Capitolo 5

Analisi dei dati di output

Una fase essenziale di ogni studio di simulazione è l'analisi dei risultati della simulazione stessa. Supponiamo di avere costruito il modello di un sistema e siano $Y_1, Y_2, \ldots, Y_m$ i dati di output che ci interessa studiare. Ad esempio Y_i rappresenti la lunghezza di una coda alla fine dell'*iesimo* intervallo di tempo in cui è stata suddivisa la giornata. Chiaramente Y_i può essere pensata come una variabile casuale. La difficoltà qui sta nel fatto che le variabili casuali $Y_1, Y_2, \ldots, Y_m$ non sono in generale indipendenti; quindi i metodi visti nei precedenti capitoli non possono essere direttamente utilizzati.

Se supponiamo però di avere effettuato n diversi *run* di simulazioni utilizzando diverse sequenze di numeri pseudocasuali, abbiamo diverse sequenze di realizzazioni delle variabili casuali $Y_1, Y_2, \ldots, Y_m$:

$$\begin{array}{ccccc} y_{11}, & \ldots, & y_{1i}, & \ldots, & y_{1m} \\ y_{21}, & \ldots, & y_{2i}, & \ldots, & y_{2m} \\ \vdots & & & & \vdots \\ y_{n1}, & \ldots, & y_{ni}, & \ldots, & y_{nm} \end{array} \tag{5.1}$$

Una sequenza $y_{1i}, y_{2i}, \ldots, y_{ni}$ può essere vista come una sequenza di realizzazioni di n variabili casuali identicamente distribuite; ciò permette l'utilizzazione delle tecniche di analisi studiate.

5.1 Analisi del transitorio

Un problema di notevole importanza in una simulazione è quello della scelta delle condizioni iniziali. A questo proposito è essenziale distinguere tra sistemi con terminazione e sistemi di cui siamo interessati al comportamento a regime.

Ad esempio lo sportello di una banca che apre alle 8,30 e chiude alle 13,30 è un tipico esempio di sistema del primo tipo. Si tratta di un sistema in cui l'andamento della coda dei clienti ha un andamento intrinsecamente variabile: si parte all'inizio con nessun cliente in coda, e alla chiusura non si accettano più clienti e si esaurisce la coda. È proprio a questa variabilità che siamo interessati.

Diverso è il caso di un sistema di produzione continuo, in cui siamo interessati ad analizzare, ad esempio, il tempo richiesto a regime perché un pezzo venga prodotto, oppure l'intervallo di tempo (sempre a regime) tra l'arrivo di un ordine e la spedizione del prodotto ordinato. In questo caso il transitorio può falsare in modo rilevante i risultati del modello.

Ad esempio, supponiamo di essere interessati al tempo medio di attesa, a regime, di fronte ad una data macchina di una linea di produzione, che indicheremo con d. Siano Y_1, Y_2, ..., Y_m, i valori del parametro che si vuole stimare ottenuti tramite una simulazione. Se indichiamo con Y la variabile casuale 'tempo di attesa a regime', abbiamo

$$d = E[Y] = lim_{j \to \infty} E[Y_j].$$

Una stima di d possiamo ottenerla usando la media campionaria

$$\overline{Y}_m = \frac{\sum_{j=1}^{m} Y_j}{m},$$

dove m è il numero di osservazioni di cui disponiamo.

Nell'effettuare la simulazione dobbiamo scegliere le condizioni iniziali del sistema. Ad esempio, se decidiamo di iniziare la simulazione col sistema scarico, sarà $Y_1 = 0$. Ciò ovviamente si riflette sulla stima ottenuta falsandola. Si potrebbe pensare di partire da una situazione il più possibile simile a quella che si ha a regime, ma questo sposta solo il problema essendo proprio questa situazione quella che noi vogliamo stimare.

Una possibile soluzione è allora quella di non considerare nella stima le prime osservazioni, quelle che sono più influenzate dalle condizioni iniziali. La media viene allora stimata dalla

$$\overline{Y}_{ml} = \frac{\sum_{j=l+1}^{m} Y_j}{m - l},$$

dove l è il numero di osservazioni che vengono scartate, quelle cioè che corrispondono alla fase del transitorio. Il problema è quello di una corretta scelta di l; infatti, un valore troppo basso rischia di portare ad una stima in cui si risente delle condizioni inziali, mentre un valore troppo alto porta a simulazioni eccessivamente costose.

Non esistono regole sicure per l'individuazione del transitorio. Una ragionevole procedura è riportata nel seguito.

1. Si effettuano n repliche della simulazione, ciascuna di lunghezza m; sia y_{ij} l'osservazione j^{esima} della i^{esima} replica.

2. Costruiamo la sequenza $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_m$, con $\overline{Y}_j = \frac{\sum_{i=1}^{n} y_{ij}}{n}$. Risulta $E[\overline{Y}_j]=E[Y_j]$, e $Var[\overline{Y}_j]=Var[Y_j]/n$.

3. Sostituiamo ora alla sequenza $\overline{Y}_1$, $\overline{Y}_2$, …, $\overline{Y}_m$, la nuova sequenza $\overline{Y}_1(k)$, $\overline{Y}_2(k)$, …, con $k \leq \lfloor m/2 \rfloor$, e

$$\overline{Y}_j(k) = \begin{cases} \frac{\sum_{h=-(i-1)}^{i-1} \overline{Y}_{i+h}}{2i-1} & , i - 1, \ldots, k \\ \frac{\sum_{h=-k}^{k} \overline{Y}_{i+h}}{2k+1} & j = k+1, \ldots, m-k. \end{cases}$$

4. Scegliere infine quel valore di l oltre il quale la sequenza $\{\overline{Y}_j(k)\}$ appare giunta a convergenza.

Si tratta, come si vede facilmente, di un approccio basato sulla ispezione da parte dell'esperto, ma dopo avere sottoposto i dati ad un trattamento che ha lo scopo fondamentale di ridurne la varianza. Tale trattamento consiste prima in una media fra dati corrispondenti nelle diverse repliche, poi nella sostituzione di ognuno dei dati così ottenuti con la media fra esso e i dati immediatamente precedenti e quelli immediatamente seguenti.

Grande cura bisognerà avere nella scelta dei parametri m, n e k. Il primo dovrà essere sufficientemente grande da risultare molto maggiore del valore aspettato per l e tale da permettere nella simulazione un numero elevato di occorrenze di tutti gli eventi, anche di quelli poco probabili. Per il secondo può essere opportuno di partire inizialmente con 5 o 10 repliche per poi aumentare tale valore se necessario. Infine k va scelto sufficientemente grande da rendere regolare il grafico della $\overline{Y}_j(k)$, ma non così grande da non permettere di distinguere bene il transitorio. Ovviamente nella scelta è fondamentale il ruolo dell'esperienza di chi effettua la simulazione.

Esercizio Una linea di produzione consiste di una cella di lavorazione ed una stazione di ispezione, in serie. I pezzi semilavorati arrivano alla cella con tempi di interarrivo esponenziali con media 1 minuto. I tempi di lavorazione nella cella sono uniformi nell'intervallo [0.65, 0.70] (in minuti). I tempi di ispezione sono uniformi in [0.75, 0.80]. Il 10% dei pezzi risultano difettosi e sono rimandati alla cella per essere lavorati di nuovo. La cella è soggetta ad interruzioni nella lavorazione a causa di guasti; l'intervallo di tempo fra un guasto ed il successivo ha legge esponenziale con media 6 ore, ed il tempo di riparazione è uniforme nell'intervallo [8, 12] (in minuti). Si voglia determinare la produzione oraria (numero di pezzi) a regime. Si usi l'approccio precedentemente descritto per l'individuazione del transitorio.

5.2 Tecniche per la riduzione della varianza

Tipico obiettivo di uno studio di simulazione è la stima di uno o più parametri. La bontà della stima che si ottiene sarà tanto migliore quanto minore sarà la varianza dello stimatore usato. Nel seguito presenteremo alcune tecniche per la riduzione della varianza.

5.2.1 Variabili antitetiche

Supponiamo di volere stimare $\theta = E[X]$, e supponiamo di avere generato due variabili casuali, X_1 e X_2, identicamente distribuite con media θ. È allora

$$Var\left[\frac{X_1 + X_2}{2}\right] = \frac{1}{4}\left(Var[X_1] + Var[X_2] + Cov[X_1, X_2]\right) \tag{5.2}$$

Se le due variabili casuali X_1 e X_2 fossero correlate negativamente, attraverso il loro uso avremmo una riduzione della varianza.

Supponiamo che la variabile casuale X di cui vogliamo stimare la media sia una funzione di m numeri casuali, uniformi in [0,1):

$$X = h(U_1, U_2, \ldots, U_m). \tag{5.3}$$

Si può allora usare X come X_1 e porre

$$X_2 = h(1 - U_1, 1 - U_2, \ldots, 1 - U_m). \tag{5.4}$$

Essendo $1-U$ anch'essa una variabile casuale uniforme in [0,1), X_2 ha la stessa distribuzione di X, ed essendo $1-U$ negativamente correlata con U, si può provare che, se h è una funzione monotona, allora anche X_1 ed X_2 sono correlate negativamente.

Esempio Consideriamo una coda, e sia D_i il tempo di attesa in coda dell'*iesimo* cliente. Supponiamo di volere stimare $\theta = E[X]$, avendo indicato con X il tempo di attesa totale dei primi n clienti:

$$X = D_1 + \cdots + D_n. \tag{5.5}$$

Indicando con T_i l'*iesimo* tempo di interarrivo e con S_i l'*iesimo* tempo di servizio. Possiamo allora scrivere

$$X = h(T_1, \ldots, T_n, S_1, \ldots, S_n), \tag{5.6}$$

dove h può ragionevolmente essere assunta monotona.

Siano F e G le distribuzioni di T e di S rispettivamente, e supponiamo di usare il metodo dell'inversione per generare tali variabili casuali a partire da $2n$ numeri casuali uniformi:

$$T_i = F^{-1}(U_i),\ S_i = G^{-1}(U_{n+i}),\ i = 1, \ldots, n. \tag{5.7}$$

Una variabile casuale "antitetica" con la stessa distribuzione di X è ottenibile effettuando una seconda simulazione usando i numeri casuali $1 - U_i$, $i = 1, \ldots, n$.

5.2.2 Condizionamento

Sia X una *v.c.* di cui si voglia stimare la media $\theta = E[X]$, e sia Y un'altra *v.c.*. Assumiano nel seguito che sia X che Y siano *v.c.* discrete; il caso di *v.c.* continue è analogo e viene lasciato per esercizio al lettore.

Definiamo ora la nuova variabile casuale Z funzione di Y:

$$\begin{aligned} Z = E[X|Y=y] &= \sum_x x P[X=x|Y=y] \\ &= \sum_x x \frac{P[X=x, Y=y]}{P[Y=y]}. \end{aligned}$$

Facciamo vedere che la media di Z è proprio il valore θ cercato:

$$\begin{aligned} E[Z] &= \textstyle\sum_y E[X|Y=y]P[Y=y] &&= \sum_y \sum_x xP[X=x,Y=y] \\ &= \textstyle\sum_x x \sum_y P[X=x,Y=y] &&= \sum_x xP[X=x] = E[X] = \theta. \end{aligned}$$

Analizziamo la varianza di Z. Abbiamo che è

$$\begin{aligned} Var[X|Y=y] &= E[(X - E[X|Y=y])^2|Y=y] \\ &= E[X^2|Y=y] - (E[X|Y=y])^2. \end{aligned}$$

Si ha allora:

$$\begin{aligned} E[Var[X|Y=y]] &= E[E[X^2|Y=y] - (E[X|Y=y])^2] \\ &= E[X^2] - E[(E[X|Y=y])^2], \\ Var[Z] &= E[Z^2] - (E[Z])^2 \\ &= E[(E[X|Y=y])^2] - (E[X])^2, \end{aligned}$$

e sommando membro a membro si ottiene

$$E[Var[X|Y=y] + Var[Z] = E[X^2] - (E[X])^2 = Var[X], \tag{5.8}$$

da cui

$$Var[Z] = Var[X] - E[Var[X|Y=y]] \leq Var[X]. \tag{5.9}$$

Quindi la *v.c.* Z ha la stessa media di X con una varianza minore (o uguale): può allora essere conveniente usare Z per la stima di θ.

Esempio Si voglia stimare la somma dei tempi di attesa in una coda

$$\theta = E[\sum_i W_i], \tag{5.10}$$

dove W_i è il tempo di attesa dell'i^{esimo} cliente. Si assume una politica di tipo *FIFO*.

Sia N_i il numero dei clienti presenti nel sistema all'istante di arrivo del cliente i^{esimo}, cioè il numero dei clienti in attesa più l'eventuale cliente che viene servito. Supponiamo anche che i tempi di servizio siano esponenziali con media μ.

Introduciamo allora la nuova *v.c.* $Z = \sum_i E[W_i|N_i]$. Essendo

$$E[W_i|N_i] = N_i\mu, \tag{5.11}$$

si ha quindi

$$\theta = E[Z] = E[\sum_i N_i\mu]. \tag{5.12}$$

Pertanto l'uso delle N_i invece delle W_i consente di ottenere uno stimatore con minore varianza e quindi più accurato.

Osserviamo che abbiamo usato la proprietà della distribuzione esponenziale per cui se il tempo di servizio è esponenziale con media μ, anche il tempo necessario per completare, a partire da un tempo fisssato (quello dell'arrivo del cliente i^{esimo}), il servizio è esponenziale con la stessa media (*assenza di memoria*[1] della distribuzione esponenziale).

[1]Una variabile casuale X, non negativa, ha la proprietà di assenza di memoria se è $P[X > s+t|X > s] = P[X > t]$, il che è vero per la distribuzione esponenziale, essendo $P[X > s+t|X > s] = \frac{P[X>s+t, X>s]}{P[X>s]} = \frac{P[X>s+t]}{P[X>s]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P[X > t]$.