



UNIVERSITÀ DI PERUGIA
Dipartimento di Matematica e Informatica



Appunti Simulazione

Autore: Chiara Luchini

Basati su:

- Slides del Prof. Sergio Tasso
- Lezioni del Prof. Sergio Tasso

Anno Accademico 2021/2022

Indice

1	Sistemi e Modelli	5
1.1	Definizione di sistemi e modelli	5
1.2	Classificazione di sistemi e di modelli	8
1.3	Ciclo dei modelli	9
1.4	Schemi di simulazione	10
1.4.1	Schemi di simulazione: processi	10
1.4.2	Schemi di simulazione: eventi	11
1.4.3	Schemi di simulazione: attività	12
1.4.4	Schemi di simulazione: "tre fasi"	12
2	Fondamenti di Calcolo delle Probabilità e Statistica	13
2.1	Variabili stocastiche	13
2.1.1	Funzione densità di probabilità	14
2.1.2	Funzione cumulativa di distribuzione	15
2.1.3	Media	15
2.1.4	Varianza	16
2.1.5	Funzione di probabilità di due variabili aleatorie	16
2.1.6	Distribuzione marginale	17
2.1.7	Indipendenza	18
2.1.8	Variabili completamente dipendenti e variabili stocasticamente correlate	18
2.1.9	Probabilità condizionali	19
2.1.10	Covarianza	19
2.1.11	Coefficiente di correlazione	20
2.2	Generazione di numeri casuali	20
2.2.1	Metodo congruente lineare	20
2.2.2	Test del Chi-quadro X^2	21
2.2.3	Test per la verifica delle proprietà di casualità	22

2.3	Generazione di distribuzioni qualsiasi	22
2.3.1	Metodo della trasformazione inversa	23
2.3.2	Metodo del rifiuto	23
2.3.3	Metodo della composizione	24
2.4	Distribuzioni	25
3	Processi stocastici	26
3.1	Processi stocastici	26
3.1.1	Spazio degli stati E	26
3.1.2	Indice tempo	27
3.1.3	Dipendenza stocastica fra variabili	27
3.1.4	Processi di Markov	28
3.1.5	Analisi stazionaria	29
3.1.6	Processo di pura nascita	32
3.1.7	Processo di Poisson	33
4	Modelli a singola coda	35
4.0.1	Notazione di Kendall	36
4.0.2	Modelli basilari di code: i sistemi M/M/1 ed M/M/m	38
4.0.3	Analisi stazionaria	39
4.0.4	Sistema M/M/ ∞ : serventi infiniti	40
4.0.5	Sistema M/M/m: serventi multipli	41
4.0.6	Sistema M/M/1/K : memoria finita	42
4.0.7	Sistema M/M/1//M: popolazione finita	42
5	Analisi di risultati e convalida in esperimenti di simulazione	43
5.1	Stima dei parametri di una distribuzione	43
5.1.1	Stima della media	43
5.1.2	Stima della varianza	44
5.2	Stima della distribuzione	44
5.2.1	Metodo del coefficiente di variazione	44
5.2.2	Metodo del goodness of fit	45
6	Esercizi	47
6.1	Esercizio su M/M/m	47
6.1.1	Soluzione	47

Elenco delle figure

1.1	Metodi di valutazione delle prestazioni.	6
1.2	Ciclo di vita di un modello.	10
2.1	Test del Chi-quadro.	22
2.2	Trasformazione inversa.	23
2.3	Metodo del rifiuto.	24
2.4	Metodo della composizione.	24
3.1	Esempio di diagramma fra gli stati di una catena di Markov a tempo discreto con matrice delle probabilità di transizione P	29
3.2	Processi di nascita-morte.	31
3.3	Diagramma di un processo di pura nascita.	33
4.1	Sistema di congestione.	35
4.2	Sistema $M/M/1$	38
5.1	Tabella esempio stima distribuzione.	45

Capitolo 1

Sistemi e Modelli

In questo Capitolo vengono introdotti i concetti e una classificazione dei sistemi e dei modelli ed il procedimento di creazione ed uso di un modello al fine di valutare le prestazioni del sistema rappresentato.

1.1 Definizione di sistemi e modelli

Un sistema è rappresentato come un insieme di componenti (elementi, entità) interdipendenti e che interagiscono per raggiungere un determinato obiettivo. Un sistema di elaborazione è composto da componenti software, hardware e firmware che permettono l'elaborazione delle informazioni eseguendo programmi di utente.

Nello studio di un sistema devono essere considerati vari fattori fra cui:

- funzionalità e correttezza;
- affidabilità;
- costo e fattori economici;
- prestazioni.

Le fasi di sviluppo e progettazione di un sistema sono:

- **fase di progettazione:** in questa fase si sceglie il progetto del sistema fra diverse configurazioni disponibili;
- **fase di dimensionamento e acquisizione:** comprende la scelta fra diversi sistemi o componenti disponibili;

- **fase di evoluzione della configurazione e del carico:** si considerano tutti gli aspetti e i problemi relativi alla modifica ed evoluzione di un sistema esistente.

Le metodologie per la valutazione delle prestazioni di sistemi possono essere suddivise in due categorie:

- tecniche di misurazione;
- tecniche modellistiche.

Le prestazioni possono essere quantificate tramite di merito o indici di prestazioni che ne descrivono l'efficienza.

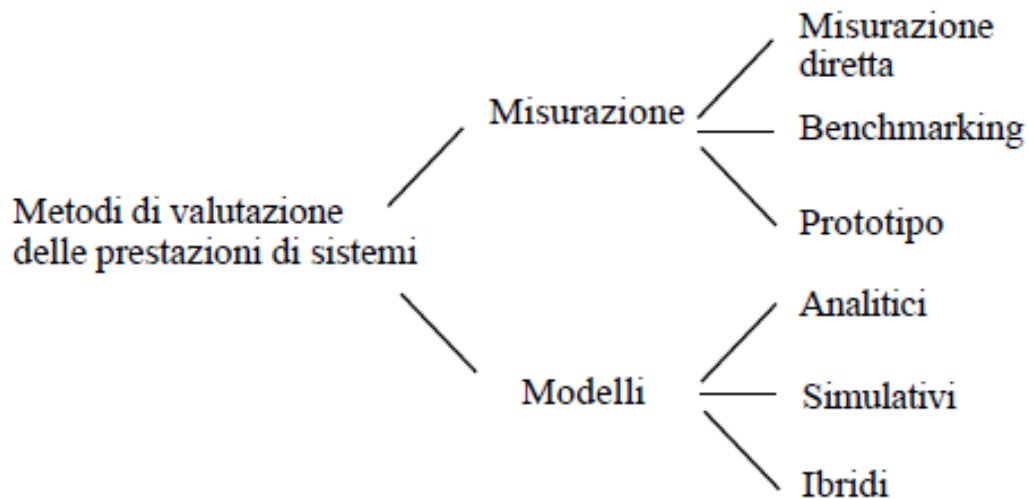


Figura 1.1: Metodi di valutazione delle prestazioni.

Come si può vedere in Figura 1.1 esistono diverse tecniche di misurazione:

- **misurazione diretta:** il sistema viene misurato utilizzando il carico reale;
- **benchmark (carico artificiale):** si ripete più volte il procedimento di misurazione e si effettuano misurazioni comparative fra diversi sistemi sotto le stesse condizioni di carico;
- **prototipazione:** se il sistema non è disponibile si può ricorrere alla costruzione di un prototipo su cui fare le misurazioni.

Definizione modello. Un **modello** è una rappresentazione astratta del sistema che include solo gli aspetti rilevanti che servono per lo studio del sistema. Esso è definito da un certo livello di astrazione, cioè il sistema viene descritto con un certo livello di dettaglio riportando solo le componenti e le interazioni fra quest'ultimi necessari allo scopo prefissato.

Tecniche modellistiche. Si possono distinguere due tipologie di tecniche di modellazione:

- **modello analitico:** variabili e parametri rappresentano le componenti e il carico del sistema, mentre le interazioni sono rappresentate tramite delle relazioni fra queste quantità;
- **modello di simulazione:** replica il comportamento dinamico di un sistema nel tempo rappresentando le componenti e le interazioni in termini di relazioni funzionali. La valutazione di quest'ultimo richiede l'esecuzione di un programma di simulazione o simulatore.

Vantaggi e svantaggi di un modello. I principali vantaggi nell'utilizzo di un modello per lo studio di un sistema sono:

1. aumento e organizzazione delle conoscenze;
2. analisi del sistema;
3. modificabilità;
4. diversi obbiettivi di studio.

Invece gli svantaggi sono i seguenti:

1. scelta del modello;
2. uso errato del modello.

I modelli che si basano su **processi stocastici** permettono di valutare la dinamica dei sistemi e in particolare le loro prestazioni e affidabilità.

I **modelli a reti di code** permettono di rappresentare **sistemi di congestione**, i quali sono sistemi formati da un insieme limitato di risorse e in cui si osserva competizione per il loro utilizzo da parte di un insieme di utenti. Alcuni esempi di questi sistemi sono i sistemi di calcolo, di comunicazione, di traffico e di produzione. La

valutazione delle prestazioni di un sistema di congestione include due differenti aspetti: dal punto di vista del sistema, si è interessati alla valutazione della utilizzazione delle risorse, mentre dal punto di vista dell'utente si valutano i tempi di attesa per l'uso delle risorse.

1.2 Classificazione di sistemi e di modelli

L'evoluzione nel tempo di un sistema è descritto dallo **stato** del sistema in ogni istante che ne rappresenta la condizione. Lo stato è espresso come **variabili di stato** che ne descrivono le entità e i loro attributi. Le attività delle componenti nel tempo e le interazioni fra queste sono descritte dalle **regole di trasformazione** fra stati. Il comportamento di un sistema nel tempo è rappresentato dalla **storia degli stati**. Le variabili di stato si distinguono in:

- **variabili endogene:** il loro cambiamento è dovuto soltanto ad attività interne al sistema. Il sistema si dice chiuso se il suo comportamento è determinato da queste variabili.
- **variabili esogene:** il loro cambiamento è influenzato dall'ambiente esterno al sistema. Il sistema si dice aperto se interagisce con l'ambiente esterno, utilizza anche variabili esogene.

I sistemi possono essere suddivisi in **continui** e **discreti** a seconda del tipo di cambiamento dei valori delle variabili di stato, ad esempio se controlliamo la temperatura in un luogo la variabile che la rappresenta è continua poiché verifichiamo i cambiamenti nel tempo.

Il modo in cui avvengono le trasformazioni fra stati determina se un sistema è **deterministico** o **stocastico**, nel primo caso le regole di trasformazione determinano il cambiamento del sistema in modo univoco mentre nel secondo si può passare da uno stato a un altro secondo diverse leggi/distribuzioni di probabilità. La natura del sistema dipende anche dalla visione dell'osservatore che è definita dagli obiettivi e dal metodo di studio. A loro volta anche i modelli possono essere distinti in queste categorie con la differenza che un modello associato un sistema non deve necessariamente corrispondergli, esempio un sistema deterministico definito da un modello stocastico. I modelli si distinguono in due principali categorie:

- modelli fisici;
- modelli simbolici: questi includono i modelli matematici e non.

1.3 Ciclo dei modelli

Il procedimento di creazione di un modello è un processo iterativo di raffinamenti successivi, vengono fatte delle assunzioni ed ipotesi da verificare e valutare. Il processo di creazione ed uso di un modello può essere schematizzato come segue:

1. **Definizione degli obiettivi:** definizione e comprensione del sistema oggetto di studio e delle sue componenti. Si stabiliscono anche i criteri di valutazione delle soluzioni e si acquisiscono i dati dal sistema misurandone anche il carico.
2. **Definizione del modello e formulazione delle assunzioni e ipotesi:** vengono identificate le componenti del modello così come le assunzioni ed ipotesi utilizzate.
3. **Parametrizzazione:** sono identificate le variabili del modello da misurare e gli strumenti di misurazione.
4. **Valutazione (soluzione) del modello e interpretazione dei risultati:** definito e parametrizzato il modello si sceglie il metodo di soluzione più appropriato.
5. **Validazione del modello e valutazione dei risultati:** si valuta il modello nella descrizione del sistema se questo risulta non soddisfacente si itera ai passi 1, per ridefinire il sistema e gli obiettivi, o 2, per modificare la definizione del modello e delle ipotesi, o 3, per ri-parametrizzare le variabili.
6. **Documentazione e analisi della sensitività:** riporta i risultati trovati compresi i dettagli relativi al modello e anche una descrizione del processo di definizione del modello.

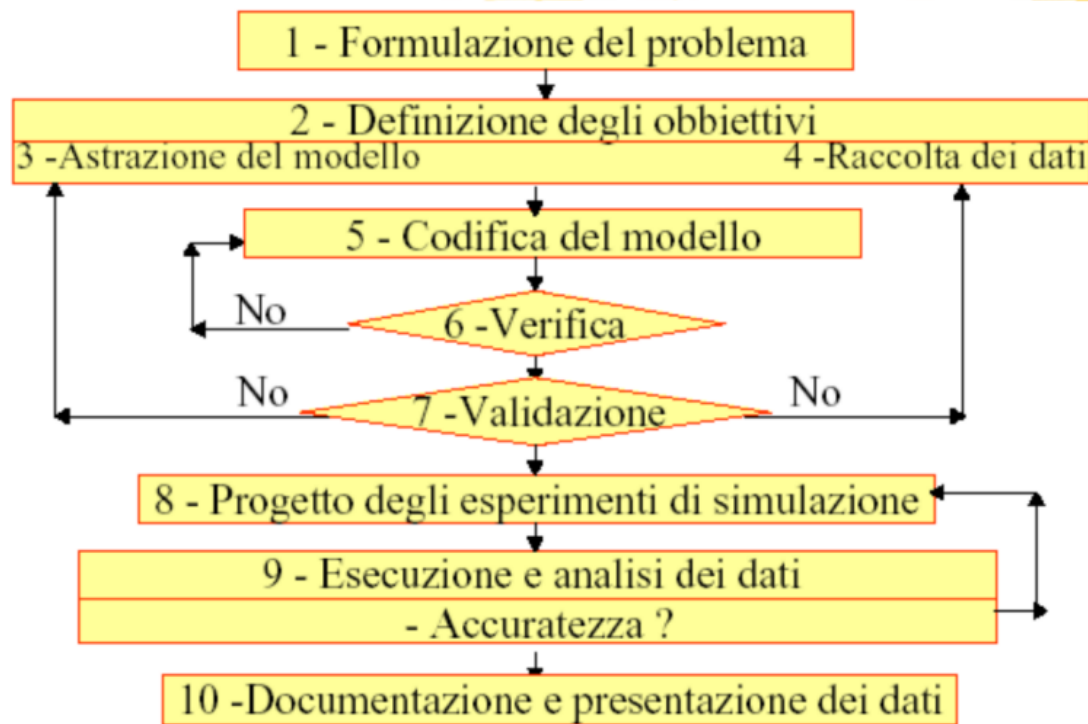


Figura 1.2: Ciclo di vita di un modello.

1.4 Schemi di simulazione

Sono delle strutture di modelli di simulazione che differiscono in base ai tre concetti di evento, attività e processo. Ne esistono quattro metodologie:

- Interazione fra processi
- Scheduling di eventi
- Scansione di attività
- Metodo a Tre fasi

1.4.1 Schemi di simulazione: processi

Il flusso di esecuzione di un processo permette di simulare il comportamento di un oggetto (entità) attraverso il sistema. L'esecuzione del processo continua fintanto che non viene bloccato, sospeso o arriva un'altra entità. Nel caso in cui il flusso viene bloccato, si avanza nel tempo fino ad arrivare al punto di inizio dell'esecuzione della prossima attività.

Esempio: scheduling dei processi In un sistema di elaborazione ogni processo viene eseguito ciclicamente supponiamo che si utilizzi la disciplina FIFO per gestire la coda di processi, i passi eseguiti sono i seguenti:

1. Esamina la coda, se vuota vai a 5
2. Preleva un job dalla coda
3. Esegui il servizio per il job
4. Vai al passo 1
5. Attendi l'arrivo del prossimo utente, quando arriva vai al passo 1

Il punto 4 è un punto di sospensione o di riattivazione del processo, al contrario nel passo 5 il processo è in stato passivo fino a che non arriva un job. In generale nella simulazione fra processi esiste una routine che gestisce i cambiamenti di stato dei processi, essa controlla due code diverse:

- **processi bloccati:** contiene gli oggetti sospesi
- **processi sospesi:** contiene gli oggetti passivi

La routine permette di riattivare un oggetto, il quale può essere:

- **attivo:** esegue azioni
- **sospeso:** ha un tempo di riattivazione
- **passivo:** non è attivo e il tempo di riattivazione dipende da un altro oggetto
- **terminato:** ha terminato il suo lavoro

1.4.2 Schemi di simulazione: eventi

Nella simulazione per scheduling di eventi si avanza il tempo simulato al tempo del prossimo evento che coincide spesso con il termine di un'attività o l'inizio. In questo caso gestiamo una lista ordinata di eventi e scheduler di eventi. Lo scheduler di eventi gestisce il tempo simulato mantenendo una lista ordinata di eventi futuri:

- **event-driven:** $clock = \text{tempo del prossimo evento in lista}$
- **unit-time:** $clock = clock + \Delta$

Inoltre esso gestisce anche la struttura dati che contiene tutti gli eventi e anche la routine evento. Quest'ultima permette di aggiornare sia la lista di variabili di stato che la lista di eventi, essa può essere di:

- **inizializzazione**: chiamata per prima e definisce le variabili di stato ed altro;
- **gestione degli eventi**;
- **trace**: serve per notificare eventi o stime a run-time;

1.4.3 Schemi di simulazione: attività

La simulazione avviene scansionando le attività ovvero descrivendone le componenti del modello. Se è applicato il meccanismo di avanzamento per intervalli fissi Δ , ad ogni avanzamento vengono esaminate tutte le attività per stabilire le condizioni di inizio e di fine. Le procedure aggiornano lo stato della singola attività. Si può anche applicare un meccanismo di avanzamento del tempo distribuito nelle procedure delle attività simile all'avanzamento per eventi.

1.4.4 Schemi di simulazione: "tre fasi"

Nella simulazione a "tre fasi" si ha un avanzamento del tempo per intervalli fissi, le varie fasi che la compongono sono:

- fase 1: avanzamento del tempo simulato;
- fase 2: rilascio risorse mantenute dalle attività che risultano terminate dopo l'avanzamento;
- fase 3: esecuzione delle attività per la quali siano ora disponibili le risorse.

Capitolo 2

Fondamenti di Calcolo delle Probabilità e Statistica

In questo Capitolo verranno riportate alcune nozioni base di Probabilità e Statistica applicabili nel campo della simulazione informatica, indispensabili per la costruzione e l'uso di modelli di simulazione stocastica.

2.1 Variabili stocastiche

Un modello di simulazione è un modello intrinsecamente stocastico. Infatti i valori usati per le variabili di ingresso, di stato iniziale e i parametri sono decisi a partire dal sistema reale attraverso delle misurazioni fisiche, ogni valore appartiene a una distribuzione di probabilità. Perciò abbiamo bisogno di ripassare alcune nozioni base della probabilità in modo da comprendere i valori estratti da queste distribuzioni.

Una variabile stocastica o aleatoria x_i rappresenta l'uscita di un'attività casuale. Essa può assumere n differenti valori dato $i = 1, 2, \dots, n$, ciascuno con una determinata probabilità $p_i(x) = p_1, p_2, \dots, p_n$. L'insieme di tutte le probabilità è detto **funzione discreta di probabilità**. Da ricordare che la somma di tutte le singole probabilità da come risultato 1, ovvero:

$$\sum_{i=1}^n p_i(x) = 1$$

Esse possono essere definite anche nel seguente modo.

Uno spazio di probabilità è una tripla (Ω, \mathcal{F}, P) dove:

- Ω è lo spazio campione, il quale è un insieme di elementi (spesso l'insieme dei possibili esiti);
- \mathcal{F} è lo spazio degli eventi, il quale è una famiglia di sottoinsiemi di Ω che ha le seguenti proprietà:
 1. $\Omega \in \mathcal{F}$,
 2. $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$,
 3. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$.
- $P : \mathcal{F} \rightarrow [0, 1]$ è la funzione di probabilità (definita come $p(x)$ in 2.1), la quale è una funzione reale con la seguenti proprietà:
 1. $P(A) \geq 0, \forall A \in \mathcal{F}$,
 2. $P(\Omega) = 1$,
 3. $(A, B \in \mathcal{F}) \wedge (A \cap B = \emptyset) \implies P(A \cup B) = P(A) + P(B)$.

Dato uno spazio di probabilità, una variabile casuale è una funzione $X : \Omega \rightarrow \mathbb{R}$ che ha la seguente proprietà, per ogni reale r , $\{\omega \in \Omega : X(\omega) \leq r\} \in \mathcal{F}$. La funzione $F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$, definita sull'insieme dei reali, è detta **funzione di distribuzione**. Le variabili stocastiche o casuali vengono utilizzate poiché nei sistemi da simulare spesso si presentano degli eventi non facilmente prevedibili a priori, come ad esempio l'arrivo di clienti ad uno sportello o la quantità di pioggia in una determinata stagione. Perciò tali fenomeni vengono rappresentati tramite queste variabili dalle quali estrarne poi la distribuzione di probabilità.

Esempio Si consideri il numero di pazienti che si presentano ad un ambulatorio tra le 9 e le 10 di mattina, e poniamo $\Omega = 0, 1, 2, \dots$, $\mathcal{F} = 2^\Omega$, e $X(\omega) = \omega$ (la funzione identità). La funzione X così definita è una variabile casuale, infatti, per ogni reale r è

$$\{\omega \in \Omega : X(\omega) \leq r\} = \{0, \dots, \lfloor r \rfloor\} \in \mathcal{F}$$

2.1.1 Funzione densità di probabilità

La **funzione densità di probabilità** $f(x)$ si utilizza per definire la probabilità di una variabile di assumere uno specifico valore tra infiniti valori, la quale è praticamente nulla, a causa del processo osservato che è continuo. Detto ciò essa si può definire come

la probabilità che $x_1 \leq x \leq x_2$, ovvero che il valore di x sia compreso in intervallo $[x_1, x_2]$, data da:

$$p(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x)dx = \int_{-\infty}^{+\infty} f(x)dx = 1$$

2.1.2 Funzione cumulativa di distribuzione

La **funzione cumulativa di distribuzione** $F(x)$ definisce la probabilità che un certo valore sia minore o uguale a x , la quale si può descrivere come:

$$F(x) = \int_{-\infty}^x f(x)dx$$

Da cui risulta che $0 \leq F(x) \leq 1$ e $p(x_1 \leq x \leq x_2) = F(x_2) - F(x_1)$.

2.1.3 Media

Per una variabile discreta ¹ si definisce il **valore atteso** o **media** $E(x)$ oppure μ_x come:

$$E(x) = \sum_{i=1}^n x_i p_i(x)$$

Si definisce media di una funzione $y = g(x)$:

$$E(g(x)) = \sum_{i=1}^n g(x_i) p_i(x)$$

Per una variabile continua si ha che la media è pari a:

$$E(x) = \int_{-\infty}^{+\infty} x f(x)dx = \lambda_1 = \lambda$$

e per una funzione continua y :

$$E(x) = \int_{-\infty}^{+\infty} g(x) f(x)dx = \int_{-\infty}^{+\infty} g(x) dF(x)$$

Altro...

¹Una variabile casuale è detta **discreta** se l'insieme di valori che può assumere è numerabile.

2.1.4 Varianza

Si definisce **varianza** di x e si indica con $\sigma^2(x)$, è la media degli scarti quadratici rispetto alla media μ_x e rappresenta una misura di dispersione di x . La sua radice quadrata, σ_x è detta **deviazione standard**. La varianza è definita come

$$\sigma^2(x) = E(x - \lambda)^2 = \int_{-\infty}^{+\infty} (x - \lambda)^2 f(x) dx$$

Da cui si ricava facilmente

$$\sigma^2(x) = E(x^2) - E^2(x)$$

Se x è una **variabile discreta** si ha

$$\sigma^2(x) = \sum_{i=1}^n (x_i - E(x))^2 \cdot p_i(x)$$

La deviazione standard viene definita come

$$\sigma(x) = \sqrt{\sigma^2(x)}$$

Teorema di Beniamé-Chebychev

Il Teorema di Beniamé-Chebychev lega la deviazione standard di x alla probabilità di deviazione dei singoli valori di x della media:

$$p(|x - \lambda| \geq k\sigma) \leq \frac{1}{k^2}$$

Ovvero, qualunque sia la forma di $f(x)$, la probabilità al di fuori di $\pm k\sigma$ è limitata, $\leq \frac{1}{k^2}$.

2.1.5 Funzione di probabilità di due variabili aleatorie

In modo analogo al precedente si definisce la funzione di distribuzione $F(X, Y)$ di due variabili aleatorie x e y

$$F(X, Y) = p(x \leq X, y \leq Y)$$

che indica la probabilità che $x \leq X$, valore prefissato, e $y \leq Y$. La suddetta funzione è detta anche **funzione di probabilità congiunta** e scritta $F(x, y)$. Da essa possiamo ottenere la definizione di **densità congiunta di probabilità**. Nel caso di variabili

discrete data l'esistenza di un insieme numerabile di punti

$$(X_1, Y_1), (X_1, Y_2), \dots, (X_2, Y_2), \dots, (X_i, Y_j)$$

con associati numeri positivi

$$p_{11}, p_{12}, \dots, p_{21}, p_{22}, \dots, p_{ij}$$

i quali soddisfano la relazione $F(X_h, Y_k) = \sum_i \sum_j p_{ij}$ con $\sum_i \sum_j p_{ij} = 1$ sommata per tutti gli i e j per i quali $X_i \leq X_h$ e $Y_j \leq Y_k$. Allora si può definire $p = (x = X_i, y = Y_j) = p_{ij}$ per tutti gli i e j per i quali esistono dei valori delle variabili x e y, assumendo $p_{ij} = 0$ altrove. Quindi la **funzione discreta di probabilità congiunta** è $f(x_i, y_j) = p_{ij}$ con distribuzione cumulativa $F(x_i, y_j)$.

Se invece la funzione $F(x, y)$ è continua si definisce la **densità congiunta**

$$f(x, y) = \frac{\lambda}{\lambda x} \frac{\lambda}{\lambda y} F(x, y)$$

e quindi

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(x, y) dx dy$$

$$p(a \leq x \leq b, c \leq y \leq d) = \int_c^d \int_a^b f(x, y) dx dy$$

$$p(X \leq x \leq X + dX, Y \leq y \leq Y + dY) = F(X, Y) dX dY$$

2.1.6 Distribuzione marginale

Si vuole determinare la probabilità $g(x)$ e $h(y)$ di una variabile x o y, data la densità congiunta $f(x, y)$ di due. In poche parole la **distribuzione marginale** di un sottoinsieme di una collezione di variabili casuali è la distribuzione di probabilità delle variabili contenute nel sottoinsieme. Ciò si interpreta dicendo che si vuole la $p(x \leq X, \text{con } y \text{ qualunque})$ simboleggiata con $F(x, \infty)$ o $F(\infty, y)$. Nel caso discreto si ha

$$F(x_h, \infty) = \sum_i \sum_j p_{ij}$$

Nel caso continuo si ha invece

$$F(x, \infty) = \sum_{-\infty}^X \sum_{-\infty}^{\infty} f(x, y) dx dy = \int_{-\infty}^X g(x) dx$$

Da cui si ha

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

dove $g(x)$ e $h(y)$ sono dette **distribuzioni marginali** di x e y .

2.1.7 Indipendenza

Due variabili aleatorie x e y sono dette **indipendenti** se, detta $f(x, y)$ la loro densità congiunta e $g(x)$ e $h(y)$ le relative distribuzioni marginali si ha:

$$f(x, y) = g(x) \cdot h(y)$$

La condizione necessaria e sufficiente alla validità della formula sopra è che $f(x, y)$ può essere fattorizzata nel prodotto di due funzioni, ovvero:

$$f(x, y) = r(x) \cdot s(y)$$

2.1.8 Variabili completamente dipendenti e variabili stocasticamente correlate

Per coppie di variabili dipendenti si ha che $f(x, y) = g(x) = h(y)$, ovvero la probabilità $f(x, y)$ è uguale a 0 solo per coppie x, y dove y è una funzione ad un solo valore di x e viceversa. In tutti i casi in cui $f(x, y)$ non è né il prodotto $g(x) \cdot h(y)$ né tale che $g(x) = h(y) = f(x, y)$ le variabili x e y si dicono stocasticamente correlate. Il coefficiente di correlazione misura il grado di correlazione il quale è pari a:

- 0 se e solo se $f(x, y) = g(x) \cdot h(y)$
- $|1|$ se e solo se $f(x, y) = g(x) = h(y)$
- $0 \leq |x| \leq 1$ in tutti gli altri casi.

2.1.9 Probabilità condizionali

La **probabilità condizionale** identifica la probabilità che una variabile x assuma il valore x_i , assunto che y assuma il valore y_j si scrive:

$$p(x = x_i | y = y_j) = \frac{p(x = x_i, y = y_j)}{p(y = y_j)} = \frac{f(x_i, y_j)}{h(y_j)} = g(x_i | y_j)$$

Si possono scrivere le funzioni densità condizionali di probabilità di x dato y e di y dato x :

$$g(x|y) = \frac{f(x, y)}{h(y)}$$
$$h(y|x) = \frac{f(x, y)}{g(x)}$$

Nel caso di variabili indipendenti di ha:

$$g(x|y) = g(x)$$
$$h(y|x) = h(y)$$

2.1.10 Covarianza

La covarianza è una misura della relazione tra due variabili x e y , essa viene definita come di seguito:

$$\text{cov}(x, y) = E(xy) - E(x)E(y)$$

Dalla definizione risulta che:

- $\text{cov}(x, y) > 0$ se, nella funzione $f(x, y)$ grandi valori di x sono associati a grandi valori di y e viceversa;
- $\text{cov}(x, y) < 0$ se piccoli valori di x sono associati a grandi valori di y e viceversa;
- $\text{cov}(x, y) = 0$ se, quando x è grande, alcuni valori di y sono grandi e alcuni piccoli.

2.1.11 Coefficiente di correlazione

Il **coefficiente di correlazione** $p(x,y)$ tra due variabili si definisce come la misura standardizzata della covarianza e si scrive come:

$$p(x,y) = \frac{cov(x,y)}{\sigma(x)\sigma(y)}$$

Esso misura il grado di dipendenza lineare tra x e y . Il suo valore è compreso tra $(-1,1)$, raggiungendo ± 1 quando esiste una dipendenza perfettamente lineare tra x e y . Al contrario se x e y sono indipendenti allora si ha $p=0$, ma il contrario non è necessariamente vero. Questo vuol dire che se due variabili hanno $p=0$ sono dette **non correlate**, ma non è detto che siano indipendenti. Perciò l'uso di p come misura dell'indipendenza deve essere limitato a quei problemi che hanno una possibile dipendenza lineare.

2.2 Generazione di numeri casuali

2.2.1 Metodo congruente lineare

Il **metodo congruente lineare** serve per generare sequenze di interi compresi tra 0 e m da utilizzare per la generazione di distribuzioni uniformi, gli obiettivi da soddisfare sono:

1. massimo periodo;
2. granularità fine;
3. efficienza di calcolo.

Questo metodo usa la seguente funzione:

$$X_{i+1} = (aX_i + c) \bmod m$$

se $c=0$ il metodo si chiama **congruente moltiplicativo**. La scelta dei parametri a = moltiplicatore e c = incremento è critica nel determinare il periodo e le doti di casualità. La funzione `rand()` implementa il metodo congruente moltiplicativo con parametri $m = 2^{31} - 1$ e $a = 75$.

2.2.2 Test del Chi-quadro X^2

Per verificare le doti di casualità di una sequenza occorre sottoporla ad una serie di test:

- Test del X^2
- Test seriale
- Test del gap

L'obiettivo di questi test è di verificare l'uniformità della distribuzione generata e la mancanza di correlazione tra numeri che si trovano ad una certa distanza k nella sequenza. Per verificare l'uniformità si calcola:

$$V = \sum_{s=1}^k \frac{(Y_s - np_s)^2}{np_s}$$

Poi si confronta il valore ottenuto con la tabella dei percentili. Il test del X^2 funziona nel seguente modo:

- si dividono i possibili valori in k categorie;
- si genera un campione di n numeri e si contano le occorrenze di valori in ciascuna categoria Y_s ;
- si calcola il valore atteso di valori in ciascuna categoria con la formula $p_s \cdot n$ dove p_s è la probabilità di estrarre un valore nella categoria s ;
- infine si calcola la sommatoria $V = \sum_{s=1}^k \frac{(Y_s - np_s)^2}{np_s}$.

Categoria	Ys	ps	n ps	(Ys - n ps)^2	(Ys - n ps)^2 / n ps
1	511	0,10	500	121	0,24
2	462	0,10	500	1444	2,89
3	382	0,10	500	13924	27,85
4	570	0,10	500	4900	9,80
5	569	0,10	500	4761	9,52
6	500	0,10	500	0	0,00
7	576	0,10	500	5776	11,55
8	411	0,10	500	7921	15,84
9	508	0,10	500	64	0,13
10	484	0,10	500	256	0,51

Figura 2.1: Test del Chi-quadro.

2.2.3 Test per la verifica delle proprietà di casualità

I valori $m = 2^31 - 1$, $a=1$ e $c=1$ garantiscono il periodo massimo e l'uniformità di distribuzione dei risultati ma non garantiscono l'assenza di correlazione fra i valori, per rilevare questo problema il test del X^2 può essere applicato a coppie di valori estratti dalla sequenza. Nell'esempio di prima si potrebbe dividere il range di valori in 5 sotto-range e formare 25 categorie corrispondenti a tutte le possibili coppie di sotto-range e la p_s si calcola come prodotto delle probabilità di generare un valore in ciascuno di essi. Altri test sono quello del **gap** ovvero si calcola la lunghezza di sequenze i cui valori sono compresi tra α e β e si confronta la loro distribuzione con quella teorica usando il chi-quadro.

2.3 Generazione di distribuzioni qualsiasi

Per la generazione di diverse distribuzioni si utilizzano i seguenti metodi:

- Metodo della trasformazione inversa
- Metodo del rifiuto
- Metodo della composizione

2.3.1 Metodo della trasformazione inversa

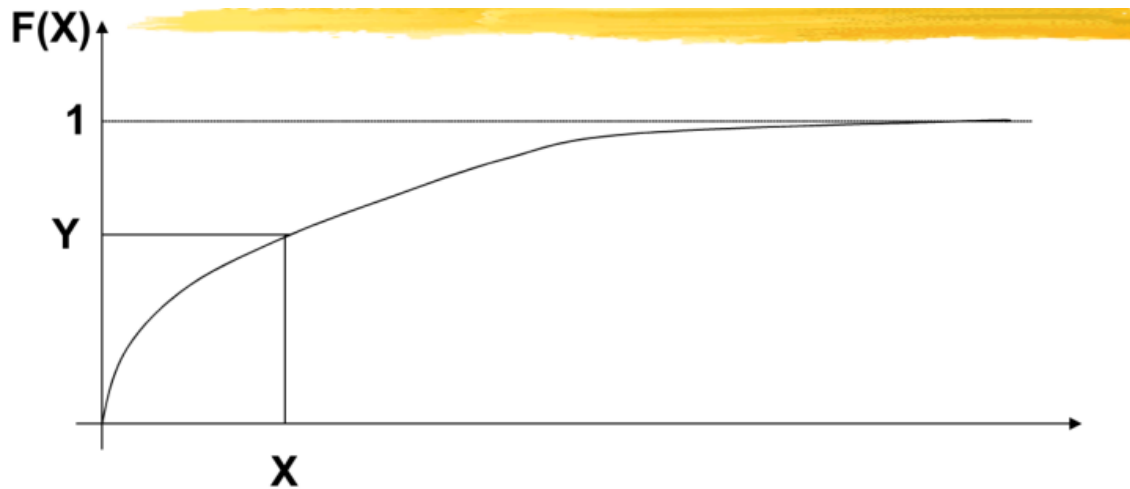


Figura 2.2: Trasformazione inversa.

Un esempio di applicazione del metodo è la generazione di variabili casuali distribuite secondo una esponenziale negativa:

$$\begin{aligned}F(x) &= 1 - e^{-\mu x} \\1 - y &= e^{-\mu x} \\ \ln(1 - y) &= \ln(e^{-\mu x}) = -\mu x \\ x &= \frac{-\ln(1 - y)}{\mu}\end{aligned}$$

Se y è distribuita come $U(0,1)$, anche $Z = 1 - Y$ ha la stessa distribuzione: quindi si può generare una sequenza distribuita secondo una esponenziale negativa semplicemente generando una sequenza con distribuzione $U(0,1)$, applicando la funzione logaritmo naturale, cambiando di segno e dividendo per μ i numeri di tale sequenza.

2.3.2 Metodo del rifiuto

Quando non è facilmente ricavabile l'inversa della cumulativa si utilizzano altri metodi come il metodo del rifiuto.

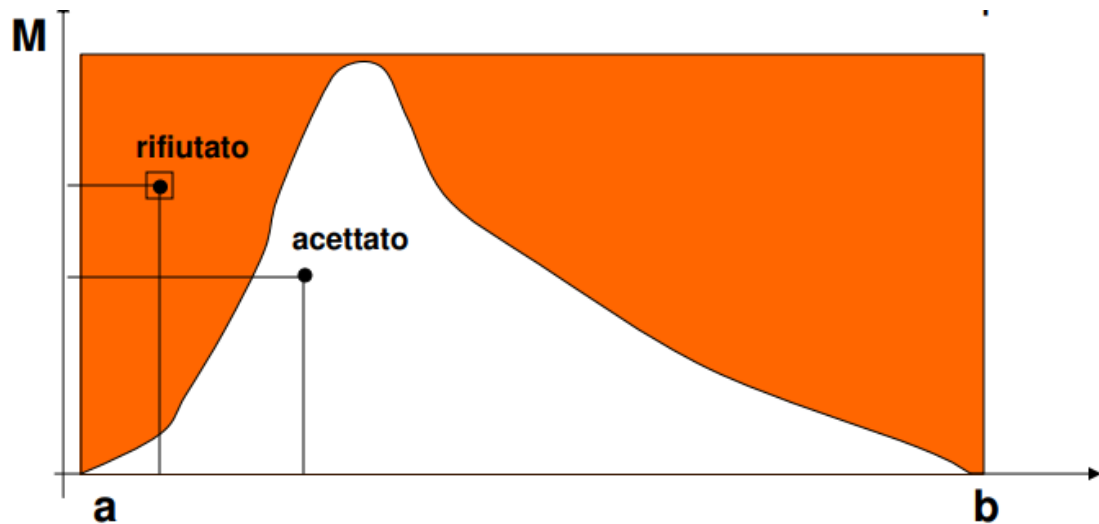


Figura 2.3: Metodo del rifiuto.

Si genera un valore y distribuito come $U(0, M)$ e un x uniformemente distribuito tra a e b , se $y \leq f(x)$ allora si accetta x come valore generato della distribuzione, sennò si ripete con altri due valori di x e y .

2.3.3 Metodo della composizione

Questa distribuzione può essere vista come risultato della composizione di 4 uniformi: $U(a_1, a_2), U(a_2, a_3), U(a_3, a_4), U(a_4, a_5)$.

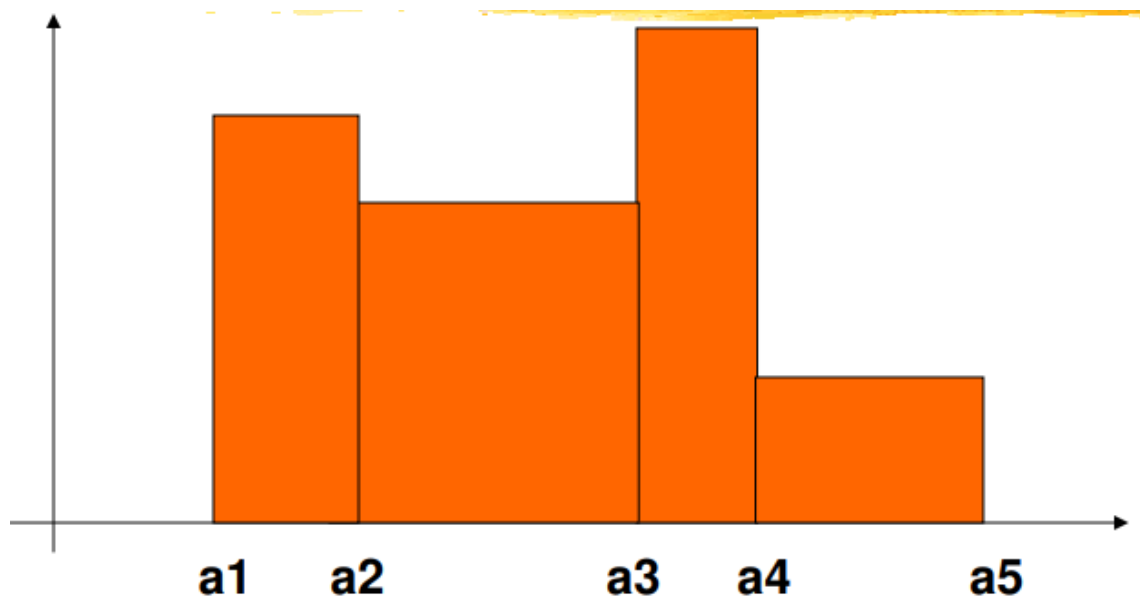


Figura 2.4: Metodo della composizione.

Siano p_1, p_2, p_3 e p_4 le aree dei quattro rettangoli: per generare un valore tratto da questa distribuzione scegliamo un rettangolo utilizzando le p_i , poi generiamo un numero uniformemente distribuito in $U(a_i, a_{i+1})$.

2.4 Distribuzioni

Capitolo 3

Processi stocastici

I processi stocastici costituiscono una classe di modelli di sistemi che permette lo studio e l'analisi dei sistemi. In questo Capitolo verranno riportate alcune definizioni e classificazioni dei suddetti processi, soffermando l'attenzione sui processi Markoviani.

3.1 Processi stocastici

Un **processo stocastico** è definito come una famiglia di variabili casuali $X = \{X(t) | t \in T\}$ definite in uno **spazio di probabilità** Ω ed indicate dal parametro t , dove t varia in un insieme T e indica il tempo. Ogni variabile casuale $X(t)$ assume valori, detti **stati**, in un insieme E detto **spazio degli stati** del processo stocastico. Essendo che ogni variabile è definita come una funzione su uno spazio campione S , allora il processo stocastico è visto come un insieme di funzioni $\{X(t, s) | t \in T, s \in S\}$. I processi stocastici possono essere classificati in base alle seguenti caratteristiche:

- spazio degli stati E
- indice tempo
- tipo di dipendenza stocastica fra le variabili casuali

3.1.1 Spazio degli stati E

Lo spazio degli stati può essere **discreto** o **continuo**. Nel primo caso il processo è detto anche catena e lo spazio E è indicato come l'insieme di interi non negativi. Nel secondo l'insieme dei valori assunti non è finito o numerabile perciò il processo è a spazio continuo.

3.1.2 Indice tempo

L'indice tempo è a sua volta **discreto** o **continuo**. Un processo a tempo discreto è anche detto sequenza stocastica ed è denotato come $\{X_n | n \in T\}$ dove T è finito o numerabile. Al contrario, se i cambiamenti avvengono in un qualsiasi istante di un tempo infinito allora si ha un processo a tempo continuo e si denota come $\{X(t) | t \in T\}$.

3.1.3 Dipendenza stocastica fra variabili

La **dipendenza stocastica** fra variabili casuali $X(t)$ per diversi valori di t è identificata dalla funzione di distribuzione di probabilità congiunta per le variabili $X = [X(t_1), X(t_2), \dots, X(t_n)]$, vedi 2.1.5, denotata come:

$$F_X(x; t) = \text{Prob}\{X(t_1) \leq x_1; X(t_2) \leq x_2; \dots; X(t_n) \leq x_n\}$$

al variare di n e di $x = [x_1, x_2, \dots, x_n]$ e $t = [t_1, t_2, \dots, t_n]$ con $x_i \in E$, $t_i \in T$ $1 \leq i \leq n$, $n \geq 1$ e $t_1 \leq t_2 \leq \dots \leq t_n$.

Classificazione in base alla dipendenza stocastica

Come accennato in precedenza è possibile classificare i processi in base alla dipendenza stocastica.

Un processo si dice **stazionario** in senso stretto se la funzione distribuzione $F_X(x; t)$ è invariante rispetto allo spostamento sull'asse del tempo T ovvero se per ogni τ costante

$$F_X(x; t + \tau) = F_X(x; t)$$

dove $t + \tau = [t_1 + \tau, t_2 + \tau, \dots, t_n + \tau]$ e per ogni x, t e n .

Un processo stocastico è **indipendente** se la funzione di distribuzione congiunta è fattorizzabile nelle singole funzioni di distribuzione marginali:

$$F_X(x; t) = \prod_{i=1}^n F_{X_i}(x_i; t_i) = \prod_{i=1}^n \text{Prob}\{X(t_i) \leq x_i\}$$

Un processo stocastico è di **rinnovamento** se le variabili casuali X_1, X_2, \dots sono indipendenti, identicamente distribuite e a valori non negativi.

3.1.4 Processi di Markov

Un processo stocastico a tempo discreto $\{X_n | n \in T\}$ è di Markov se la probabilità di stato al tempo $n+1$ dipende soltanto dallo stato al tempo attuale, n , e non dalla storia precedente, ovvero se per ogni $n > 0$ e per ogni valore degli stati $j, i_0, i_1, i_2, \dots, i_n \in E$ si ha:

$$Prob\{X_{n+1} = j | X_0 = i_0; X_1 = i_1; \dots; X_n = i_n\} = Prob\{X_{n+1} = j | X_n = i_n\}$$

Un processo stocastico a tempo continuo $\{X_n | n \in T\}$ è di Markov se per ogni sequenza di valori $t_0 \leq t_1 \leq \dots \leq t_n < t$ la distribuzione di probabilità di $X(t)$ condizionata a $X(t_0), X(t_1), \dots, X(t_n)$ dipende soltanto dallo stato $X(t_n)$, per ogni $n > 0$ e per ogni valore degli stati $j, i_0, i_1, i_2, \dots, i_n \in E$ si ha:

$$Prob\{X(t) = j | X(t_0) = i_0; X(t_1) = i_1; \dots; X(t_n) = i_n\} = Prob\{X(t) = j | X(t_n) = i_n\}$$

Per un processo di Markov a tempo discreto il tempo di permanenza in uno stato è un v.c. **geometrica** mentre per un processo di Markov a tempo continuo è un v.c. **esponenziale**. Questo vincolo deriva dalla proprietà di assenza di memoria indotta sulla v.c. dal processo Markoviano. Come è noto le uniche distribuzioni che godono della proprietà di assenza di memoria sono, rispettivamente, la geometrica fra le distribuzioni discrete e l'esponenziale fra le continue. Se si assume che la distribuzione di probabilità non sia geometrica o esponenziale allora si ha un processo stocastico di semi-Markov, in questo caso le transizioni di tempo vanno ad istanti di tempo presi arbitrariamente.

Un p.s. di Markov a spazio discreto E è anche detta **catena di Markov**. Una catena di Markov è di **nascita-morte** (birth-death) se le uniche transizioni possibili che il processo può effettuare sono soltanto fra stati vicini. Nel caso di tempo discreto, se lo stato $X_n = i$ allora lo stato al tempo successivo può soltanto essere $X_{n+1} \in i - 1, i, i + 1$.

Se lo spazio degli stati E è a tempo discreto allora $\{X_n | n \in N\}$ è una catena di Markov a tempo discreto, $X_n = j$ indica che il processo si trova nello stato j al tempo n , $n \in N$, $j \in E$. Nella formula della probabilità condizionata al membro destro:

$$Prob\{X_{n+1} = j | X_n = i\}$$

è detta probabilità di transizione al tempo n dallo stato i allo stato j , $i, j \in E$, $n \geq 0$. Una catena di Markov si dice omogenea se le probabilità di transizione sono

indipendenti dal tempo n . La probabilità di transizione dallo stato i allo stato j è definita come:

$$p_{ij} = Prob\{X_{n+1} = j | X_n = i\}$$

$\forall i, j \in E, \forall n \geq 0$.

Una catena di Markov può essere analizzata sia in un tempo finito (**analisi transiente**, non importante per esame) sia a lungo termine (**analisi stazionaria**).

3.1.5 Analisi stazionaria

La descrizione del processo stocastico in condizioni asintotiche, per il tempo n che tende ad infinito, si esprime in termini di distribuzione di probabilità di stato. Sia $\pi_j = \lim_{n \rightarrow \infty} \pi_j(n) = Prob\{X = j\}$ la probabilità di osservare lo stato j in condizioni di equilibrio, $\forall j \in E$. In forma compatta indichiamo con $\pi = [\pi_0, \pi_1, \pi_2, \dots, \pi_n]$ il vettore delle probabilità stazionarie di stato.

Classificazione degli stati

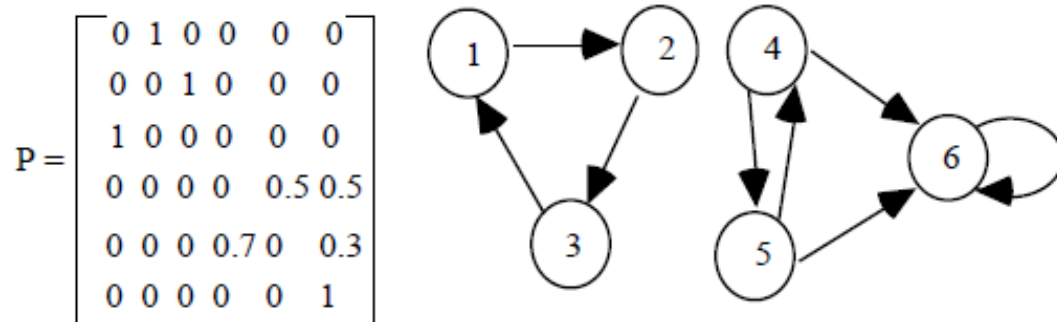


Figura 3.1: Esempio di diagramma fra gli stati di una catena di Markov a tempo discreto con matrice delle probabilità di transizione P .

In Fig.3.1 il sottoinsieme $E' = \{1, 2, 3\}$ è chiuso, gli stati 1, 2 e 3 sono ricorrenti periodici di periodo 3. Gli stati 4 e 5 sono transienti e lo stato 6 è assorbente.

Teorema La distribuzione di probabilità stazionaria π_i di una catena di Markov omogenea, irriducibile e aperiodica ¹ esiste sempre ed è indipendente dallo stato iniziale. Inoltre

¹Uno stato ricorrente è **aperiodico** se $g_i = 1$ ed è periodico se $g_i > 1$.

- se gli stati sono transienti ² o ricorrenti ³ nulli allora $\pi_j = 0, \forall j \in E$, e non esiste distribuzione stazionaria;
- se gli stati sono ricorrenti non-nulli allora $\pi_j > 0, \forall j \in E$, e la distribuzione stazionaria π è l'unica soluzione non negativa del sistema lineare

$$\pi = \pi P$$

con la condizione di normalizzazione $\sum j \pi_j = 1$.

Teorema Per una catena di Markov a tempo continuo omogenea, irriducibile ed aperiodica esiste sempre la probabilità stazionaria di stato $\pi_j = \lim_{t \rightarrow \infty} \pi_j(t), j \in E$, ed è indipendente dallo stato iniziale. La soluzione stazionaria si ottiene allora imponendo $\frac{d\pi(t)}{dt} = 0$ da cui, dalla equazione $\frac{d\pi(t)}{dt} = \pi(t)Q$ deriva

$$\pi Q = 0$$

che, insieme alla condizione di normalizzazione $\sum i \pi_i = 1$, determina univocamente la soluzione.

Il sistema di equazioni 3.1.5 è detto anche sistema di **equazioni di bilanciamento globale** in quanto la singola equazione relativa allo stato i può essere riscritta come:

$$\pi_i \sum_{j \neq i} q_{ij} = \sum_{j \neq i} \pi_j q_{ji}$$

Questo vuol dire che la probabilità di entrare in un stato i è uguale alla probabilità di uscire da quest'ultimo. Tale equazione si può interpretare come bilanciamento fra il flusso totale uscente dallo stato i (membro sinistro) e il flusso totale entrante in i e proveniente da tutti gli altri stati $j \neq i$ (membro destro).

Al contrario il bilanciamento locale è soddisfatto da una catena di Markov se sono uguali il flusso di probabilità da uno stato i a uno stato j e viceversa, ovvero se per ogni $i, j \in E$:

$$\pi_i q_{ij} = \pi_j q_{ji}$$

²Uno stato i è **transiente** se vi è probabilità non nulla di non ritorno allo stato i .

³Uno stato $i \in E$ è **ricorrente** se il processo, iniziando nello stato i , torna allo stato i con certezza (con probabilità 1).

Processi di nascita-morte

In una catena di Markov di nascita-morte (birthdeath), con spazio degli stati $E=N$, le uniche transizioni fra stati ammesse sono dallo stato i agli stati $i-1$, i , $i+1$, $\forall i \in E$.

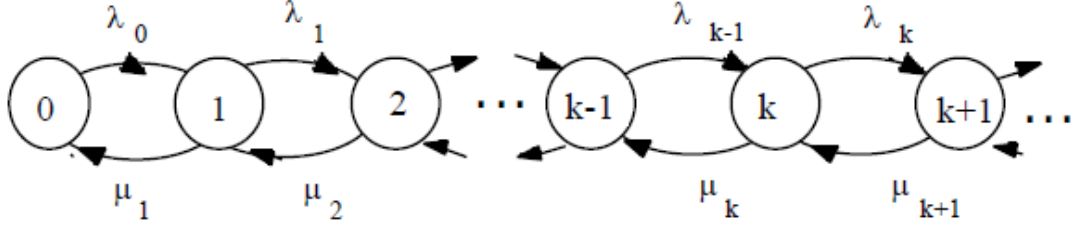


Figura 3.2: Processi di nascita-morte.

Consideriamo un processo di nascita-morte a tempo continuo. In tal caso definiamo la matrice $Q = [q_{ij}]$ delle velocità di transizione fra stati $i, j \in E$, come:

$$q_{ii+1} = \lambda_i \quad i \geq 0 \quad (3.1)$$

$$q_{ii+1} = \mu_i \quad i \geq 0 \quad (3.2)$$

$$q_{00} = -\lambda_0 \quad (3.3)$$

$$q_{ii} = -(\lambda_i + \mu_i) \quad i \geq 1 \quad (3.4)$$

$$q_{ij} = 0 \quad \text{per } |i - j| > 1 \quad (3.5)$$

Il processo è omogeneo perché questi tassi dipendono solo dallo stato e non dal tempo. Il sistema di equazioni $\frac{d\pi(t)}{dt} = \pi(t)Q$ relativo alla distribuzione di probabilità di stato $\pi(t)$ al tempo $t > 0$ per un processo di nascita e morte assume la forma seguente:

$$\frac{d\pi_0(t)}{dt} = \mu_1\pi_1(t) - \lambda_0\pi_0(t) \quad (3.6)$$

$$\frac{d\pi_k(t)}{dt} = \lambda_{k-1}\pi_{k-1}(t) + \mu_{k+1}\pi_{k+1}(t) - (\lambda_k + \mu_k)\pi_k(t) \quad k > 0 \quad (3.7)$$

$$(3.8)$$

Questo sistema di equazioni descrive il comportamento del processo in funzione del tempo e può essere risolto assumendo una distribuzione di probabilità di stato iniziale, riguarda la parte dell'analisi transiente.

Se il processo di nascita e morte è irriducibile la probabilità stazionaria di stato π si ottiene dalla soluzione del sistema di equazioni di bilanciamento globale che può

essere riscritto come:

$$\mu_1\pi_1 = \lambda_0\pi_0 \quad (3.9)$$

$$\lambda_{k-1}\pi_{k-1} + \mu_{k+1}\pi_{k+1} = (\lambda_k + \mu_k)\pi_k \quad k > 0 \quad (3.10)$$

$$(3.11)$$

Questo sistema può essere facilmente risolto ad esempio per sostituzione. Infatti dalla prima equazione si ricava

$$\pi_1 = \pi_0 \frac{\lambda_0}{\mu_1}$$

e per sostituzione, dalla seconda equazione

$$\pi_2 = \pi_1 \frac{\lambda_1}{\mu_2} = \pi_0 \frac{\lambda_1\lambda_0}{\mu_2\mu_1}$$

e, per sostituzioni successive, dall'equazione k-sima

$$\pi_k = \pi_{k-1} \frac{\lambda_{k-1}}{\mu_k} = \pi_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}$$

per $k > 0$.

Condizione di stazionarietà Una condizione necessaria e sufficiente perché il processo di nascita e morte sia ergodico (irriducibile, aperiodico, ricorrente non nullo) è che esista un $k_0 > 0$ tale che $\forall k > k_0$ sia $\lambda_k < \mu_k$, ovvero $\rho = \frac{\lambda}{\mu} < 1$. Questo vuol dire che se la velocità degli arrivi è maggiore di quella delle partenze allora il sistema è congestionato.

3.1.6 Processo di pura nascita

Considerando il particolare processo di nascita e morte in cui i tassi di morte sono nulli ($\mu_k = 0, k > 0$), definiamo un processo di **pura nascita**. Assumiamo inoltre che il tasso di nascita sia costante, indipendente dallo stato: $\lambda_k = \lambda, k \geq 0$.

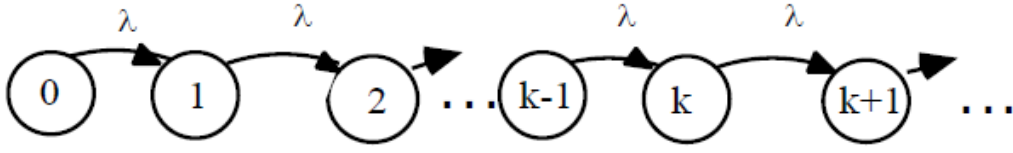


Figura 3.3: Diagramma di un processo di pura nascita.

Assumendo la condizione iniziale di sistema vuoto, $\pi_0(0) = 1, \pi_k(0) = 0, \forall k \neq 0$, si ricava facilmente la soluzione transiente:

$$\pi_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad k \geq 0, t \geq 0 \quad (3.12)$$

che riconosciamo, in accordo alla formula 3.1.7, come la distribuzione di Poisson di parametro λt , di media e varianza λt .

3.1.7 Processo di Poisson

Indichiamo con $N(t)$, $t \geq 0$, il numero di eventi osservati in un intervallo di tempo $[0, t]$, nel caso di un sistema aperto esso rappresenta il numero di arrivi nel suddetto intervallo (**processo di arrivo**). Inoltre sia X_i la v.c. che indica il tempo che intercorre fra l'arrivo dell'($i-1$)-esimo job e l'arrivo dell' i -esimo job al sistema, per $i > 1$, con X_1 indichiamo l'arrivo del primo job. Se le v.c. X_i , $i > 0$, sono indipendenti ed identicamente distribuite, allora il processo stocastico $\{N(t) | t \geq 0\}$ è di rinnovamento. Inoltre se le v.c. X_i , $i > 0$, sono i.i.d. con distribuzione esponenziale di parametro λ il quale indica il tasso di arrivo, allora il processo stocastico $\{N(t) | t \geq 0\}$ è di Poisson. La distribuzione di probabilità di Poisson con parametro λt è data dalla seguente formula:

$$\begin{aligned} Prob\{N(t) = i\} &= e^{-\lambda t} \frac{(\lambda t)^i}{i!} \\ E[N(t)] &= Var[N(t)] = \lambda t \end{aligned}$$

con $i \geq 0$. Il processo di Poisson può essere definito come il processo stocastico che indica il numero di eventi osservati in un intervallo di tempo $[0, t]$, e che soddisfa le seguenti condizioni:

- il numero di eventi in intervalli disgiunti sono stocasticamente indipendenti;
- il numero di eventi in un intervallo di tempo dipende solo dalla sua lunghezza e non dall'istante iniziale;

- per un valore piccolo di t e per una costante positiva λ si ha che:
 - $\text{Prob} \{ \text{un evento in un intervallo di ampiezza } t \} = \lambda t + o(t)$
 - $\text{Prob} \{ \text{nessun evento in un intervallo di ampiezza } t \} = o(t)$
 - $\text{Prob} \{ \text{più di un evento in un intervallo di ampiezza } t \} = o(t)$ dove $o(t) = \lim_{t \rightarrow \infty} [o(t)/t] = 0$.

Dalla seconda proprietà deriva che

$$\text{Prob}\{k \text{ eventi nell'intervallo } (s, s+t)\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

In questo caso se il tempo di interarrivo è esponenziale allora il processo di arrivo è di Poisson e viceversa. Indichiamo con la X la variabile casuale tempo di interarrivo con distribuzione cumulativa $A(t) = \text{Prob}\{X \leq t\}$ allora possiamo scrivere che

$$A(t) = \text{Prob}\{N(t) = 0\} = 1 - e^{-\lambda t}, \text{ con } t > 0$$

Composizione di processi di Poisson

La sovrapposizione o composizione di n processi di Poisson indipendenti rispettivamente di parametro $\lambda_1, \lambda_2, \dots, \lambda_n$ è anch'esso un processo di Poisson di parametro $\lambda_1 + \lambda_2 + \dots + \lambda_n$.

Decomposizione di processi di Poisson

Dato un processo di Poisson $\{N(t) | t \geq 0\}$ di parametro λ , assumiamo che sia decomposto in n processi, selezionando ogni uscita del processo secondo una distribuzione di probabilità di Bernoulli ad n uscite, con probabilità $p_i, 1 \leq i \leq n$. Allora gli n processi di uscita $\{N_i(t) | t \geq 0\}$ sono indipendenti e sono anch'essi processi di Poisson rispettivamente con parametro $\lambda p_i, 1 \leq i \leq n$.

Capitolo 4

Modelli a singola coda

Un modello a coda è costituito dalla rappresentazione di un sistema di congestione in cui utenti che provengono da una popolazione si mettono in coda per ottenere il servizio richiesto da un insieme di risorse o server. Una coda è una linea di attesa per un servizio. L'insieme formato da coda e server è detto **centro di servizio**.

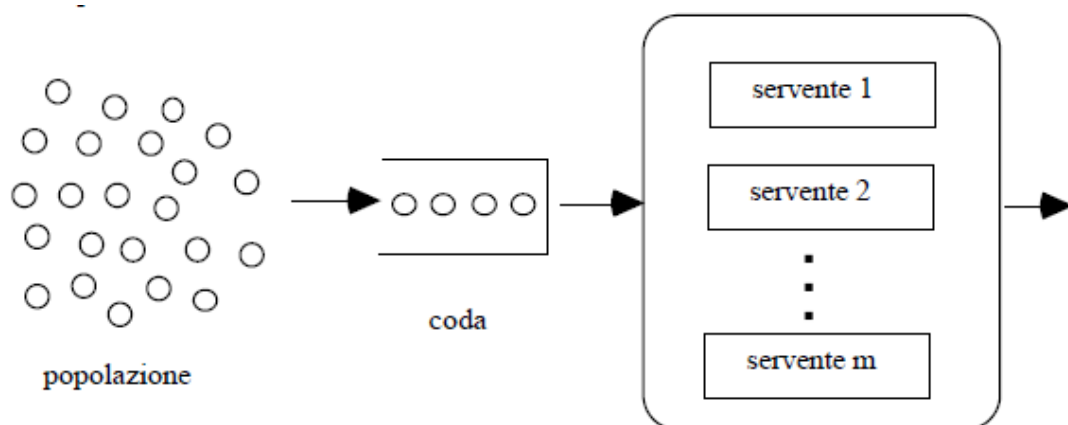


Fig. 3.1 - Sistema di congestione -

Figura 4.1: Sistema di congestione.

Un sistema a coda è costituito da alcune grandezze che sono:

- Δ tempo di interarrivo;
- w numero di utenti in coda e t_w tempi di attesa in coda, tempo che intercorre fra l'arrivo di un utente e l'istante in cui entra in servizio;
- s numero di utenti in coda e t_s tempo di servizio, tempo fra inizio e completamento del servizio;

- q numero di utenti nel sistema e t_q tempo di risposta, tempo fra arrivo e partenza dal sistema dello stesso utente.

Per definizione $q = w + s$ e le variabili di tempi sono legati dalla seguente relazione $t_q = t_w + t_s$.

Fra queste tipicamente alcune sono parametri del sistema, quali t_s e Δ mentre le altre sono oggetto di analisi e di valutazione.

Processo di arrivo Assumiamo che il tempi di interarrivo Δ siano variabili casuali statisticamente indipendenti con la stessa distribuzione di probabilità.

Domanda di servizio, tasso di servizio e tempo di servizio La quantità di servizio richiesta da un utente ad un centro di servizio è detta domanda di servizio ed è espressa in unità di tempo. La velocità o tasso di servizio di un servente è una caratteristica del servente, ovvero della risorsa, ed è espressa in unità di servizio per unità di tempo. Il tempo di servizio è la il rapporto fra la domanda di servizio e la velocità di servizio espresso in unità di tempo.

Disciplina di servizio Una **disciplina di servizio** è un algoritmo di ordinamento degli utenti in coda in base al quale viene selezionato l'utente da servire, ovvero l'ordine con cui estrarre gli utenti dalla coda. Alcuni esempi sono FIFO che serve gli utenti in ordine di arrivo e LIFO che serve gli utenti in ordine inverso al tempo di arrivo. La disciplina di servizio RAND determina l'utente da estrarre dalla coda e servire in modo casuale, secondo la distribuzione uniforme discreta. Le **discipline a priorità** determinano l'ordine di servizio degli utenti sulla base di priorità assegnate ad esse. La disciplina round robin è un esempio di disciplina che dipende dalla quantità di servizio già assegnato. Tale algoritmo assegna il servente ciclicamente agli utenti per un intervallo massimo di tempo detto quanto.

4.0.1 Notazione di Kendall

Per descrivere e definire i modelli a coda singola Kendall ha introdotto la seguente notazione:

$$A/B/c/n/p/Z$$

dove i simboli denotano:

- A : distribuzione del tempo di interarrivo;

- B : distribuzione del tempo di servizio t_s ;
- c : numero di serventi m;
- n : dimensione della coda ovvero la capacità del centro, q;
- p : dimensione della popolazione;
- Z : disciplina di servizio.

Tale notazione si semplifica in A/B/c quando coda e popolazione sono illimitate e la disciplina di servizio è FIFO, ovvero se $n = p = \infty$ e Z=FIFO. D denota la distribuzione deterministica o costante, M la distribuzione esponenziale negativa, H_h l'iperesponenziale ad h stadi, E_k l'Erlangiana a k stadi e G la generale. Il comportamento del sistema nel tempo può essere analizzato in un tempo finito. In tal caso si parla di **analisi transiente**. Quando il sistema raggiunge le condizioni di equilibrio in un tempo che tende ad infinito, ovvero se il sistema è stabile si effettua **l'analisi stazionaria**. Tale analisi porta alla valutazione degli indici di prestazione del sistema di code, in particolare di:

- numero di utenti nel sistema q;
- numero di utenti in coda w;
- tempo di risposta t_q ;
- tempo di attesa t_w ;
- utilizzazione U;
- throughput X.

Le prime quattro sono variabili casuali di cui si valuta la distribuzione di probabilità, e/o i momenti. Possiamo scrivere immediatamente le seguenti relazioni fra i valori medi:

$$E[q] = E[w] + E[s] \quad (4.1)$$

$$E[t_q] = E[t_w] + E[t_s] \quad (4.2)$$

Una relazione fondamentale nella teoria delle code è rappresentato dal **teorema di Little** il quale stabilisce che il numero medio di utenti in un sistema è uguale al prodotto fra il throughput e il tempo medio di risposta, ovvero:

$$E[q] = X E[t_q] \quad (4.3)$$

$$E[w] = X E[t_w] \quad (4.4)$$

4.0.2 Modelli basilari di code: i sistemi M/M/1 ed M/M/m

M/M/1 costituisce la base per la definizione di modelli più complessi a struttura reticolare e per lo sviluppo gerarchico di modelli a diversi livelli di astrazione. Nella notazione di Kendall il sistema di code M/M/1 denota un sistema aperto formato da un singolo centro di servizio con:

- distribuzione del tempo di interarrivo esponenziale, di parametro λ ,
- tempo di servizio degli utenti indipendente e con identica distribuzione esponenziale di parametro μ ,
- singolo servente.

La disciplina di servizio è FIFO e sia la memoria che la popolazione sono infinite.



Figura 4.2: Sistema M/M/1.

Per le ipotesi di esponenzialità dei tempi di interarrivo e di servizio, gli unici eventi possibili che si osservano in un intervallo di tempo d_t quando nel sistema vi sono k utenti sono i seguenti:

1. nessun arrivo e nessun completamento di servizio (permanenza nello stato k);
2. un arrivo e nessun completamento di servizio (transizione dallo stato k allo stato $k+1$ con probabilità λdt);
3. nessun arrivo e un completamento di servizio (transizione dallo stato k allo stato $k-1$ con probabilità μdt);

4. uno o più arrivi e uno o più completamenti di servizio (transizione con probabilità trascurabile rispetto alle altre transizioni per l'ipotesi di esponenzialità).

4.0.3 Analisi stazionaria

Definiamo l'intensità di traffico del sistema come rapporto fra tempo medio di servizio e tempo medio di interarrivo, denotato da $\rho = \lambda/\mu$. La condizione di stazionarietà del sistema M/M/1 richiede che $\rho < 1$, ovvero che il tasso di arrivo al sistema sia minore del tasso di servizio, $\lambda < \mu$. La distribuzione di probabilità di stato in equilibrio si ricava :

$$\pi_0 = 1 - \rho \quad (4.5)$$

$$\pi_k = \rho^k \pi_0 = \rho(1 - \rho) \quad k > 0 \quad (4.6)$$

la distribuzione di probabilità del numero di utenti nel sistema in condizioni stazionarie è geometrica a ragione ρ .

Il numero medio di utenti nel sistema, che denotiamo con $N=E[q]$, si può immediatamente ricavare come:

$$N = \frac{\rho}{1 - \rho} \quad (4.7)$$

e la varianza del numero di utenti, denotata da $\text{Var}[q]$, è

$$\text{Var}[q] = \frac{\rho}{(1 - \rho)^2} \quad (4.8)$$

Per la relazione (5.1) ricaviamo immediatamente il numero medio di utenti in coda, denotato da $W=E[w]$, poiché $E[s] = \rho$ ed $E[q]=N$:

$$W = N - \rho = \frac{\rho^2}{1 - \rho} \quad (4.9)$$

Il tempo medio di risposta, denotato da $R=E[tq]$, si ricava dal teorema di Little che per il sistema M/M/1 assume la forma $N = \lambda R$:

$$R = \frac{\frac{1}{\mu}}{1 - \rho} \quad (4.10)$$

Infine il tempo medio di attesa in coda, $T_w = E[t_w]$ si ottiene o dal teorema di Little come $W = \lambda T_w$ o dalla relazione (3.4) ovvero $R = T_w + T_s$, dove $T_s = \frac{1}{\mu}$ denota

il tempo medio di servizio, da cui si ricava

$$T_w = \frac{\frac{\rho}{\mu}}{1 - \rho} \quad (4.11)$$

L'utilizzazione è esprimibile per il sistema M/M/1 come:

$$U = 1 - \pi_0 = \rho \quad (4.12)$$

Dalla distribuzione di probabilità (5.6) possiamo calcolare la probabilità con cui si osservano almeno k utenti nel sistema in condizioni di equilibrio:

$$Prob\{\text{numero di utenti nel sistema} \geq k\} = \sum_{i \geq k} \pi_i = (1 - \rho) \sum_{i \geq k} \rho^i = \rho^k \quad (4.13)$$

4.0.4 Sistema M/M/ ∞ : serventi infiniti

Un modello M/M/ ∞ permette di rappresentare un sistema nel quale ad ogni arrivo vi è sempre un servente libero. In un sistema di questo tipo non si forma quindi mai coda, poiché ogni utente riceve immediatamente servizio dall'istante in cui arriva al sistema. Il processo stocastico $q(t) \mid t > 0$ associato al sistema M/M/ ∞ è un processo di nascita e morte con tassi di nascita $\lambda_k = \lambda, k \geq 0$, e con tassi di morte $\mu_k = k\mu, k \geq 1$. La condizione di stazionarietà è certamente sempre soddisfatta quindi il sistema è sempre stabile. Esiste la distribuzione stazionaria del numero di utenti nel sistema, che coincide con il numero di serventi occupati da cui si ricava:

$$\pi_k = \frac{\rho^k}{k!} e^{-\rho} \quad k \geq 0 \quad (4.14)$$

dove $\rho = \lambda/\mu$. Si ricava immediatamente il numero medio di utenti nel sistema

$$N = \rho \quad (4.15)$$

Dato che nel sistema M/M/ ∞ gli utenti non sperimentano mai coda, il tempo medio di risposta coincide con il tempo medio di servizio:

$$R = T_s = \frac{1}{\mu} \quad (4.16)$$

4.0.5 Sistema M/M/m: serventi multipli

Il sistema aperto M/M/m è caratterizzato da un processo di arrivo Poissoniano di λ parametro, distribuzione del tempo di servizio esponenziale di parametro μ e un numero di serventi $m > 0$. Come in M/M/1 il tasso medio di servizio dipende dallo stato k secondo la seguente funzione $\mu(k) = k\mu, 0 \leq k \leq m, e \mu(k) = m\mu, k > m$. Definiamo l'intensità di traffico come $\rho = \lambda/m\mu$. In questo caso esiste la distribuzione stazionaria del numero di utenti nel sistema :

$$\pi_k = \frac{(m\rho)^k}{k!} \pi_0 \quad 0 \leq k \leq m \quad (4.17)$$

$$\pi_k = \frac{m^m \rho^k}{m!} \pi_0 \quad k > m \quad (4.18)$$

$$\pi_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!} \cdot \frac{1}{1-\rho} \right]^{-1} \quad (4.19)$$

Il numero medio di serventi occupati $E[s]$ è dato da

$$E[s] = \sum_{k=0}^{m-1} k\pi_k + \frac{m\pi_m}{1-\rho} = m\rho = \frac{\lambda}{\mu} \quad (4.20)$$

dove ρ rappresenta l'utilizzazione di ogni singolo servente. Il numero medio di utenti nel sistema è dato da

$$N = m\rho + \pi_m \frac{\rho}{(1-\rho)^2} \quad (4.21)$$

$$W = \pi_m \frac{\rho}{(1-\rho)^2} \quad (4.22)$$

Il tempo medio di risposta si ottiene applicando il teorema di Little $N = \lambda R$ ottenendo:

$$R = \frac{1}{\mu} + \frac{\pi_m}{(m\mu(1-\rho)^2)} \quad (4.23)$$

Similmente, otteniamo il tempo medio di attesa:

$$T_W = \frac{\pi_m}{m\mu(1-\rho)^2} \quad (4.24)$$

La probabilità che un utente in arrivo trovi tutti i serventi occupati e quindi sperimenti

coda è data dalla formula di Erlang-C:

$$Prob\{\text{coda}\} = \sum_{k=m}^{\infty} \pi_k = \pi_0 \frac{(m\rho)^m}{m!} \cdot \frac{1}{1-\rho} \quad (4.25)$$

Questa formula, è stata applicata per lo studio di sistemi di traffico telefonico per valutare la probabilità che una chiamata (un utente) trovi tutti i canali di comunicazione (i serventi) occupati.

Vedere esercizio 6.1.

4.0.6 Sistema M/M/1/K : memoria finita

Consideriamo il sistema M/M/1 nel quale sono ammessi al più K utenti. Il processo di arrivo è Poissoniano di parametro λ , ma un utente che arrivando trova il sistema completo, cioè con K utenti già presenti, non viene accettato e viene perso. Il processo stocastico $q(t)|t > 0$ associato al sistema M/M/1/k è un processo di nascita e morte con spazio degli stati finito $E = \{0, 1, \dots, K\}$ e tassi di nascita $\lambda_k = \lambda, 0 \leq k < K$, e tassi di morte $\mu_k = \mu, 1 \leq k < K - 1$ e $\lambda_k = 0$ e $\mu_k = 0$ altrove. La condizione di stazionarietà è certamente verificata, poiché il processo è finito e irriducibile e quindi ergodico. La distribuzione stazionaria del numero di utenti nel sistema definita dalle formule:

$$\pi_k = \frac{1-\rho}{1-\rho^{K+1}} \rho^k \quad 0 \leq k \leq K \quad (4.26)$$

$$\pi_k = 0 \quad k > K \quad (4.27)$$

Nel caso particolare di K=1 lo spazio è formato da due soli stati e si ricava

$$\pi_0 = \frac{\mu}{\lambda + \mu} \quad (4.28)$$

$$\pi_1 = \frac{\lambda}{\lambda + \mu} \quad (4.29)$$

4.0.7 Sistema M/M/1//M: popolazione finita

Consideriamo il sistema M/M/1 assumendo che gli utenti provengano da una popolazione finita, di dimensione M>0.

Capitolo 5

Analisi di risultati e convalida in esperimenti di simulazione

5.1 Stima dei parametri di una distribuzione

5.1.1 Stima della media

Sia data una popolazione la cui distribuzione $f(x)$ è stazionaria (cioè costante nel tempo), con media $E(x) = \eta$ e varianza $\sigma^2(x) = \sigma^2$. Facciamo n osservazioni indipendenti $x_1, x_2, x_3, \dots, x_n$ che chiamiamo **campione**. La media del campione è

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

è detta **media campionaria**. Allora

$$E(\bar{x}) = \eta$$

$$\sigma^2 = \frac{\sigma^2}{n}$$

La costante $1-\alpha$, usualmente espressa in percentuale, è detta di livello di confidenza per la stima η e l'intervallo

$$\bar{x} \pm \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}$$

è detto **intervallo di confidenza**. Ad esempio con un livello di confidenza pari a

95% allora

$$1 - \alpha = 0,95, \alpha = 0,05, \frac{\alpha}{2} = 0,025, 1 - \frac{\alpha}{2} = 0,975$$

Dalla tavola di distribuzione si ricava che per $F(u_{\alpha}) = 0,975, u_{\alpha} = 1,96$. Ciò vuol dire che su 100 campioni, ciascuno di n osservazioni, in 95 di essi l'intervallo $\bar{x} \pm 1,96 \frac{\sigma}{\sqrt{n}}$ conterrà il valore η . Nella pratica la varianza σ^2 della popolazione non è nota. In tal caso, la si sostituisce con la varianza del campione, definita come:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

e la variabile standard

$$t = \frac{\bar{x} - \eta}{s/\sqrt{n}}$$

5.1.2 Stima della varianza

Il metodo indiretto per la stima di $\sigma^2(x)$ è quello che usa la relazione $\sigma^2(x) = E(x^2) - E^2(x)$ e passa dalla stima di $E(x)$ a quella di $\sigma^2(x)$. Se ho la media la calcolo al quadrato e ho il secondo termine dall'equazione $\sigma^2(x) = E(x^2) - E^2(x)$, poi avendo tutti i campioni li elevo al quadrato e calcolo la media di questi, in questo modo si trova la varianza.

5.2 Stima della distribuzione

5.2.1 Metodo del coefficiente di variazione

Si definisce coefficiente di variazione di una distribuzione teorica il rapporto tra la deviazione standard e la media:

$$\nu = \frac{\sigma}{\eta}$$

Se $\nu = 0$ allora la distribuzione è una costante(deterministica). Esso definisce una misura di come i valori della variabile sono dispersi intorno alla media. Per le

distribuzioni viste sinora abbiamo:

$$\nu(\text{esponenziale}) = \frac{1/\lambda}{1/\lambda} = 1$$

$$\nu(\text{Erlang} - k) = \frac{1/(\sqrt{k}\lambda)}{1/\lambda} = \frac{1}{\sqrt{k}} \quad k = 1, 2, \dots$$

5.2.2 Metodo del goodness of fit

Il numero delle osservazioni deve essere maggiore di 30, sennò si usa il metodo del Kolmogorov-Smirnov. Supponiamo di aver fatto un campione di n osservazioni a partire da una popolazione continua reale e averne tracciato il relativo istogramma. Il procedimento che andremo a effettuare con una distribuzione normale può essere applicato ad altri tipi di distribuzioni continue di cui sia nota l'espressione della funzione cumulativa.

Esempio dal Capitolo 5

I dati di questo esempio sono i seguenti.

Tabella 5.2

Categorie	Intervalli	Quantità osservate f_i	Quantità teoriche F_i	$\frac{(f_i - F_i)^2}{F_i}$
1	80-90	8	5,2	1,50
2	90-100	15	13,1	0,29
3	100-110	21	20,9	0,00
4	110-120	23	23,5	0,01
5	120-130	16	18,6	0,36
6	130-140	9	10,4	0,19
7	140-150	8	5,5	1,14
		<u>100</u>	<u>97,2</u>	<u>3,49</u>
$\bar{x} = 113,3$		$s = 16,65$		

Figura 5.1: Tabella esempio stima distribuzione.

Innanzitutto bisogna controllare il numero di osservazioni se è maggiore di $n > 30$ allora possiamo applicare il metodo, inoltre suddividendo i campioni in categorie dobbiamo controllare che ci siano almeno 5 $f_i > 5$ osservazioni per categoria. Calcoliamo la media del campione \bar{x}

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{85 \cdot 8 + 95 \cdot 15 + 105 \cdot 21 + 115 \cdot 23 + 125 \cdot 16 + 135 \cdot 9 + 145 \cdot 8}{100} = 113,3$$

Calcoliamo la deviazione standard dalla varianza s

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \cdot [(85 \cdot 8)^2 + (95 \cdot 15)^2 + (105 \cdot 21)^2 + (115 \cdot 23)^2 + (125 \cdot 16)^2 + (135 \cdot 9)^2 + (145 \cdot 8)^2] = 16,65}$$

Poi facciamo il rapporto fra le due ovvero

$$\nu = \frac{16}{113}$$

Essendo che è una normale il rapporto è un numero lontano da 1, al contrario di una esponenziale negativa.

Capitolo 6

Esercizi

6.1 Esercizio su M/M/m

La clinica oculistica dell'Ospedale offre ogni mercoledì pomeriggio dei test gratuiti della vista. Ci sono 3 oculisti in contemporanea. Un test impiega, in media, 20 minuti e il tempo reale è stato riscontrato essere approssimativamente distribuito in modo esponenziale intorno a questa media. I clienti arrivano in accordo ad un processo di Poisson con un tasso medio di 6 all'ora e i pazienti sono serviti seguendo una politica FIFO.

I dirigenti dell'Ospedale sono interessati a conoscere:

1. Quale è il numero medio di persone in attesa?
2. Quale è il tempo medio speso da un paziente nella clinica?
3. Quale è la percentuale media del tempo in cui i dottori non lavorano?
4. Quale è la frazione di tempo nella quale un cliente che arriva trova almeno 1 dottore libero.

6.1.1 Soluzione

Troviamo alcuni parametri, $m=3$ $\mu = \frac{1}{20}/min = \frac{20}{60} = \frac{1}{3}/ora$ questo vuol dire che in un ora vengono effettuate 3 visite, $\lambda = 6/ora$ quindi

$$\rho = \frac{\lambda}{\mu m} = \frac{6}{3 \cdot 3} = \frac{2}{3} < 1$$

1. Il numero medio di persone in attesa è

$$\begin{aligned}
 W &= \pi_m \frac{\rho}{(1-\rho)^2} = \frac{4}{27} \cdot \frac{2/3}{(1-2/3)^2} = \frac{8}{9} \\
 \pi_0 &= \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!} \cdot \frac{1}{1-\rho} \right]^{-1} = \left[\sum_{k=0}^2 \frac{(3 \cdot 2/3)^k}{k!} + \frac{(3 \cdot 2/3)^3}{3!} \cdot \frac{1}{1-2/3} \right]^{-1} = \\
 &\left[\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + 4 \right]^{-1} = [9]^{-1} = \frac{1}{9} \\
 \pi_k &= \frac{(3 \cdot 2/3)^3}{3!} \frac{1}{9} = \frac{8}{54} = \frac{4}{27}
 \end{aligned}$$

2. Il tempo medio speso da un paziente nella clinica è:

$$R = \frac{1}{\mu} + \frac{\pi_m}{(m\mu(1-\rho)^2)} = \frac{1}{\frac{1}{20}} + \pi_m$$