UNIVERSITÀ DEGLI STUDI ROMA TRE
FACOLTÀ DI SCIENZE M.F.N.

Sintesi della tesi di Laurea in Matematica
di
Elisa Ciccone

Prestazioni di code infinite alimentate da traffico esattamente autosimile

Relatore
Prof. Gianpaolo Scalia Tomba

Il Candidato

Il Relatore

ANNO ACCADEMICO 2000 - 2001
FEBBRAIO 2002

Classificazione AMS: 60g8, 60k25;
Parole chiave: Processi autosimili, Teoria delle code.

Indice

Capitolo 1

Introduzione

Una rete, o network di telecomunicazione, è una struttura di collegamento tra sistemi diversi in grado di generare e scambiarsi dei segnali; un esempio comune è la rete telefonica.
Differenti da un punto di vista strutturale e funzionale sono le reti di telecomunicazione digitali: esse sfruttano, infatti, una modalità di trasmissione chiamata *a commutazione di pacchetto*.
Il traffico è composto dal flusso di pacchetti che vengono trasmessi da un punto all'altro del network; componenti fondamentali delle reti sono i *buffer*: essi sono in grado di memorizzare determinate quantità di pacchetti, in modo da consentirne l'elaborazione e l'indirizzamento; tale quantità viene detta *capacità del buffer* e il tempo di elaborazione, detto anche *tempo di servizio*, è fisso.
Una problematica fondamentale nel campo dell'ingegneria delle telecomunicazioni è il dimensionamento dei buffer; tale campo di ricerca è in piena espansione, sopratutto a causa del continuo evolversi delle reti.
Le analisi svolte hanno mostrato come il traffico nelle reti di telecomunicazione abbia due fondamentali caratteristiche:

1. una struttura di correlazione con dipendenza lungo-range,

2. una proprietà di invarianza di scala detta autosimilarità.

Il nostro obiettivo è quello di costruire modelli che siano in grado di descrivere le proprietà del traffico e analizzarne il comportamento nella rete e nei buffer. In questa ottica si sono rivelati degli strumenti particolarmente utili la teoria dei processi stocastici e la teoria delle code.
Un modello parsimonioso che descrive in modo soddisfacente le proprietà osservate è il Moto Browniano Frazionario, che utilizza un unico parametro H,

detto parametro di Hurst; tale processo è l'unico processo gaussiano, esattamente autosimile, a incrementi stazionari; il processo dei suoi incrementi si chiama Rumore Gaussiano Frazionario.
È stato studiato in particolare da I.Norros un sistema di coda con processo in arrivo

$$A(t) = mt + \sqrt{am}Z(t)$$

dove $Z(t), t \in (-\infty, \infty)$ è il Rumore Gaussiano Frazionario (moto browniano frazionario standard).
Tali studi hanno evidenziato come questi sistemi di code abbiano caratteristiche molto differenti da quelli con processi in ingresso markoviani: è stato infatti dimostrato che l'andamento della funzione di distribuzione della coda è Weibulliano,

$$P(V > x) \sim e^{-\gamma x^{\beta}} \quad \text{per } x \to \infty,$$

dove γ è una costante e $\beta = 2 - 2H \in (0, 1)$, mentre nel caso di processi in ingresso markoviani è asintoticamente esponenziale,

$$P(V > x) \sim e^{-\eta x} \quad \text{per } x \to \infty,$$

con V la dimensione della coda e η un parametro chiamato tasso di decadimento asintotico.
Non esistono allo stato attuale altri risultati analitici di rilievo e dunque risulta evidente l'importanza delle simulazioni come strumento di indagine.
Diversi algoritmi sono stati implementati per generare dei campioni di FGN e tra questi il più preciso è l'algoritmo di discretizzazzione (dFGN); tale algoritmo è risultato particolarmente utile, in questo lavoro, per implementare delle simulazioni di code con FBM come traffico di ingresso al buffer.
Da queste simulazioni si è riscontrato come le proprietà di correlazione a lungo raggio non siano le sole a condizionare in modo determinante la coda.
Particolari tipologie di traffico, come il traffico video, non sono infatti descritte in modo soddisfacente da processi esattamente autosimili come il FBM, essi presentano delle correlazioni a corto raggio molto forti e sono asintoticamente autosimili.
Si può pensare allora di modellizzarli utilizzando dei processi che presentano delle proprietà di scala differenti a seconda della scala di tempo che si considera.
Si può ad esempio pensare a dei processi che siano autosimili di parametro H_1 su un dominio di scale temporali limitato da un parametro indicato con r, e autosimili di parametro H_2 su un dominio di scale temporali maggiori di r.

Capitolo 2

Teoria dei processi stocastici e teoria delle code

2.1 Modelli di code

I processi stocastici sono modelli matematici di fenomeni aleatori che si evolvono nel tempo.
Nel nostro lavoro saranno utili le seguenti classi di processi stocastici:

Definizione 2.1. *Un processo stocastico* $X = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in T}, \{X_t\}_{t\in T}, P)$ *a valori nello spazio topologico* $(E, \mathcal{B}(E))$ *si dice:*
stazionario *se le sue distribuzioni finito dimensionali sono invarianti per traslazioni temporali.*
(X *è stazionario se* $\mu_{t_1,\dots,t_n}(A) = \mu_{t_1+h,\dots,t_n+h}(A)$ *per ogni* $t_1 < \ldots < t_n, h \in T$ *e per ogni insieme misurabile* A*).*
a incrementi stazionari *se la distribuzione congiunta degli incrementi* $(X_{t2} - X_{t1}), \ldots, (X_{tn} - X_{tn-1})$ *è la stessa di* $(X_{t2+h} - X_{t1+h}), \ldots,$
$\ldots (X_{tn+h} - X_{t(n-1)+h})$
a incrementi indipendenti *se per ogni* $t_1 < \ldots < t_n \in T$ *le variabili* $(X_{t2} - X_{t1}), \ldots, \ldots (X_{tn} - X_{tn-1})$ *sono mutuamente indipendenti.*
a incrementi indipendenti e stazionari *se soddisfa entrambi i requisiti della definizione*
processo di Markov *se*

$$P(X(t+s) \in A | \mathcal{F}_{\leq t}) = P(X(t+s) \in A | X(t)) \tag{2.1}$$

per ogni $s, t \geq 0$ *e* $A \in \mathcal{B}(E)$*.*

Un modello di accodamento è un modello usato per descrivere sistemi di code, ovvero processi in cui dei *clienti* (pacchetti) arrivano, aspettano il loro turno per essere serviti e, quindi, si allontanano.
In una coda, *l'arrivo (nel sistema) di un cliente*, ovvero l'istante in cui un cliente si accoda agli, eventuali, altri clienti, è casuale; inoltre, si fa riferimento al ritardo di trasmissione di un pacchetto, come al tempo che il cliente deve attendere per essere servito oppure, più semplicemente, come al *tempo di servizio.*
Una coda è descritta dalle seguenti funzioni:

- $N(t) = \#$ clienti nel sistema di coda all'istante di tempo t.
- $T_i =$ tempo trascorso nel sistema di coda dall'i-esimo cliente.

Solitamente, queste funzioni non sono note e devono essere stimate, quindi, siamo interessati a ciascun valore che ognuna delle funzioni appena descritte assume in media ($\bar{N}$ e $\bar{T}$).
Indichiamo con λ e δ, rispettivamente, il tasso degli arrivi e delle partenze nel sistema di coda; allora il risultato seguente mette in relazione $\bar{N}$ e $\bar{T}$:

Teorema 2.1. ***(di Little)*** *Dato un sistema stabile ($\delta = \lambda$), allora il numero medio di clienti nel sistema è dato dalla seguente relazione $\bar{N} = \lambda\bar{T}$*

Di seguito, andremo a vedere diversi tipi di sistemi di accodamento; in generale per indicare un modello di coda si usa un'etichetta del tipo $A/B/X/Y/Z$ dove:

1. Con A si caratterizza il processo stocastico secondo cui si succedono gli *arrivi* dei clienti.
2. Con B si descrive il processo dei *tempi di servizio.*
3. Con X si intende il numero (intero, possibilmente infinito) di servitori.
4. Con Y si descrive la capacità della *fila d'attesa*, cioè il massimo numero di clienti che possono trovare posto nella fila d'attesa (escludendo quindi i clienti già in servizio). Si può avere $Y = \infty$.
5. Con Z si indica la disciplina di coda, ovvero la regola che viene seguita per scegliere il prossimo cliente da servire tra quelli nella fila d'attesa.

Il parametro Y si indica solo nel caso in cui sia diverso da ∞, il parametro Z si indica solo nel caso in cui sia diverso da $FIFO$ (First In First Out).
Definiamo ora un **Processo di Poisson:**

Definizione 2.2. *Un processo di Poisson* $\{N_t, t \geq 0\}$, *è l'insieme delle variabili aleatorie definite su uno spazio di probabilità di Poisson* $(\Omega, \mathcal{A}, \mathcal{P})$, *dove*

$$\mathcal{P}(N_t = k) = \frac{(\lambda t)^k \cdot e^{-\lambda t}}{k!}, \text{ per un fissato valore di } \lambda. \tag{2.2}$$

È possibile caratterizzare un processo di Poisson come segue:

Proposizione 2.2. *Sia* $\{N_t, t \geq 0\}$ *un processo, allora risulta che:*
$\{N_t, t \geq 0\}$ *è un processo di Poisson* $\Longleftrightarrow$

1. $N_b - N_a$ *ha la stessa distribuzione di* N_{b-a} *(incrementi stazionari);*
2. $[s, t] \cap [u, v] = \varnothing \Rightarrow N_t - N_s$ *e* $N_v - N_u$ *indipendenti.*

Definiamo inoltre le variabili aleatorie W_k e T_j rispettivamente, *tempo di attesa per il k-esimo arrivo* e *j-esima lacuna*:

- $W_k =$ k-esimo arrivo $= \omega_k$
- $T_j = \omega_j - \omega_{j-1}$

La distribuzione di W_k è la funzione definita come segue:

$$P(W_k \leq t) = \begin{cases} 0 & t < 0 \\ 1 - \sum_{r=0}^{k-1} \frac{(\lambda t)^r \cdot e^{-\lambda t}}{r!} & t \geq 0 \end{cases} \tag{2.3}$$

In particolare, la distribuzione di W_1 è la *distribuzione esponenziale* di parametro λ; la distribuzione di T_1 è *esponenziale* dello stesso parametro, in quanto $T_1 = W_1 - W_0 = W_1$; la distribuzione delle altre lacune, in virtù del fatto che sono mutuamente indipendenti ed equidistribuite, allora $P(T_j \leq t) = P(T_1 \leq t)$ e quindi è anch'essa *esponenziale*.
La generica variabile aleatoria T_j appena definita su $(\Omega, \mathcal{A})$ è perciò solitamente denotata con T ed è generalmente detta *tempo di interarrivo*.
Un' importante proprietà di un processo di Poisson è la proprietà di *memoryless* del tempo di interarrivo che può essere espressa come segue:

$$P(T > r + t | T > t) = P(T > r)$$

Un processo di Poisson è generalmente considerato un buon modello per descrivere il processo originato da diversi processi di arrivo di pacchetti, indipendenti e equidistribuiti anche se i singoli processi non sono tutti descrivibili attraverso un processo di Poisson.

Analizziamo adesso i diversi sistemi di coda.
Modello di coda M/M/1: L'M/M/1 è un sistema di coda in cui gli arrivi si susseguono secondo un Processo di Poisson, cioè la distribuzione di ciascuna statistica di arrivo $T_i, i \in \{1, 2, \ldots\}$, è la distribuzione esponenziale di parametro λ, la distribuzione di ciascuna statistica di servizio $S_i, i \in \{1, 2, \ldots\}$, è la distribuzione esponenziale di parametro μ, il # server è 1, la capacità della fila di attesa è infinita e la disciplina della coda è $FIFO$; le statistiche di arrivo e le statistiche di servizio sono mutuamente indipendenti.
Dalla definizione di M/M/1, si capisce che:

- La prima M denota il fatto che ogni statistica di arrivo gode della proprietà di memoryless.
- La seconda M denota il fatto che ogni statistica di partenza gode della proprietà di memoryless.
- L'1 denota il fatto che il # di server è 1.

Modello di coda M/M/m: La più semplice generalizzazione del modello di coda M/M/1 consiste nel considerare un numero arbitrario di server (tipicamente maggiore di 1), il modello di coda che si ottiene è l'M/M/m.
Anche in questo caso la distribuzione degli arrivi è esponenziale di parametro λ, la distribuzione dei tempi di servizio è esponenziale di parametro μ, il # di server è m e si suppone che la disciplina della fila sia $FIFO$.
Modelli di coda M/E2/1 e M/H2/1: Nella coda $M/E_2/1$ gli arrivi seguono un processo di Poisson di parametro λ, mentre i tempi di servizio hanno densità di probabilità di tipo E_2 [1] con media $1/\mu$, il servizio è composto da due stadi separati esponenziali in serie, i due stadi non sono due servitori indipendenti.
Quando invece i tempi di servizio sono descritti da una variabile aleatoria H_2 [2] è possibile rappresentare il servizio mediante due stadi in parallelo statisticamente indipendenti tra loro, ciascuno con tempi di servizio distribuiti secondo densità di probabilità esponenziali.
Modello di coda M/G/1: Nelle code $M/G/1$ i tempi di servizio sono descritti da distribuzioni arbitrarie.

[1] In generale una v.a. Erlang-n (E_n) ha espressione $f_X^{(n)}(t) = \frac{(n\mu)^n t^{n-1}}{(n-1)!} e^{-n\mu t}, t \geq 0$

[2] In generale una v.a. Iperesponenziale H_n è del tipo $f_Y(t) = \sum_{i=1}^{n} \alpha_i \mu_i e^{-\mu_i t}, t \geq 0$ con $\sum_{i=1}^{n} \alpha_i = 1$

Capitolo 3

Autosimilarità: processi autosimili gaussiani e non gaussiani.

3.1 Processi α-stabili

La teoria delle distribuzioni stabili è stata sviluppata negli anni '20 e '30 principalmente da P.Levy e A.Khinchine; si possono dare quattro definizioni equivalenti di distribuzione stabile.
Una variabile aleatoria X ha una **distribuzione stabile** se:

1. per ogni coppia di numeri A e B positivi, esistono un numero positivo C e un numero reale D tali che

   $$AX_1 + BX_2 \stackrel{d}{=} CX + D \tag{3.1}$$

   dove X_1 e X_2 sono copie indipendenti di X e dove $\stackrel{d}{=}$ denota uguaglianza in distribuzione.

2. per ogni $n \geq 2$ esistono un numero positivo C_n e un numero reale D_n tali che

   $$X_1 + X_2 + \ldots + X_n \stackrel{d}{=} C_n X + D_n \tag{3.2}$$

   dove $X_1, X_2, \ldots, X_n$ sono copie indipendenti di X.

3. possiede un dominio di attrazione, cioè, se esiste una successione $Y_1, Y_2, \ldots$ di variabili aleatorie i.i.d. e una successione di numeri positivi d_n e di

numeri reali a_n, tali che

$$\frac{Y_1 + Y_2 + \ldots + Y_n}{d_n} + a_n \xrightarrow{d} X \tag{3.3}$$

dove $\xrightarrow{d}$ denota convergenza in distribuzione.

4. esistono parametri reali $0 < \alpha \leq 2, -1 \leq \beta \leq 1$, e μ, tali che la sua funzione caratteristica ha la seguente forma:

   (3.4)
   $$E[\exp{(i\theta X)}] = \begin{cases} \exp\{-\sigma^\alpha|\theta|^\alpha(1 - i\beta(sign\theta)\tan\frac{\pi\alpha}{2}) + i\mu\theta\} & \text{se } \alpha \neq 1 \\ \exp\{-\sigma|\theta|(1 + i\beta\frac{2}{\pi}(sign\theta)\ln|\theta|) + i\mu\theta\} & \text{se } \alpha = 1 \end{cases}$$

Il numero α è detto *indice di stabilità* o *esponente caratteristico.*
Una variabile aleatoria stabile con indice α verrà chiamata α-stabile.
Le distribuzioni stabili verranno denotate con $S_\alpha(\sigma, \beta, \mu)$ e con SαS le distribuzioni stabili simmetriche.
Il parametro μ è detto *parametro di shift*, il parametro σ è detto *parametro di scala*, il parametro β è invece detto *parametro di skewness* e misura in qualche modo le proprietà di simmetria della distribuzione: la distribuzione $S_\alpha(\sigma, \beta, \mu)$ si dirà *spostata a destra* se $\beta > 0$, *spostata a sinistra* se $\beta < 0$.
Una delle più importanti proprietà delle distribuzioni stabili è quella che riguarda il comportamento asintotico delle code della distribuzione $P(X > \lambda)$ e $P(X < -\lambda)$ per $\lambda \to \infty$: si ha che per $X \sim S_\alpha(\sigma, \beta, \mu)$ con $0 < \alpha < 2$, le code hanno un andamento del tipo $\lambda^{-\alpha}$, implicando dunque una molto maggiore variabilità; per una variabile stabile è cioè molto più probabile che per una variabile gaussiana assumere valori lontani dalla media.
Il comportamento asintotico è dunque una caratteristica delle distribuzioni stabili largamente usata nelle applicazioni. Una distribuzione con questa proprietà è detta *a coda pesante.*
Un ulteriore caratteristica interessante è che per $X \sim S_\alpha(\sigma, \beta, \mu)$ con indice di stabilità $0 < \alpha < 2$, si ha:

$$\begin{array}{ll} E[|X|^p] < \infty & \text{per } 0 < p < \alpha \\ E[|X|^p] = \infty & \text{per } p \geq \alpha \end{array} \tag{3.5}$$

I processi stocastici α-stabili sono dei processi che hanno distribuzioni finito dimensionali α-stabili. Le loro marginali hanno quindi varianza infinita per $0 < \alpha < 2$ e ciò li rende degli ottimi candidati per modellare fenomeni reali con comportamenti erratici.
I processi α-stabili si possono rappresentare anche in termini di integrali stocastici α-stabili, che sono degli integrali di funzioni deterministiche rispetto a misure aleatorie α-stabili.

3.2 Processi autosimili

I processi autosimili sono dei processi stocastici con la proprietà di essere invarianti in distribuzione rispetto a determinati cambiamenti di scala temporale e spaziale.
Nel contesto delle telecomunicazioni tali processi si sono rivelati degli ottimi modelli in grado di descrivere in modo parsimonioso le proprietà di autosimilarità del traffico che sono state osservate, con l'uso di un solo parametro indicato con la lettera H e chiamato *parametro di Hurst*.
Un processo stocastico a valori reali $\{X(t), t \in T\}$ è **autosimile** con indice $H > 0$ (H-as) se per ogni $a > 0$ le distribuzioni finito dimensionali di $\{X(at), t \in T\}$ sono identiche alle distribuzioni finito dimensionali di $\{a^H X(t), t \in T\}$, cioè se per ogni $d \geq 1$, $t_1, \dots, t_d \in T$ e ogni $a > 0$

$$(X(at_1), \dots, X(at_d)) \stackrel{d}{=} (a^H X(t_1), \dots, a^H X(t_d)) \tag{3.6}$$

Il parametro di Hurst è chiamato anche *coefficiente di scala* o *indice di autosimilarità* ed è un numero non negativo, l'insieme dei possibili valori che H può assumere è limitato dall'esistenza dei momenti: infatti, se $\{X(t), t \in T\}$ è H-asis (autosimile a incrementi stazionari) e $P(X(1) \neq 0) > 0$, allora la relazione $E[|X(1)|^\gamma] < \infty$ implica che

$$\begin{array}{ll} 0 < H < 1/\gamma & \text{se } 0 < \gamma < 1 \\ 0 < H \leq 1 & \text{se } \gamma \geq 1 \end{array}$$

Sottolineiamo ora la differenza tra:
processo esattamente autosimile: se $X(t) \stackrel{d}{=} m^{1-H} X^{(m)}(t)$, per ogni $m > 0$, dove $X^{(m)}(t) = \frac{1}{m} \sum_{i=(t-1)m+1}^{tm} X(i)$,
processo asintoticamente autosimile: se la relazione precedente è verificata per $m \to \infty$,
processo esattamente autosimile del secondo ordine: se per ogni m intero si ha $Var(X^{(m)}) = \sigma^2 m^{-\beta}$, con $\beta \in (0, 1)$,
processo asintoticamente autosimile del secondo ordine: se la relazione precedente è verificata per $m \to \infty$.
Il moto browniano è un esempio di processo autosimile ad incrementi stazionari con indice H = 1/2.

3.3 Moto browniano frazionario, rumore gaussiano frazionario, simulazione FGN

Per un fissato $\alpha \in (0,2)$ esistono diversi processi α-stabili H-asis.
Nel caso gaussiano $\alpha = 2$, $0 < H \leq 1$, invece, esiste un unico processo H-asis, il **moto browniano frazionario** (FBM), ed è denotato $\{B_H(t), t \in \mathbb{R}\}$. Esso è importante non solo perchè è l'unico processo gaussiano H-asis, o perchè diversi processi α-stabili H-asis si riducono ad esso nel caso $\alpha = 2$, ma anche per il suo largo uso nelle applicazioni, in particolare nel contesto della *dipendenza a lungo-range* che esamineremo in seguito.
La successione degli incrementi del FBM

$$Y_j = B_H(j+1) - B_H(j), \quad j = \ldots, -1, 0, 1, \ldots$$

è per definizione una successione stazionaria, essa è chiamata **rumore gaussiano frazionario** (FGN).
Il rumore gaussiano frazionario ha diverse proprietà interessanti: è una successione gaussiana stazionaria con media zero e varianza $E[{Y_j}^2] = E[{B_H}^2(1)] = {\sigma_0}^2$.
Vediamo ora l'implementazione di un algoritmo di simulazione di campioni finiti di rumore gaussiano frazionario.
Sono stati studiati molti esempi di tali algoritmi simulativi, tra i più usati risultano il Random Midpoint Displacement(RMD) ed il metodo di discretizzazione dFGN, quest'ultimo implementato per la prima volta da B.Mandelbrot [13]; è stato scelto il secondo per generare i campioni di traffico successivamente utilizzati nelle simulazioni.
Tale algoritmo produce delle tracce di FGN, il processo degli incrementi del FBM; quest'ultimo è dunque ottenibile sommando tali incrementi sull'intervallo desiderato.
Per simulare il FGN si approssima la sua rappresentazione integrale.
Abbiamo precedentemente visto come la ricerca delle possibili distribuzioni limite per somme di variabili aleatorie indipendenti e identicamente distribuite, porti alla definizione delle leggi α-stabili, che generalizzano il problema originario di convergenza alla distribuzione normale.
In questo contesto vediamo come i processi autosimili siano l'analogo delle distribuzioni α-stabili e come dunque possano essere visti come limiti di determinate classi di processi stocastici.
Nell'insieme di tali processi limite, il moto browniano frazionario ha un'importanza primaria, paragonabile a quella che ha la distribuzione gaussiana nell'insieme delle distribuzioni α-stabili.

Capitolo 4

Analisi dei dati e dipendenza lungo range

4.1 Modelli LRD, analisi dei dati

Descriviamo ora le caratteristiche fondamentali dei processi stocastici che presentano una forte correlazione tra elementi anche molto distanti tra loro, si dirà che tali processi presentano una struttura di correlazione a lungo raggio (LRD).
Prendiamo in considerazione i processi stazionari con media e varianza finite: sia $X(1), \ldots, X(n)$ un campione di osservazioni, v.a. i.d, questa è una serie stazionaria, dunque $E[X(t)] = \mu$ e la varianza di $\bar{X} = n^{-1} \sum_{i=1}^{n} X(i)$ è uguale a

$$Var(\bar{X}) = n^{-2} \sum_{i,j=1}^{n} \gamma(i,j) = n^{-2} \sigma^2 \sum_{i,j=1}^{n} r(i,j)$$

dove indichiamo con

$$\gamma(i,j) = Cov(X(i), X(j)) = E[(X(i) - E[X(i)])(X(j) - E[X(j)])]$$

$$r(i,j) = \frac{\gamma(i,j)}{\sigma^2}, \quad \sigma^2 = Var(X(i)) \quad \text{per } i = 1, \ldots, n$$

rispettivamente l'autocovarianza e l'autocorrelazione del processo.
Se $\sum_{i \neq j}^{n} r(i,j) = 0$, allora

$$Var(\bar{X}) = n^{-1} \sigma^2 \tag{4.1}$$

le $X(1), \dots, X(n)$ sono mutuamente scorrelate, se invece $\sum_{i \neq j}^{n} r(i,j) \neq 0$, allora

$$Var(\bar{X}) = n^{-1}[1 + \delta_n(r)]\sigma^2 \tag{4.2}$$

con il coefficiente correttivo non nullo

$$\delta_n(r) = n^{-1} \sum_{i \neq j}^{n} r(i,j)$$

La varianza di $\bar{X}$ è in entrambi i casi proporzionale a n^{-1}, ciò è vero per ogni successione di osservazioni per cui $\delta(r) = \lim_{n \to \infty} \delta_n(r)$ esiste, è finito ed è maggiore di -1. In questo caso si ha asintoticamente

$$Var(\bar{X}) \sim n^{-1}[1 + \delta(r)]\sigma^2 = n^{-1}c(r)\sigma^2 \tag{4.3}$$

I processi di Markov, ad esempio, presentano questo comportamento; nel caso della dipendenza a lungo-range invece il comportamento della varianza di $\bar{X}$ ha un decadimento a zero più lento

$$Var(\bar{X}) \sim n^{-\alpha}c(r)\sigma^2 \tag{4.4}$$

per una certa costante $\alpha \in (0,1)$ e dove $c(r) = \lim_{n \to \infty} n^{\alpha - 2} \sum_{i \neq j}^{n} r(i,j)$.
Partendo dalla forma di $c(r)$ e di $\delta_n(r)$ si ricava che al crescere della dimensione n del campione

$$\sum_{k=-(n-1)}^{n-1} r(k) \sim cost \cdot n^{1-\alpha},$$

ed essendo $\alpha < 1$ si ha che

$$\sum_{k=-\infty}^{\infty} r(k) = \infty \tag{4.5}$$

dunque si può dire che le correlazioni decadono a zero così lentamente che non sono sommabili.
Abbiamo la seguente situazione con $0 < H < 1$:

- per $1/2 < H < 1$ si ha che il processo $X(t)$ ha memoria lunga:

$$\sum_{k=-\infty}^{\infty} r(k) = \infty,$$

- per $H = 1/2$ le osservazioni sono scorrelate,
- per $0 < H < 1/2$ si ha che

$$\sum_{k=-\infty}^{\infty} r(k) = 0.$$

Vi sono diversi metodi euristici di stima del parametro H che controlla la memoria lunga, il più conosciuto è la statistica R/S; un altro metodo molto usato è il grafico di $Var(\bar{X}(n))$ su n scala logaritmica detto *variance-time plot* [12].
Consideriamo il variance-time plot: esso è stato usato per la stima di H di quasi tutte le tracce reali a disposizione, come di quelle simulate. Esso sfrutta la proprietà della varianza della media campionaria

$$Var(X^{(m)}) = Var(\bar{X}(m)) \sim cm^{2H-2}$$

(dalla (4.4) sostituendo $H = 1 - \alpha/2$) dove $c > 0$.
Facendo dunque un grafico di $Var(X^{(m)})$ su m scala logaritmica, si ottiene una retta di pendenza $\beta = 2H - 2$; tale pendenza viene stimata con il metodo di regressione lineare basato sui minimi quadrati a partire da un certo valore di k.
Nel caso di pendenza corto range o di indipendenza, la pendenza finale è $\beta = 2\frac{1}{2} - 2 = -1$; più il processo è correlato più la pendenza sarà minore.
Nella Figura 4.1 mostriamo i variance-time plot di tracce di FGN simulate. Per quanto riguarda la prima traccia FGN-$H = 0.83$ la pendenza della retta è sempre negativa ed è $\beta \sim -0.34$ e dunque H stimato risulta ~ 0.8, per la seconda FGN-$H = 0.50$ invece, la pendenza della retta è $\beta \sim -1$ e dunque H stimato risulta $\sim 1/2$.
Vogliamo sottolineare che il variance-time plot è utile anche ad un altro scopo, oltre a stimare il parametro H, è cioè possibile valutare attraverso di esso se un dato processo è esattamente autosimile del secondo ordine o se lo è solo asintoticamente.
Diciamo allora che se il variance-time plot di una certa serie risulta approssimativamente rettilineo per ogni valore di k, allora tale serie si può considerare generata da un processo esattamente autosimile del secondo ordine, se invece il variance-time plot presenta un andamento curvilineo che solo successivamente si attesta su una retta, diremo che la serie sarà generata da un processo asintoticamente autosimile del secondo ordine; per le due serie di FGN simulate notiamo quindi come il loro variance-time plot sia coerente con il fatto che esse sono esattamente autosimili del secondo ordine.

Figura 4.1: **Variance-time plot**

Qualitativamente diverso è il variance-time plot di una traccia video, tale comportamento diverso è imputabile alla presenza di forti correlazioni a corto raggio, esso è mostrato nella Figura 4.2 La traccia non può essere considera-

Figura 4.2: **Variance-time plot traccia video**

ta esattamente autosimile, ma solo asintoticamente autosimile: si individua infatti una parte finale del variance-time plot con andamento rettilineo che corrisponde alla zona asintotica.

Capitolo 5

Sistemi di code con input autosimili e analisi delle prestazioni

5.1 Modello di coda con ingresso FBM

Studiamo ora dei sistemi di code per i quali i processi in ingresso sono dei processi stocastici autosimili con proprietà di dipendenza a lungo-range.
Allo stato attuale, i risultati riguardanti tali code sono relativamente pochi ed incompleti e si basano per lo più su approssimazioni e proprietà asintotiche.
Presentiamo il modello analizzato da I.Norros [2], che rappresenta il solo esempio di modello di coda con processo in ingresso avente dipendenza a lungo-range, per il quale abbiamo dei risultati analitici di una certa importanza.
Si considera come processo di ingresso al sistema un moto browniano frazionario di parametro $1/2 \leq H \leq 1$, cioè un processo esattamente autosimile con una funzione di autocorrelazione non sommabile (LRD); il tempo di servizio viene considerato deterministico.
Consideriamo $V(t)$ la variabile aleatoria rappresentante il numero di clienti nel sistema in condizioni di equilibrio, in altre parole la lunghezza della coda nell'istante t in condizioni di stazionarietà.
Tale quantità aleatoria può essere definita anche nel modo seguente:

$$V(t) = \sup_{s \leq t}(A(t) - A(s) - C(t-s)), \quad t \in (-\infty, \infty) \tag{5.1}$$

dove $A(t)$ è il seguente processo

$$A(t) = mt + \sqrt{am}Z(t), \quad t \in (-\infty, \infty) \tag{5.2}$$

con $Z(t)$ un moto browniano frazionario standard, con parametro di Hurst $H \in (1/2, 1)$.
Questo rappresenta il modello per un sistema di coda con ingresso $A(t)$, che costituisce il processo di arrivo cumulativo fino al tempo t; esso è caratterizzato da quattro parametri, m, a, H, C, con la seguente interpretazione:

1. $m > 0$ rappresenta la *media* del processo in igresso;
2. $a > 0$ parametro detto *coefficente di varianza*, rappresenta il rapporto tra la varianza e la media del processo in ingresso;
3. $H \in [1/2, 1)$ è il *parametro di Hurst* di $Z(t)$;
4. $C > m$ è la *capacità di servizio*, rappresenta il numero medio di clienti che possono essere serviti nell'unità di tempo ed è una costante.

Teorema 5.1. *Sia $V(t)$ il processo stazionario definito nella* (5.1)*, si scelga l'unità di tempo in modo tale che $C = 1$, si ha allora che*

$$P(V(t) > x) \geq \Phi\left(\frac{1}{\sqrt{am}}\left(\frac{1-m}{H}\right)^H \left(\frac{x}{1-H}\right)^{1-H}\right) \tag{5.3}$$

dove $\Phi(y) = P(Z(1) > y)$ è la funzione di sopravvivenza della distribuzione gaussiana standard.

Dimostrazione : Essendo $V(t)$ stazionario si ha che

$$P(V(t) > x) = P(V(0) > x) = P(\sup_{t \geq 0}(-A(s) + s) > x)$$

Operando quindi un'inversione dei tempi $-s = t$, segue che

$$P(\sup_{-s \geq 0}(A(-s) - (-s)) > x) = P(\sup_{t \geq 0}(A(t) - t) > x)$$

essendo le distribuzioni marginali di $A(-s)$ e $-A(s)$ identiche. Si ottiene dunque

$$\begin{aligned} P(V(t) > x) &\geq \max_{t \geq 0} P(A(t) > t + x) = \max_{t \geq 0} P\left(Z(t) > \frac{(1-m)t + x}{\sqrt{am}}\right) \\ &= \max_{t \geq 0} P\left(Z(1) > \frac{(1-m)t + x}{\sqrt{am} \cdot t^H}\right) = \max_{t \geq 0} \Phi\left(\frac{(1-m)t + x}{\sqrt{am} \cdot t^H}\right); \end{aligned}$$

dove si è sfruttato il fatto che $Z(t)$ è esattamente autosimile .
Dalle proprietà di $\Phi(y)$, che è monotona decrescente in y, si ha che

$$\max_{t\geq 0}\Phi\left(\frac{(1-m)t+x}{\sqrt{am}\cdot t^H}\right)=\Phi\left(\min_{t\geq 0}\frac{(1-m)t+x}{\sqrt{am}\cdot t^H}\right)$$

Differenziando si vede che tale massimo si ha per

$$t=\frac{Hx}{(1-H)(1-m)}$$

quindi sostituendo questo valore nella formula di sopra, si ottiene la (5.3). $\square$
La (5.3) rappresenta una stima dal basso per la funzione $P(V>x)$, valida in generale quando il traffico in ingresso è esattamente autosimile.
Per esso si può ottenere una forma più utile nella pratica, usando la seguente approssimazione asintotica

$$\Phi(y)\sim\frac{1}{\sqrt{2\pi}(1+y)}\exp\left(-y^2/2\right)\sim\exp\left(-y^2/2\right).$$

Si ottiene dunque la seguente stima prevista dal modello per la funzione $P(V>x)$

(5.4)

$$P(V(t)>x)\geq\exp\left(-\left[\frac{1}{2am(1-H)^2}\left(\frac{(1-m)(1-H)}{H}\right)^{2H}\right]x^{2(1-H)}\right)$$

Il comportamento della coda della funzione di distribuzione di V avrà allora un andamento che nelle migliori delle ipotesi sarà Weibulliano, cioè del tipo

$$P(V>x)\sim\exp\left(-\gamma x^{\beta}\right),\ \text{con } \beta\leq 1.$$

L'espressione ottenuta (5.4) nel caso browniano $H=1/2$, $a=1$ si riduce alla distribuzione esponenziale

$$P(V>x)\sim\exp\left(-2\frac{1-m}{m}x\right)$$

che risulta coincidere con l'approssimazione asintotica, nel caso di traffico pesante, per il sistema $M/D/1$ che abbiamo visto in precedenza.
Due caratteristiche fondamentali sono:

- la distribuzione della lunghezza della coda nel buffer è invariante rispetto a cambiamenti di scala temporale, in altre parole $V(\alpha t)$ ha la stessa distribuzione di $V(t)$; tale proprietà deriva dall'ipotesi fatta di stazionarietà;
- $Z(\alpha t)$ ha la stessa distribuzione di $\alpha^H Z(t)$; tale proprietà deriva dall'autosimilarità del processo di arrivo.

5.2 Simulazioni, ipotesi di modello per il traffico asintoticamente autosimile

Analizziamo infine sperimentalmente, le proprietà del modello di coda discusso precedentemente, effettuando cioè delle simulazioni di code con processi in ingresso esattamente autosimili (FBM).
Descriviamo dunque l'ambiente di simulazione utilizzato: si considera una coda, o buffer, con capacità infinita, un singolo server e con processo in ingresso la serie di dati in esame.
In queste simulazioni verrà considerato un metodo di trasmissione dei dati detto ATM: nei buffer ATM le tracce in ingresso sono suddivise in celle di dimensione costante, ogni cella è composta da 53 bytes e ogni 8 celle vengono aggiunti 5 bytes detti di *overhead* contenenti delle informazioni sull'indirizzamento, utilizzate dal protocollo di trasmissione AAL.
Le celle ATM generate dalle tracce verranno poi servite dal buffer ATM con tempo di servizio determinstico e con disciplina FIFO (First In First Out); nel caso in cui i canali in ingresso siano N e in un determinato istante ci siano due o più celle in entrata contemporaneamente, viene assegnata a tali celle una priorità di servizio in modo casuale; il carico richiesto nella simulazione verrà poi ottenuto variando la capacità in uscita.
Possiamo ora verificare l'andamento Weibulliano della distribuzione asintotica della coda, previsto dal modello discusso, per il quale (5.2) è il traffico in arrivo, confrontandolo con la distribuzione empirica di una coda generata da una simulazione in cui il processo in ingresso è un moto browniano frazionario.
La figura 5.1 mostra il grafico di $\log P(V > x)$ su x, per $P(V > x)$ ottenuta in un caso della simulazione e nell'altro dall'andamento asintotico.
È stato considerato un carico $\rho = 0.75$; l'approssimazione sembra essere sufficentemente accurata.
Tra gli altri effetti che questo modello prevede, vi è il peggioramento delle prestazioni in coda all'aumentare del parametro di Hurst H per il processo in ingresso e all'aumentare del carico ρ.
In generale comunque, esistono diverse caratteristiche del traffico in entrata in un buffer che possono condizionare le prestazioni.
Le proprietà di correlazione, ad esempio, sono importanti su tutte le scale di tempo; si è riscontrato infatti come le correlazioni a corto range abbiano anch'esse un impatto non trascurabile sulle prestazioni, almeno nei casi in cui esse siano sufficentemente forti.
È stato rilevato che il traffico video, asintoticamente autosimile, è caratterizzato da correlazioni a corto range molto forti.

Figura 5.1: **Simulazione a carico 0.75**

È allora necessario un modello diverso di FBM per questo tipo di traffico, si potrebbe pensare quindi ad un modello che sia autosimile su due domini di scale temporali diverse, cioè

$$X(\alpha t) \stackrel{d}{=} \begin{cases} \alpha^{H_1} X(t) & \text{per } 0 < \alpha \leq r \\ \alpha^{H_2} X(t) & \text{per } \alpha > r \end{cases} \tag{5.5}$$

dove $r > 0$ è il parametro che definisce l'ampiezza dell'insieme di scale temporali su cui il processo è autosimile di parametro H_1, mentre H_2 è il parametro di autosimilarità asintotico, su scale temporali molto grandi.

Bibliografia

[1] K.Park, W.Willinger *Self-Similar Network Traffic and Performance evaluation*, Jhon Wiley & Sons, 2000

[2] I.Norros *A storage model with self-similar input*, Queueing Systems vol 16, 1994, 387-396

[3] E.Castamagna, G.Iacovoni, M.Isopi *How self-similar processes persistence determines the queue length distribution moments*, Proc. of Tenth IEEE TYRRENIAN Workshop, Ischia, 1998, 383-392

[4] W.Willinger, M.S.Taqqu, W.E.Leland, D.V.Wilson *On the self-similar nature of Ethernet traffic*, IEEE/ACM Trans. on Networking, 1994, 1-15

[5] W.Willinger, M.S.Taqqu, W.E.Leland, D.V.Wilson *Self-similarity in high-speed packet traffic analysis and modeling of Ethernet traffic measurement*, Sta. Sci. 1995, 67-85

[6] O.Narayan *Exact asymptotic queue length distribution for Fractional Brownian Traffic*, Advances in Performance Analisys, vol 1, 1998

[7] S. Hatangadi *Communication Networks*, Lecture N6, 1998

[8] D.Gross, C.Harris *Fundamentals of Queueing Theory*, New York, Wiley, 1974

[9] J.Lamperti *Stochastic Processes: A Survey of the Mathematical Theory*, New York, Springer-Verlag

[10] S.N.Ethier, T.G.Kurtz *Markov Processes characterization and convergence*, New York, John Wiley & Sons

[11] J.Jacod, A.N.Shiryaev *Limit Theorems for Stochastic Processes*, Springer Verlag, 1987

[12] J.Beran *Statistics for Long Memory Processes*, New York: Chapman and Hall, 1994

[13] B.Mandelbrot *A fast fractional gaussian noise generator*, W. Resources n 3, 1971, 543-553

[14] W.Willinger, M.S.Taqqu, R.Sherman *Proof of a fundamental result in self-similar traffic modelling*, Computer Communication Review vol 26, 1997, 5-23

[15] G.Samorodnitsky, M.S.Taqqu *Stable Non-Gaussian Processes: Stochastic Models with infinite Variance*, New York: Chapman and Hall, 1994

[16] J.W.Lamperti *Semi-stable stochastic processes*, Trans. Am. Math. Soc. 104, 1962, 62-78

[17] J.Beran, W.Willinger, M.Taqqu, R.Sherman *Long-range dependence in VBR video traffic*, IEEE Trans. Comm., 1995, 1566-1579

[18] I.P.Cornfield, Y.G.Sinai, S.V.Fomin *Ergodic Theory*, Springer 1982