrambe le simulazioni). Devo sapere il legame della misura di performance con le variabili casuali in ingresso e la "risposta" deve essere la stessa nelle due alternative di voto e confronto altrimenti non so come controllare la simulazione.

- Da quello appena descritto dobbiamo controllare la generazione dei numeri casuali; avere lo stesso numero casuale che ha generato la stessa variabile casuale.

Ad ogni variabile casuale deve corrispondere un unico valore di un unico numero casuale e ci deve essere un legame **monotono** tra variabile e numero.

Usando lo stesso numero nelle due simulazioni ottengo un evento ~~casuale~~ sincronizzato e correlato.



## Sincronizzazione

Perché AV lavori bene,  $R_1$  ed  $R_2$  devono essere sincronizzati. Anche qui valgono le stesse voluzioni del CRN.

## CONTROL VARIATES (CV)

Anche i CV considerano sistemi singoli e usano la correlazione (positiva o negativa) per ridurre la varianza dello stimatore.

Cerco le **VARABILI DI CONTROLLO** che sono quelle legate alle nostre misure di performance e di cui conosciamo il legame (utilizzo  $\leftrightarrow$  throughput: se aumenta uno, aumenta l'altro). Usiamo queste variabili perché ci dicono qualcosa sulle stime che stiamo facendo della nostra misura di performance.

Poniamo  $X$  essere una variabile casuale in output e supponiamo di volerla stimare  $\mu = E[X]$ . Sia  $Y$  un'altra variabile casuale, correlata con  $X$  di cui conosciamo il valore atteso  $\nu = E[Y]$ . Se nella simulazione  $Y > \nu$ , è ragionevole pensare che  $X > \mu$  e così possiamo aggiustare, aumentando o diminuendo il valore in quel modo.

La  $Y$  è la variabile di controllo.

## Fondamento logico

Sia  $d$  una costante con lo stesso segno della correlazione tra  $X$  e  $Y$ . Lo **stimatore controllato** è definito come

$$X_c = X - d(Y - \nu)$$

Questo segue direttamente che  $E[X_c] = \mu$ . ad esempio  $X_c$  è uno stimatore senza bias di  $X$  e

$$\text{Var}[X_c] = \text{Var}[X] + d^2 \text{Var}[Y] - 2d \text{Cov}[X, Y]$$



costruire un sistema parallelo o quello desiderato, da cui dico l' $Y$  su cui posso costruire lo stimatore controllato, aggiungo un costo computazionale (forse basso perché il sistema è più semplice, ma devo mantenere la sincronizzazione).

### Determining $d^*$

Dipendendo dalla natura di  $Y$ , il valore  $\text{Var}[Y]$  potrebbe essere ignoto e sicuramente lo è anche  $\text{Cov}[X, Y]$ . Uno dei metodi più semplici per stimare  $d^*$  è rimpiazzare  $\text{Var}[Y]$  e  $\text{Cov}[X, Y]$  con degli stimatori.

$$\hat{C}_{xy}(n) = \frac{\sum_{j=1}^n [X_j - \bar{X}(n)][Y_j - \bar{Y}(n)]}{n-1}$$

$$\hat{d}^*(n) = \frac{\hat{C}_{xy}(n)}{S_y^2(n)}$$

$$\hat{X}_c^*(n) = \bar{X}(n) - \hat{d}^*(n) [\bar{Y}(n) - \mu]$$

$\hat{d}^*$  è un numero casuale e generalmente non è indipendente da  $\bar{Y}(n)$ .

Questo porta direttamente al fatto che, in generale,  $\hat{X}_c^*(n)$  non è uno stimatore senza bias di  $\mu$ .

Quando si hanno tante variabili di controllo bisogna fare tutte le derivate parziali rispetto le varie  $d$  e vengono fuori formule analoghe e quelle che si ottengono applicando il metodo dei minimi quadrati con la regressione.

Per questa ragione, lo **multiple control variables** è anche chiamato **regression sampling**.

### IMPORTANCE SAMPLING (IS)

Differentemente da tutte le tecniche precedenti che cercano di indurre una correlazione di qualche o qualunque



dall'assunzione  $h$  e distribuzione  $F$  alla funzione  $(h \cdot f)/g$  con distribuzione  $g$  perché mi aspetto che sia, variando la  $x$ , quasi costante.

Se esistesse una  $g$  per cui

$$g(x) = \frac{h(x) f(x)}{\mu}$$

si sarebbe ridotta a 0 la varianza.

Come si sceglie  $g$ ?

### Scegliere $g$

La scelta della funzione  $g(x)$  è il vero problema per una significativa riduzione della varianza. Se  $g(x)$  è scelta male, infatti, non c'è garanzia di ridurre la varianza che potrebbe, altrimenti, crescere.

Con questa scelta cambiamo le probabilità delle singole  $x$  in modo da avere un valore atteso che non cambi rispetto l'originario ma che presenti una variabilità diversa.

Scegliere  $g(x)$  non è facile ed esistono solo delle regole generali:

- $g(x) \geq 0$  ogni volta che  $h(x) \neq 0$ ;
- deve essere vicino ad  $h(x) f(x)$  in modo che il rapporto risulti quasi costante;
- deve essere facile simulare valori della  $g$ ;
- dovrebbe essere facile trovare la densità di  $g(x)$  per ogni valore di  $x$ .