

Università degli Studi di Pisa
Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Scienze dell'Informazione



Tesi di Laurea

Modelli di simulazione per l'analisi di servizi ambulatoriali di un ospedale

Candidati

Mauro Deidda

Marco Piero Franco Mele

Relatore

Prof. Giorgio Gallo

Correlatore

Dr. Tommaso D'Angelo

Controrelatore

Prof. Carlo Montangero

Anno Accademico 1998-1999

Ai miei genitori e mia sorella.

Marco

*A mia madre, ad Alessandra e alla mia famiglia,
per avermi sostenuto e incoraggiato con affetto.*

Mauro

Indice

I	Centro Prelievi	13
1	Concettualizzazione del modello	14
1.1	Analisi del sistema	15
1.2	Modello concettuale	18
1.2.1	Entità	19
1.2.2	Eventi	20
1.2.3	Attività	21
1.2.4	Variabili in input	21
1.2.5	Diagrammi dei cicli di attività	23
2	Modello dei dati in input	27
2.1	Raccolta dei dati	29
2.2	Identificazione delle distribuzioni	33
2.2.1	Istogrammi	33
2.2.2	Scelta della famiglia di distribuzioni	35
2.2.3	Diagrammi Quantile-Quantile	35
2.3	Stima dei parametri	36
2.3.1	Media campionaria e varianza campionaria	37
2.3.2	Stimatori di massima verosimiglianza	37

2.4	Valutazione della bontà delle scelte fatte sulle distribuzioni	39
2.4.1	Test χ^2	40
2.4.2	Test di Kolmogorov-Smirnov	42
2.5	Un esempio	46
2.6	Test di Kruskal-Wallis	51
2.7	Risultati delle analisi	53
2.7.1	Tempi di servizio	53
2.7.2	Processo di arrivo	56
2.7.3	Altri risultati	57
3	Implementazione	59
3.1	Scelta del software	59
3.1.1	Modellazione	61
3.1.2	Animazione	62
3.1.3	Capacità statistiche	62
3.1.4	Esecuzione	63
3.1.5	Presentazione degli output	63
3.2	<i>ExtendTM</i>	64
3.2.1	L'ambiente di modellazione di Extend	65
3.2.2	Esempio di modello	66
3.2.3	Personalizzazione di Extend	70
3.2.4	Librerie Standard, BPR e Manufacturing	74
3.3	Implementazione del modello del Centro Prelievi	74
3.3.1	Accettazione	77
3.3.2	Pagamento ticket	85
3.3.3	Prelievi	85

3.3.4	Interfaccia utente	90
4	Verifica, calibrazione e validazione	96
4.1	Verifica del modello di simulazione	98
4.2	Calibrazione e validazione	99
4.2.1	Sviluppo del modello con una alta <i>face validity</i>	100
4.2.2	Validazione delle assunzioni del modello	101
4.2.3	Validazione delle trasformazioni ingresso-uscita	101
4.2.4	Validazione ingresso-uscita con dati di ingresso storici	115
4.3	Qualche prova con il modello	120
II	Reparto di Gastroenterologia	123
5	Formulazioni e modelli concettuali	124
5.1	Centro antidiabetico	124
5.2	Ambulatorio di Gastroenterologia e Day Hospital	131
5.2.1	Ambulatorio di Gastroenterologia	131
5.2.2	Day Hospital	137
5.3	Gastrosocopia	142
	Bibliografia	153

Ringraziamenti

Desideriamo ringraziare sentitamente il Prof. Giorgio Gallo per la fiducia, i preziosi consigli e la disponibilità che ha voluto offrirci, dal quale abbiamo appreso tanto.

Ringraziamenti cordiali vanno inoltre al Dr. Tommaso D'Angelo per averci dato la possibilità di realizzare questa tesi e per l'interesse con il quale ha seguito il suo sviluppo.

Un ringraziamento particolare va indirizzato al Dr. Renato Colombai per la pazienza e l'entusiasmo avuti durante le lunghe discussioni che sono state di grande aiuto nella ricerca di soluzioni per lo svolgimento del lavoro per il Centro Prelievi.

Siamo grati al Dr. Andrea Geracitano che ha creduto all'utilità di questa tesi e ha contribuito notevolmente alla sua realizzazione.

Ringraziamo il Dr. Capria, il Dr. Metrangolo ed il Dr. Tumino per la disponibilità e cortesia avute durante lo studio degli ambulatori del Reparto di Gastroenterologia.

Un grazie al Dott. Roberto Aringhieri per gli utili consigli che ha voluto darci.

Doverosi ringraziamenti vanno alle allieve infermiere professionali che hanno svolto brillantemente il gravoso compito della rilevazione dei tempi, malgrado i loro impegni di studio, al personale della Direzione Sanitaria, del Reparto di Gastroenterologia e del Centro Prelievi.

Introduzione

Nel corso dell'ultimo decennio si sono sviluppati approcci, metodi e tecniche per la *reingegnerizzazione dei processi*. Il concetto di *Business Process Reengineering* (BPR), nasce come una metodologia per modificare o riprogettare radicalmente i processi fondamentali di un'organizzazione con l'obiettivo di ottenere una più alta produttività, costi più bassi, una maggiore efficienza e una maggiore soddisfazione degli utenti destinatari del servizio o del prodotto. Per *processo* (o “processo di servizio”, cioè teso all'erogazione di servizi), si intende un insieme di attività tra loro interrelate, svolte da una o più unità organizzative, finalizzate alla realizzazione di un risultato definito e misurabile (prodotto/servizio) che trasferisce valore al fruitore del servizio (il cliente). Uno strumento potente che gli amministratori possono usare come ausilio per prendere decisioni nella fase di reingegnerizzazione è la simulazione.

La simulazione di un processo consiste nel fare degli esperimenti con un modello, che è la rappresentazione astratta e semplificata di un sistema reale, per determinare come il sistema risponde a modifiche della sua struttura, senza correre i rischi che si avrebbero se gli esperimenti fossero condotti direttamente sul sistema reale, con l'ulteriore vantaggio che il tempo può essere compresso o espanso permettendo di accelerare o rallentare il fenomeno studiato. Nel contesto della reingegnerizzazione, la simulazione aiuta gli amministratori a verificare se la riprogettazione dei processi è fattibile. Per capire se i nuovi processi sono realizzabili, può essere costruito un modello del nuo-

vo sistema, focalizzando l'attenzione sui suoi punti critici, determinando quali sono i potenziali colli di bottiglia e individuando le risorse necessarie per fornire prodotti o servizi di buona qualità. Una volta che il modello è stato realizzato, si possono condurre analisi approfondite per studiare le prestazioni del prototipo del sistema (per esempio il tempo necessario per servire un cliente), confrontandole con quelle del sistema attuale. In sostanza, la simulazione fornisce un mezzo per valutare diversi scenari derivanti dall'applicazione di nuove strategie gestionali, per fissare i valori delle variabili che li influenzano e per valutare l'impatto di queste variabili sul sistema.

La tesi nasce dall'esigenza della Direzione Sanitaria della Azienda Ospedaliera Pisana, di analizzare la struttura organizzativa dei servizi ambulatoriali con l'obiettivo di migliorare la qualità dei servizi erogati e ottimizzare l'uso delle risorse impiegate.

Il lavoro è iniziato con l'analisi degli ambulatori del reparto di Gastroenterologia. Il primo passo è stato la costruzione del modello concettuale, rappresentato attraverso diagrammi dei cicli di attività che forniscono la rappresentazione grafica del comportamento di tutti i soggetti coinvolti, per esempio medici e utenti. Per fare questo è stata necessaria una raccolta di informazioni, che permettessero di capire il funzionamento dei vari ambulatori, realizzata mediante interviste al personale. Abbiamo inoltre integrato le informazioni in nostro possesso con quelle acquisite tramite l'osservazione diretta, ottenendo in questo modo conferma della validità delle interviste. Tutto questo ha comportato dei costi abbastanza elevati in termini di tempo perché sono stati necessari ripetuti incontri con medici e infermieri del reparto, sia per raccogliere le informazioni di cui avevamo bisogno che per avere la loro validazione sul modello in fase di sviluppo.

Non è stato possibile completare lo studio degli ambulatori di Gastroenterologia perché, durante lo svolgimento del lavoro, si sono presentati problemi organizzativi nel Centro Prelievi e la Direzione Sanitaria ha ritenuto, data l'urgenza, più opportu-

no analizzare quest'ultimo. Nonostante il modello del Reparto di Gastroenterologia sia incompleto, potrebbe essere utilizzato come punto di partenza per un eventuale proseguimento dello studio.

L'obbiettivo del modello di simulazione del Centro Prelievi era quello di fornire alla Direzione Sanitaria uno strumento che permettesse di studiare gli effetti di diverse politiche di gestione del personale, dei macchinari e delle altre infrastrutture, sui tempi di attesa degli utenti. Il tempo che un utente spende in attesa di ricevere l'erogazione di un servizio, rappresenta uno degli indici di valutazione della sua qualità e quindi, bisognerebbe adottare delle politiche che permettano di ottenere livelli accettabili dei tempi di attesa. In pratica, la Direzione aveva la necessità di uno strumento d'aiuto per stabilire le risorse da impiegare, con particolare attenzione al personale, per abbassare i tempi di attesa e aumentare l'indice di gradimento dell'utenza.

Il processo di sviluppo del modello è stato svolto attraverso una sequenza di passi con un ordinamento logico non sempre sequenziale perché, in diversi casi, si è presentata la necessità di dover rianalizzare il lavoro svolto nelle fasi precedenti per il fatto che, man mano che l'analisi procedeva, aumentavano le informazioni in nostro possesso e questo ci consentiva di incrementare l'accuratezza del modello.

Il primo passo, come per gli ambulatori di Gastroenterologia, è stato quello della raccolta delle informazioni che ci permettessero di formulare il problema e di sviluppare il modello concettuale preliminare, che si è gradualmente raffinato anche attraverso i ripetuti confronti con gli esperti del Centro Prelievi. Abbiamo cioè cercato di inferire quali fossero le caratteristiche essenziali del problema, i parametri che caratterizzavano il sistema e abbiamo quindi elaborato il modello concettuale al quale sono state apportate modifiche successive, fino al raggiungimento di una buona approssimazione. Si noti che, in questa fase, non è necessario rappresentare il sistema in modo speculare, ma piuttosto bisogna cercare di coglierne l'essenza.

Dopo aver stabilito quali fossero i dati di interesse, è iniziata la loro raccolta in collaborazione con il personale della direzione sanitaria dell'ospedale Santa Chiara di Pisa, il personale del centro prelievi e un gruppo di allieve infermiere professionali. Tutti i dati raccolti sono stati analizzati per determinare quali fossero le variabili stocastiche e deterministiche di ingresso del sistema. Per ogni variabile aleatoria di ingresso, è stata scelta una distribuzione di probabilità che la potesse rappresentare e sono stati stimati i suoi parametri.

Naturalmente, le ipotesi fatte sulle leggi di probabilità dovevano essere validate mediante analisi statistiche. La fase di analisi della bontà delle assunzioni fatte sulle distribuzioni, è particolarmente delicata perché per ogni prodotto software, ma per i modelli di simulazione in particolare, vale la cosiddetta “regola GIGO” (*Garbage-In-Garbage-Out*). In altre parole, anche se la struttura di un modello è valida, se gli si danno in ingresso dati raccolti o analizzati in modo non appropriato, le risposte del modello di simulazione molto probabilmente non saranno corrette e il loro uso potrebbe causare dei danni. Per questo motivo abbiamo ritenuto opportuno curare in modo particolare l'analisi dei dati, dedicandole una cospicua quantità di tempo.

Sono state eseguite innanzitutto delle procedure euristiche, cioè istogrammi e grafici Quantile-Quantile, per determinare le famiglie di distribuzioni che potessero rappresentare i dati. Siccome entrambe le procedure sfruttano delle rappresentazioni grafiche, le distribuzioni scelte mediante confronti visivi, perciò in modo soggettivo, avrebbero potuto non rappresentare i dati adeguatamente. È stato perciò conveniente fare delle analisi statistiche oggettive, per avere una risposta analitica sulla bontà delle scelte fatte, utilizzando i test di *Goodness-of-fit* di Kolmogorov-Smirnov e χ^2 . Tutto il lavoro per definire il modello dei dati in ingresso è stato svolto utilizzando il foglio elettronico Microsoft Excel.

Per l'implementazione del modello è stato scelto Extend, una applicazione *general-*

purpose, grafica, orientata alla simulazione discreta e continua con un ambiente di sviluppo integrato, disponibile per i sistemi operativi Microsoft Windows e Macintosh. La scelta è ricaduta su questo pacchetto perché, pur avendo un costo abbastanza contenuto, possedeva delle buone caratteristiche tra le quali la flessibilità di modellazione, la possibilità di visualizzare animazioni del modello, capacità statistiche, il controllo delle esecuzioni del modello, la generazione di tracce relative alle esecuzioni e la presentazione degli output. Anche se con il pacchetto sono fornite diverse librerie di moduli predefiniti, in diverse occasioni si è presentata la necessità di dover costruire dei nuovi blocchi o personalizzare blocchi esistenti, usando il linguaggio ModL dell'ambiente di sviluppo di Extend.

La verifica, che consiste nel controllare che il modello si comporti nel modo voluto, è stata fatta con tecniche che includevano l'analisi delle tracce, l'osservazione delle animazioni e il controllo dei valori delle variabili di uscita al variare dei valori di quelle di ingresso.

La fase successiva è stata quella di validazione e calibrazione. Si dovevano confrontare le variabili di uscita del modello di simulazione con quelle del sistema reale, per vedere se il modello replicava abbastanza fedelmente il comportamento del sistema. Si osserva che il processo di validazione inizia già in fase di costruzione del modello perché si deve cercare di svilupparlo con un alta *face validity*, cioè bisogna fare in modo che esteriormente il modello appaia ragionevole alle persone che hanno familiarità con il sistema. Uno dei passi della validazione è quello della validazione delle trasformazioni ingresso-uscita, cioè il modello è visto come una "scatola nera" che prende dei valori in ingresso e restituisce dei risultati che devono essere confrontati con quelli del sistema e dai quali si dovrebbero discostare il meno possibile. Le trasformazioni ingresso-uscita sono state realizzate in due modi: con dati di ingresso generati artificialmente in base alle leggi di probabilità adottate e con i dati di ingresso storici ottenuti dall'osservazio-

ne del sistema reale. In entrambi i casi sono state valutate le risposte con l'applicazione del test *t*-Student.

Dopo la validazione sono state fatte alcune prove con il modello di simulazione del sistema attuale, modificando alcuni parametri d'ingresso. Esperimenti più approfonditi dovrebbero essere pianificati con i destinatari del modello.

Organizzazione della Tesi

La Tesi è organizzata nel modo seguente:

- Nella Sezione *Alcuni concetti di simulazione*, si fa una breve rassegna sui sistemi e sul processo di sviluppo di progetti di simulazione.

Il seguito della dissertazione è diviso in due parti:

Parte I : Centro Prelievi

- Nel Capitolo 1 si studia il funzionamento del Centro Prelievi, definendo le entità che caratterizzano il sistema e le loro interazioni e si definisce il modello concettuale.
- Nel Capitolo 2 si descrive la fase della raccolta dei dati di interesse e si fa una descrizione accurata sui metodi di analisi utilizzati per validare il modello dei dati di ingresso. Viene illustrato, attraverso un esempio, il procedimento usato per la costruzione del modello dei dati di ingresso e si forniscono i risultati.
- Il Capitolo 3 è dedicato all'implementazione del modello di simulazione. Si descrivono, innanzitutto, le principali caratteristiche che i linguaggi di simulazione dovrebbero avere e si fa una presentazione dello strumento di simulazione Extend, usato per l'implementazione del modello del Centro Prelievi.

- Nel Capitolo 4 si discutono i metodi per la verifica, la calibrazione e la validazione di un modello di simulazione, dedicando una considerevole parte ai metodi statistici, agli errori ed alla potenza dei test. Vengono infine presentati i risultati.

Parte II : Reparto di Gastroenterologia

Questa parte è composta dal Capitolo 5 che descrive il modello concettuale degli ambulatori del reparto di Gastroenterologia studiati.

Alcuni concetti di simulazione

Il termine *simulazione* si riferisce alla realizzazione della rappresentazione di qualche sistema di qualsiasi grandezza e complessità. In ogni caso, l'idea chiave della simulazione è la realizzazione di un modello che approssima il sistema, con lo scopo di analizzare e capire il suo comportamento, sperimentando decisioni o azioni alternative.

Un *sistema* è un insieme di entità o componenti che interagiscono per raggiungere qualche obiettivo. I sistemi che si incontrano comunemente possono essere ospedali, check-in degli aeroporti, sistemi di telecomunicazioni come le reti telefoniche, autostrade e così via. Questi sistemi sono grandi e complessi ed è difficile e costoso fare degli esperimenti direttamente sul campo.

Diamo ora alcune definizioni formali dei componenti che caratterizzano un sistema:

Definizione 1 (entità) *Le entità sono gli elementi del sistema che deve essere simulato e possono essere identificate singolarmente e gestite dal programma di simulazione.*

Il *sistema* può così essere visto come un insieme di entità in relazione tra loro, che cooperano per produrre dei risultati. In un sistema governato, questa cooperazione è controllata e coordinata [Pid84]. Le interazioni tra le entità definiscono il comportamento del sistema. Le entità che rimangono nel sistema per tutta la durata della simulazione, sono dette *permanenti*. Quelle che passano nel sistema e che, dopo averlo abbandonato, non sono più interessanti per la simulazione, sono definite entità *tempora-*

nee. Un'altra classificazione delle entità può essere fatta in base al loro comportamento e permette di definire le entità come *attive* o *passive*. Generalmente, i server sono entità attive ed i clienti entità passive.

Definizione 2 (classi) *Le entità sono identificabili singolarmente, ma possono anche essere raggruppate in classi di entità omogenee.*

Definizione 3 (attributi) *Ogni entità può avere uno o più attributi che forniscono informazioni aggiuntive riguardanti le entità stesse.*

Gli attributi sono utili per diversi scopi. Ad esempio, potrebbero essere usati per controllare il comportamento delle entità, oppure per controllare le discipline di eventuali code.

Definizione 4 (insiemi) *Anche se le entità sono organizzate permanentemente in classi, durante la simulazione possono cambiare stato¹ e questi stati possono essere rappresentati come insiemi.*

Gli insiemi possono essere usati come indicatori dello stato del sistema in un certo istante di tempo. In certi casi si possono pensare gli insiemi come code nelle quali le entità aspettano che succeda qualcosa. Queste code sono gestite con specificate discipline, la più comune delle quali è la FIFO².

Definizione 5 (evento) *È un istante di tempo in cui si verifica un cambiamento di stato significativo per il sistema.*

Il fatto che un cambiamento di stato sia significativo, dipende dal contesto e dagli obiettivi della simulazione.

¹Lo stato di una entità è dato dai valori dei suoi attributi

²First In First Out, il primo arrivato è il primo ad essere servito.

Definizione 6 (attività) *Le entità si muovono da insiemi ad insiemi a causa delle operazioni in cui sono coinvolte. Le operazioni e le procedure che iniziano al verificarsi di un evento sono conosciute come attività.*

Un *modello* è una rappresentazione astratta e semplificata di un sistema e delle interazioni tra i suoi componenti fondamentali. Un modello *stocastico* è un modello il cui comportamento non può essere previsto con certezza, ma è soggetto alla casualità. La maggior parte dei modelli che sono analizzati con la simulazione, sono stocastici e quindi la loro descrizione include la definizione di distribuzione di probabilità per i valori delle variabili aleatorie. Per esempio, in un sistema di comunicazioni, il numero di pacchetti da trasmettere e gli istanti di tempo in cui sono trasmessi sono variabili aleatorie.

I modelli di simulazione sono usati per analizzare una grande varietà di sistemi. La simulazione è una metodologia che non è dedicata ad una specifica area, ma può essere applicata a ogni sistema attraverso metodi di modellazione.

Passi di uno studio di simulazione

Un progetto di simulazione, con l'obiettivo di analizzare un sistema, può essere diviso nei seguenti passi logici:

1. **Definizione del problema.** Si definiscono in modo chiaro gli obiettivi dello studio, in modo da sapere quale scopo bisogna raggiungere, assicurandosi di avere a disposizione le risorse necessarie (strumenti software e persone) per svolgere il lavoro. Bisogna anche determinare quali sono i confini del sistema da rappresentare e il suo funzionamento.

2. **Concettualizzazione del modello.** Si deve sviluppare un modello grafico, oppure con pseudocodice, per definire le componenti, le variabili, e le interazioni logiche che costituiscono il sistema.
3. **Definizione del modello dei dati di ingresso.** Bisogna identificare e raccogliere i dati di ingresso necessari. I dati raccolti devono essere analizzati per determinare la loro natura (stocastica o deterministica) ed eventualmente scegliere, dopo aver testato la validità delle decisioni prese, le distribuzioni di probabilità che li rappresentino.
4. **Implementazione.** Si implementa il modello con un appropriato linguaggio di simulazione.
5. **Verifica.** Riguarda la costruzione del modello corretto e si esegue confrontando il modello concettuale e il modello di simulazione che rappresenta questi concetti. Si deve perciò controllare se il programma è implementato correttamente e se i parametri di ingresso e la struttura logica del modello sono rappresentati fedelmente.
6. **Calibrazione e validazione.** Anche la *validazione* riguarda la costruzione di un modello corretto e serve per determinare se il modello è una accurata rappresentazione del sistema reale. La validazione è di solito usata per realizzare, attraverso la *calibrazione* del modello, un processo iterativo di confronto tra il comportamento del modello e quello del sistema attuale. Eventuali discrepanze vengono usate per migliorare il modello stesso.
7. **Progetto degli esperimenti.** Devono essere determinate le alternative da simulare. Spesso le decisioni riguardanti le alternative da simulare possono essere

in funzione delle esecuzioni delle simulazioni precedenti già completate e analizzate. Quindi, si confrontano i diversi risultati per vedere come conviene configurare il sistema sotto studio.

8. **Documentazione.** Se il programma deve essere usato da persone diverse dai suoi sviluppatori è necessario che capiscano come il programma lavora. È quindi utile allegare una esauriente documentazione sul suo funzionamento.

Parte I

Centro Prelievi

Capitolo 1

Concettualizzazione del modello

Il primo passo di uno studio di simulazione è l'analisi del sistema nel mondo reale e la definizione dell'ambiente del modello in base ai suoi obiettivi. Attraverso l'osservazione diretta e il coinvolgimento delle persone che hanno familiarità con il sistema si individuano le componenti principali che lo caratterizzano e i suoi confini. È importante sottolineare che il livello di dettaglio dell'analisi del sistema dipende dallo scopo che si vuole raggiungere e quindi si deve cercare di capire quali sono gli elementi di interesse e quali particolari tralasciare. Perciò, dopo aver raccolto più informazioni possibili, si definisce una rappresentazione grafica del sistema che serve come base per lo sviluppo del modello di simulazione. L'importanza della rappresentazione grafica è data dal fatto che illustra le caratteristiche fondamentali del modello di simulazione come imitazione della struttura e delle operazioni del sistema reale. Inoltre, è uno straordinario strumento di comunicazione tra gli sviluppatori del modello e suoi utenti che dovrebbero partecipare alla realizzazione.

Nel seguito del capitolo, si analizzerà la struttura del Centro Prelievi cercando di capire quali sono le sue componenti fondamentali, il comportamento delle entità individuate e le loro interazioni.

Dopo aver introdotto i diagrammi dei cicli di attività, si fornirà la rappresentazione grafica del modello concettuale del Centro Prelievi.

1.1 Analisi del sistema

Il Centro Prelievi è aperto tutti i giorni dalle ore 7:15 alle ore 9:30 esclusi domenica e festivi.

Nel sistema sono individuabili sei classi di **entità** : pazienti, distributori numeri, addetti accettazione, macchine automatiche riscossione ticket, servizi igienici e la classe degli infermieri addetti al prelievo.

Le attività svolte nel centro sono di due tipi : i prelievi e l'accettazione di campioni. Entrambe le attività citate devono essere precedute da una operazione di accettazione che viene svolta in appositi sportelli.

I pazienti che accedono al centro prelievi possono essere divisi in cinque categorie che sono state stabilite per esigenze di ordinamento delle file di attesa agli sportelli d'accettazione, ai box prelievi e per distinguere i pazienti in base ai diversi tipi di servizi richiesti. La suddivisione è la seguente :

- *urgenti:*

Questi pazienti si possono ulteriormente classificare in : trapiantati, pazienti sottoposti a terapia di coagulazione, pazienti affetti da patologie oncologiche, assunzioni S.Chiara e medicina preventiva.

- *preospedalizzazioni:*

Sono quei pazienti che devono fare gli esami necessari per il ricovero.

- *solo consegna campioni:*

Sono coloro che devono solo consegnare dei campioni di materiale organico su cui devono essere fatte delle analisi.

- *normali*

- *impegnative multiple:*

È una categoria di pazienti che ha lo stesso comportamento dei pazienti normali con la differenza che presentano un numero di impegnative maggiore di uno.

La Figura 1.2 a pag. 23 descrive il ciclo delle attività dei pazienti.

Quando un paziente arriva al Centro, per prima cosa, deve ritirare il numero d'ingresso che gli viene consegnato da una infermiera ed attende il suo turno nella sala d'aspetto. La fila d'attesa è gestita con una politica a priorità. Le priorità vengono date ai pazienti secondo la seguente tabella nella quale al valore più basso corrisponde la priorità più alta:

Tipo paziente	Priorità
Urgenti	0
Preospedalizzazioni	1
Consegna campioni	1
Normali	2
Impegnative multiple	2

Tabella 1.1: Priorità all'accettazione

La priorità più alta è assegnata alla categoria degli urgenti che quindi passano in testa alla coda. A seguire, in ordine di priorità, vengono le categorie delle preospedalizzazioni e consegna campioni, che hanno uguale priorità, ed infine i pazienti normali e i pazienti

con impegnative multiple. A parità di priorità viene seguita una politica FIFO. Questa disciplina è stata decisa per abbassare il tempo di attesa sia dei pazienti che hanno particolari problemi di salute (urgenti) che di quelli per i quali i tempi di servizio sono particolarmente bassi (preospedalizzazioni all'accettazione e consegna campioni ai box).

Nel momento in cui arriva il suo turno, il paziente viene chiamato dall'infermiera addetta a regolare l'afflusso delle persone agli sportelli d'accettazione. Le file agli sportelli, che sono quattro, sono gestite in modo da avere in coda non più di cinque persone oltre a quella che è servita e la loro gestione segue una politica FIFO. Il paziente chiamato viene indirizzato allo sportello che ha la fila più corta e lì aspetta il suo turno. Una volta arrivato allo sportello, l'infermiera addetta esegue le operazioni di accettazione e stampa le etichette da applicare alle provette che conterranno i campioni da analizzare. Finita questa fase i pazienti normali, quelli che hanno impegnative multiple e quelli che devono solo consegnare i campioni sono ulteriormente suddivisi in base al fatto che debbano pagare il ticket oppure no. Quelli che devono pagarlo, prima di andare ai box prelievi, si mettono in coda alle macchinette automatiche di riscossione, mentre gli altri utenti vanno direttamente a fare la fila per l'accesso ai box. Le macchinette sono due, mentre la fila è unica e gestita con politica FIFO. Dopo aver pagato il ticket e aver preso la ricevuta di versamento, i pazienti paganti raggiungono le code per i prelievi. I box sono otto e ciascuno ha una propria fila di attesa.

Nel momento in cui un paziente arriva nel corridoio, sul quale si affacciano le varie porte di accesso ai box, controlla la lunghezza della coda di ogni box e normalmente sceglie la fila più corta. Anche in questo caso le code sono gestite con una politica a priorità per cui ciascun tipo di paziente ha una certa priorità per l'accesso ai box come si può vedere nella tabella delle priorità ai box:

Tipo paziente	Priorità
Consegna campioni	0
Urgenti	1
Preospedalizzazioni	2
Normali	2
Impegnative multiple	2

Tabella 1.2: Priorità ai box

Osservando la tabella risulta che i pazienti che passano per primi sono quelli che devono consegnare i campioni in quanto l'operazione di consegna può essere svolta in pochissimo tempo anche nel caso in cui il box sia già occupato da un'altra persona.

Dopo il prelievo del sangue i pazienti che devono consegnare anche i campioni, senza averli portati già pronti da casa, hanno la necessità di dover usufruire dei servizi igienici. Tali pazienti al loro ritorno passano in testa alla fila e consegnano le provette.

1.2 Modello concettuale

I *Diagrammi dei cicli di attività* sono un modo di rappresentare le interazioni tra le entità e sono particolarmente utili per sistemi in cui sono presenti delle code. In molti casi questi diagrammi non sono in grado di rappresentare appieno la complessità del sistema che deve essere simulato, ma forniscono uno "scheletro" che può essere sviluppato in seguito. I diagrammi rappresentano la storia della vita di ogni classe di entità e mostrano graficamente le loro interazioni. Ogni classe di entità ha un ciclo di vita che consiste in una serie alternata di stati attivi e stati passivi.

Generalmente, in uno *stato attivo* sono coinvolte entità cooperanti che appartengono

a classi diverse. La durata di uno stato attivo può sempre essere determinata in anticipo attraverso l'esame dei campioni di una appropriata distribuzione di probabilità nel caso in cui il modello di simulazione sia stocastico¹.

Diversamente, uno *stato passivo* coinvolge entità non cooperanti appartenenti a classi differenti ed è di solito lo stato in cui le entità aspettano il verificarsi di un certo tipo di evento. Gli stati passivi possono essere visti come insiemi o code ed il tempo di permanenza delle entità, in tali stati, non può essere stabilito in anticipo in quanto dipende dal comportamento di altre entità.

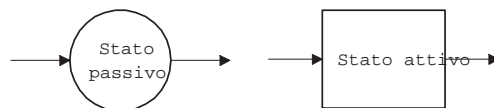


Figura 1.1: Simboli per i diagrammi dei cicli di attività

A questo punto possiamo schematizzare la descrizione del funzionamento del Centro Prelievi fatta in precedenza.

1.2.1 Entità

- Pazienti:
 - Urgenti;
 - Preospedalizzazioni;
 - Solo consegna campioni:
 - con pagamento ticket;
 - senza pagamento ticket;

¹Non si può prevedere in anticipo quale sarà l'esatto comportamento del sistema

- Normali:
 - con pagamento ticket;
 - senza pagamento ticket;
- Impegnative multiple:
 - con pagamento ticket;
 - senza pagamento ticket;
- Infermiere:
 - Addette accettazione;
 - Addette box;
- Macchine automatiche riscossione ticket;
- Servizi igienici;
- Distributore numeri.

1.2.2 Eventi

- Arrivo paziente;
- Ritiro numero;
- Inizio accettazione;
- Fine accettazione;
- Inizio pagamento ticket;
- Fine pagamento ticket;

- Inizio prelievo;
- Fine prelievo.

1.2.3 Attività

1. Arrivo paziente:

Il paziente arriva e fa la fila per ritirare il numero d'ingresso.

2. Ritiro numero:

Il paziente ritira il numero e si reca in sala d'attesa.

3. Accettazione:

Quando è il suo turno il paziente si reca allo sportello, vengono svolte le pratiche d'accettazione e, se deve pagare il ticket, va a fare la fila alle macchine automatiche di riscossione, altrimenti va a fare la fila ai box.

4. Pagamento ticket:

I pazienti che devono pagare il ticket lo fanno alle macchinette automatiche e dopo vanno a fare la fila ai box prelievi.

5. Prelievo / Consegna campioni:

Vengono svolte le attività di prelievo e/o consegna campioni. Nel caso in cui, dopo aver subito il prelievo, il paziente debba usufruire dei servizi igienici vi si reca, altrimenti abbandona il Centro Prelievi.

6. Servizi igienici.

1.2.4 Variabili in input

- Processo di arrivo;

-
- Probabilità di arrivo dei diversi tipi di pazienti;
 - Tempi di interservizio accettazione;
 - Tempi di servizio accettazione:
 - normali, consegna campioni, urgenti;
 - preospedalizzazioni;
 - multiple;
 - Tempo di servizio pagamento ticket;
 - Tempi di interservizio box;
 - Tempi di servizio box:
 - prelievi;
 - consegna campioni;
 - Numero sportelli accettazione aperti;
 - Numero box aperti;
 - Numero macchinette attive.

1.2.5 Diagrammi dei cicli di attività

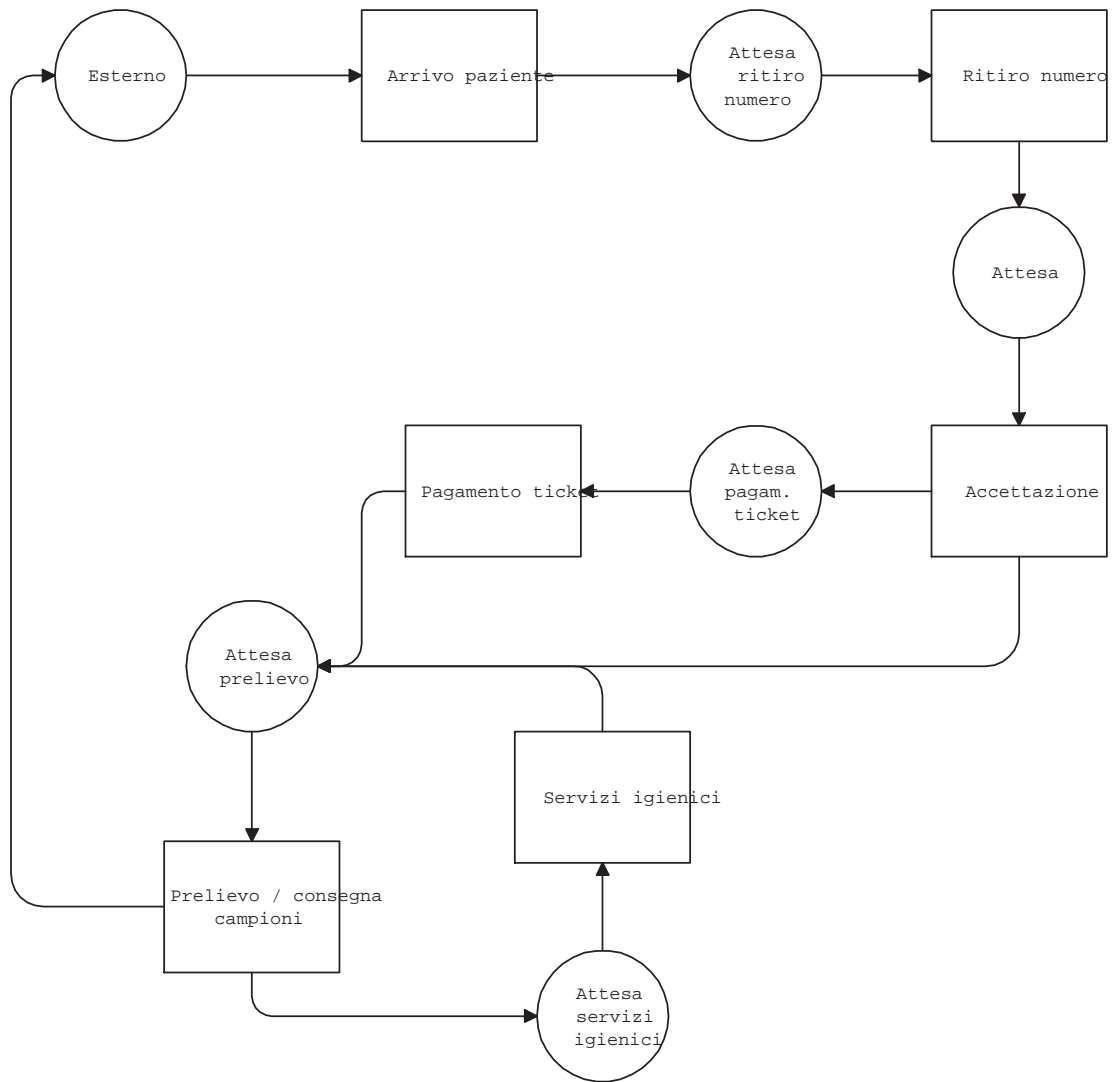


Figura 1.2: Diagramma dei cicli di attività Pazienti

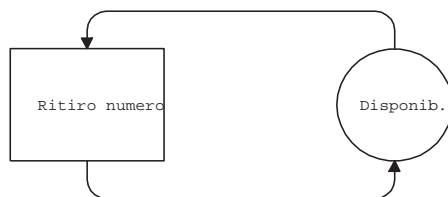


Figura 1.3: Diagramma dei cicli di attività Distributore Numeri

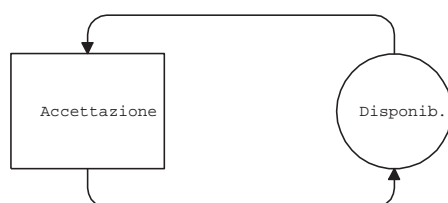


Figura 1.4: Diagramma dei cicli di attività Sportello Accettazione

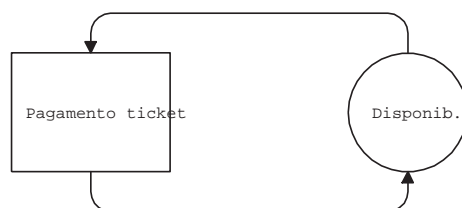


Figura 1.5: Diagramma dei cicli di attività Macchina Automatica Pagamento Ticket

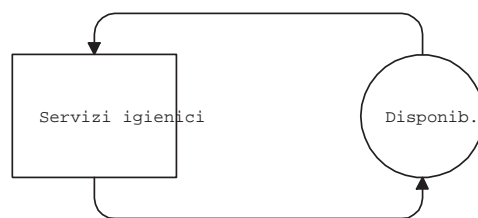


Figura 1.6: Diagramma dei cicli di attività Servizi Igienici



Figura 1.7: Diagramma dei cicli di attività Box Prelievi

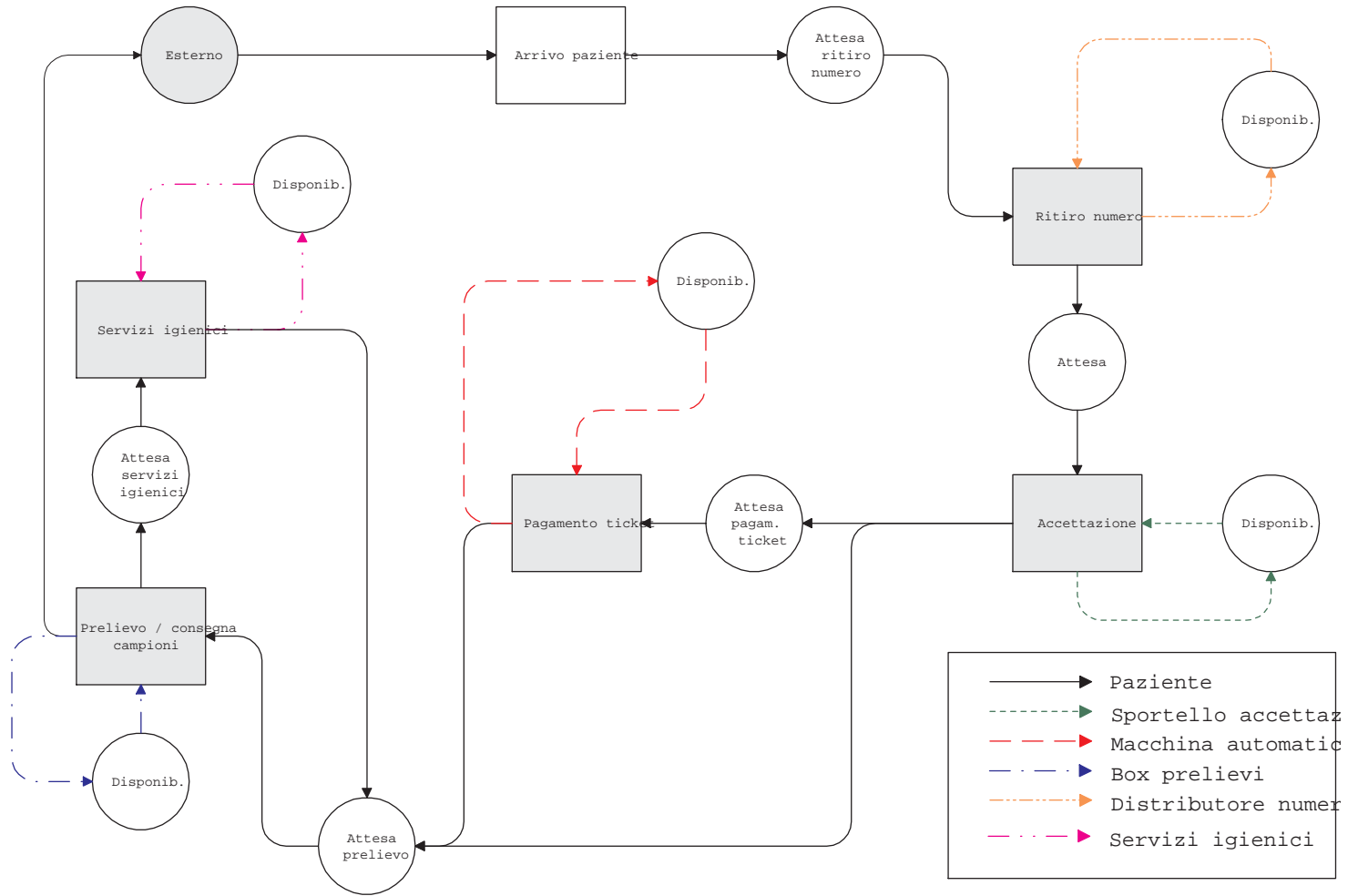


Figura 1.8: Diagramma dei cicli di attività Centro Prelievi

Capitolo 2

Modello dei dati in input

I dati in input forniscono la forza motrice per il modello di simulazione. Per esempio, in un sistema di code gli input tipici sono : i tempi di interarrivo e i tempi di servizio.

Nell'applicazione della simulazione nel mondo reale, tuttavia, il problema più importante, sia dal punto di vista del tempo richiesto che delle risorse impiegate, è determinare le appropriate distribuzioni per i dati in input. Inoltre, bisogna tenere conto del fatto che un modello inappropriato dei dati in input porterà ad output degli esperimenti di simulazione inadeguati.

Nello sviluppo del modello dei dati in input abbiamo seguito 4 passi fondamentali:

1. *Raccolta dei dati di interesse dal sistema reale:* questo richiede un tempo sostanziale e risorse non indifferenti, da parte del committente, in termini di tempo e persone impiegate per la sua realizzazione.
2. *Identificazione delle distribuzioni di probabilità per rappresentare i dati in input:* quando i dati sono disponibili questo passo tipicamente inizia con l'analisi delle distribuzioni delle frequenze e/o costruendo degli istogrammi sui dati in questione. Sulla base delle distribuzioni delle frequenze e delle informazioni sulla

struttura del processo di input vengono ipotizzate le distribuzioni di probabilità per i dati. In molti casi pratici alcune distribuzioni teoriche forniscono una buona approssimazione per i dati; in caso contrario si può scegliere una distribuzione empirica.

3. *Scelta dei parametri per determinare le specifiche istanze delle famiglie di distribuzione:* se esiste la disponibilità dei dati tale stima si può ottenere dai dati stessi.
4. *Valutazione della bontà delle scelte fatte sulle distribuzioni e dei parametri associati:* questa analisi viene fatta informalmente attraverso metodi grafici e formalmente con l'utilizzo di test statistici. Se la scelta fatta non è soddisfacente si ritorna al passo 2 e si ipotizza una famiglia di distribuzioni diversa ripetendo la procedura. Se diverse iterazioni di questa procedura falliscono, allora si potrebbe optare per la forma di distribuzione empirica.

Nelle sezioni che seguono descriveremo il lavoro svolto per costruire il modello dei dati. Nel prossimo paragrafo si illustra la fase di pianificazione e di raccolta dei dati. Si noti che questa fase non è totalmente disgiunta dalle altre perché da una parte può essere considerata come un completamento del modello concettuale mentre dall'altra incorpora una analisi preliminare, che sarà successivamente ampliata e completata, delle leggi di distribuzione che regolano le variabili di ingresso stocastiche. Nelle Sezioni 2.2, 2.3 e 2.4 si fa una panoramica della teoria relativa ai metodi ed ai test statistici usati per la costruzione e la validazione del modello dei dati in input. Per ogni variabile stocastica di ingresso è stato fatto uno studio completo ripetendo i passi 2, 3 e 4 precedentemente accennati. Attraverso l'esempio dello studio di una variabile di ingresso (Sezione 2.5), più precisamente la variabile relativa ai tempi di interarrivo del lunedì, illustreremo in dettaglio lo studio che è stato fatto per ciascuna variabile. Si noti che,

soprattutto nella fase iniziale di definizione dei modelli, molte scelte sono soggettive e molto è lasciato all'intuito ed alla esperienza dell'analista in quanto non esistono dei procedimenti rigorosi e ben delineati da seguire. Questo porta, nella fase di verifica delle scelte fatte, a continui cambiamenti delle ipotesi sulle leggi che regolano il modello sia in termini di famiglie di distribuzioni che di parametri. Considerata l'importanza di costruire un buon modello dei dati, la parte di lavoro che è stata svolta per ottenere i risultati, riportati nella parte finale del Capitolo, è stata molto lunga e laboriosa.

2.1 Raccolta dei dati

Questa parte rappresenta un aspetto molto importante per la soluzione del problema reale ed è uno dei problemi di difficile soluzione nel campo della simulazione.

Anche quando la struttura del modello di simulazione è valida, se i dati sono raccolti senza cura, analizzati in maniera non appropriata o non sono abbastanza rappresentativi per il sistema che si sta studiando, l'output degli esperimenti di simulazione risulterà non veritiero e se usato per prendere delle decisioni importanti, oppure per decidere sulle politiche di gestione in una azienda, potrà arrecare dei danni e far aumentare i costi.

Nel nostro caso abbiamo eseguito la raccolta dei dati in collaborazione con il personale della Direzione Sanitaria dell'Ospedale Santa Chiara di Pisa, il personale del centro prelievi e un gruppo di allieve infermiere professionali.

I dati sono stati raccolti per un periodo di due settimane consecutive e sono stati utilizzati per la scelta delle distribuzioni di probabilità per il modello dei dati in input. A distanza di un mese si è provveduto ad un'ulteriore raccolta e questi campioni di dati sono stati utilizzati per la convalida del modello di simulazione.

Tuttavia, come normalmente accade, questa fase ha evidenziato alcuni problemi: i

tempi di interarrivo dei pazienti, pur risultando omogenei, seguivano un legge di distribuzione che variava da periodo a periodo nell'arco della stessa giornata e da giorno a giorno nell'arco della settimana. I tempi di servizio degli sportelli dell'accettazione sono stati classificati in base al numero degli esami richiesti ed alla tipologia del paziente. In più, è stato necessario capire quale fosse la proporzione di pazienti che pagavano il ticket perché ciò comportava un'ulteriore operazione di pagamento alle macchinette automatiche e conseguentemente abbiamo dovuto stimare la distribuzione del tempo di servizio per questa operazione. Sono state raccolte ulteriori informazioni riguardo all'urgenza o no dei pazienti perché le urgenze avevano priorità agli sportelli dell'accettazione e ai box prelievi. Inoltre, è stato necessario stimare la proporzione di pazienti che facevano esami per il ricovero in ospedale perché questi pazienti avevano priorità all'accettazione.

Nel nostro lavoro abbiamo cercato di seguire i suggerimenti di [BCN96] e i passi fondamentali sono stati i seguenti:

- Parte del tempo è stato speso preliminarmente per analizzare il problema e conseguentemente pianificare le operazioni di raccolta dei dati. Ciò è stato fatto con la direzione sanitaria, con il personale del centro prelievi e con le allieve infermiere. Partendo da una bozza primaria di piano di raccolta, dopo diversi incontri, si è arrivati alla stesura del piano finale dal quale è venuta fuori la scheda di rilevamento dei tempi a pag. 32. Il rilevamento veniva fatto nel seguente modo: nel momento in cui un utente accedeva ai locali del Centro Prelievi gli veniva consegnata una scheda, che doveva portare con sé per tutto il tempo di permanenza, nella quale il personale addetto annotava i tempi e le altre informazioni di interesse.
- Mentre la raccolta dei tempi procedeva abbiamo fatto le prime analisi sui dati

raccolti cercando di individuare le distribuzioni che potevano rappresentarli. Inoltre, abbiamo eseguito un'ulteriore analisi per verificare se i dati raccolti fossero utili e sufficienti per il modello di simulazione.

- Si è cercato di capire se era possibile combinare insieme di dati tra loro, ossia abbiamo controllato il grado di omogeneità dei diversi campioni di dati. L'analisi sull'omogeneità è stata fatta relativamente a periodi di tempo successivi nell'arco della stessa giornata e durante lo stesso periodo in giorni diversi (per i tempi di interarrivo). L'analisi è stata fatta inizialmente confrontando le medie dei campioni e dopo è stata eseguita un'analisi più accurata utilizzando i Diagrammi Q-Q e alcuni test statistici (Kruskal-Wallis) per determinare l'equivalenza delle distribuzioni.

CENTRO PRELIEVI**S. CHIARA**

Data _____ n° ingresso _____ ore _____ min _____ sec _____

ACCETTAZIONE

Sportello n° 1 2 3 4

Operazioni sportello: INIZIO ore _____ min _____ sec _____ FINE ore _____ min _____ sec _____

TIPO DI IMPEGNATIVA urgente (*trapiantati, pazienti oncologici,
pazienti che consegnano solo campioni*)
§ preospedalizzazione multipla standardPagamento ticket: Si NoCASSA

Numero casse 5 6

Operazioni di pagamento: INIZIO ore _____ min _____ sec _____ FINE ore _____ min _____ sec _____

PRELIEVO

Numero box 1 2 3 4 5 6 7 8

Solo consegna campioni Priorità al box prelievo Si No

Operazione di prelievo: INIZIO ore _____ min _____ sec _____

Fine operazioni di prelievo ore _____ min _____ sec _____

Campione urina INIZIO ore _____ min _____ sec _____ FINE ore _____ min _____ sec _____

2.2 Identificazione delle distribuzioni

Dopo la raccolta dei dati siamo passati alla loro analisi per stabilire se appartenessero a qualche famiglia di distribuzione nota. Questo lavoro è stato fatto inizialmente con dei metodi grafici e più precisamente con l'utilizzo degli istogrammi e dei Diagrammi Q-Q.

Ipotesi fondamentale per tutto ciò che verrà trattato nel seguito di questo capitolo è che i dati osservati siano *indipendenti e identicamente distribuiti* (IID). In caso contrario molti dei metodi illustrati non sarebbero applicabili.

2.2.1 Istogrammi

La distribuzione delle frequenze, o istogramma, è utile per identificare la forma della curva della densità della distribuzione dei dati osservati X_1, X_2, \dots, X_n IID.

Un istogramma [LK91] è costruito nel modo seguente:

1. Si divide l'intervallo in cui cadono i valori dei dati rilevati in k sottointervalli adiacenti disgiunti $[b_0, b_1), [b_1, b_2), \dots, [b_{k-1}, b_k)$ che dovrebbero avere tutti la stessa ampiezza $\Delta b = b_j - b_{j-1}$.
2. I valori delle ascisse del grafico sono quelli degli estremi dei sottointervalli b_0, b_1, \dots, b_k .
3. Si determinano le frequenze per ogni intervallino e per $j = 1, 2, \dots, k$ si definisce q_j come la proporzione degli X_i che cadono nell'intervallo $[b_{j-1}, b_j)$.
4. A questo punto si determina la funzione:

$$h(x) = \begin{cases} 0 & \text{se } x < b_0 \\ q_j & \text{se } b_{j-1} \leq x < b_j \quad \text{per } j = 1, 2, \dots, k \\ 0 & \text{se } b_k \leq x \end{cases} \quad (2.1)$$

5. Si traccia il grafico della funzione $h(x)$.

Il numero dei sottointervalli dipende dalla dimensione del campione osservato e in [HM90] viene suggerito un numero approssimativamente pari alla radice quadrata della dimensione del campione. Se i sottointervalli fossero troppo grandi l'istogramma sarebbe grossolano e la sua forma non riuscirebbe a far capire a quale distribuzione appartengono i dati. Nel caso in cui i sottointervalli fossero troppo piccoli l'istogramma sarebbe irregolare e non rappresenterebbe bene i dati.

Il grafico di h , che è costante a tratti, è quindi confrontato con i grafici delle densità di varie distribuzioni, sulla base della forma, ignorando i fattori di scala e di posizione. Se vogliamo vedere il motivo della somiglianza della forma di h con quella della vera densità f dei dati definiamo X variabile aleatoria con densità f , cosicché X è distribuita identicamente alle X_i . Allora, per ogni j fissato ($j = 1, 2, \dots, n$) si ha

$$P\{b_{j-1} \leq X \leq b_j\} = \int_{b_{j-1}}^{b_j} f(x) dx = \Delta b f(y) \quad (2.2)$$

per qualche particolare numero $y \in (b_{j-1}, b_j)$. La prima uguaglianza della (2.2) deriva dalla definizione di variabile aleatoria continua e la seconda dal teorema del valore medio. Quindi, la probabilità di X nel j -esimo intervallo è approssimata da q_j che corrisponde al valore $h(y)$. Perciò,

$$h(y) = q_j \approx \Delta b f(y) \quad (2.3)$$

cosicché $h(y)$ è approssimativamente proporzionale a $f(y)$ da cui deriva che h ed f hanno all'incirca la stessa forma.

2.2.2 Scelta della famiglia di distribuzioni

Lo scopo per cui si costruisce un istogramma è quello di cercare di ricavare una funzione densità di probabilità nota.

La scelta della famiglia di distribuzioni, per rappresentare i dati campionati, viene fatta tenendo conto del contesto e della forma degli istogrammi. Per esempio, normalmente i tempi di interarrivo seguono un legge di probabilità esponenziale, i tempi di servizio seguono delle leggi gamma e così via.

2.2.3 Diagrammi Quantile-Quantile

L'utilizzo degli istogrammi è stato il punto di partenza per la selezione della famiglia di distribuzioni più rappresentativa per i campioni di dati sotto analisi. Tuttavia, gli istogrammi presentano dei limiti riguardo alla bontà della scelta fatta. Quando il campione dei dati è poco numeroso il grafico risulta, normalmente, abbastanza frastagliato. Inoltre, è molto importante la scelta dell'ampiezza dei sottointervalli ed anche quando tale scelta è stata fatta accuratamente, in alcuni casi, risulta difficile capire quale distribuzione di probabilità possa rappresentare il campione di dati. Quindi, per cercare di eliminare questi problemi, abbiamo fatto uso di un ulteriore strumento: i Diagrammi Quantile-Quantile (Diagrammi Q-Q) .

Sia X una variabile aleatoria ed F la sua funzione di distribuzione. Il q -quantile di X è il valore x_q tale che $F(x_q) = P\{X \leq x_q\} = q$, con $0 < q < 1$ e quindi risulta $x_q = F^{-1}(q)$.

Dato $\{x_1, x_2, \dots, x_n\}$, campione di osservazioni della variabile X , si ordinano i valori osservati in ordine crescente e li si denota con $\{y_j, j = 1, 2, \dots, n\}$, dove $y_1 \leq y_2 \leq \dots \leq y_n$. Sia j il numero d'ordine, ossia $j = 1$ per il valore più piccolo e $j = n$ per

il più grande. Il *Diagramma Q-Q* si basa sul fatto che y_j è una stima del $(j - 1/2)/n$ quantile di X . In altre parole,

$$y_j \cong F^{-1} \left(\frac{j - \frac{1}{2}}{n} \right)$$

Supponiamo che la funzione F sia una possibile rappresentazione della funzione di distribuzione di X . Se F appartiene alla famiglia appropriata di distribuzioni, allora il grafico delle coppie $\left(y_j, F^{-1} \left(\frac{j - \frac{1}{2}}{n} \right) \right)$ sarà approssimativamente una linea dritta e se i valori dei suoi parametri sono appropriati, allora la linea avrà coefficiente angolare uguale a 1. Se F e G sono due funzioni di distribuzione è chiaro che $F = G$ se e solo se i quantili di F sono gli stessi dei corrispondenti quantili di G . Se x_q e y_q sono rispettivamente i q -quantili di F e G il grafico dei punti (x_q, y_q) , per diversi valori di q , sarà una retta passante per l'origine con pendenza unitaria per cui $x_q = y_q$, per ogni q . Nel caso in cui F e G differiscano solo per scala e posizione, allora sarà possibile trovare due numeri reali γ e β ($\beta > 0$), tali che $G(x) = F((x - \gamma)/\beta)$ per ogni x . In questo caso è facile vedere che, per ogni q , $y_q = \gamma + \beta x_q$ e il grafico delle coppie (x_q, y_q) produrrà una linea dritta con coefficiente angolare non necessariamente uguale a 1 e che non dovrà per forza passare per l'origine.

Se la distribuzione assunta è inappropriata i punti del grafico devieranno da una linea dritta in maniera sistematica.

La decisione di rigettare o accettare il modello ipotizzato è soggettiva.

2.3 Stima dei parametri

Il passo successivo alla scelta della famiglia di distribuzioni, che rappresenti adeguatamente i dati, consiste nella stima dei parametri della distribuzione.

2.3.1 Media campionaria e varianza campionaria

In moltissimi casi, per la stima dei parametri delle distribuzioni ipotizzate, si usano la media campionaria e la varianza campionaria.

Sia X_1, X_2, \dots, X_n un campione di osservazioni di taglia n IID, la *media campionaria* è definita da

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (2.4)$$

e la *varianza campionaria* S^2 da

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (2.5)$$

2.3.2 Stimatori di massima verosimiglianza

Uno stimatore è una funzione numerica dei dati. Esistono diverse vie per specificare la forma di uno stimatore per un particolare parametro di una data distribuzione e la sua qualità. In questa parte si analizzeranno gli *Stimatori di Massima Verosimiglianza* (SMV).

Gli SMV sono particolarmente usati per diverse ragioni:

1. Questi stimatori hanno delle proprietà non offerte da altri metodi di stima: stimatori dei minimi quadrati, stimatori corretti e il metodo dei momenti.
2. L'uso degli SMV è molto importante quando viene utilizzato il test χ^2 del *goodness-of-fit* (Sezione 2.4.1).
3. L'idea sulla quale si basano questi stimatori è abbastanza intuitiva.

Il concetto di SMV è facilmente comprensibile nel caso discreto. Supponiamo di aver ipotizzato, per i nostri dati, una distribuzione discreta con un parametro θ non noto. Sia

$p_\theta(x)$ la funzione densità di probabilità per questa distribuzione. Assumendo che i dati osservati X_1, X_2, \dots, X_n siano IID definiamo la *funzione di massima verosimiglianza* $L(\theta)$ come:

$$L(\theta) = p_\theta(X_1)p_\theta(X_2) \cdots p_\theta(X_n) \quad (2.6)$$

La (2.6) è la funzione di probabilità congiunta (essendo le X_i indipendenti) e ci dà la probabilità di ottenere i dati osservati nel caso in cui θ fosse il valore del parametro sconosciuto. Quindi lo *stimatore di massima verosimiglianza*, che denotiamo con $\hat{\theta}$, è il valore di θ che massimizza la funzione $L(\theta)$ per cui dovrà essere $L(\hat{\theta}) \geq L(\theta), \forall \theta$. Cosicché $\hat{\theta}$ è il “miglior rappresentante” per i nostri dati.

Nel caso continuo gli SMV non hanno una spiegazione semplice ed intuitiva per il fatto che la probabilità di una variabile aleatoria continua, in un punto fissato, è uguale a zero. Tuttavia, gli SMV per il caso continuo sono definiti in modo analogo al caso discreto. Se denotiamo con $f_\theta(x)$ la funzione densità ipotizzata, con θ parametro sconosciuto, la *funzione di massima verosimiglianza* è data da

$$L(\theta) = f_\theta(X_1)f_\theta(X_2) \cdots f_\theta(X_n) \quad (2.7)$$

Lo stimatore di massima verosimiglianza $\hat{\theta}$, del parametro θ , è il valore di θ che massimizza $L(\theta)$ per tutti i possibili valori di θ .

Gli stimatori di massima verosimiglianza hanno diverse proprietà statistiche alcune delle quali sono:

- Per la maggior parte delle distribuzioni lo SMV è unico, cosicché $L(\hat{\beta})$ è *strettamente* più grande di $L(\beta)$ per ogni altro valore di β .
- Sebbene gli SMV non siano necessariamente corretti, in generale, la distribuzione asintotica (per $n \rightarrow \infty$) di $\hat{\theta}$ ha media uguale a θ .

- Gli SMV sono invarianti, cioè se per qualche funzione h si ha $\phi = h(\theta)$, allora lo SMV di ϕ è $h(\hat{\theta})$.
- Gli SMV sono asintoticamente normalmente distribuiti, cioè

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \nu(\theta))$$

dove $\nu(\theta) = -1/E(d^2l/d\theta^2)$ (la media è calcolata rispetto alle X_i assumendo che le X_i abbiano la distribuzione ipotizzata), $l(\theta) = \ln L(\theta)$ e $\xrightarrow{\mathcal{L}}$ denota la convergenza in legge.

- Gli stimatori di massima verosimiglianza sono fortemente consistenti, vale a dire che $\lim_{n \rightarrow +\infty} \hat{\theta} = \theta$.

2.4 Valutazione della bontà delle scelte fatte sulle distribuzioni

Dopo aver ipotizzato che una certa famiglia di distribuzioni possa rappresentare i nostri dati ed averne stimato i parametri dobbiamo esaminare se la distribuzione in questione è in accordo con i dati osservati X_1, X_2, \dots, X_n . Supponiamo che la distribuzione ipotizzata sia \hat{F} . La verifica della bontà delle scelte fatte può essere effettuata attraverso dei test statistici sull'ipotesi nulla:

$$H_0 : \text{Le } X_i \text{ sono variabili aleatorie IID con funzione di distribuzione } \hat{F}$$

Questi test sono chiamati *goodness-of-fit* e forniscono una valutazione oggettiva delle scelte. Per ogni test deve essere fissato un livello α che rappresenta la probabilità

di rigettare l'ipotesi nulla sapendo che tale ipotesi è vera. Formalmente:

$$\alpha = P(\text{rigettare } H_0 | H_0 \text{ vera})$$

e spesso per α vengono scelti i valori 0.01 oppure 0.05. Comunque bisogna tenere sempre in mente che non esiste una singola distribuzione corretta in una applicazione reale, perciò non bisogna essere schiavi del verdetto fornito dai test. Di notevole importanza è la taglia del campione dei dati. Infatti, se il campione è poco numeroso, allora sarà molto probabile accettare qualsiasi distribuzione candidata. Al contrario, se il campione è abbastanza numeroso, il verdetto del test del *goodness-of-fit*, con molta probabilità, sarà quello di rigettare ogni distribuzione candidata a poter rappresentare i dati. Quindi la scelta di rigettare una distribuzione ipotizzata dovrebbe essere presa solo se c'è forte evidenza in favore di questa scelta.

2.4.1 Test χ^2

Una procedura, per testare le ipotesi che un campione aleatorio di rango n di una variabile aleatoria X abbia una specifica distribuzione di probabilità, è il *test* χ^2 di *goodness-of-fit*. Il test formalizza l'idea intuitiva del confronto degli istogrammi forniti dai dati con la curva di una funzione densità di probabilità ipotizzata.

Siano $\{X_1, X_2, \dots, X_n\}$ i dati osservati e assumiamo che le X_j siano variabili aleatorie IID. Il test è applicabile sia nel caso continuo che in quello discreto ed è valido per campioni abbastanza numerosi sotto l'ipotesi che i parametri siano stati stimati con gli stimatori di massima verosimiglianza.

Si divide l'intero intervallo della distribuzione ipotizzata in k intervalli adiacenti $[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$. Si definisce N_i come la frequenza osservata delle

variabili X_j nell' i -esimo intervallo $[a_{i-1}, a_i)$ per $i = 1, 2, \dots, k$. Si noti che $\sum_{j=1}^n X_j = n$. La frequenza attesa per ogni classe i è np_i , dove p_i è la probabilità teorica ipotizzata associata all' i -esimo intervallo. Nel caso continuo,

$$p_i = \int_{a_{i-1}}^{a_i} \hat{f}(x) dx$$

dove \hat{f} è la densità della distribuzione ipotizzata. Per variabili discrete

$$p_i = \sum_{\{j: a_{i-1} \leq x < a_i\}} \hat{p}(x_j)$$

dove \hat{p} è la funzione di probabilità ipotizzata.

Alla fine la statistica è data da

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

Si può dimostrare che χ_0^2 converge in legge (per $n \rightarrow \infty$) a una distribuzione χ^2 con $k - s - 1$ gradi di libertà dove s rappresenta il numero di parametri della distribuzione ipotizzata stimati attraverso il campione statistico.

Definiamo $\chi_{1-\alpha, k-s-1}^2$ *valore critico*, ossia quel valore tale che

$$P\{\chi_{k-s-1}^2 \leq \chi_{1-\alpha, k-s-1}^2\} = 1 - \alpha$$

L'ipotesi nulla H_0 verrà rigettata se $\chi_0^2 > \chi_{1-\alpha, k-s-1}^2$.

Purtroppo, non esiste un metodo per determinare le p_i in modo tale da massimizzare la potenza del test. Tuttavia è raccomandabile scegliere l'ampiezza degli intervalli $[a_{j-1}, a_j)$ per $j = 1, 2, \dots, k$ tali che $p_1 = p_2 = \dots = p_k$ o all'incirca uguali e inoltre dovrà essere $np_j \geq 5$ per $j = 1, 2, \dots, k$. Nel caso discreto è abbastanza difficile

raggruppare i dati adiacenti in modo tale da ottenere le p_j uguali tra di loro e quindi i dati verranno raggruppati in maniera tale che $np_j \geq 5$. Nel continuo, nei casi in cui la funzione di distribuzione sia facilmente invertibile, esiste un metodo abbastanza semplice che ora andremo a descrivere.

Se $p_1 = p_2 = \dots = p_k$, allora $1 = \sum_{j=1}^k p_j = kp_j$ e quindi $p_j = \frac{1}{k}$ per $j = 1, 2, \dots, k$.

Dovendo essere $np_j \geq 5$, sostituendo p_j con $\frac{1}{k}$, otteniamo: $\frac{n}{k} \geq 5$ e di conseguenza $k \leq \frac{n}{5}$, ossia $\frac{n}{5}$ rappresenta il limite superiore per il numero di intervalli.

Una volta determinato k , avendo stabilito che tutte le p_j debbano essere uguali tra loro, avremo $p_j = \frac{1}{k}$.

A questo punto siamo in grado di determinare i valori degli estremi degli intervalli a_1, \dots, a_{k-1} . Essendo

$$F(a_v) = \sum_{j=1}^v p_j = \sum_{j=1}^v \frac{1}{k} = \frac{v}{k}$$

dove F è la funzione di distribuzione e $1 \leq v \leq k - 1$, avremo

$$a_v = F^{-1}\left(\frac{v}{k}\right)$$

2.4.2 Test di Kolmogorov-Smirnov

Il *test di Kolmogorov-Smirnov* (K-S) confronta una funzione di distribuzione empirica con la funzione di distribuzione ipotizzata \hat{F} . Questo test presenta dei vantaggi rispetto al test χ^2 a causa del fatto che non occorre in alcun modo raggruppare i dati ed inoltre non ci sono le limitazioni riguardo alla dimensione minima del campione ed al fatto che il test χ^2 è valido solo in senso asintotico. Uno degli svantaggi è che K-S non è applicabile nel caso discreto, se non in casi molto particolari. Un'altra limitazione del test K-S è che la sua formulazione iniziale è stata fatta per distribuzioni in cui tutti i parametri sono noti, cioè non era possibile utilizzare i parametri stimati dai dati.

Tuttavia, molti analisti applicano il test per distribuzioni con parametri stimati anche se in effetti il test non sarebbe valido, ma gli effetti di questa errata applicazione non sono stati ben capiti. Recentemente il test K-S è stato esteso per essere utilizzato nel caso di parametri stimati dai dati con distribuzioni ipotizzate normali, esponenziali e Weibull.

Per definire la statistica del test dobbiamo, innanzitutto, definire una *funzione di distribuzione empirica*:

$$F_n(x) = \frac{\#\{X_i : X_i \leq x\}}{n}, \forall x \in \mathfrak{R}$$

dove X_i , $i = 1, 2, \dots, n$ sono i dati osservati ordinati in ordine crescente, quindi risulta che $F_n(X_i) = i/n$.

La *statistica* D_n del test K-S è definita come la più grande distanza tra $F_n(x)$ e $\hat{F}(x)$. Formalmente:

$$D_n = \sup_x \{|F_n(x) - \hat{F}(x)|\}$$

D_n può essere calcolato nel seguente modo:

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - \hat{F}(X_i) \right\}, \quad D_n^- = \max_{1 \leq i \leq n} \left\{ \hat{F}(X_i) - \frac{i-1}{n} \right\}$$

e alla fine si mette

$$D_n = \max \{D_n^+, D_n^-\}$$

Chiaramente alti valori di D_n indicano che l'ipotesi nulla H_0 è da rigettare. Il valore numerico del *punto critico* $d_{n,1-\alpha}$ dipende da come è stata specificata la distribuzione ipotizzata e ci sono diversi casi:

1. *Tutti i parametri di \hat{F} sono noti*, cioè non è stato stimato dai dati nessun parame-

tro della distribuzione ipotizzata. In questo caso il valore D_n non dipende dalla distribuzione \hat{F} , naturalmente sotto l'ipotesi che \hat{F} sia continua e può essere utilizzata un'unica tabella di punti critici che sono validi per qualsiasi distribuzione ipotizzata. Stephens [Ste74] fornisce uno schema approssimato per eliminare la dipendenza del punto critico da n correggendo la statistica nel modo seguente:

$$D_{adjusted} = \left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}} \right) D_n$$

quindi, l'ipotesi H_0 verrà rigettata se $D_{adjusted} > c_{1-\alpha}$. I valori di $c_{1-\alpha}$ possono essere ricavati dalla tabella 2.1.

2. *Distribuzione $N(\mu, \sigma^2)$ con μ e σ^2 non noti.* Se stimiamo μ con \bar{X} (media campionaria) e σ^2 con S^2 (varianza campionaria) e consideriamo \hat{F} funzione di ripartizione della distribuzione $N(\bar{X}, S^2)$, allora la statistica corretta è

$$D_{adjusted} = \left(\sqrt{n} + 0.01 + \frac{0.85}{\sqrt{n}} \right) D_n$$

in modo analogo al caso precedente, l'ipotesi H_0 verrà rigettata se $D_{adjusted} > c'_{1-\alpha}$.

3. *Distribuzione $Expo(\lambda)$ con λ non noto.* Il parametro λ viene stimato con lo stimatore di massima verosimiglianza $\frac{1}{\bar{X}}$. Definiamo \hat{F} funzione di ripartizione della distribuzione $Expo\left(\frac{1}{\bar{X}}\right)$. La statistica corretta è definita come:

$$D_{adjusted} = \left(\sqrt{n} + 0.26 + \frac{0.5}{\sqrt{n}} \right) \left(D_n - \frac{0.2}{n} \right)$$

e in questo caso il punto critico viene denotato con $c''_{1-\alpha}$.

Caso	$D_{adjusted}$	$(1 - \alpha)$				
		0.850	0.900	0.950	0.975	0.990
Con parametri noti	$\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) D_n$	1.138	1.224	1.358	1.480	1.628
$N(\bar{X}, S^2)$	$\left(\sqrt{n} + 0.01 + \frac{0.85}{\sqrt{n}}\right) D_n$	0.775	0.819	0.895	0.955	1.035
$Expo\left(\frac{1}{\bar{X}}\right)$	$\left(\sqrt{n} + 0.26 + \frac{0.5}{\sqrt{n}}\right) \left(D_n - \frac{0.2}{n}\right)$	0.926	0.990	1.094	1.190	1.308

Tabella 2.1: Tabella dei valori critici $c_{1-\alpha}$, $c'_{1-\alpha}$, $c''_{1-\alpha}$

2.5 Un esempio

Trattiamo un esempio che mostrerà l'analisi dei tempi di interarrivo, riportati nella Tabella 2.2, rilevati in una giornata. Si inizia costruendo l'istogramma il quale ci permetterà di capire se la sua forma somiglia alla funzione densità di qualche legge nota. I tempi, ordinati in ordine crescente, assumono valori nell'intervallo $A = [0.0833, 4.8033]$.

0,0833	0,1	0,1166	0,1166	0,1166	0,1166	0,1333	0,15
0,1666	0,1666	0,1666	0,1833	0,1833	0,1833	0,1833	0,1833
0,1833	0,2	0,2166	0,2166	0,2166	0,2166	0,2166	0,2333
0,2333	0,25	0,25	0,25	0,2666	0,2666	0,2833	0,2833
0,2833	0,2833	0,2833	0,3	0,3	0,3	0,3166	0,3166
0,35	0,3833	0,3833	0,3833	0,3833	0,4666	0,4666	0,4833
0,4833	0,5	0,5	0,5	0,5166	0,5666	0,5833	0,6
0,6	0,6333	0,6666	0,6666	0,6666	0,7	0,7	0,7333
0,75	0,7666	0,7666	0,7833	0,8	0,8166	0,8166	0,8166
0,85	0,9	0,9333	0,95	0,9666	1	1,0333	1,05
1,05	1,3	1,3333	1,35	1,3666	1,3666	1,3666	1,4
1,4166	1,5	1,5333	1,55	1,7166	1,7166	1,75	1,8333
1,8333	1,8666	1,8666	1,95	1,9833	2,0333	2,3333	2,3666
2,7666	2,9333	3,0166	3,1833	3,55	3,5833	4,7333	

Tabella 2.2: Tabella tempi (minuti) ordinati in ordine crescente

Dividiamo l'intervallo A in 11 sottointervalli $[b_0, b_1)$, $[b_1, b_2)$, \dots , $[b_{10}, b_{11})$, ciascuno di ampiezza $\Delta b = 0.429091$, dove $b_0 = 0.0833$ e $b_{11} = 4.8033$. Applicando l'equazione (2.1) determiniamo la funzione $h(x)$ e dalla (2.3) deriva $f(x) = \frac{h(x)}{\Delta b}$. Con

i valori ottenuti, illustrati nella Tabella 2.3, costruiamo l'istogramma della $f(x)$, riportato nella Figura 2.1 a pagina 48, che evidenzia una certa somiglianza con la curva della densità esponenziale.

i	b_i	q_i	$\frac{h(x_i)}{\Delta b}$	$Expo\left(\frac{1}{X}\right)$
1	0,512424242	52	1,091769736	0,948019279
2	0,941515152	23	0,482898152	0,557843301
3	1,370606061	12	0,251946862	0,328251919
4	1,79969697	8	0,167964575	0,193153386
5	2,228787879	7	0,146969003	0,113657311
6	2,657878788	2	0,041991144	0,06687941
7	3,086969697	3	0,062986716	0,039353873
8	3,516060606	1	0,020995572	0,023157013
9	3,945151515	2	0,041991144	0,013626289
10	4,374242424	0	0	0,008018122
11	4,803333333	1	0,020995572	0,004718106

Tabella 2.3: Valori per la costruzione dell'istogramma

Dall'analisi dei tempi risulta che il minimo tempo osservato è $b_0 = 0.0833$ e conseguentemente l'istogramma appare traslato a destra. A causa della somiglianza dell'istogramma con la curva della densità esponenziale è stato deciso di ipotizzare una *distribuzione esponenziale traslata* con $\hat{\gamma} = \frac{n}{n+1}b_0 = 0.0826$ dove $n = 111$ è il numero di tempi osservati e la distribuzione ipotizzata sarà

$$\hat{f}(x) = \begin{cases} \hat{\lambda}e^{-\hat{\lambda}(x-\hat{\gamma})} & , \text{ se } x > \hat{\gamma} \\ 0 & , \text{ altrimenti} \end{cases}$$

Il parametro $\hat{\lambda}$ della distribuzione viene stimato come $\hat{\lambda} = \frac{1}{\bar{Z}}$, dove

$$\bar{Z} = \sum_{i=1}^n \frac{(X_i - \hat{\gamma})}{n} = 0.8091$$

è la media campionaria dei dati traslati.

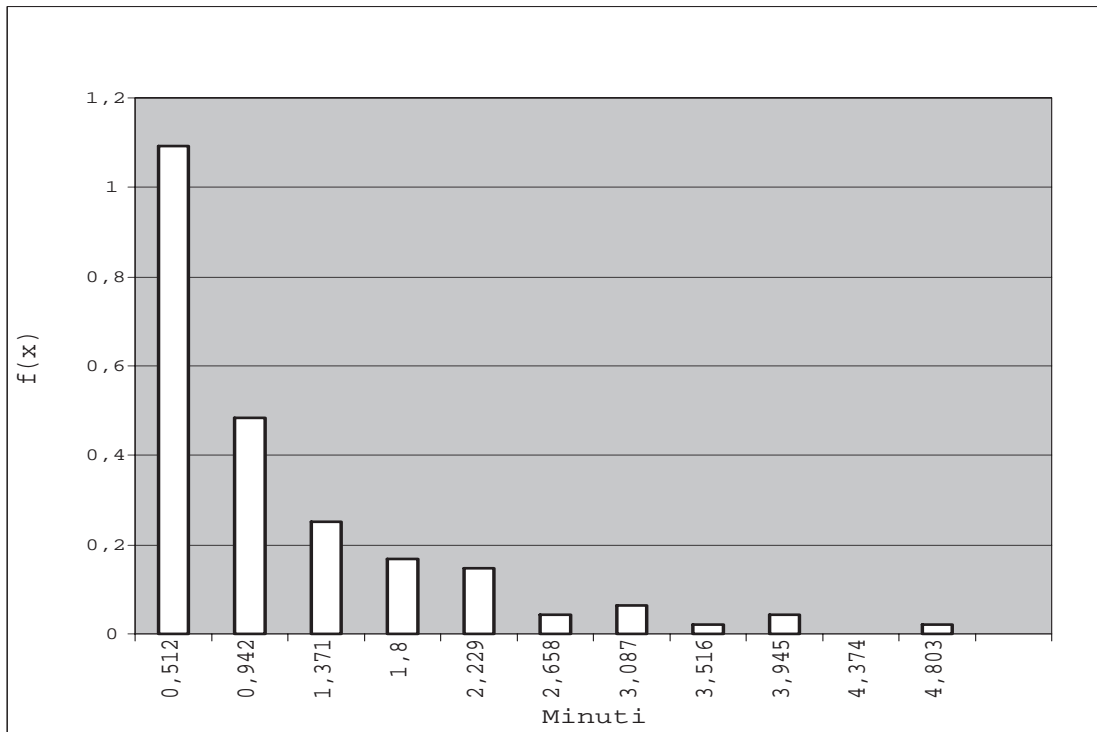


Figura 2.1: Istogramma dei tempi di interarrivo

La figura 2.2 mostra il confronto tra l'istogramma e il grafico della \hat{f} . Dalla sua osservazione si può concludere che potrebbe essere ragionevole ipotizzare la famiglia delle distribuzioni esponenziali per rappresentare i nostri dati.

A questo punto possiamo utilizzare il metodo dei grafici Q-Q per vedere se l'ipotesi fatta sulla famiglia di distribuzioni è buona.

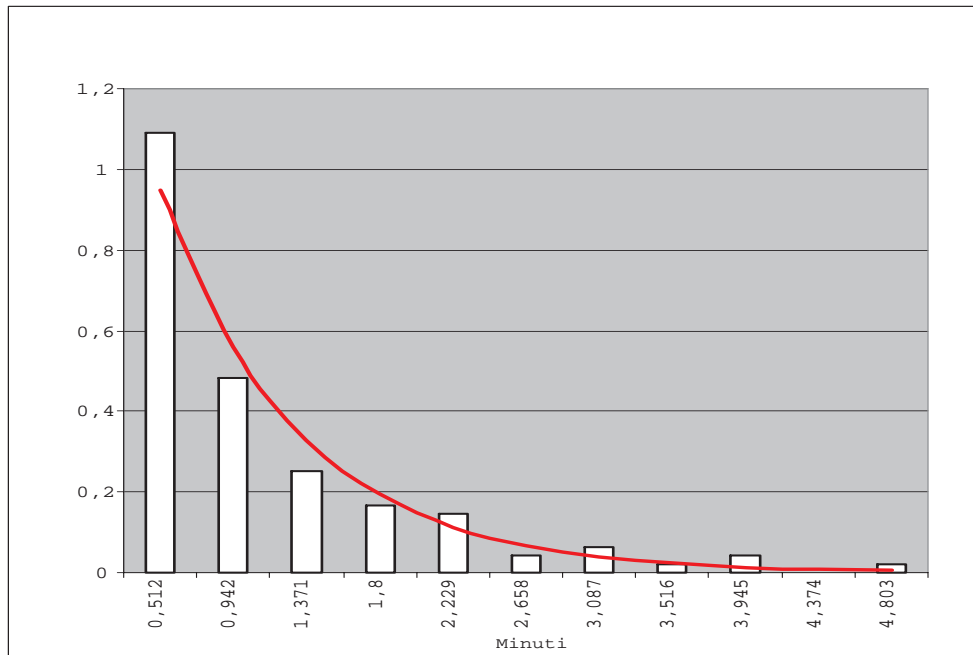


Figura 2.2: Istogramma dei tempi di interarrivo e grafico di \hat{f}

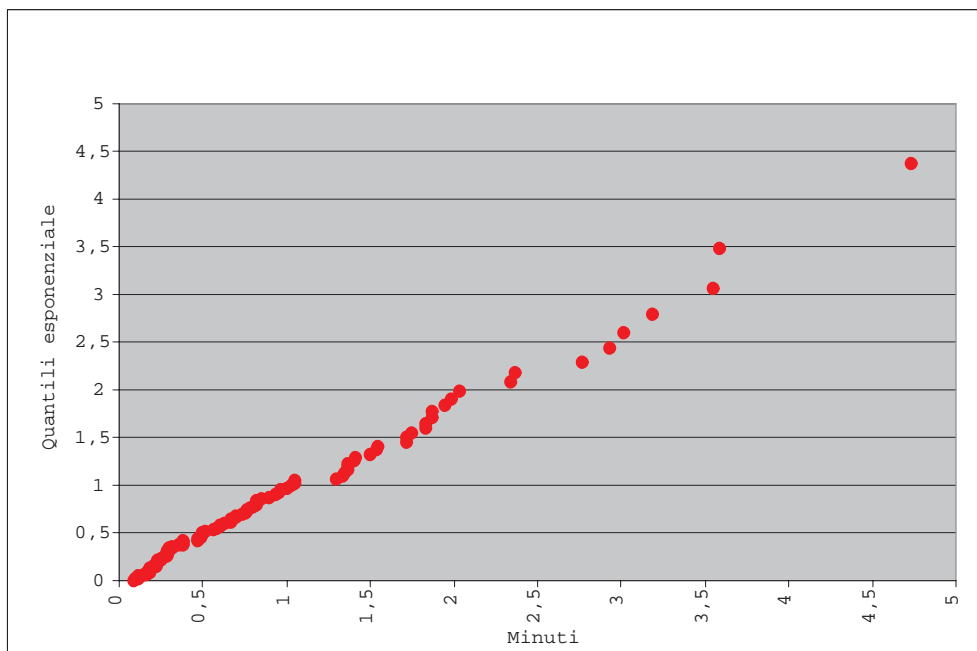


Figura 2.3: Grafico Quantile-Quantile per i tempi di interarrivo

Come si può notare, dalla Figura 2.3, i punti del grafico cadono all'incirca su una linea retta e ciò conferma, in una certa misura, l'ipotesi che la distribuzione del campione segua la legge \hat{f} .

i	a_i	N_i	$\frac{(N_i - Np_i)^2}{Np_i}$	Np_i
1	0,041504095	6	0,036486486	5,55
2	0,08525272	5	0,054504505	5,55
3	0,131502591	7	0,378828829	5,55
4	0,180557153	10	3,568018018	5,55
5	0,232778655	10	3,568018018	5,55
6	0,288604407	3	1,171621622	5,55
7	0,348569055	4	0,432882883	5,55
8	0,413335808	4	0,432882883	5,55
9	0,483741277	4	0,432882883	5,55
10	0,560861744	5	0,054504505	5,55
11	0,646114464	5	0,054504505	5,55
12	0,741418896	9	2,144594595	5,55
13	0,84946615	2	2,270720721	5,55
14	0,974197552	7	0,378828829	5,55
15	1,121723487	0	5,55	5,55
16	1,30228064	6	0,036486486	5,55
17	1,535059295	5	0,054504505	5,55
18	1,863142383	7	0,378828829	5,55
19	2,424004127	5	0,054504505	5,55
20	∞	7	0,378828829	5,55
			$\chi_0^2 = 21,43243243$	

Tabella 2.4: Test χ^2 per i tempi di interarrivo

Passiamo adesso ai test di *goodness-of-fit* per valutare formalmente la bontà delle scelte fatte. In tabella 2.4 è riportato il risultato del test χ^2 . Il test è stato fatto sui tempi traslati con $k = 20$ intervalli in cui $a_0 = 0$ e $a_{20} = \infty$. Gli estremi degli intervalli possono essere rilevati dalla tabella. Siccome abbiamo stimato due parametri, cioè $\hat{\lambda}$ e $\hat{\gamma}$, avremo $s = 2$ e la statistica χ_0^2 converge in legge a una distribuzione χ^2 con $k - s - 1 = 17$ gradi di libertà. Posto $\alpha = 0.1$ abbiamo $21.432 = \chi_0^2 < \chi_{0.9,17}^2 = 24.769$.

Per cui l'ipotesi viene accettata con un intervallo di confidenza pari a $100(1 - \alpha)\% = 90\%$.

Anche il test di Kolmogorov-Smirnov ha accettato le ipotesi. Infatti, la statistica $D_{111} = 0.10916$ ha fornito un valore

$$D_{adjusted} = \left(\sqrt{111} + 0.26 + \frac{0.5}{\sqrt{111}} \right) \left(D_{111} - \frac{0.2}{111} \right) = 0.9778$$

che è minore del valore $c''_{0.9} = 0.990$ rilevato dalla tabella 2.1 a pag. 45.

2.6 Test di Kruskal-Wallis

Supponiamo di avere i seguenti insiemi di dati:

$$\begin{array}{cccc} X_{1,1}, & X_{1,2}, & \dots, & X_{1,n_1} \\ X_{2,1}, & X_{2,2}, & \dots, & X_{2,n_2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m,1}, & X_{m,2}, & \dots, & X_{m,n_m} \end{array}$$

e di essere interessati nel testare l'ipotesi nulla H_0 che tutte le $n = \sum_{i=1}^m n_i$ variabili aleatorie siano indipendenti e abbiano la stessa distribuzione (IID). Tutto questo può essere fatto con il test di Kruskal-Wallis (*multisample rank test*) [Ros97]. Per testare l'ipotesi H_0 ordiniamo gli n valori distinti $X_{1,1}, \dots, X_{1,n_1}, X_{2,1}, \dots, X_{2,n_2}, \dots, X_{m,1}, \dots, X_{m,n_m}$ in ordine crescente e denotiamo con $r_{i,j}$ la *posizione* della variabile $X_{i,j}$ nell'ordinamento, con $1 \leq i \leq m$ e $1 \leq j \leq n_i$. Definiamo il *rango* $R_i = \sum_{j=1}^{n_i} r_{i,j}$ come la somma delle posizioni degli elementi dell' i -esimo insieme. Siccome, sotto l'ipotesi H_0 , abbiamo che la probabilità che una generica variabile $X_{i,j}$ cada nella k -esima posizione

dell'ordinamento globale è $\frac{1}{n}$, avremo che

$$E[R_i] = n_i \frac{(n+1)}{2}$$

Usando la precedente formula possiamo ricavare la statistica per il test

$$R = \frac{12}{n(n+1)} \sum_{i=1}^m \frac{(R_i - E[R_i])^2}{n_i}$$

Quindi, un valore piccolo di R indica che l'ipotesi H_0 è buona e viene accettata mentre nel caso di valori di R grandi la si rigetta.

Se denotiamo con $y = R$ il valore ottenuto, il p -valore della statistica è dato da

$$p\text{-valore} = P_{H_0} \{R \geq y\}$$

Questo valore può essere approssimato usando il risultato che, per grandi valori di n_i , R segue approssimativamente una distribuzione χ^2 con $m-1$ gradi di libertà. Quindi, se $R = y$

$$p\text{-valore} = P_{H_0} \{ \chi_{m-1}^2 \geq y \} \quad (2.8)$$

e l'ipotesi H_0 è accettata se $p\text{-valore} > \alpha$, con α ampiezza della regione critica.

Quando le $X_{i,j}$ non sono tutte distinte, per il calcolo del p -valore si utilizza sempre la (2.8), ma cambia il calcolo di R , cioè la posizione $r_{i,j}$ delle $X_{i,j}$ aventi uguale valore viene calcolata come la media delle loro posizioni nell'ordinamento globale.

2.7 Risultati delle analisi

I risultati che seguono sono stati ottenuti applicando, per ogni variabile aleatoria di ingresso del modello, la metodologia descritta nell'esempio della sezione 2.5, ossia costruendo gli istogrammi, analizzando i grafici Q-Q, ipotizzando una distribuzione di probabilità e stimando i suoi parametri con gli stimatori di massima verosimiglianza ed infine applicando i test. Tutte le analisi sui dati sono state condotte utilizzando Microsoft Excel.

2.7.1 Tempi di servizio

Dall'analisi sui tempi di servizio sono emersi i risultati riportati nella Tabella 2.5 a pag. 54. Per testare il modello statistico supposto, relativo alle leggi di probabilità che regolano ciascun tipo di servizio, è stato applicato il test Chi-quadro. Inoltre, per i tempi di accettazione delle preospedalizzazioni, per i quali la distribuzione ipotizzata è una legge normale, è stato possibile applicare il test di Kolmogorov-Smirnov perché avevamo gli strumenti per calcolare il valore del suo punto critico (si veda la sezione 2.4.2).

Gli intervalli sono stati definiti in quanto, dalla osservazione dei dati raccolti, si poteva vedere che i tempi di servizio non scendevano sotto una certa soglia (min) ed erano limitati superiormente (max). I valori del minimo e del massimo degli intervalli sono stati stimati dai dati campionati nel modo seguente:

$$\min = \frac{n}{n+1} X_{\min} \quad \max = \frac{n+1}{n} X_{\max}$$

dove n è la dimensione del campione, con X_{\min} e X_{\max} rispettivamente il minimo e il massimo valore dei tempi osservati.

Legge		Test		
<i>Tipo</i>	<i>Parametri</i>	<i>Intervallo min max</i>	<i>Kolmogorov - Smirnov D adjusted</i> $100(1 - \alpha^*)\%$	<i>Chi quadro di libertà'</i> k $100(1 - \alpha^*)\%$
Accettazione normali, consegna campioni, urgenti				
gamma	α β 1.543756294 0.540041941	0.6989899 4.90708092	- -	35 43.8179%
Accettazione preospitalizzazioni				
normale	μ σ^2 1.169230769 0.447992087	0.1822549 2.66568047	0.72417766 < 85%	30 63.41678%
Accettazione multiple				
lognormale	μ σ^2 0.350088919 0.661379611	0.75833333 5.83003663	- -	13 82.92434%
prelievi				
gamma	α β 2.088243778 1.226387556	0.94900731 10.8770408	- -	30 74.51709%
macchinette ticket				
gamma	α β 3.706715912 0.342122909	0.1826809 4.48261905	- -	30 60.37498%

Tabella 2.5: Risultati delle analisi dei tempi di servizio

Giorno	Fascia oraria		Legge		Test					
	Inizio	Fine	Tipo	Parametro (λ)	Shift (γ)	<i>Kolmogorov - Smirnov</i> $D_{adjusted}$	$100(1 - \alpha^*)\%$	<i>Chi quadro</i> k	<i>Gradi di liberta'</i>	
<i>lun</i>	7.20.00	9.00.00	<i>Expo</i>	1.235861	0.082589	0.977770	< 90%	20	79.2469%	17
	7.20.00	8.10.00	<i>Expo</i>	1.041736	-	0.427221	< 85%	8	55.4926%	6
	8.10.00	9.00.00	<i>Expo</i>	1.393470	-	0.836646	< 85%	10	69.0772%	8
<i>mer</i>	7.20.00	8.10.00	<i>Expo</i>	1.546592	0.131282	0.378191	< 85%	10	34.7663%	7
	8.10.00	8.30.00	<i>Expo</i>	3.274306	0.081666	0.819073	< 85%	10	83.2356%	7
	8.30.00	9.00.00	<i>Expo</i>	1.308171	0.161616	0.635674	< 85%	10	<i>dati insu. ff.</i>	
<i>gio</i>	7.20.00	8.05.00	<i>Expo</i>	1.58191	0.131073	0.479128	< 85%	10	11.8205%	7
	8.05.00	8.30.00	<i>Expo</i>	3.137962	0.114655	1.068030	< 95%	11	55.4284%	8
	8.30.00	9.00.00	<i>Expo</i>	0.7	-	0.504881	< 85%	10	<i>dati insu. ff.</i>	
<i>ven</i>	7.20.00	7.55.00	<i>Expo</i>	1.131805	-	0.702073	< 85%	16	79.3839%	14
	7.55.00	8.20.00	<i>Expo</i>	1.986008	0.049418	0.876003	< 85%	15	88.0690%	12
	8.20.00	9.00.00	<i>Expo</i>	0.909090	-	0.516945	< 85%	14	52.3299%	12
<i>sab</i>	7.20.00	8.00.00	<i>Expo</i>	1.094284	0.082264	0.903499	< 85%	15	80.5168%	12
	8.00.00	8.30.00	<i>Expo</i>	1.262173	0.098484	0.86712	< 85%	13	44.8816%	10
	8.30.00	9.00.00	<i>Expo</i>	1.003170	-	0.697003	< 85%	12	76.2517%	10
<i>lun, ... , sab</i>	9.00.00	9.15.00	<i>Expo</i>	0.556483	-	0.901295	< 85%	10	80.9377%	8

Tabella 2.6: Risultati delle analisi dei tempi di interarrivo

La raccolta dei tempi di servizio ai box, degli utenti che consegnavano i campioni, ha presentato notevoli difficoltà operative perché si sovrapponevano con i prelievi. Come conseguenza della difficile raccolta di questi dati, oltre al fatto che il numero di utenti che dovevano solo consegnare i campioni era basso, non è stato possibile applicare un test statistico oggettivo e abbiamo stabilito che il tempo di consegna dei campioni dovesse seguire una legge uniforme con valori compresi tra 10 e 30 secondi. Questa scelta è stata avallata dagli esperti del Centro Prelievi.

2.7.2 Processo di arrivo

È stato fatto inizialmente uno studio per stabilire se i processi di arrivo dei diversi giorni della settimana provenissero dalla stessa popolazione. Come si può vedere dalla Tabella 2.6 a pag. 55, i processi di arrivo differiscono da giorno a giorno. È stato ipotizzato che fattori esterni come condizioni metereologiche, particolari periodi dell'anno ecc., non influenzino i tassi di arrivo degli utenti. Gli arrivi costituiscono la realizzazione di un processo stocastico a tempo continuo e stati discreti. Nell'analisi delle singole giornate si è presentato il problema di dover verificare se i processi fossero stazionari, cioè si è dovuto vedere se la velocità di arrivo (λ) delle persone nei diversi momenti della giornata, rimaneva costante. Ciò non si è verificato, cioè è risultato che il fattore λ dipendeva dal tempo e questo ci ha portato a dover dividere le singole giornate in fasce orarie stabilendo, per ciascuna fascia, un proprio processo di arrivo stazionario.

Dalle osservazioni dei dati raccolti risultava che il numero delle persone, all'apertura del Centro Prelievi, variava da un minimo di 10 a un massimo di 20. Abbiamo ipotizzato che tale variabile fosse regolata da una distribuzione uniforme discreta nell'intervallo $[10, 20]$:

$$p(x) = \begin{cases} \frac{1}{20-10+1} & \text{se } x \in \{10, 11, \dots, 20\} \\ 0 & \text{altrimenti} \end{cases} \quad (2.9)$$

dove x indica il numero di pazienti già in coda al momento dell'apertura.

2.7.3 Altri risultati

Un'ulteriore variabile di ingresso, da tenere in considerazione, è quella che regola il tipo di utente. Come già detto nel Capitolo 1 abbiamo a che fare con sette diversi tipi di utenti.

Facendo il rapporto tra il numero di pazienti osservati di ciascun tipo ed il numero totale delle osservazioni abbiamo stimato la probabilità che un utente arrivato appartenga a una delle sette categorie e i risultati ottenuti sono riportati nella seguente tabella:

	<i>preospedalizzazioni</i>	<i>normali</i>	<i>consegna campioni</i>	<i>impegnative multiple</i>	<i>urgenti</i>
<i>senza ticket</i>	0.17509728	0.3628405	0.06420233	0.057393	0.08365759
<i>con ticket</i>	—	0.20233463	0.02626459	0.02821012	—

Tabella 2.7: Probabilità di arrivo dei diversi tipi di utenti

Dall'osservazione dei tempi raccolti è emerso che tra un servizio e il successivo, sia all'accettazione che ai box, intercorreva un certo periodo di tempo. Quindi, per rappresentare in modo accurato il sistema è stato deciso di prevedere, prima dell'inizio di ogni

servizio, un tempo di transizione sia per l'accettazione (Tempo di interaccettazione) che per i box (Tempo di interbox).

L'analisi statistica dei tempi di interaccettazione non ci ha permesso di trovare alcun modello statistico teorico che potesse rappresentare i dati, per cui l'unica scelta possibile è stata quella di rappresentarli con una distribuzione empirica derivante dai dati stessi.

Per i tempi di interbox abbiamo verificato, attraverso i test, che sono regolati da una legge gamma con parametri $\alpha = 1.767322$ e $\beta = 0.188813$.

Capitolo 3

Implementazione

3.1 Scelta del software

Una decisione importante nello sviluppo di un modello di simulazione è la scelta del linguaggio da usare per la sua implementazione [LK91] [NHP98]. Una scelta non appropriata può causare l'insuccesso dell'intero progetto. Saranno di seguito elencati alcuni vantaggi che si hanno nel programmare il modello di simulazione con un linguaggio di simulazione piuttosto che con un linguaggio *general-purpose* come per esempio Pascal, C, oppure Basic:

- I linguaggi di simulazione forniscono automaticamente le caratteristiche più importanti necessarie per la programmazione di un modello con una conseguente riduzione del tempo di sviluppo.
- Forniscono un ambiente di lavoro naturale per la modellazione della simulazione. I loro costrutti di base sono più aderenti alla simulazione rispetto a quelli dei linguaggi *general-purpose*.

- I modelli, quando sono scritti in un linguaggio di simulazione, sono più facilmente modificabili.
- Forniscono una migliore rilevazione degli errori perché i più comuni tipi di errore sono identificati e controllati automaticamente. Dato che vengono scritte poche linee di codice la possibilità di fare un errore, probabilmente, sarà piccola.

D'altra parte, molti modelli di simulazione vengono scritti usando linguaggi *general-purpose*. Alcuni vantaggi di questa scelta sono i seguenti:

- Molti modellisti conoscono dei linguaggi *general-purpose*, ma non conoscono linguaggi di simulazione.
- Molti linguaggi *general-purpose* sono disponibili sulla maggior parte dei calcolatori elettronici, ma questo può non essere vero per un particolare linguaggio di simulazione.
- Un programma efficiente, scritto in FORTRAN o in C, può richiedere un tempo di esecuzione inferiore rispetto a quello richiesto dal corrispondente programma scritto in un linguaggio di simulazione. Questo perché un linguaggio di simulazione è progettato per modellare un vasta varietà di sistemi con un certo insieme di blocchi costruttivi predefinito, mentre un programma scritto con un linguaggio *general-purpose* può essere adattato alla particolare applicazione. Queste considerazioni iniziano a perdere di importanza a causa della sempre crescente potenza di calcolo dei moderni sistemi di calcolo.
- I linguaggi *general-purpose* hanno una flessibilità maggiore rispetto a certi linguaggi di simulazione.

- Il costo del software potrebbe essere più basso (ma non necessariamente il costo del progetto).

I linguaggi di simulazione hanno un certo numero di elementi di modellazione comuni come entità, attributi, code e risorse.

Ci sono, tuttavia, delle caratteristiche aggiuntive che dovrebbero essere disponibili in un pacchetto di simulazione e che analizziamo nelle sezioni seguenti.

3.1.1 Modellazione

Una delle caratteristiche più importanti per un pacchetto di simulazione è la *flessibilità di modellazione*, cioè deve permettere di implementare modelli di natura diversa senza dover essere costretti ad approssimazioni eccessive che fornirebbero dei modelli che potrebbero non essere totalmente validi. Inoltre, dovrebbe esserci la possibilità di modificare il codice dei blocchi costruttivi esistenti o di crearne di nuovi adatti alle proprie esigenze. Le entità dovrebbero avere *attributi generali* che dovrebbero poter essere contestualmente modificati. Sarebbe utile poter importare codice scritto in un linguaggio diverso da quello proprio del pacchetto, per esempio scritto in C++ o Java. La *facilità di sviluppo del modello* è un'altra importante caratteristica che permette di ridurre il tempo di costruzione del progetto. La *massima dimensione del modello* potrebbe essere importante in caso di modelli molto complessi. Sarebbe inoltre auspicabile che il pacchetto abbia la più ampia *portabilità* in modo da poter essere eseguito su macchine con architetture e sistemi operativi diversi. Dovrebbe anche avere la capacità di *combinare simulazioni discrete e continue*.

3.1.2 Animazione

La facilità con cui si usa l'animazione è il principale motivo della crescente popolarità dei modelli di simulazione. In una animazione gli elementi chiave di un sistema (per esempio macchine e parti) sono rappresentati a video con icone che cambiano forma e posizione quando c'è un cambiamento di stato nella simulazione. Così si può vedere graficamente l'evoluzione del sistema nel tempo. Naturalmente, le animazioni fanno aumentare il tempo di esecuzione della simulazione.

La ragione principale della popolarità delle animazioni è la loro abilità nel comunicare l'essenza del modello, a manager o altre persone chiave del progetto, incrementando in grande misura la credibilità del modello.

Altri potenziali benefici portati dalle animazioni sono:

- rilevare gli errori in un programma di simulazione (*debugging*);
- dimostrare che un modello di simulazione non è valido;
- suggerire dei miglioramenti nelle procedure o nei controlli di un sistema;
- capire il comportamento dinamico di un sistema;
- addestrare il personale.

3.1.3 Capacità statistiche

Dato che, molti sistemi del mondo reale hanno un comportamento casuale, un pacchetto di simulazione deve avere buone capacità statistiche. In generale, ogni sorgente di un sistema stocastico (tempi di interarrivo, tempi di servizio, ecc.) deve essere modellata da una distribuzione di probabilità e quindi un pacchetto di simulazione dovrebbe contenere un vasta varietà di distribuzioni standard (per esempio esponenziale, gamma,

triangolare), essere in grado di usare distribuzioni basate sui dati osservati dal sistema (distribuzioni empiriche) e avere generatori di numeri casuali multipli per facilitare il confronto tra sistemi diversi.

3.1.4 Esecuzione

Uno strumento di simulazione che ha la caratteristica delle esecuzioni multiple fornisce strumenti per eseguire automaticamente il modello diverse volte cambiando i semi dei generatori ad ogni esecuzione. L' output di run multipli dovrebbe poter essere memorizzato in un file. La capacità di resettare dà all'utente la possibilità di reimpostare le statistiche del modello durante la simulazione. Un altro aspetto importante per la valutazione del software è la possibilità di poter far partire la simulazione da uno stato non vuoto, specificando i valori iniziali delle variabili e degli attributi e determinando la situazione iniziale delle entità, delle code e delle risorse. Il controllo della velocità di esecuzione del modello è un'altra caratteristica gradita che potrebbe ridurre il tempo di esecuzione stesso.

3.1.5 Presentazione degli output

Un pacchetto di simulazione dovrebbe fornire la possibilità di memorizzare gli indici di valutazione delle prestazioni standard come fattore di utilizzo, dimensione delle code, tempi di attesa ecc. Inoltre, dovrebbe dare la possibilità di fare dei rapporti su dati costruiti su misura per il fatto che gli indici standard non sono facilmente comprensibili dai non addetti. Altra caratteristica importante è la possibilità di ottenere delle visualizzazioni grafiche (istogrammi, grafici a barre, ecc.) dei valori assunti da variabili di interesse sia per singole esecuzioni della simulazione che per esecuzioni multiple. I

dati in uscita dovrebbero poter essere reindirizzati come ingressi per altri programmi (fogli elettronici o pacchetti statistici).

3.2 *ExtendTM*

Uno degli strumenti di simulazione che possiede la maggior parte delle caratteristiche elencate precedentemente è *ExtendTM*, della Imagine That, Inc. Per questo motivo è stato deciso il suo utilizzo per la realizzazione del modello di simulazione.

Extend è una applicazione *general-purpose*, grafica, orientata alla simulazione discreta e continua con un ambiente di sviluppo integrato. *Extend* è disponibile per i sistemi operativi Microsoft Windows e Macintosh.

Per molti anni c'è stata un dicotomia nel software di simulazione tra i linguaggi di simulazione e i simulatori. I linguaggi erano visti come qualcosa di più potente e generale, mentre i simulatori erano caratterizzati dalla loro facilità d'uso, ma erano generalmente limitati a un campo specifico di applicazione. *Extend* fa da ponte tra questi due tipi di programmi costituendo un programma flessibile e di semplice uso. Esso può essere visto come:

- uno strumento di simulazione *stand-alone* che può essere usato per creare modelli discreti e continui senza programmare;
- un pacchetto di simulazione con un ambiente di creazione dove le interfacce dei modelli possono essere create con facilità, senza dover programmare, con un accrescimento della produttività e semplicità d'uso;
- un ambiente di sviluppo per costruire modelli personalizzati di particolari sistemi. L'ambiente di programmazione permette al progettista di creare un simulatore per ogni campo di applicazione.

3.2.1 L'ambiente di modellazione di Extend

Prima di vedere come Extend può essere usato per costruire modelli è utile capire il suo ambiente di modellazione. I modelli di Extend sono costruiti con *blocchi* iconici appartenenti a librerie. Ogni blocco descrive un passo in un processo oppure un calcolo. La finestra di dialogo associata ad ogni blocco permette di definire il comportamento del blocco stesso e di riportare i risultati ad esso associati. Ogni *libreria* rappresenta un raggruppamento di blocchi con caratteristiche simili come *Discrete Event*, *Plotters*, *Electronics* o *Business Process Reengineering*. I blocchi sono messi nel foglio di lavoro del modello trascinandoli dalla finestra della libreria. Viene quindi stabilito un flusso tra i blocchi.

Ci sono due tipi di *flussi logici* tra i blocchi di Extend. Il primo tipo di flusso è quello degli *items* che rappresentano gli oggetti che si muovono nel sistema. Gli items possono avere associati attributi e priorità. Esempi di items sono pazienti o pacchetti di informazioni. Il secondo tipo di flusso logico è quello dei *valori* che cambieranno, con il trascorrere del tempo, durante l'esecuzione della simulazione. I valori rappresentano un singolo numero. Esempi di valori sono il numero di items in una coda, il numero generato da un generatore casuale o il numero di items che sono usciti dal sistema.

Ogni blocco ha dei connettori che sono i suoi punti di interfaccia. La Figura 3.1 mostra i simboli dei connettori dei valori e degli items.

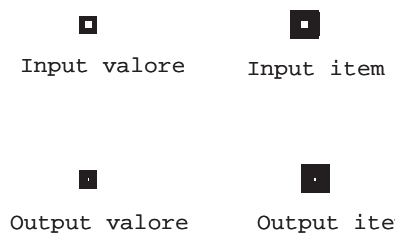


Figura 3.1: Connettori di Valori e Item

Le linee di connessione sono usate per specificare il flusso logico da un connettore ad un altro. Le linee doppie rappresentano le connessioni degli items e le singole le connessioni dei valori.

3.2.2 Esempio di modello

Nel seguente esempio consideriamo un sistema a singola coda e singolo server, come mostrato nella Figura 3.2.

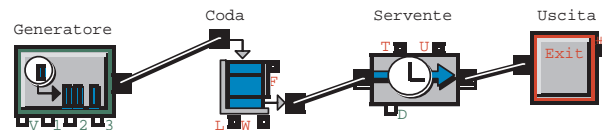


Figura 3.2: Un modello Singolo Server Singola Coda

Il blocco a sinistra è un generatore che periodicamente crea degli oggetti. In seguito, è collegata una coda FIFO che trattiene gli oggetti finché non arriva la richiesta dal blocco successivo, l'*activity Delay*, che rappresenta il server con una capacità di servizio limitata ad un oggetto per volta e per un periodo di tempo fissato. L'ultimo blocco è il blocco *Exit* che rimuove gli oggetti dal sistema.

Tempo di servizio casuale

Supponiamo che il tempo di servizio possa essere rappresentato da una specifica distribuzione casuale. Questo può essere modellato, attraverso la connessione dell'output di un blocco *Input Random Number* al connettore *delay* (D) del blocco che rappresenta il server, come mostrato in Figura 3.3.

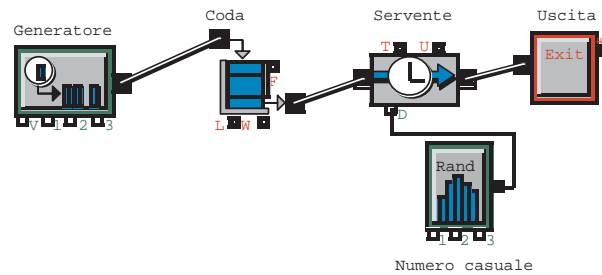


Figura 3.3: Modello con Tempo di Servizio Casuale

Ogni volta che un item entra nel blocco *Activity Delay* quest'ultimo chiede al blocco *Input Random Number* un nuovo tempo di servizio. Il generatore di numeri casuali, ad ogni richiesta, genera un nuovo tempo in accordo con la distribuzione definita nella sua finestra di dialogo.

Output grafici

Per mostrare graficamente le metriche del modello si può aggiungere un blocco *Discret Event Plotter*. In questo esempio (Figura 3.4) il *Plotter* visualizzerà graficamente il numero di items in attesa nella coda nel tempo. Per fare questo basta connettere il connettore di ingresso dei valori del *Plotter* con il connettore di uscita del valore della lunghezza della coda (L).

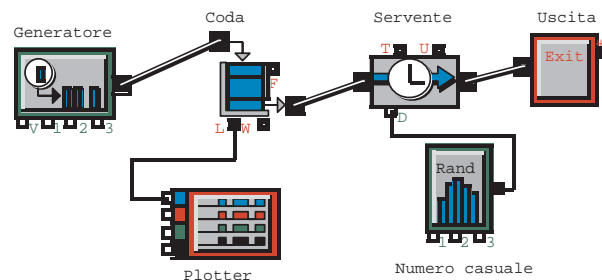


Figura 3.4: Discret Event Plotter Aggiunto al Modello

Attributi

Assumiamo che il nostro sistema offra due tipi di servizi A e B e che il tempo di servizio dipenda dal tipo di servizio. Per differenziare i due tipi di servizio richiesto possiamo aggiungere agli oggetti un attributo. Usando un blocco *Set Attribute* possiamo aggiungere un attributo chiamato “Tipo” ad ogni oggetto e il suo valore viene settato a 0 (A) oppure a 1 (B), in modo casuale, usando un nuovo generatore di numeri casuali come mostrato in Figura 3.5. Quando un oggetto lascia la coda e arriva al blocco servernte possiamo leggere l’attributo “Tipo” usando il blocco *Get Attribute* e convertire questo valore in un altro, che rappresenta la media del tempo del servizio richiesto, attraverso il blocco *Conversion Table*. Il valore della media può essere mandato in ingresso al blocco *Input Random Number* che è a sua volta collegato con il servernte.

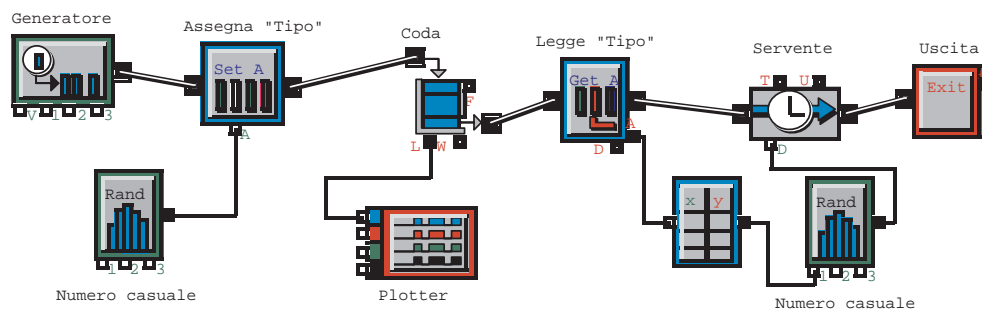
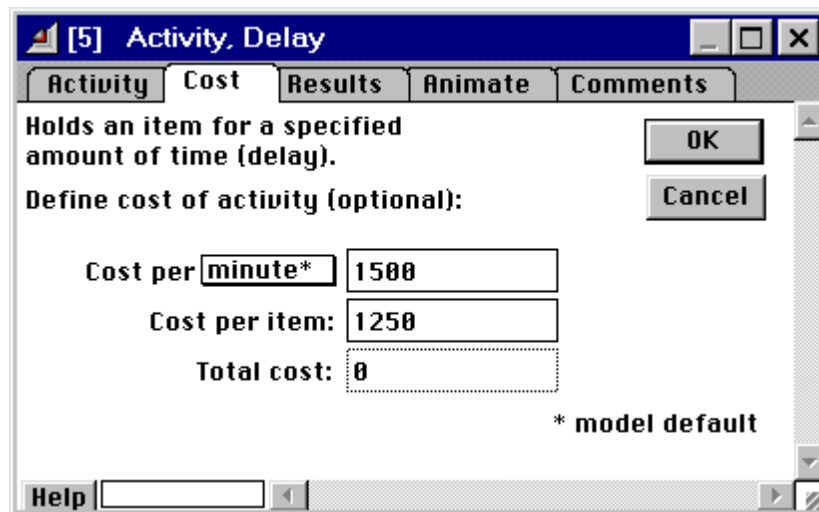


Figura 3.5: Assegnazione dell’Attributo “Tipo”

Calcolo dei costi delle attività

Ora che abbiamo costruito il modello di base, potremmo voler calcolare il costo medio per ogni servizio. Questo si può calcolare attraverso la finestra di dialogo del blocco *Activity Delay* (Figura 3.6) nella quale è possibile settare il costo di utilizzo del

servente per unità di tempo e/o il costo il costo per item. A partire da questi valori, viene calcolato in automatico il costo totale oppure il costo medio per ogni tipo di servizio.



Field	Value
Cost per minute*	1500
Cost per item:	1250
Total cost:	0

Figura 3.6: Tabella dei Costi del Blocco Activity Delay

Comunicazione Interprocesso

Il termine comunicazione interprocesso descrive l'azione di comunicazione di due applicazioni e di condivisione di dati con un'altra. Supponiamo di voler usare, come dati in ingresso per un modello Extend, dati provenienti da un foglio di lavoro di Excel. Dopo aver copiato i dati dal foglio di lavoro si può selezionare un parametro in Extend e scegliere *Past Link* dalla voce del menù *Edit*. I dati saranno copiati nei campi dei parametri e sarà creato un collegamento dinamico tra il parametro e le specificate celle di Excel. Se nel foglio di lavoro cambiano i valori, automaticamente saranno aggiornati in Extend. Un collegamento di questo tipo può essere creato anche anche per l'output.

I blocchi e le funzioni sono forniti di capacità di comunicazione interprocesso aggiuntive come esecuzione di macro oppure per controllare Extend da un'altra applicazione.

Analisi della sensitività

Supponiamo di voler determinare la sensitività del nostro sistema alla variazione del tempo di interarrivo. Per fare questo possiamo eseguire l'analisi della sensitività sul parametro della media degli interarrivi del blocco generatore di items. Selezionando il parametro e scegliendo *Sensitize Parameter* dal menù *Edit* possiamo definire come deve variare il parametro da un run ad un altro. I parametri della simulazione, come numero di run e il tempo di fine simulazione, possono essere specificati nella finestra di dialogo *Setup* sotto la voce *Run* del menù. Facendo variare i tempi di interarrivo e confrontando i risultati di diverse esecuzioni della simulazione possiamo capire quanto il nostro sistema è sensibile alla variazione dei tassi di arrivo.

Risultati del modello

Una volta che è completata l'esecuzione della simulazione, i risultati della simulazione sono riportati all'interno dei blocchi. Aprendo le finestre di dialogo dei vari blocchi, si possono osservare le informazioni relative al run (Figura 3.7).

3.2.3 Personalizzazione di Extend

La discussione fatta finora ha illustrato la natura altamente grafica e interattiva di Extend. Inoltre, può anche essere personalizzato in base alla specifica applicazione. Possono essere usate interfacce, componenti e grafici per adattare il modello ad una

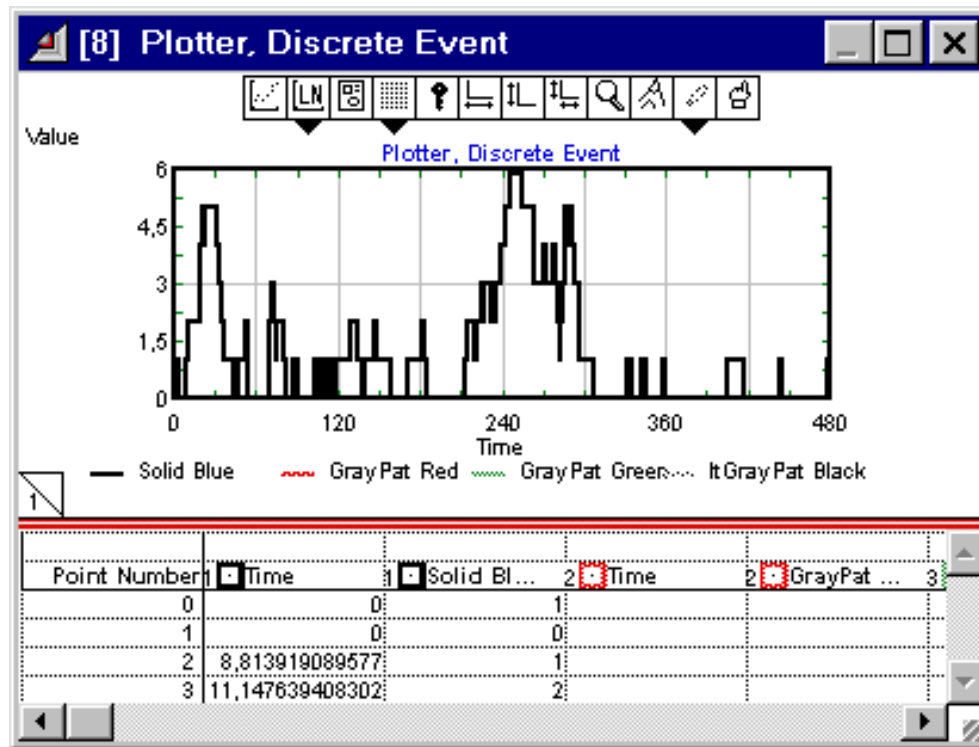


Figura 3.7: Grafico della Lunghezza della Coda

specifica area di applicazione. L'aspetto più evidente di un modello è l'interfaccia utente e attraverso questa il modellista può rendere il modello più aderente al sistema che si sta rappresentando oltre che più amichevole per il suo futuro utente. Le sezioni seguenti descriveranno alcuni degli strumenti di Extend che permettono di personalizzare i modelli.

Animazione

L'animazione è un potente strumento di presentazione e di *debugging* che può far crescere fortemente la chiarezza del modello. In Extend le icone animate si muovono da un blocco all'altro rappresentando il flusso degli oggetti nel sistema. Gli utenti possono scegliere le icone animate predefinite oppure crearne di nuove. Inoltre, possono essere aggiunte animazioni personalizzate, *flashing pictures*, testo, indicatori di livello

e filmati Quick Time.

Modellazione gerarchica

La gerarchia permette ai modelli di essere suddivisi in componenti logiche o sottomodelli. Una singola icona descrittiva può rappresentare un sottomodello. Aprendo un blocco gerarchico, comparirà una nuova finestra nella quale sarà visualizzato il sottomodello. Questo può semplificare notevolmente la rappresentazione di un modello e permette all'utente di nascondere o mostrare i dettagli del modello in modo contestuale. Selezionando un gruppo di blocchi e scegliendo *Make Selection Hierarchical* dal menù *Model*, si può incapsulare una parte del modello in un blocco gerarchico. Questo blocco può essere salvato in una libreria e usato nuovamente in altri modelli. Il numero di livelli gerarchici è illimitato.

Clonazione di finestre di dialogo e il *notebook*

Come sottolineato nelle sezioni precedenti i parametri di ingresso e uscita associati al modello si trovano nelle finestre di dialogo degli appropriati blocchi. Mentre questo fornisce un'associazione intuitiva tra le metriche del sistema e i costrutti usati per modellarle, può essere difficile cercare un dato specifico quando si lavora con modelli molto grandi con diversi livelli di gerarchia. Un modo efficace per risolvere questo problema è quello di usare il *notebook* e di clonare al suo interno le interfacce dei dati di interesse. Con il *notebook* si può creare un'interfaccia personalizzata evidenziando parametri critici e risultati. Il *notebook* è una finestra separata associata ad ogni modello. Inizialmente il *notebook* è un foglio di lavoro vuoto nel quale si possono inserire testo, figure e cloni. I cloni sono collegamenti diretti alle finestre di dialogo dei parametri e possono essere creati selezionando *Cloning Tool* dalla barra degli strumenti e trascinando l'oggetto dalla finestra di dialogo al *notebook*. Una volta che un clone è

creato, ogni modifica al clone si riflette immediatamente nel blocco e viceversa. Perciò non è necessario accedere alla finestra di dialogo del blocco per modificare un parametro di ingresso o per osservare le modifiche di un risultato.

Sviluppo di blocchi

L'ambiente di sviluppo dei blocchi è una delle più potenti caratteristiche di Extend. Mentre la maggior parte degli utenti di Extend trova che i blocchi precostruiti siano sufficienti per le proprie necessità, l'ambiente di sviluppo dei blocchi fornisce uno strumento agli utenti per espandere le capacità di modellazione in modo da poter elaborare compiti inusuali o altamente specializzati. L'architettura aperta di Extend permette di accedere alla struttura di ogni blocco costruito con Extend. Aprendo la struttura è possibile modificare l'icona, la finestra di dialogo, il testo dell'*help* e il codice del blocco. Si possono modificare l'interfaccia e la funzionalità di ciascun blocco oppure creare dei blocchi nuovi.

ModL è il linguaggio di programmazione che definisce la logica di ogni blocco di Extend. Il linguaggio è simile al linguaggio di programmazione C. Questo permette agli utenti di Extend, che hanno familiarità con il C, di essere in grado di usare facilmente il ModL. Ogni blocco ha al suo interno una serie di *message handlers* e procedure che forniscono i metodi di interfaccia primari con gli altri blocchi. I *message handlers* sono funzioni che vengono invocate in risposta ad eventi del sistema. Le procedure sono funzioni che vengono chiamate attraverso i *message handlers*. In Extend è disponibile un buon insieme di procedure di supporto. Alcuni esempi sono: operazioni matematiche (min, max), operazioni su matrici, integrazione, animazione, manipolazione di stringhe e così via. Una finestra di dialogo può includere *check boxes*, *radio buttons*, bottoni, caselle di testo, campi numerici e altro. Il programmatore può decidere che ciascuno di questi oggetti possa essere solo visualizzabile o modificabile. Inoltre, ogni parametro

può essere trattato come una variabile di un programma. Un esempio è il blocco *Input Random Number* dove i nomi dei parametri vengono visualizzati in base alla distribuzione che è stata scelta.

3.2.4 Librerie Standard, BPR e Manufacturing

Le librerie standard di Extend includono funzioni per la modellazione della simulazione discreta, visualizzazione di risultati, calcoli generici, progetti elettronici, comunicazioni interprocesso e utilità. Per la modellazione della simulazione discreta le librerie standard più comunemente usate sono la *Discret Event*, *Generic* e *Plotter*. In aggiunta, come *optional*, possono essere incluse le librerie *Business Process Reengineering* (BPR) e *Manufacturing* (MFG). Queste sono librerie per la simulazione discreta che forniscono blocchi aggiuntivi da usare congiuntamente alle librerie standard.

La libreria BPR è progettata per modellare processi aziendali. Ogni blocco della libreria BPR usa una terminologia economica. Alcuni esempi di blocco sono: Transazione, Decisione, Lavoro. I blocchi sono rappresentati con simboli standard dei diagrammi di flusso che rendono i modelli facilmente comprensibili da coloro che non hanno familiarità con Extend.

La libreria Manufacturing è orientata alla modellazione industriale. Essa include costrutti come Nastri Trasportatori, AGV, lottizzazione flessibile e così via.

3.3 Implementazione del modello del Centro Prelievi

In questa sezione descriviamo l'implementazione del modello di simulazione del Centro Prelievi realizzata con il *tool* di simulazione Extend descritto nella sezione precedente.

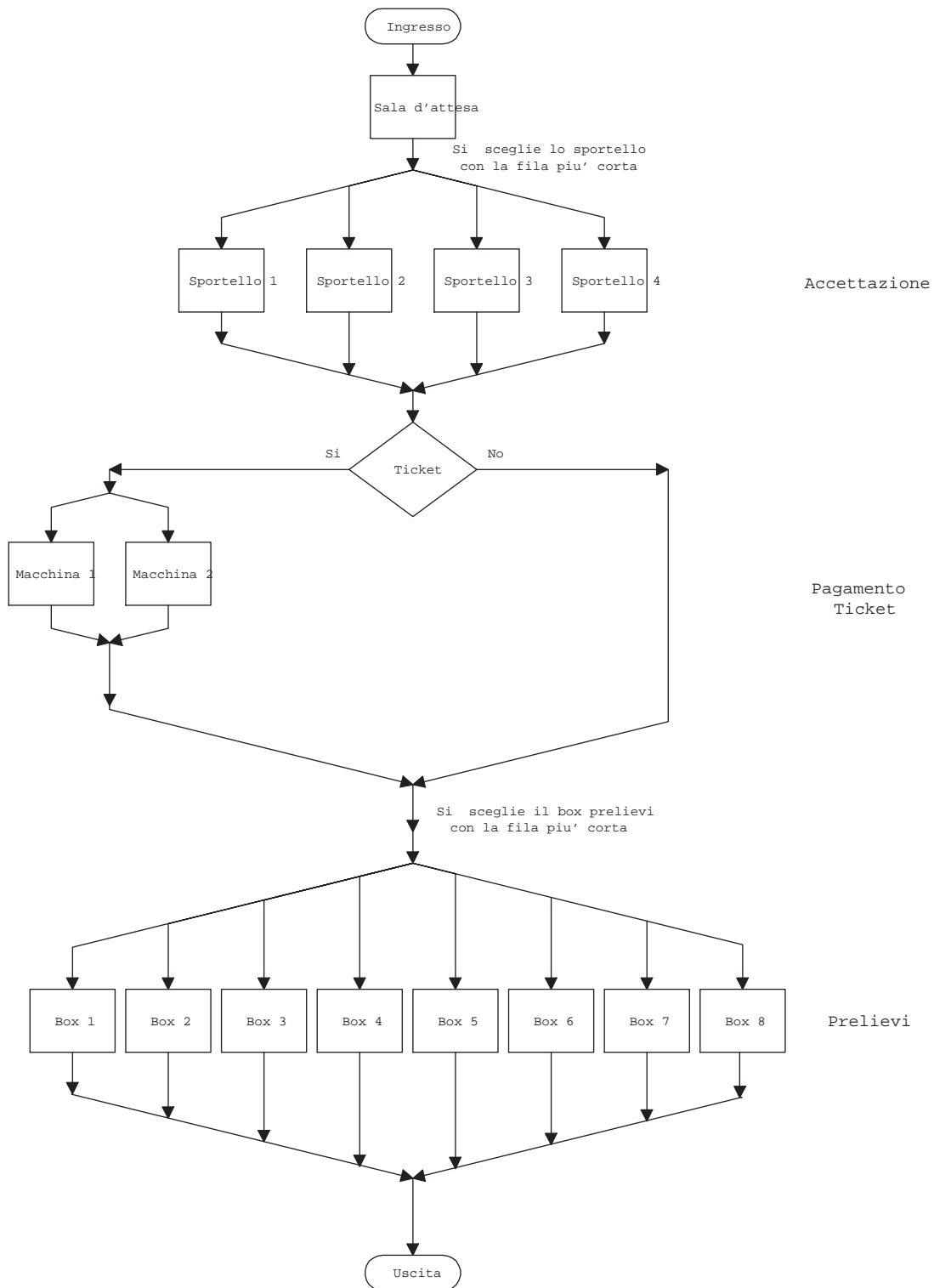


Figura 3.8: Rappresentazione grafica del modello

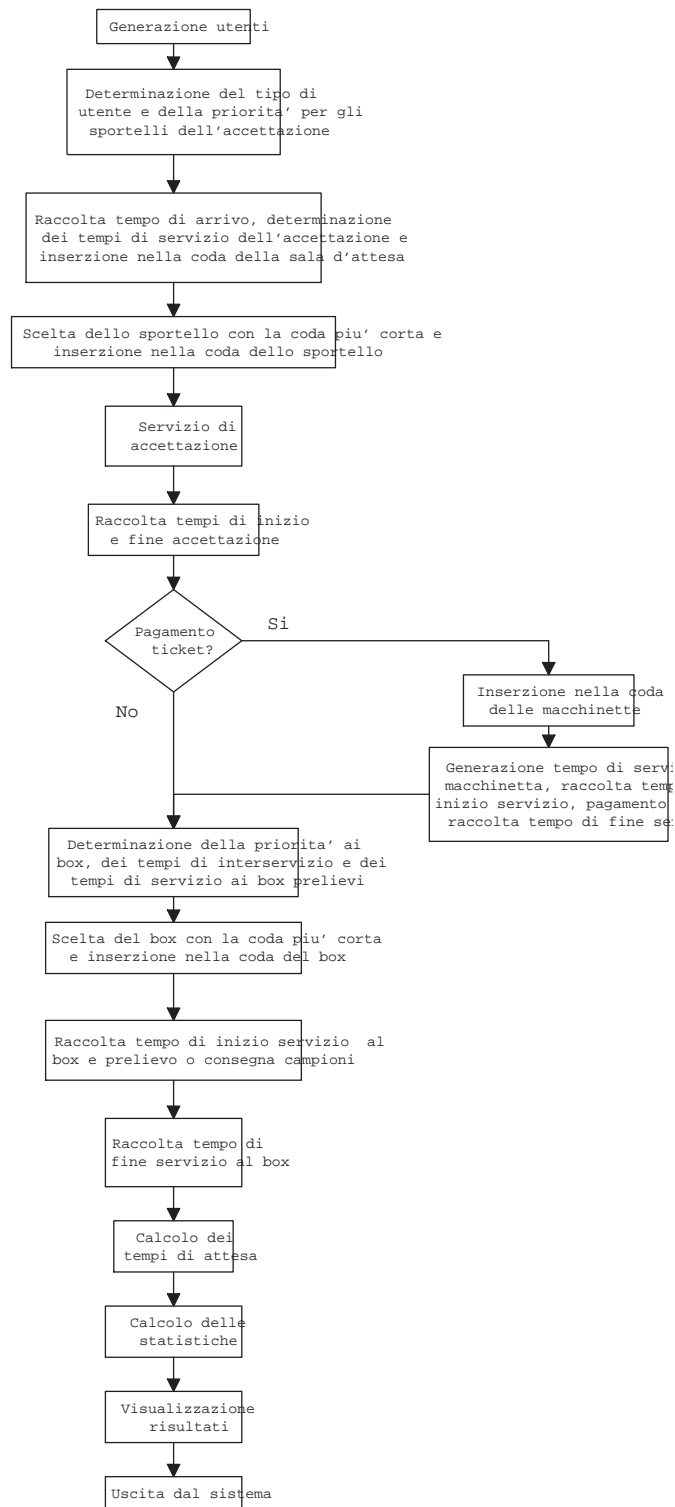


Figura 3.9: Diagramma a blocchi del modello di simulazione

La realizzazione del modello è stata fatta seguendo una metodologia *bottom-up*, cioè abbiamo implementato separatamente i sottosistemi che in seguito sono stati integrati tra loro per formare il sistema globale. Nella fattispecie, abbiamo individuato tre sottosistemi principali: Accettazione, Pagamento Ticket e Prelievi come descritto in Figura 3.8 che mostra, oltre alla struttura del sistema, anche i possibili percorsi di un generico utente. La descrizione in linguaggio naturale è stata data nella Sezione 1.1. La struttura del programma di simulazione è descritta nella Figura 3.9 nella quale ogni blocco descrive una o più azioni svolte dal simulatore. Nelle sezioni seguenti descriveremo in dettaglio l'implementazione dei tre sottosistemi principali.

3.3.1 Accettazione

In questo sottosistema la prima fase è quella della generazione degli utenti, realizzata nel blocco denominato “Processo di arrivo” (Figura 3.10).

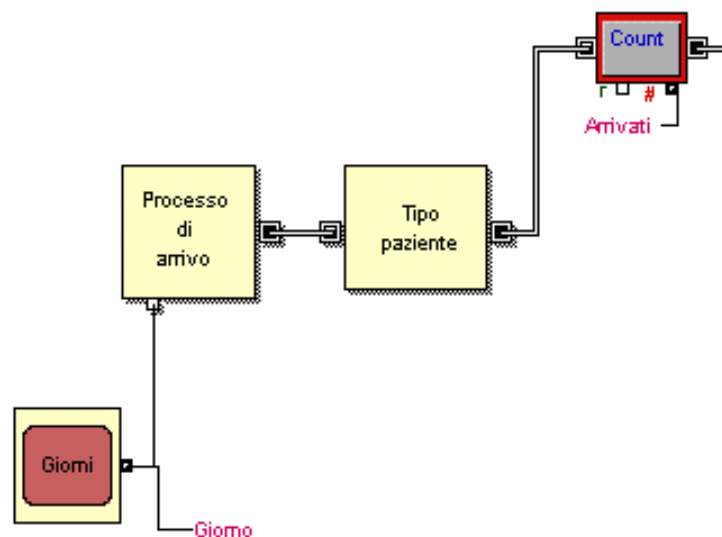


Figura 3.10: Generazione utenti

Quest'ultimo è un blocco gerarchico e la sua struttura è mostrata in Figura 3.11. I

due blocchi gialli, nella parte superiore della figura, sono i generatori di items. Quello più in alto genera gli utenti, che sono già presenti al momento dell'apertura del Centro, secondo una distribuzione di probabilità uniforme tra 10 e 20. L'altro genera gli items con tempi di interarrivo in accordo con la Tabella 2.6.

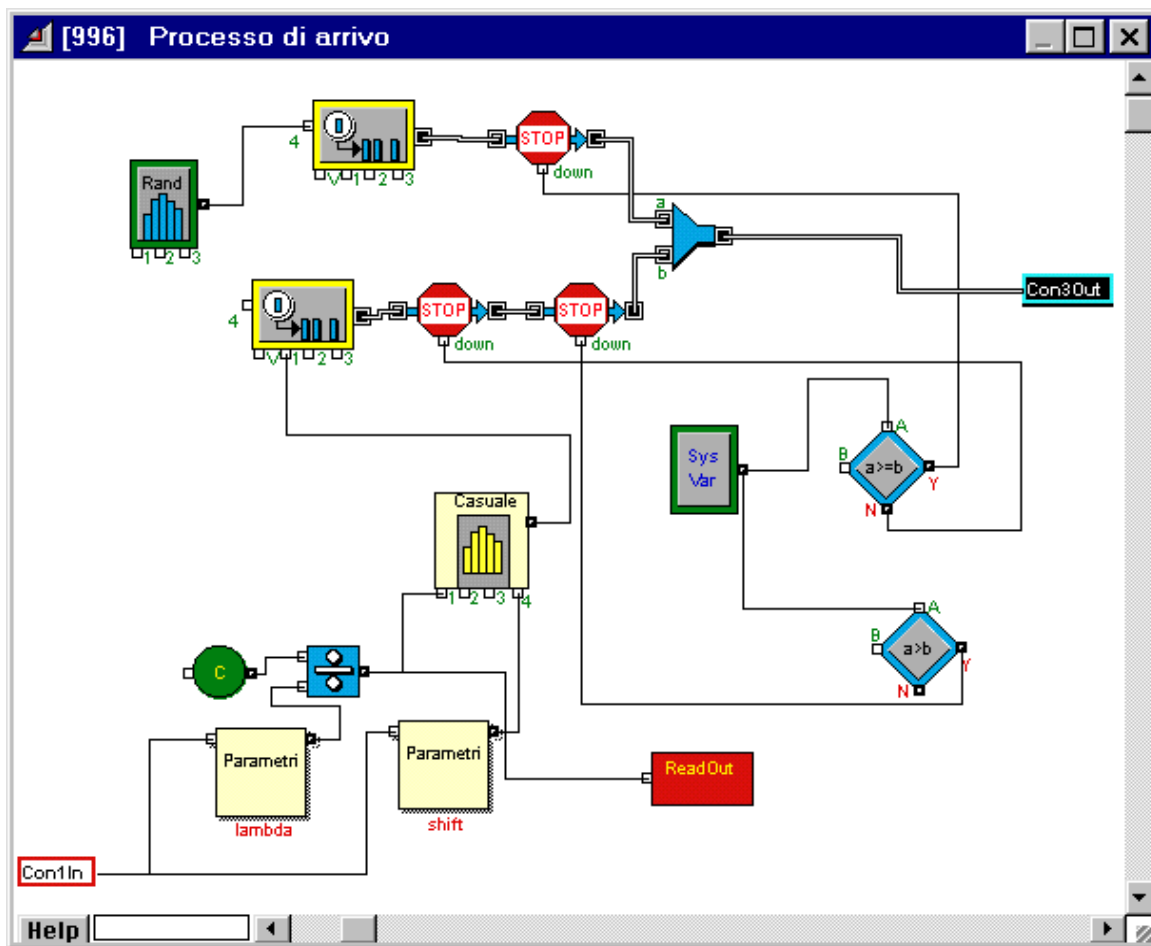


Figura 3.11: Processo di arrivo

I parametri da passare al generatore di numeri casuali, che determina i tempi di interarrivo, sono scelti in base al giorno della settimana che viene ricevuto in input attraverso il connettore “Con1In” al quale arriva il segnale dal blocco “Giorni” della Figura 3.10. Per soddisfare esigenze specifiche del modello è stato necessario modificare

il codice di alcuni blocchi delle librerie fornite insieme al programma e in alcuni casi abbiamo costruito dei blocchi *ex novo* (blocchi colorati di giallo).

Successivamente (Figura 3.10), si determina il tipo di paziente in accordo alla Tabella 2.7.

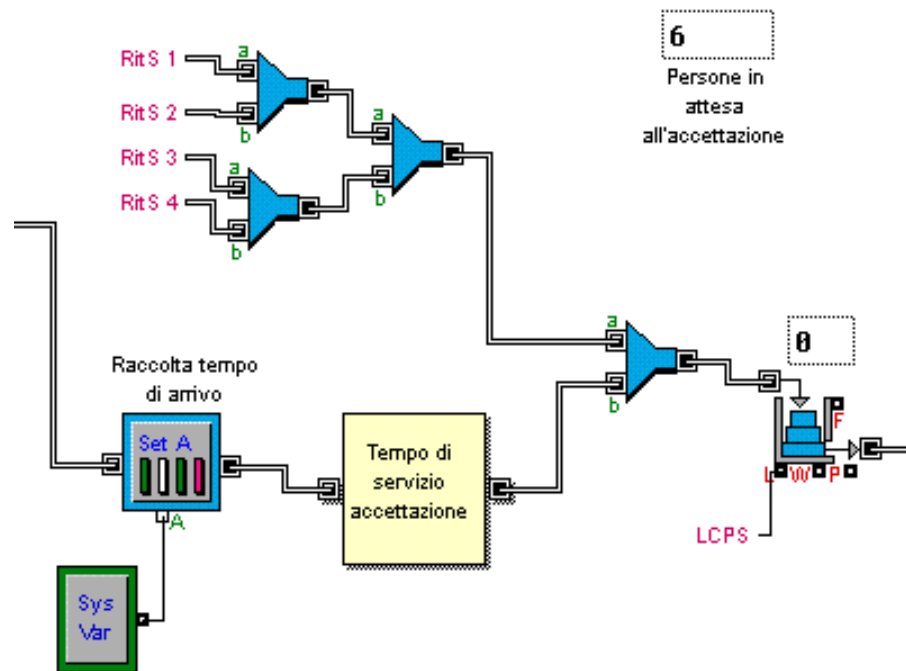


Figura 3.12: Tempi di servizio e coda sala d'attesa

Una volta che sono stati generati gli utenti ed è stato determinato il loro tipo, si passa alla fase successiva che consiste nel raccogliere il loro tempo di arrivo che viene assegnato all'oggetto come attributo. Questo attributo verrà usato per calcolare alcuni valori di interesse (tempo di attesa all'accettazione, tempo di permanenza). Dalla Figura 3.12 si può vedere che subito dopo vengono stabiliti i tempi di servizio all'accettazione per i pazienti prima che questi vengano inseriti nella coda a priorità che rappresenta la sala d'attesa. Nella parte in alto a destra della figura è presente una casella nella quale è visualizzata la somma del numero di persone in sala d'attesa e in coda agli sportelli d'accettazione. I connettori di ingresso di items Rit S 1, ...,

Rit S 4 servono per ridistribuire le persone nelle varie code quando viene chiuso uno sportello che ha una coda non vuota. Prima della redistribuzione si dà a questi items una priorità più alta rispetto alle persone che ancora stanno in sala d'attesa.

La Figura 3.13 rappresenta la struttura del blocco gerarchico “Tempo di servizio accettazione” della figura precedente. Il tempo di servizio viene assegnato ai vari items come attributo usando un generatore di numeri casuali differente a seconda del tipo di utente considerato. Le distribuzioni di probabilità di ogni generatore derivano dalla Tabella 2.5.

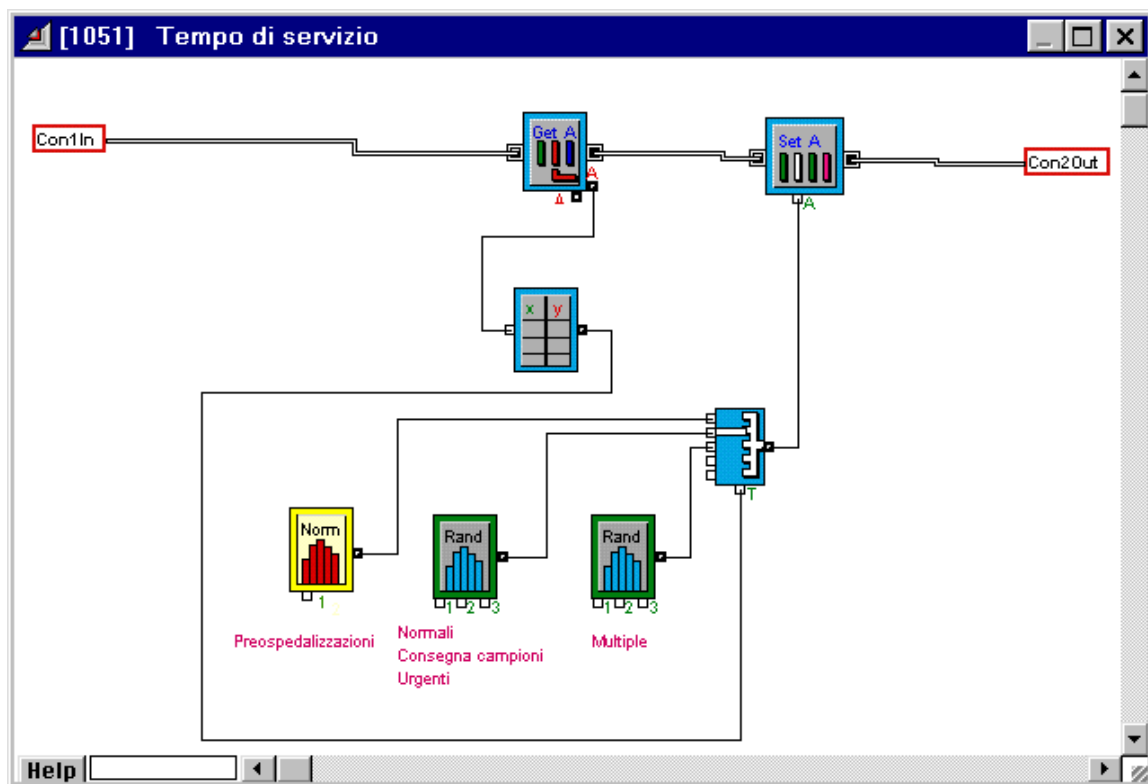


Figura 3.13: Tempo di servizio accettazione

Il modulo del generatore di numeri casuali in basso a sinistra, relativo al tempo di servizio per le preospedalizzazioni, è stato costruito per realizzare una distribuzione normale tronca che non era fornita dalle librerie standard.

Nella figura seguente (Figura 3.14) l'instradatore (blocco gerarchico) sceglie la coda più corta, tra quelle degli sportelli aperti, nella quale viene inserito l'item che rappresenta l'utente considerato. Ricordiamo che la capacità massima delle code degli sportelli, gestite con politica FIFO, è di 5 persone.

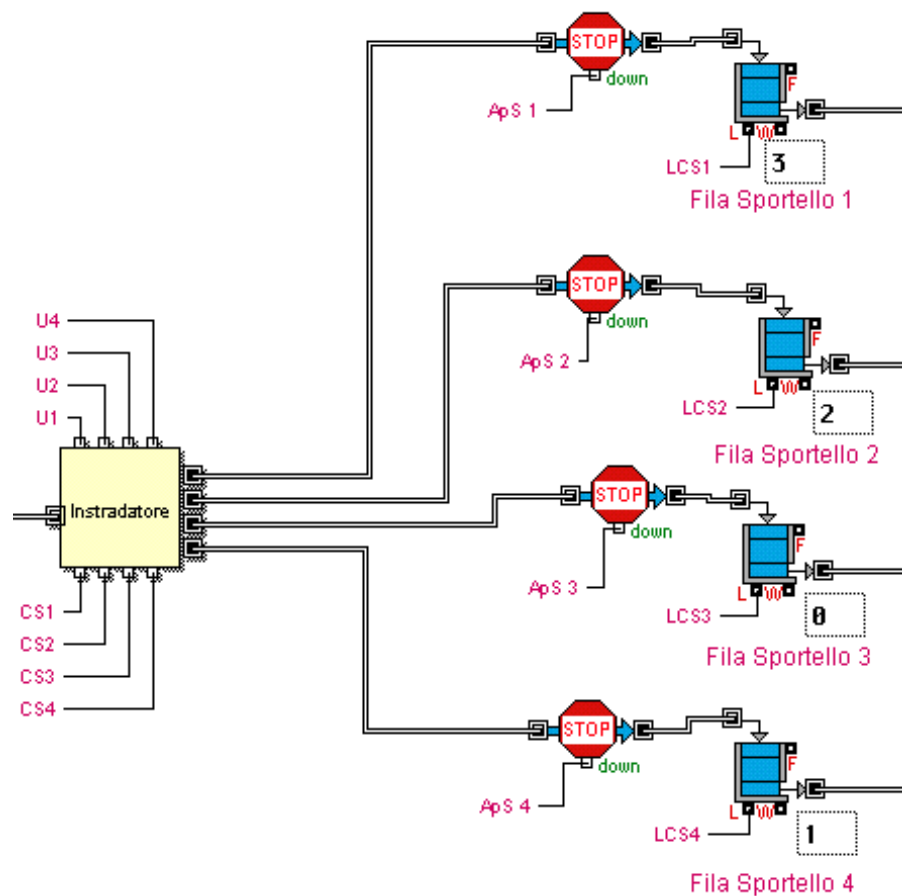


Figura 3.14: File sportelli accettazione

Quando uno sportello è disponibile l'utente viene estratto dalla coda corrispondente e dopo un periodo di transizione (Intersportello) può iniziare la fase di accettazione. I tempi di intersportello seguono una distribuzione empirica ricavata dai dati raccolti. I quattro sportelli della accettazione sono rappresentati nella Figura 3.15. I blocchi nella parte destra della figura modellano l'attività accettazione e trattengono gli items che

gli passano attraverso per un periodo di tempo simulato pari al valore dell'attributo "Tempo di servizio accettazione" che è stato assegnato all'oggetto in precedenza. In caso di chiusura anticipata di uno sportello, i blocchi a sinistra si occupano di riassegnare la priorità agli utenti della sua coda e di inviarli ai connettori Rit S 1, ..., Rit S 4 della Figura 3.12.

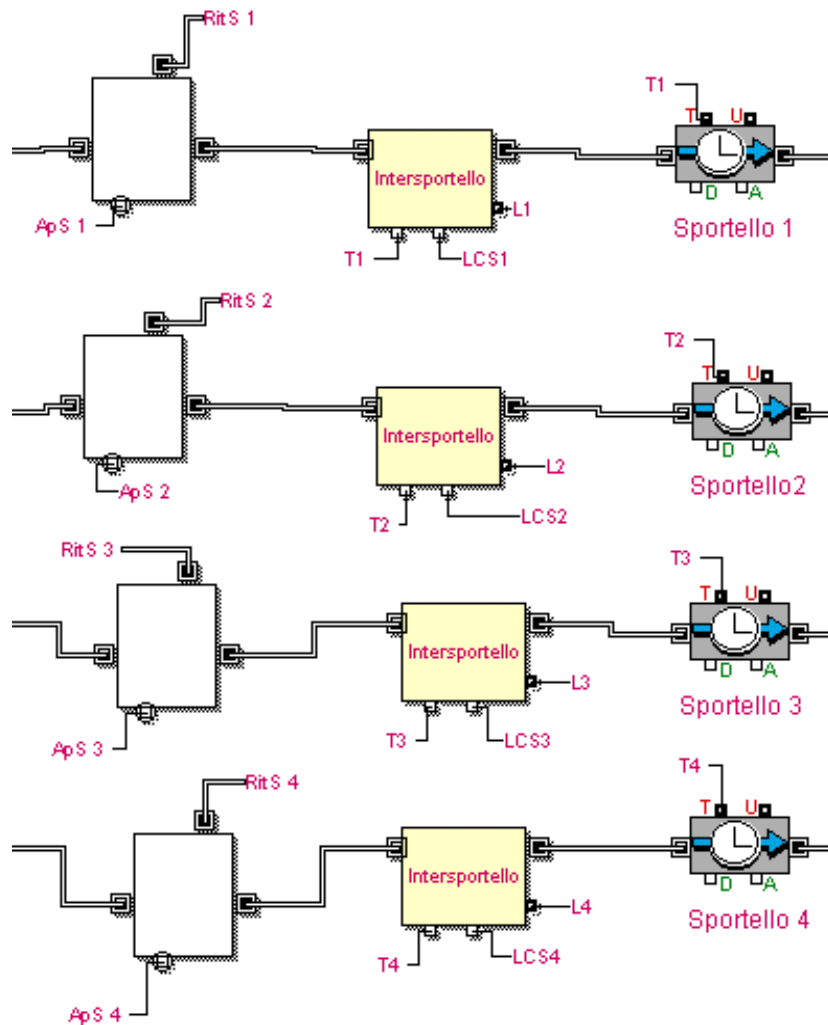


Figura 3.15: Sportelli accettazione

Appena finito il servizio di accettazione il generico utente attraversa uno dei quat-

tro percorsi della Figura 3.16 nei quali viene calcolato il tempo di inizio accettazione dell'utente facendo una sottrazione tra il tempo simulato corrente ed il valore dell'attributo "Tempo di servizio accettazione". Il risultato è quindi assegnato a un attributo dell'oggetto. A questo punto, all'interno dell'instradatore, viene assegnato anche il tempo di fine accettazione e la fase di accettazione può essere considerata conclusa. L'instradatore controlla anche l'attributo "Tipo paziente" dell'item e nel caso in cui debba pagare il ticket lo indirizza verso le macchine automatiche di riscossione, in caso contrario lo invia ai box.

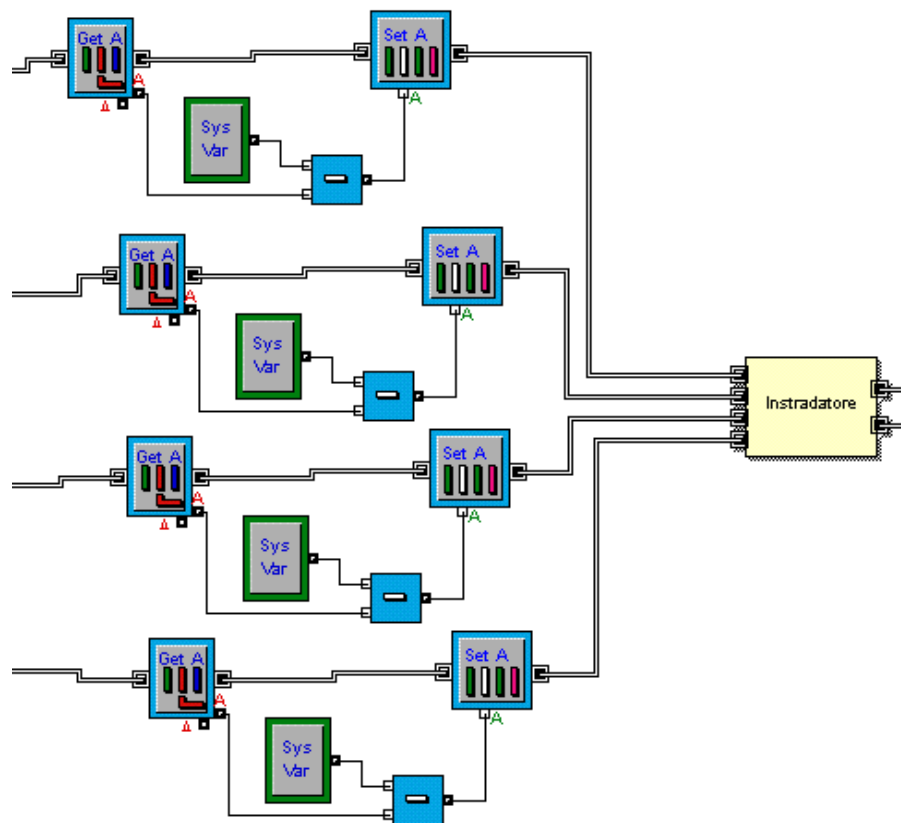


Figura 3.16: Raccolta tempi di inizio accettazione

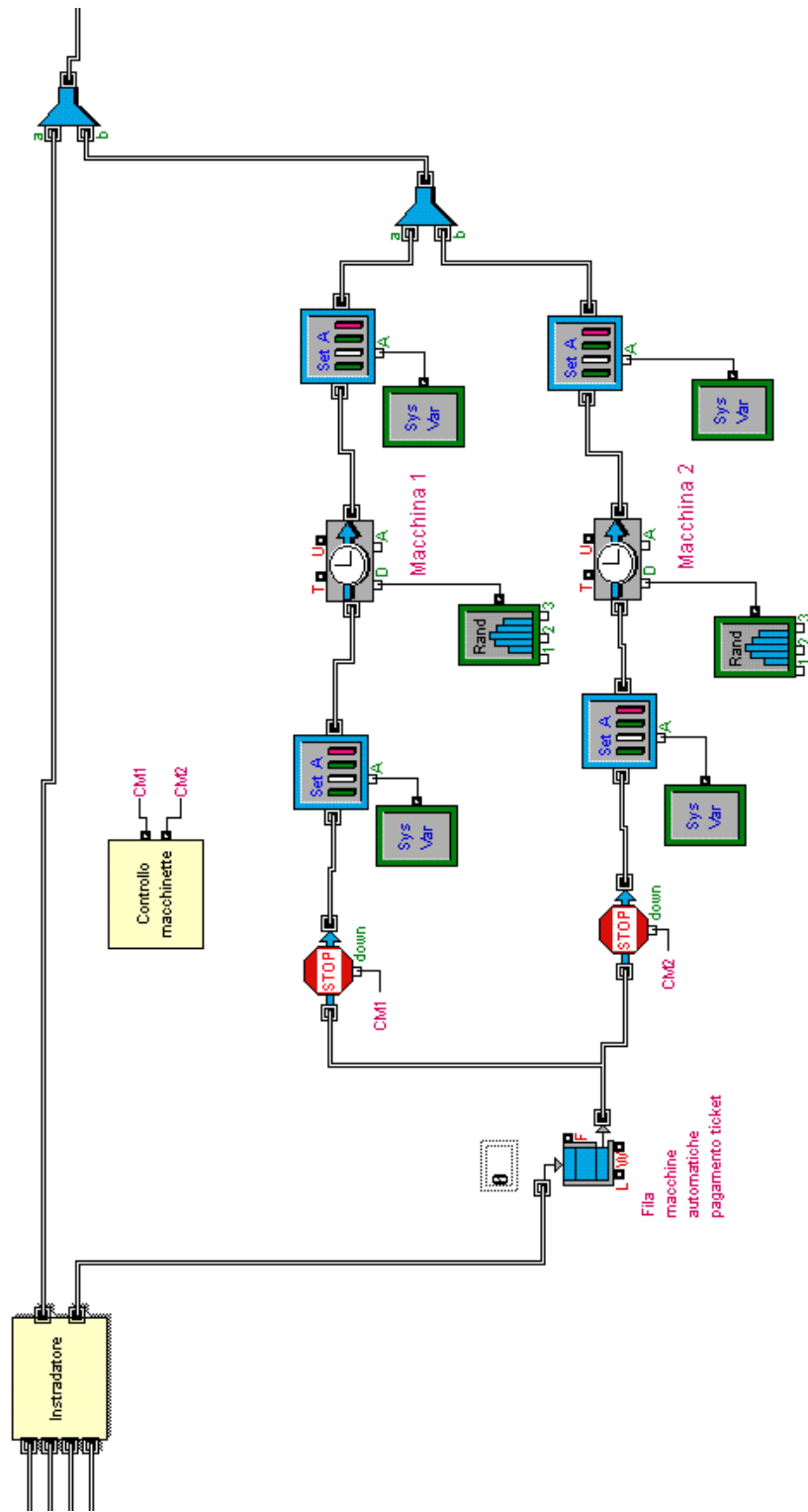


Figura 3.17: Pagamento ticket

3.3.2 Pagamento ticket

Come già accennato nella sezione precedente, gli utenti che devono pagare il ticket si dirigono a fare la fila alle macchine automatiche di riscossione. La Figura 3.17 illustra tale sottosistema. Gli utenti vengono inseriti nella coda FIFO comune alle due macchine e appena una delle due è disponibile l'item viene estratto dalla coda, viene assegnato l'attributo "Inizio pagamento ticket" e inizia il servizio. Il servizio è simulato dal blocco *Activity Delay* e il tempo di servizio viene passato dal generatore di numeri casuali ad esso collegato. I tempi di servizio seguono una legge di probabilità di tipo Gamma (Tabella 2.5). Una volta terminato il pagamento si raccoglie il tempo di fine servizio e l'item viene indirizzato alla zona dei box prelievi.

3.3.3 Prelievi

Prendiamo ora in considerazione il sottosistema dei box prelievi. Gli utenti provengono dalle macchinette di riscossione ticket oppure direttamente dall'accettazione.

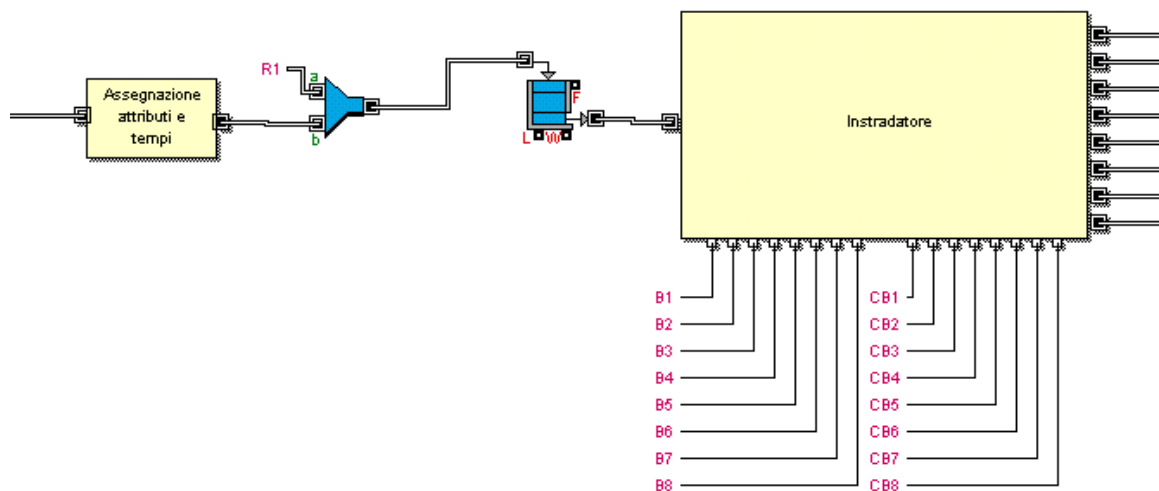


Figura 3.18: Tempi di servizio box

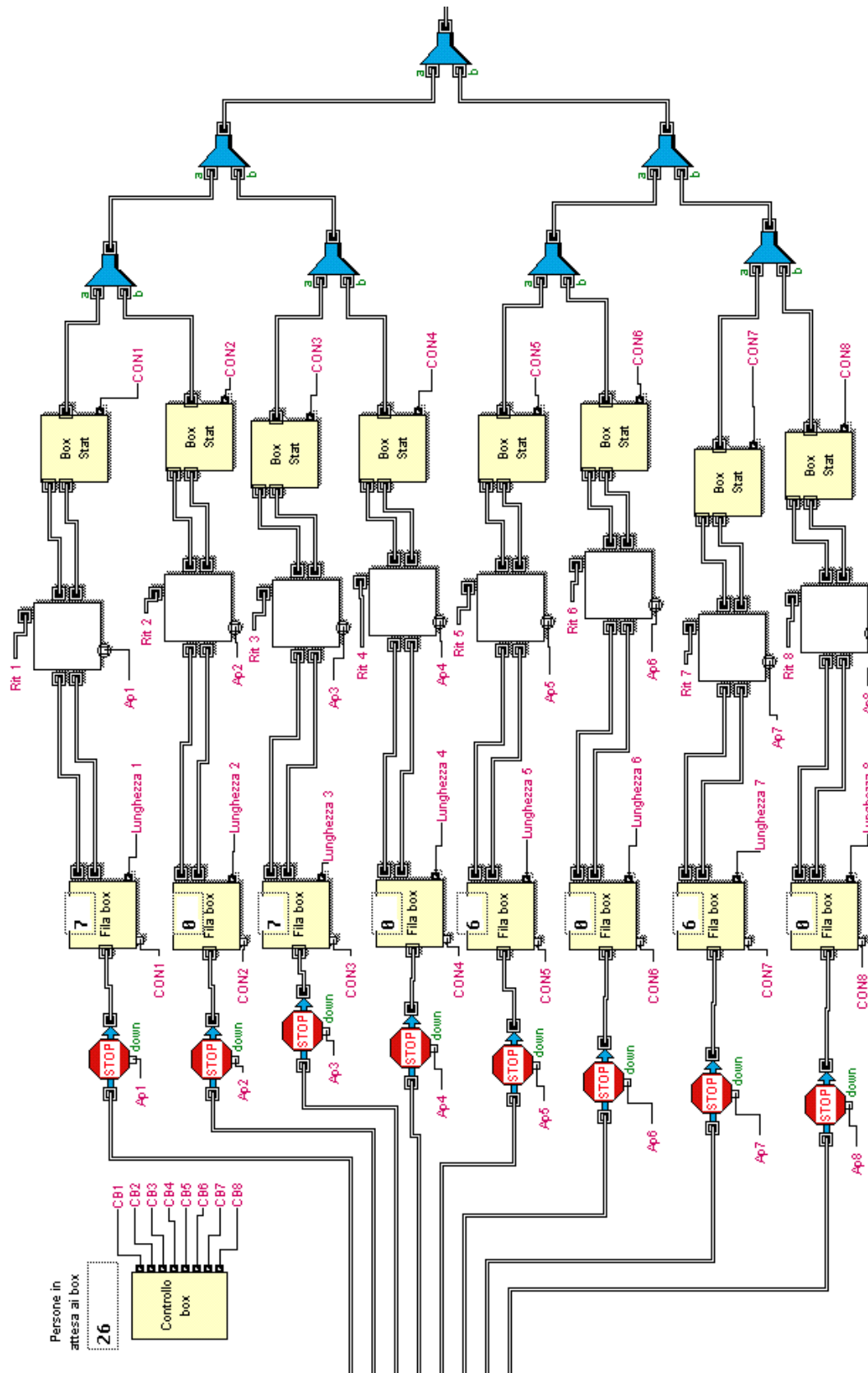


Figura 3.19: Zona prelievi

Prima che l'instradatore scelga il box aperto con la fila più corta (Figura 3.18) ad ogni item vengono assegnati i tempi di interservizio e di servizio ai box oltre al valore della priorità per la fila.

La Figura 3.19 illustra gli otto box prelievi con le relative file d'attesa. I blocchi "Fila box" sono dei blocchi gerarchici la cui struttura è quella di Figura 3.20. In questo blocco per prima cosa si controlla l'attributo "Tipo paziente" dell'item e se deve solo consegnare i campioni viene indirizzato alla fila 3. Questo perché chi deve solo consegnare i campioni può farlo anche se il box è occupato per un prelievo, per cui ci sono due file: una per le consegne dei campioni e una per i prelievi. La fila 1 è una fila a priorità, mentre la fila 2 è una fila con capacità unitaria nella quale è inserito il paziente in testa alla coda, che non viene mai superato da un altro utente con priorità superiore.

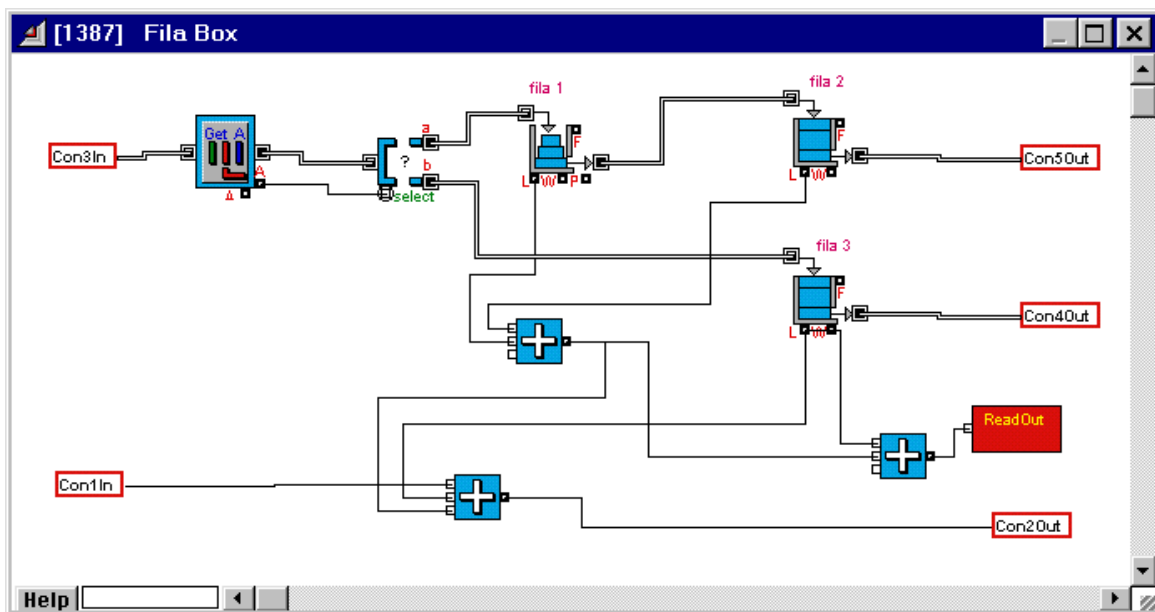


Figura 3.20: Fila box

Sempre facendo riferimento alla Figura 3.19, vediamo che il blocco successivo è quello che si occupa di distribuire gli utenti che sono in fila in un box nel caso in cui

quest'ultimo chiuda per qualche motivo. Dopodiché troviamo il box vero e proprio. Anche questo è un blocco gerarchico (Figura 3.21). Qui troviamo due percorsi: quello più in basso è per le consegne dei campioni, mentre quello in alto per tutti gli altri. In entrambi i percorsi viene settato l'attributo relativo alla raccolta del tempo di inizio del servizio al box.

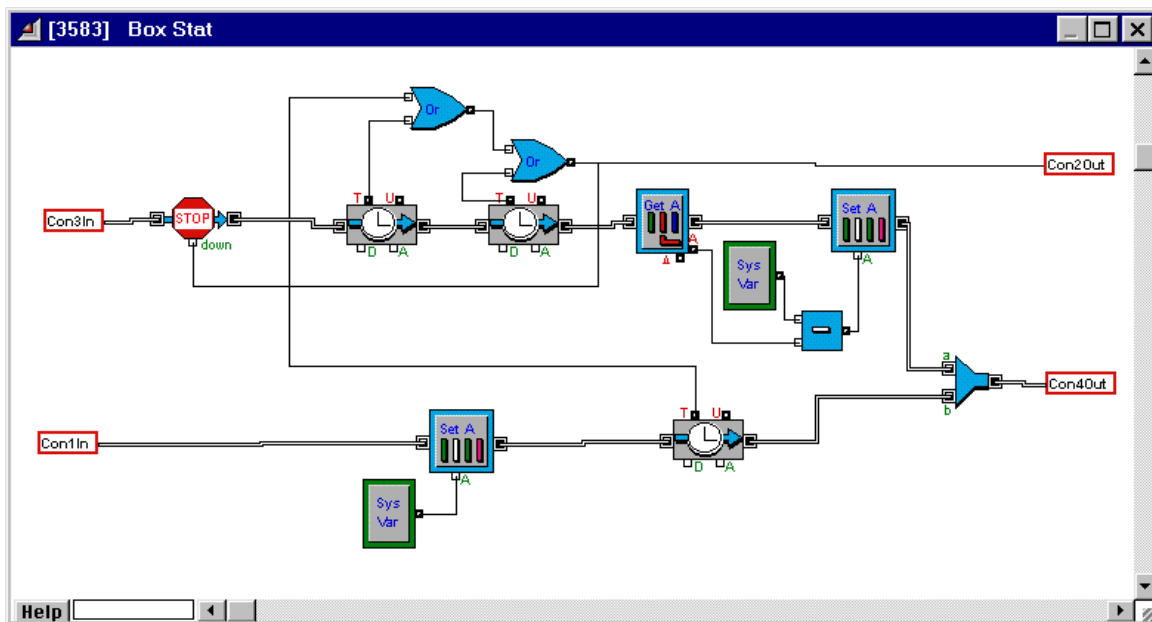


Figura 3.21: Box prelievi

Una volta terminato il servizio al box l'utente può uscire dal sistema. Prima della sua uscita rimane da settare il tempo di uscita, calcolare i tempi di attesa (Figure 3.22 e 3.23) oltre al calcolo delle loro statistiche (media, deviazione standard e varianza) sia per la simulazione corrente che per simulazioni multiple (Figura 3.24). Alla fine di ogni run vengono memorizzati in un file il numero delle persone servite e le statistiche dei tempi d'attesa.

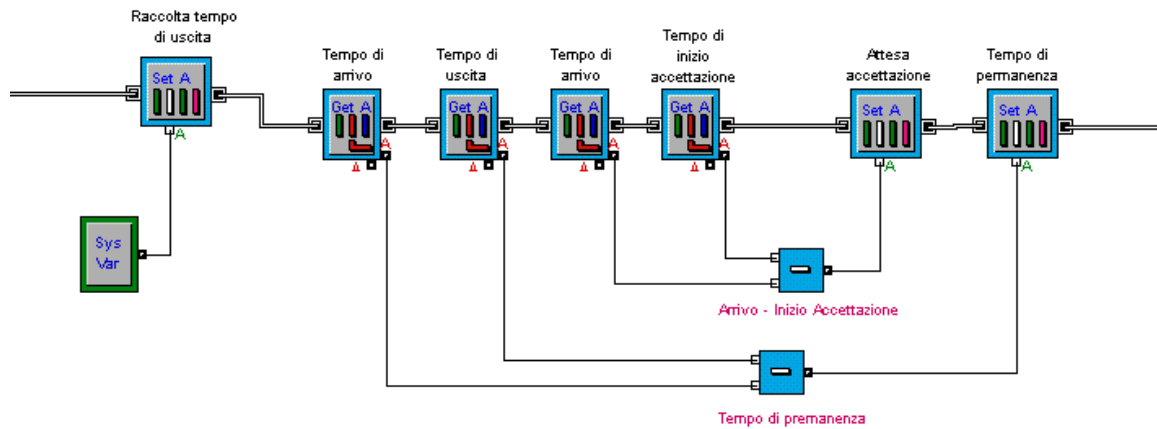


Figura 3.22: Calcolo tempi di attesa accettazione e permanenza

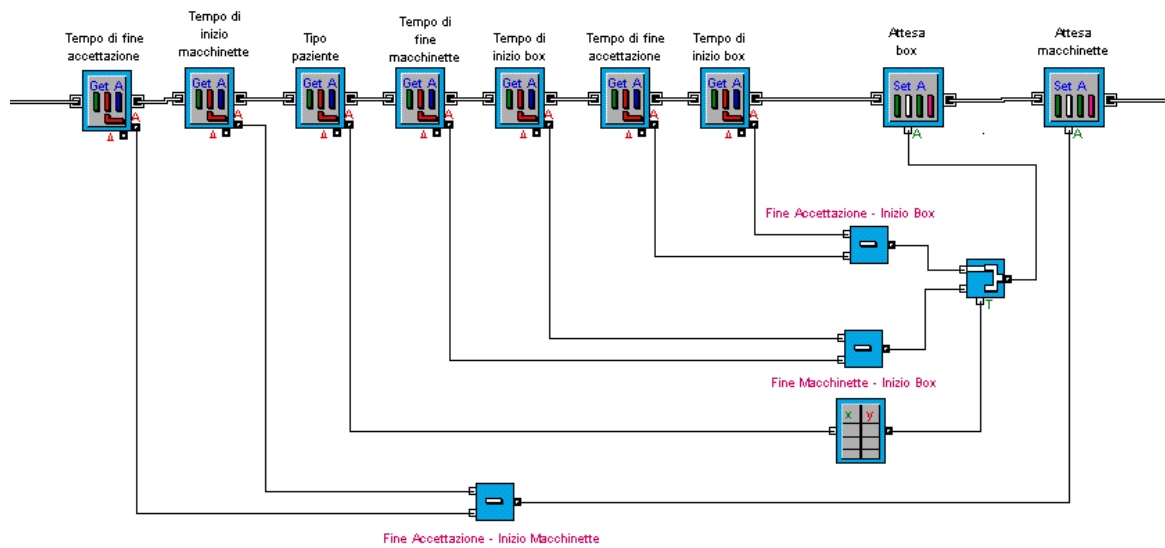


Figura 3.23: Calcolo tempi di attesa box e macchinette

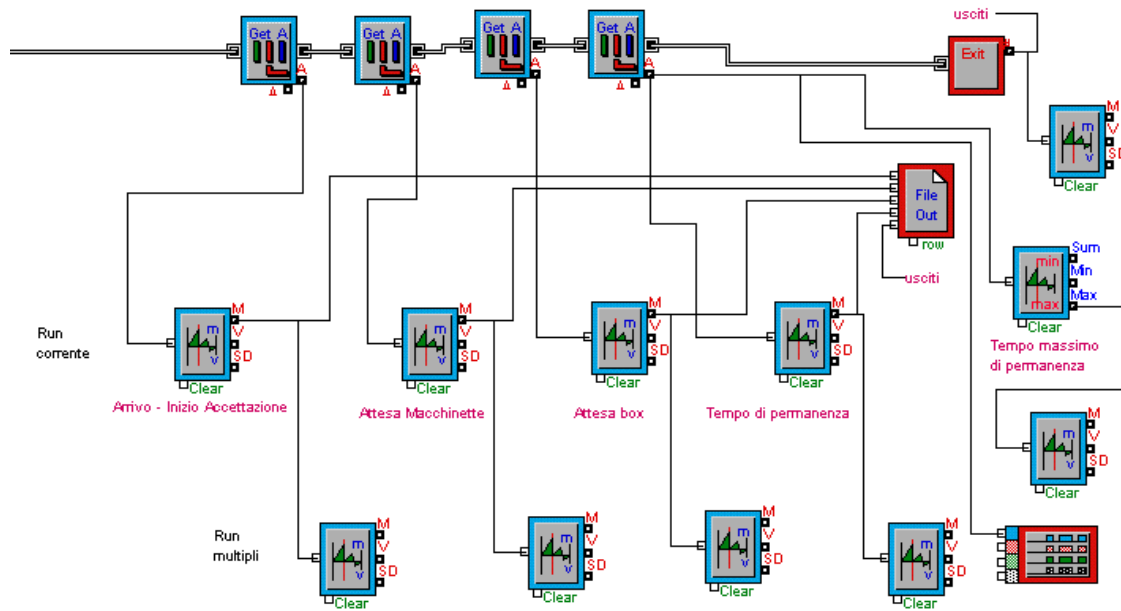


Figura 3.24: Calcolo delle statistiche e uscita dal sistema

3.3.4 Interfaccia utente

L'interfaccia è strutturata in due sezioni. La prima riguarda il controllo dei parametri di ingresso come giorno della settimana, apertura degli sportelli di accettazione e box prelievi e accensione delle macchine per il pagamento del ticket (Figura 3.25). Questi sono i parametri che possono essere modificati da chi si occupa della gestione del Centro Prelievi. Per esempio, si potrebbe voler studiare il comportamento del sistema con diverse configurazioni del numero di sportelli e box prelievi aperti. I suddetti parametri possono essere modificati durante tutto il periodo di esecuzione della simulazione. Le aperture e le chiusure degli sportelli e dei box possono essere regolate anche attraverso opportune tabelle, cioè per ogni giorno della settimana può essere deciso, prima di far partire la simulazione, a quali ore gli sportelli dell'accettazione e i box prelievi debbano aprire o chiudere. Durante l'esecuzione della simulazione di un generico giorno gli

sportelli e i box possono aprire e chiudere diverse volte.

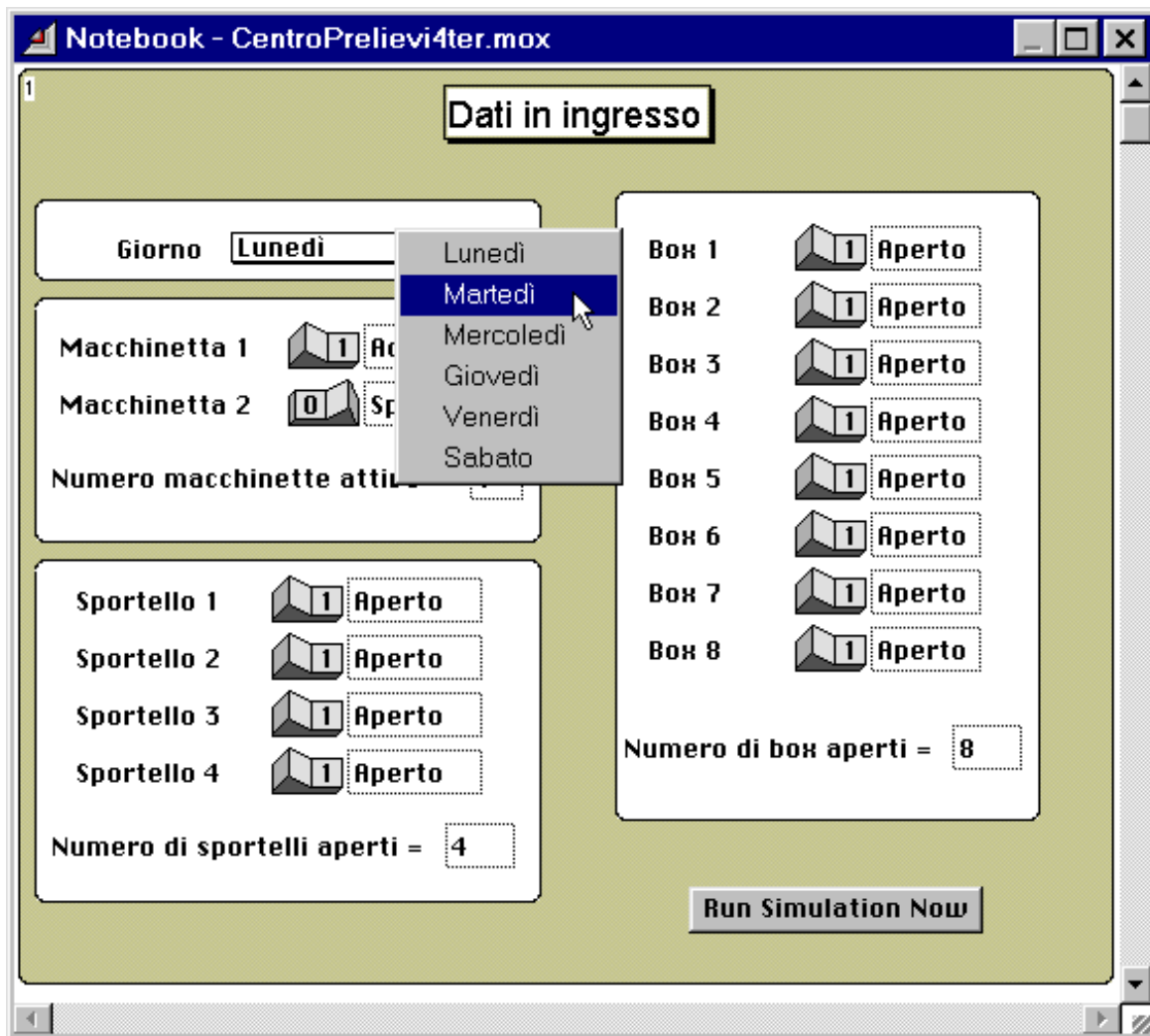


Figura 3.25: Dati in ingresso

Non è stato inserito nell'interfaccia il controllo delle altre variabili in input del modello, cioè quelle relative al processo di arrivo e ai tempi di servizio, in quanto sono variabili non controllabili. Per esempio, il tasso di arrivo degli utenti non può essere modificato dal gestore del sistema, ma solo da un cambiamento della popolazione dei possibili utenti. Tuttavia il modello fornisce anche la possibilità di studiare il sistema al variare dei parametri non controllabili. Una volta che sono stati fissati i parametri,

cliccando sul tasto *Run Simulation Now* appare la finestra di dialogo di Figura 3.26 mediante la quale si possono settare ulteriori parametri della simulazione.

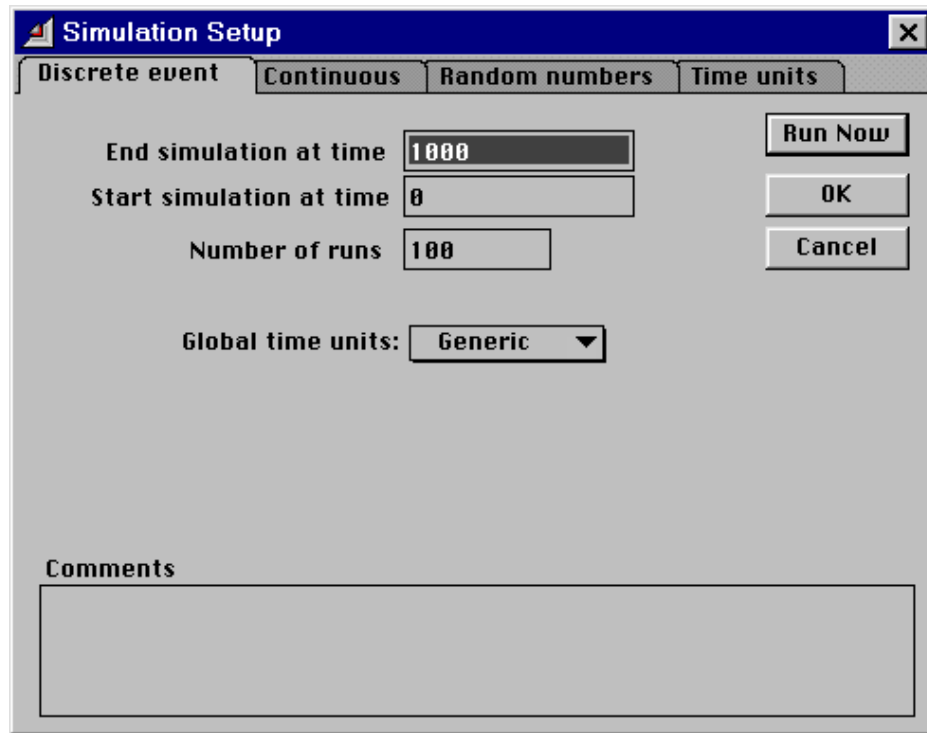


Figura 3.26: Simulation setup

La seconda sezione dell'interfaccia utente, illustrata nelle figure che seguono, riguarda i risultati della simulazione. Per ogni singolo run vengono visualizzati i valori medi dei tempi d'attesa nei diversi sottosistemi, il tempo medio di permanenza nel sistema, il tempo massimo di permanenza e il numero di utenti serviti. Nel caso di run multipli vengono visualizzate le medie dei risultati raccolti nei singoli run della simulazione. Inoltre, vengono visualizzati i grafici del numero di persone in attesa di fare l'accettazione, del numero di persone che attendono nella zona dei prelievi e il grafico del tempo di permanenza nel sistema.

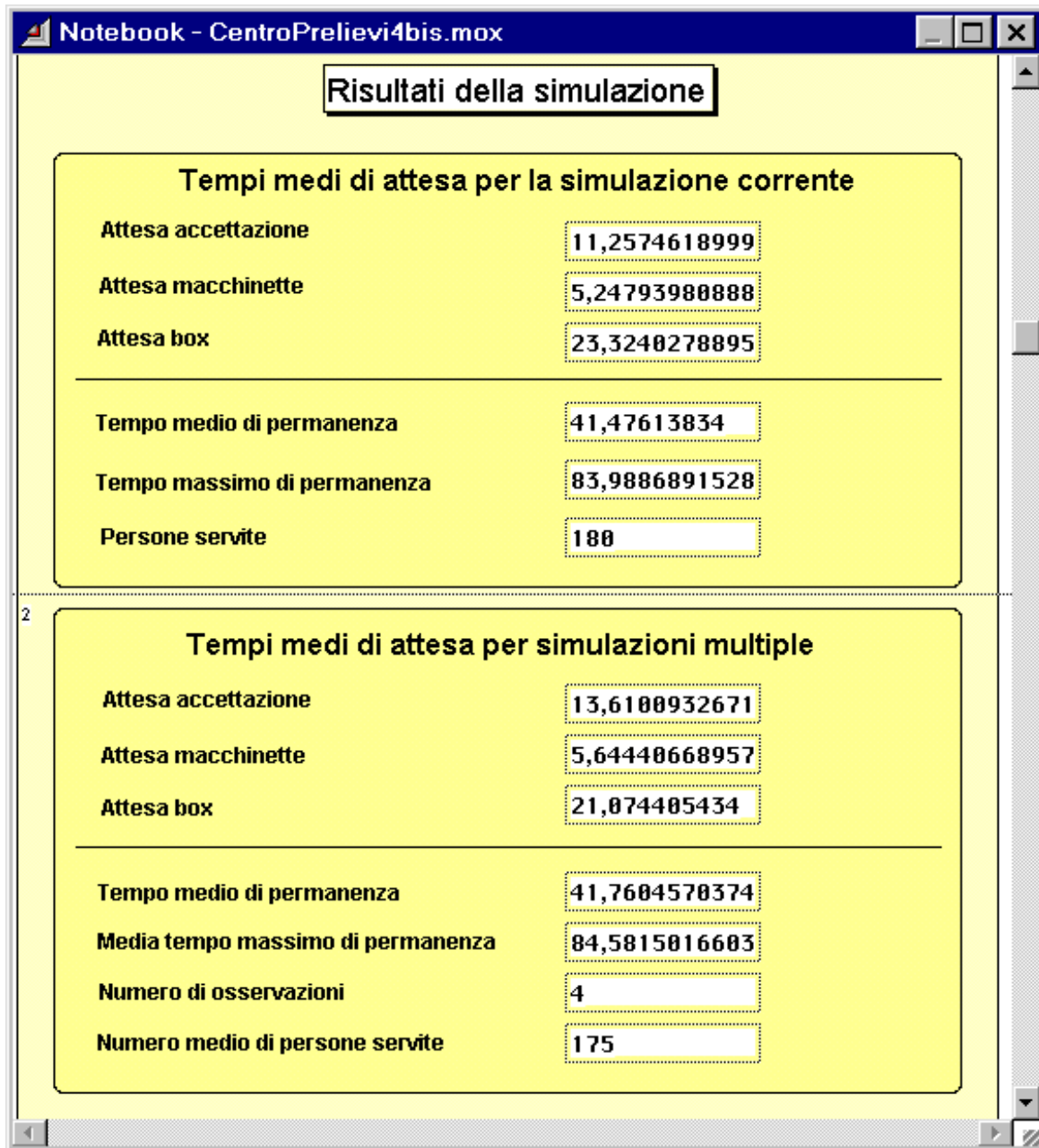


Figura 3.27: Risultati della simulazione

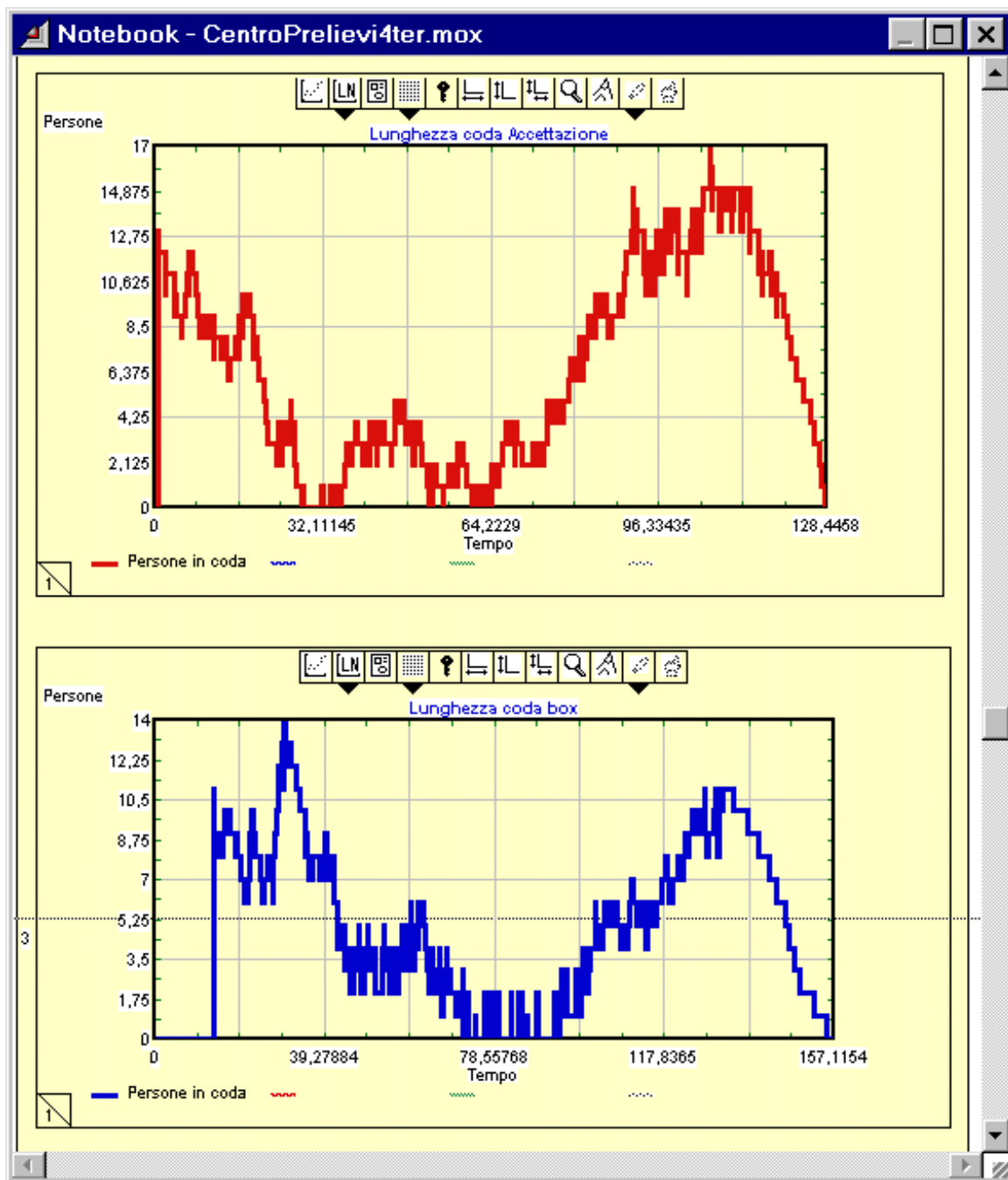


Figura 3.28: Grafici della lunghezza delle code di attesa

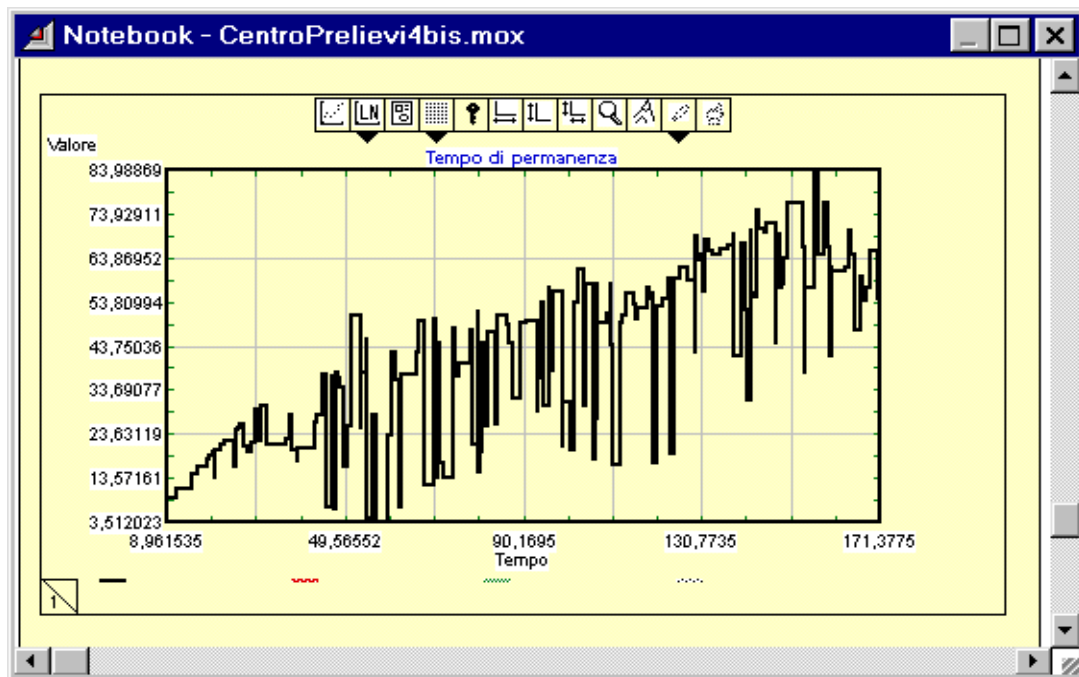


Figura 3.29: Grafico del tempo di permanenza

Capitolo 4

Verifica, calibrazione e validazione

Un passo fondamentale che deve affrontare lo sviluppatore di un modello di simulazione consiste nella verifica e validazione. Gli utenti e gli amministratori del sistema sotto analisi che utilizzano i risultati del modello di simulazione come assistenza per prendere decisioni, accettano con un certo grado di scetticismo la validità del modello.

L'obbiettivo del processo di validazione è duplice:

- produrre un modello che rappresenti il comportamento del sistema abbastanza fedelmente in modo che possa essere usato come sostituto del sistema attuale allo scopo di fare degli esperimenti con il sistema stesso;
- accrescere la credibilità del modello in modo che possa essere usato da amministratori o da coloro che prendono decisioni.

La validazione non dovrebbe essere vista come un insieme di procedure isolate successive allo sviluppo del modello, ma piuttosto come parte integrante della modellazione. Concettualmente, tuttavia, i processi di verifica e validazione consistono nelle seguenti componenti:

1. La *verifica* riguarda la costruzione del modello corretto e si esegue confrontando il modello concettuale e il modello di simulazione che rappresenta questi concetti. Si deve perciò controllare se il programma è implementato correttamente e se i parametri di ingresso e la struttura logica del modello sono rappresentati fedelmente.
2. Anche la *validazione* riguarda la costruzione di un modello corretto e serve per determinare se il modello è una accurata rappresentazione del sistema reale. La validazione è di solito usata per realizzare, attraverso la *calibrazione* del modello, un processo iterativo di confronto tra il comportamento del modello e quello del sistema attuale. Eventuali discrepanze vengono usate per migliorare il modello stesso.

Nei paragrafi successivi descriveremo i metodi che sono stati usati nei processi di verifica e validazione. Alcuni metodi sono dei confronti soggettivi e informali, mentre altri sono formali procedure statistiche. L'uso di queste ultime implica questioni relative all'analisi dei dati in uscita, cioè si analizzano i dati prodotti dalla simulazione e si inferisce dai dati il comportamento del sistema reale.

La costruzione del modello è stata realizzata attraverso passi successivi.

Il *primo passo* è stato l'osservazione del sistema reale, delle interazioni tra le sue varie componenti e la raccolta dei dati relativi al suo comportamento. Le sole osservazioni però, raramente forniscono una conoscenza sufficiente del comportamento del sistema, quindi è stato necessario intervistare persone che avevano familiarità col sistema, o con suoi sottosistemi, per trarre vantaggio dalle loro conoscenze specifiche. Mentre procedeva lo sviluppo del modello sono sorte nuove domande e si è tornati diverse volte a questo passo per capire l'effettiva struttura del modello e il suo comportamento.

Il *secondo passo* è consistito nella costruzione del modello concettuale, cioè è stata

fatta una serie di assunzioni sulla struttura e sulle componenti del sistema oltre alle ipotesi sui valori dei parametri dei dati in ingresso. La *validazione concettuale* è il confronto tra il sistema reale e il modello concettuale.

Il *terzo passo* è stato la traduzione del modello concettuale in una forma comprensibile dal computer, cioè l'implementazione del modello con il linguaggio di simulazione Extend. In pratica, la costruzione del modello non è stato un processo lineare come quello descritto nei tre punti precedenti, ma siamo stati obbligati a tornare diverse volte su questi passi nelle fasi di costruzione, validazione e verifica ed è stato necessario fare continui confronti tra il modello concettuale e il sistema reale con conseguenti modifiche del modello per migliorare la sua accuratezza.

4.1 Verifica del modello di simulazione

Come accennato precedentemente, la verifica del modello serve per assicurarsi che la rappresentazione computerizzata rifletta accuratamente il modello concettuale. Spesso il modello concettuale implica un certo grado di astrazione delle effettive operazioni del sistema con conseguenti semplificazioni sul suo comportamento.

Banks, Carson e Nelson [BCN96] hanno dato alcuni suggerimenti che abbiamo usato nel processo di verifica. Sono stati fatti diagrammi di flusso che includevano ogni azione, logicamente possibile, che il sistema poteva fare al verificarsi di ogni tipo di evento. È stato esaminato accuratamente l'output del modello, per verificare la sua ragionevolezza al variare dei parametri di ingresso, attraverso le statistiche stampate dal programma di simulazione. Siccome il programma forniva delle animazioni, abbiamo verificato che ciò che veniva visualizzato fosse una imitazione del sistema reale. Un altro strumento essenziale durante lo sviluppo del modello è stato il *debugger* che ci ha assistito nel rilevare e correggere gli errori. Le interfacce grafiche sono state di

grande aiuto per la verifica e la validazione in quanto sono essenzialmente una forma di autodocumentazione e semplificano il lavoro di comprensione del modello.

Una tecnica che è stata utile nella fase di verifica è stata quella dell'uso della traccia. In generale, una *traccia* è una stampa dettagliata che fornisce il valore di ogni variabile del programma, o di un sottoinsieme, ogni volta che una di queste variabili cambia valore. Perciò, la traccia di una simulazione non è niente più di una stampa dettagliata dello stato del modello di simulazione al variare del tempo.

4.2 Calibrazione e validazione

La calibrazione e la validazione [BCN96] [LK91], sebbene concettualmente distinte, sono di solito condotte simultaneamente. La validazione è il processo globale di confronto del modello e del suo comportamento con il sistema reale. La calibrazione è il processo iterativo che consiste, attraverso confronti del modello con il sistema reale, nel fare degli aggiustamenti al modello, nel confrontare il modello revisionato con la realtà e così via. Questo confronto è attuato da una varietà di test, alcuni soggettivi e altri oggettivi. I test soggettivi di solito coinvolgono le persone che sono a conoscenza di uno o più aspetti del sistema. Quelli oggettivi richiedono sempre dati sul comportamento del sistema oltre ai corrispondenti dati prodotti dal modello. Quindi, vengono eseguiti uno o più test statistici per confrontare degli aspetti dell'insieme dei dati del sistema con gli stessi aspetti dell'insieme dei dati del modello. Questo processo iterativo, di confronto tra modello e sistema con conseguenti revisioni del modello concettuale per accomodare ogni deficienza percepita del modello, è condotto fino a quando il modello viene giudicato sufficientemente accurato.

La validazione non porta a risultati assoluti; nessun modello è mai totalmente rappresentativo per il sistema in analisi. Inoltre ogni revisione del modello implica

costi in termini di tempo e lavoro. Perciò, bisogna arrivare a un compromesso tra i vantaggi che derivano da un livello più elevato di accuratezza del modello, che non è garantito, e gli svantaggi causati dall'incremento del costo del lavoro di validazione. Se il livello di accuratezza ideale non può essere ottenuto a causa di risorse limitate, deve essere abbassato il livello di accuratezza oppure il modello deve essere abbandonato. A volte la verifica di un modello è così difficile che può essere un problema intrattabile.

Per la validazione del nostro modello, abbiamo seguito l'approccio a tre fasi proposto da Naylor e Finger [NF67] che può essere concretizzato nei passi seguenti:

1. costruire il modello con una alta *face validity*;
2. validare le assunzioni del modello;
3. confrontare le trasformazioni ingresso-uscita del modello con le corrispondenti trasformazioni ingresso-uscita del sistema reale.

Nelle tre seguenti sezioni analizzeremo in dettaglio ognuno dei tre passi.

4.2.1 Sviluppo del modello con una alta *face validity*

L'obiettivo primario, nel corso del primo passo della validazione, è stato quello di sviluppare un modello con una alta *face validity* ossia un modello che, esteriormente, dovesse apparire ragionevole alle persone che hanno familiarità con il sistema da modellare. Gli utenti potenziali del modello sono stati coinvolti nella sua costruzione, dalla sua concettualizzazione alla sua implementazione, per assicurarci che fosse costruito con un alto grado di realismo attraverso assunzioni ragionevoli sulla struttura del sistema e sui dati. Anche gli utenti potenziali e le persone che conoscono il sistema hanno valutato la ragionevolezza dell'output del modello e sono stati d'aiuto per identificare i difetti del modello.

4.2.2 Validazione delle assunzioni del modello

Le assunzioni del modello ricadono in due classi generali: assunzioni strutturali e assunzioni sui dati. Le assunzioni strutturali riguardano questioni relative al modo di operare del sistema e di solito implicano semplificazioni ed astrazioni della realtà e sono state verificate attraverso l'osservazione del sistema oltre che attraverso discussioni con i suoi amministratori circa le politiche di gestione.

Le assunzioni sui dati dovrebbero essere basate sulla raccolta di dati affidabili e su analisi statistiche corrette dei dati. L'affidabilità dei dati è stata verificata sia attraverso la consultazione degli amministratori che attraverso test statistici, come ampiamente illustrato nel Capitolo 2.

4.2.3 Validazione delle trasformazioni ingresso-uscita

L'analisi essenziale di un modello, di fatto il solo test oggettivo completo del modello, è l'abilità del modello di predire il comportamento futuro del sistema reale quando i dati d'ingresso del modello sono uguali ai dati d'ingresso reali e quando una politica implementata nel modello è implementata allo stesso modo nel sistema. Inoltre, se il livello di qualche variabile d'ingresso (per esempio il tasso d'arrivo degli utenti) dovesse crescere o decrescere, il modello dovrebbe predire accuratamente ciò che succederebbe nel sistema reale in circostanze simili. In altre parole, la struttura del modello dovrebbe essere abbastanza accurata da fare buone previsioni, non solo per un insieme di dati d'ingresso, ma per tutti gli insiemi di dati di ingresso.

In questa fase del processo di validazione il modello è visto come una trasformazione ingresso-uscita. Cosicché il modello accetta valori dei parametri di ingresso e li trasforma in valori in uscita che misurano le prestazioni del sistema. È questa corrispondenza che doveva essere validata.

Illustreremo ora il procedimento usato per la validazione ingresso-uscita.

Le variabili di ingresso possono essere divise in due gruppi: variabili incontrollabili X e variabili controllabili o decisionali D . Le variabili di uscita saranno denotate con Y e quindi il nostro sistema può essere visto come una funzione di trasformazione che prende in ingresso X e D producendo in uscita Y , cioè

$$(X, D) \xrightarrow{f} Y$$

oppure

$$f(X, D) = Y$$

Qui f denota la trasformazione che è dovuta alla struttura del modello.

Le variabili decisionali sono sotto il controllo dell'amministrazione, mentre le variabili incontrollabili non lo sono. Per esempio, la frequenza di arrivo può cambiare nel tempo, ma questo cambiamento si verifica per fattori esterni non soggetti a un controllo da parte dell'amministrazione.

Nel nostro modello le *variabili incontrollabili* X sono:

- *tempi di servizio* (Tabella 2.5);
- *processo di arrivo* degli utenti (Tabella 2.6 e Formula 2.9);
- *tempi di interservizio* agli sportelli ed ai box (vedere pag. 58);
- *tipo di paziente* (Tabella 2.7).

Le *variabili decisionali* D sono:

- *apertura sportelli accettazione*: le aperture degli sportelli, usate per la calibrazione, sono impostate con i valori ricavati dalla rilevazione dei tempi nell'arco di

una settimana e tali valori sono riportati nella Tabella 4.1. Come si può vedere, per ogni giorno della settimana, ciascuno sportello apre e chiude a determinati istanti di tempo. I valori sono espressi come la differenza in minuti tra l'apertura dello sportello (o la chiusura) e l'orario di apertura del Centro Prelievi (7:15).

- *apertura box prelievi*: i valori di queste variabili usati per la calibrazione sono riportati in Tabella 4.2. Anche per queste variabili valgono le considerazioni fatte nel punto precedente.
- *accensione macchinette*: una macchinetta accesa per tutto il periodo di apertura del Centro Prelievi.

Le *variabili di uscita* Y sono:

- *tempo medio di attesa all'accettazione*: $Y_{1,i}$ dove $i = 1, \dots, 100$ indica l' i -esimo run della simulazione;
- *tempo medio di attesa ai box prelievi*: $Y_{2,i}$ dove $i = 1, \dots, 100$;
- *tempo medio di permanenza*: $Y_{3,i}$ dove $i = 1, \dots, 100$;
- *tempo massimo di permanenza*.

		Sport.1	Sport.2	Sport.3	Sport.4
Lunedì	<i>Apri</i>	3.667	6.333	0.033	17.883
	<i>Chiude</i>			13	32.333
Martedì	<i>Apri</i>	17.367	1.883	2.233	85
	<i>Chiude</i>			38.2	
Mercoledì	<i>Apri</i>	2.85	2		3.35
	<i>Chiude</i>				69.833
	<i>Apri</i>				119.217
	<i>Chiude</i>				
Giovedì	<i>Apri</i>	2.633	3.1	10.5	
	<i>Chiude</i>			64.05	
Venerdì	<i>Apri</i>	3.333	2.85		5
	<i>Chiude</i>	105			88.5
Sabato	<i>Apri</i>	0.833	2.983		89.45
	<i>Chiude</i>				

Tabella 4.1: Aperture sportelli accettazione

Affinché fosse possibile la validazione della trasformazione ingresso-uscita del modello del Centro Prelievi, dovevano essere disponibili dati del sistema reale da confrontare con qualche variabile d'uscita del modello. Le risposte del sistema sono state raccolte durante lo stesso periodo in cui sono stati raccolti i dati di ingresso.

Abbiamo ritenuto, vista la struttura del sistema, che le variabili interesse per la validazione fossero:

- *tempo medio di attesa all'accettazione Z_1 ;*
- *tempo medio di attesa ai box prelievi Z_2 ;*
- *tempo medio di permanenza Z_3 .*

	Box 1	Box 2	Box 3	Box 4	Box 5	Box 6	Box 7	Box 8
Lunedì	<i>Apri</i> <i>Chiude</i>	13.5 40 98.733	30 85.967	11.083	9.983			
Martedì	<i>Apri</i> <i>Chiude</i>	13.267	56.733	5.717 59.267	4.717	21.667		
Mercoledì	<i>Apri</i> <i>Chiude</i>	9.617 142.167	57.817	73.55	5.717	3.767		
Giovedì	<i>Apri</i> <i>Chiude</i>	27.8	40.667	14.5	5.517	7.283		
Venerdì	<i>Apri</i> <i>Chiude</i>	29 110.567	20.883 66.65	49.667 97.133	7 7.517	13.367 109.433		
Sabato	<i>Apri</i> <i>Chiude</i>	20.567	28.383 104.483	12.183	7.5	9.3		

Tabella 4.2: Aperture box prelievi

Raccogliere i dati di ingresso e le risposte nello stesso periodo di tempo era importante perché se fossero stati raccolti in periodi diversi, per esempio con tassi di arrivo degli utenti differenti, ci sarebbero state delle incongruenze.

Nella Tabella 4.3 sono riportati i valori, espressi in minuti, delle risposte del sistema, cioè *Media attesa accettazione* (Z_1), *Media attesa box* (Z_2) e *Media tempo di permanenza* (Z_3) che, come detto prima, sono calcolati sui valori rilevati direttamente dal sistema reale. Precisiamo che Z_1 , Z_2 e Z_3 , usate come confronto con le corrispondenti variabili di uscita del modello, sono le medie campionarie dei tempi di attesa di ogni utente servito in una particolare giornata e sono calcolate sui dati raccolti (Sezione 2.1) che sono riportati nella directory /Allegati/DatiRaccolti/ del CD-ROM.

<i>Giorno</i>	<i>Media attesa accettazione (Z_1)</i>	<i>Media attesa box (Z_2)</i>	<i>Media tempo di permanenza (Z_3)</i>
<i>Lunedì</i>	8.1667	8.7	22.3
<i>Martedì</i>	8.2667	8.25	22.5333
<i>Mercoledì</i>	12.3667	16.3	34.6667
<i>Giovedì</i>	12.5333	8.4833	26.9333
<i>Venerdì</i>	5.0333	7.3667	18.25
<i>Sabato</i>	10.3667	4.3	21.1333

Tabella 4.3: Risposte del sistema

Per ogni giorno della settimana sono state eseguite 100 simulazioni generando casualmente, con le distribuzioni specificate nel paragrafo 2.7, i valori delle variabili di ingresso e fissando i valori delle variabili decisionali con quelli delle Tabelle 4.1 e 4.2.

Ci si attendeva che i valori medi dei tempi di attesa del modello fossero vicini alle risposte del sistema. Tuttavia, non potevamo aspettarci che i valori di ingresso generati fossero esattamente uguali a quelli osservati dal sistema reale, ma ci si attendeva che essi rispondessero al modello statistico degli ingressi osservati nella realtà. Per esempio, ci aspettavamo che i valori di Y_3 del tempo medio di permanenza generati dalla simulazione fossero consistenti con la variabile del sistema Z_3 .

Per ogni giorno della settimana abbiamo eseguito molteplici serie di replicazioni statisticamente indipendenti. L'indipendenza era garantita dal fatto che il simulatore generava i numeri usando semi indipendenti per ciascuna replicazione. Dopo ogni serie di replicazioni venivano fatti dei controlli sui valori in uscita del modello per vedere se fossero sufficientemente vicini a quelli di risposta del sistema reale. I controlli sulle replicazioni iniziali venivano fatti senza l'applicazione di alcun test statistico, ma si decideva a occhio se andassero bene oppure no. Nei casi in cui i valori risultavano molto lontani tra loro si rianalizzava il modello, sia dal punto di vista strutturale che da quello dei dati di ingresso, cercando di capire la causa degli errori e si apportavano le opportune modifiche. Quando abbiamo reputato che i valori fossero abbastanza vicini abbiamo condotto rigorosi test statistici (test t di Student) in modo da avere una valutazione oggettiva sulla validità del modello. In qualche caso, non si ottenevano subito delle risposte positive e quindi si apportavano degli ulteriori aggiustamenti. In seguito illustreremo in dettaglio, attraverso un esempio, in che modo abbiamo condotto il t -test su ogni variabile di uscita usata per la validazione. La fase appena descritta è stata iterata per ogni variabile di uscita rilevante per la validazione.

Nella directory `/Allegati/Validazione` del CD-ROM sono riportati i risultati delle 100 replicazioni della simulazione del modello finale e nella Tabella 4.4 i valori delle medie e deviazioni standard campionarie dei risultati.

<i>Giorno</i>	<i>Attesa accettazione</i>		<i>Attesa box</i>		<i>Tempo di permanenza</i>	
	<i>Media</i> (\bar{Y}_1)	<i>Dev.Stand.</i> (S_1)	<i>Media</i> (\bar{Y}_2)	<i>Dev.Stand.</i> (S_2)	<i>Media</i> (\bar{Y}_3)	<i>Dev.Stand.</i> (S_3)
<i>Lunedì</i>	8.0928	4.156	8.7528	2.7658	22.7619	4.9919
<i>Martedì</i>	8.0973	3.468	8.0344	3.2836	22.1059	5.417
<i>Mercoledì</i>	12.0885	3.9088	16.9260	5.0267	35.0245	5.9806
<i>Giovedì</i>	12.9439	4.3901	8.3353	2.9911	27.3053	5.1588
<i>Venerdì</i>	4.9805	2.0325	7.0892	3.2453	18.0372	4.4222
<i>Sabato</i>	10.5256	4.4525	4.29	0.9938	20.7139	4.6224

Tabella 4.4: Risposte del modello finale

Vediamo formalmente il procedimento che è stato fatto, per testare la consistenza tra le risposte del sistema e le corrispondenti risposte del modello, attraverso l'esempio per il tempo di permanenza del lunedì.

Bisogna fare un test statistico sull'ipotesi nulla

$$H_0 : E(Y_3) = Z_3 = 22.3 \text{ minuti} \quad (4.1)$$

contro

$$H_1 : E(Y_3) \neq Z_3 = 22.3 \text{ minuti} \quad (4.2)$$

dove Z_3 è il tempo medio di permanenza rilevato dal sistema reale e Y_3 è la corrispondente risposta di una replicazione del modello di simulazione.

Se H_0 non viene rigettata allora, sulla base di questo test, non ci sono ragioni per considerare non valido il modello. Viceversa, se H_0 è rigettata, allora viene rigettata l'attuale versione del modello e bisogna cercare di migliorarlo. Il test statistico in

questione è il t -test che è condotto nel modo seguente:

Si sceglie un livello di significatività α e una dimensione n del campione. Nel nostro caso

$$\alpha = 0.05, \quad n = 100$$

Calcoliamo la media campionaria \bar{Y}_3 e la deviazione standard campionaria S_3 dei risultati delle n replicazioni

$$\bar{Y}_3 = \frac{1}{n} \sum_{i=1}^n Y_{3,i} = 22.7619 \text{ minuti}$$

e

$$S_3 = \left[\frac{\sum_{i=1}^n (Y_{3,i} - \bar{Y}_3)^2}{n-1} \right]^{1/2} = 4.9919 \text{ minuti}$$

dove $Y_{3,i}$, $i = 1, \dots, 100$ hanno valori come riportato nel file **Attese_lun.xls** che si trova nella directory **/Allegati/Validazione** del CD-ROM.

Dalla tavola dei punti percentuale della distribuzione t di Student ricaviamo che per un test bilaterale, come quello della Equazione 4.1, abbiamo $t_{\alpha/2, n-1} = 1.984$.

Calcoliamo la statistica del test

$$t_0 = \frac{\bar{Y}_3 - \mu_0}{S_3/\sqrt{n}} \tag{4.3}$$

dove μ_0 è il valore specificato nell'ipotesi nulla H_0 , ossia $\mu_0 = 22.3 \text{ minuti}$, cosicché

$$t_0 = \frac{22.7619 - 22.3}{4.9919/\sqrt{100}} = 0.925$$

Per il test bilaterale se $|t_0| > t_{\alpha/2, n-1}$ si rigetta H_0 altrimenti si accetta. Siccome, in questo caso, $|t_0| = 0.925 \leq t_{\alpha/2, n-1} = 1.984$ non rigettiamo e così sperimentalmente

accettiamo come valido il modello.

Ricordiamo che quando testiamo l'ipotesi H_0 il suo rigetto è una conclusione forte perché

$$P(\text{rigettare } H_0 | H_0 \text{ vera}) = \alpha \quad (4.4)$$

e abbiamo scelto il livello di significatività α piccolo ($\alpha = 0.05$).

L'equazione 4.4 dice che la probabilità di fare un errore nel rigettare H_0 , quando H_0 è di fatto vera, è bassa. Cioché è piccola la probabilità di dichiarare non valido un modello che invece lo è.

Le assunzioni che devono essere verificate per un t -test sono che le osservazioni $(Y_{3,i})$ siano normalmente distribuite. Vediamo perché nel nostro caso le assunzioni sono verificate:

1. L'osservazione i -esima $Y_{3,i}$ è il tempo medio di permanenza di tutti gli utenti del Centro Prelievi durante l' i -esima esecuzione della simulazione e così, per effetto del teorema limite centrale, è ragionevole assumere che ogni osservazione $Y_{3,i}$ è approssimativamente normalmente distribuita dato che il numero di utenti su cui è basata è abbastanza grande (mediamente 135).
2. Le osservazioni $Y_{3,i}$ sono statisticamente indipendenti per costruzione per il fatto che, per ogni replicazione, sono stati scelti semi casuali indipendenti.
3. La statistica t_0 , calcolata con l'equazione 4.3, è approssimativamente distribuita come la distribuzione t con $n - 1$ gradi di libertà anche quando le $Y_{3,i}$ non sono esattamente normalmente distribuite e così i valori critici della tavola dei punti percentuale della distribuzione t di Student possono essere usati con fiducia.

L'accettazione dell'ipotesi nulla H_0 deve essere considerata un conclusione debole se la stima della potenza del test non è risultata alta, cioè vicina a 1. La *potenza del test* è

la probabilità di rilevare una deviazione da H_0 quando di fatto tale deviazione esiste. Nel contesto della validazione la potenza del test è la probabilità di rilevare non valido un modello. La potenza può anche essere espressa come 1 meno la probabilità di un errore di tipo II, o β , dove

$$\beta = P(\text{errore di tipo II}) = P(\text{non rigettare } H_0 | H_1 \text{ vera})$$

è la probabilità di accettare il modello come valido quando non lo è.

Per considerare il non rigetto di H_0 come una conclusione forte si dovrebbe avere un valore di β piccolo. Il valore della probabilità β dipende dalla dimensione n del campione e dalla reale differenza tra $E(Y_3)$ e μ_0 , cioè da

$$\delta = \frac{|E(Y_3) - \mu_0|}{\sigma}$$

dove σ è la deviazione standard di un singolo $Y_{3,i}$ ed è stimata da S_3 .

La Figura 4.1¹ riporta i grafici delle curve caratteristiche operative (OC) che danno la probabilità di errore di tipo II $\beta(\delta)$ rispetto a δ per diverse dimensioni n del campione e $\alpha = 0.05$. Supponiamo di voler rigettare H_0 (validità del modello), con probabilità di almeno 0.9, se la media reale del tempo di permanenza del modello differisce dal tempo medio di permanenza del sistema per almeno il 15% del valore di quest'ultimo.

Allora δ è stimato da

$$\hat{\delta} = \frac{|E(Y_3) - \mu_0|}{S_3} = \frac{3.345}{4.9919} = 0.67$$

¹Fonte: C.L. Ferris, F.E. Grubbs, and C.L. Veaver, "Operating Characteristics for the Common Statistical Tests of Significance", *Annals of Mathematical Statistics*, Giugno 1946

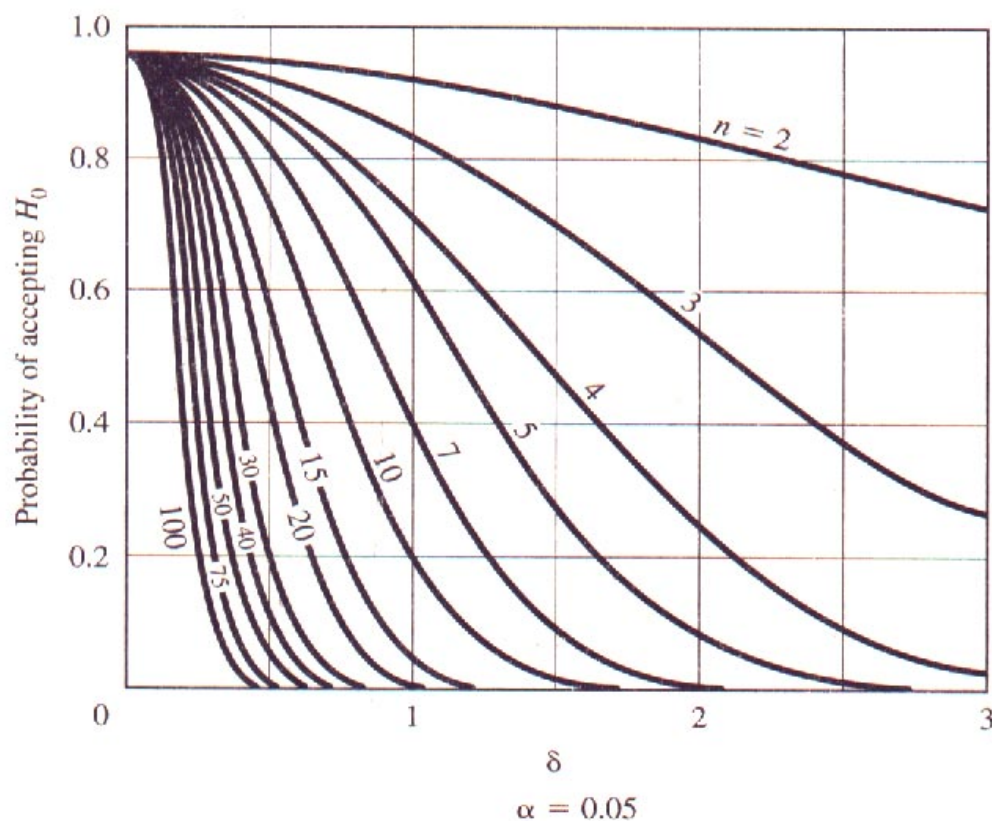


Figura 4.1: Curve caratteristiche operative il t test bilaterale per diversi valori della dimensione del campione n

Per il test in questione, usando la curva per $n = 100$ della Figura 4.1, risulta

$$\beta(\hat{\delta}) = \beta(0.67) \cong 0 \quad , \text{ per } n = 100$$

Noi volevamo $\beta(\hat{\delta}) \leq 0.1$ per avere una potenza di almeno 0.9 ed essendo $\beta(\hat{\delta}) \cong 0$ abbiamo che la probabilità di accettare H_0 (validità del modello), quando di fatto il modello non è valido ($|E(Y_3) - \mu_0| = 3.345 \text{ minuti}$), è $\beta \cong 0$ da cui segue che il test ha una potenza che si avvicina fortemente a 1.

In generale, è sempre meglio controllare l'errore di tipo II, o β , specificando una

differenza critica δ e scegliendo una dimensione del campione facendo uso della appropriata curva OC.

<i>Giorno</i>	t_{0,Y_1}	t_{0,Y_2}	t_{0,Y_3}
<i>Lunedì</i>	-0.1777	0.1908	0.9254
<i>Martedì</i>	-0.4883	-0.6566	-0.789
<i>Mercoledì</i>	-0.7112	1.2455	0.5984
<i>Giovedì</i>	0.9351	-0.495	0.7211
<i>Venerdì</i>	-0.2598	-0.855	-0.4813
<i>Sabato</i>	0.357	-0.1002	-0.9074

Tabella 4.5: Statistiche per il t -test

Nello stesso modo in cui abbiamo condotto i test per i tempi di permanenza Y_3 del lunedì abbiamo fatti i test per gli altri parametri di uscita importanti per la validazione, cioè per i tempi medi di attesa all'accettazione Y_1 e per i tempi medi di attesa ai box Y_2 . I risultati del calcolo delle statistiche t_{0,Y_i} , $i = 1, \dots, 3$ sono riportati nella Tabella 4.5.

Ricordiamo ancora che per $\alpha = 0.05$ e $n = 100$ si ha un valore critico $t_{\alpha/2, n-1} = 1.984$. Confrontando tale valore con i moduli dei valori della tabella (deve valere $|t_{0,Y_i}| \leq t_{\alpha/2, n-1}$) si può concludere che il modello del Centro Prelievi è valido dal punto di vista della trasformazione ingresso-uscita.

Nella Tabella 4.6 sono riportati i valori $\hat{\delta}_i$, $i = 1, \dots, 3$ calcolati per determinare la potenza dei test.

<i>Giorno</i>	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$
<i>Lunedì</i>	0.2947	0.4718	0.67
<i>Martedì</i>	0.3575	0.3769	0.6239
<i>Mercoledì</i>	0.4746	0.4864	0.8695
<i>Giovedì</i>	0.4282	0.4254	0.7831
<i>Venerdì</i>	0.3715	0.3405	0.619
<i>Sabato</i>	0.3492	0.649	0.6858

Tabella 4.6: Valori $\hat{\delta}$

Utilizzando il grafico delle curve operative della Figura 4.1 si può facilmente verificare che l'errore di tipo II (β) è sempre inferiore 0.1 (diverse volte è molto vicino a zero) e conseguentemente i test hanno una potenza superiore a 0.9. I calcoli delle potenze sono stati fatti considerando

$$|E(Y_i) - \mu_0| = \frac{15}{100}Z_i \quad , i = 1, 2, 3$$

Ricapitolando, nel contesto della validazione del modello l'errore di tipo I è il rigetto di un modello valido ed è facilmente controllabile specificando un livello di significatività α piccolo ($\alpha = 0.2, 0.1, 0.05, 0.01$). L'errore di tipo II è l'accettazione di un modello come valido quando effettivamente non è tale. Per n (dimensione del campione) fissato, l'incremento di α farà decrescere β che rappresenta la probabilità di un errore di tipo II. Una volta che α è impostato e la differenza critica da rilevare ($|E(Y_3) - \mu_0|$) è selezionata, il solo modo per decrementare β consiste nell'incremento della dimensione del campione. L'errore di tipo II è il più serio tra i due tipi di errore ed è quindi

importante progettare gli esperimenti della simulazione in modo da controllare il rischio di accettare un modello che non è valido.

4.2.4 Validazione ingresso-uscita con dati di ingresso storici

Quando come dati di ingresso vengono usati dati generati artificialmente, come è stato fatto nella sezione precedente, ci si aspetta che il modello produca sequenze di eventi che siano compatibili, ma non identici, con le sequenze di eventi che si verificano nel sistema reale durante il periodo di raccolta dei dati.

Un'alternativa alla generazione dei dati d'ingresso è l'uso della registrazione storica dei dati per guidare il modello di simulazione e quindi confrontare i dati di uscita del modello con i corrispondenti dati di uscita del sistema. Per implementare questa tecnica i dati di ingresso sono stati memorizzati in diversi file e sono stati da lì recuperati nel momento in cui si presentava la necessità. In particolare, i dati storici da dare in ingresso al modello ($i = 1, \dots, n$, con n numero di utenti serviti in una giornata) sono:

- Tempo di arrivo $T_{1,i}$;
- Tipo paziente $T_{2,i}$;
- Tempo di interservizio accettazione $T_{3,i}$;
- Tempo di servizio accettazione $T_{4,i}$;
- Tempo di servizio macchinette $T_{5,i}$;
- Tempo di interservizio box $T_{6,i}$;
- Tempo di servizio box $T_{7,i}$.

I dati storici relativi a una settimana, usati per la validazione, sono riportati nei file della directory /Allegati/ValStorica del CD-ROM. Questi dati vengono usati nel modo seguente: al tempo $T_{1,i}$ viene fatto arrivare l' i -esimo utente e al tempo $T_{1,i+1}$ arriva il successivo. Ogni utente, dopo che gli è stato impostato l'attributo che specifica il suo tipo con il valore $T_{2,i}$, viene inserito in coda. Se il cliente i -esimo iniziasse il servizio all'accettazione al tempo t_i il suo servizio terminerebbe al tempo $t_i + T_{4,i}$. Tutta la simulazione va avanti senza che sia generato alcun numero casuale. Quando si usa questa tecnica si spera che la simulazione duplichi, nel modo più aderente possibile, gli eventi importanti che si verificano nel sistema reale. Nel modello del Centro Prelievi $T_{1,i}$, $T_{2,i}$, ..., $T_{7,i}$ duplicano esattamente ciò che è successo nel sistema reale nell'arco di una giornata. Se il modello è sufficientemente accurato, allora i tempi di attesa all'accettazione, ai box e i tempi di permanenza prodotti dal modello, dovrebbero essere vicini a quelli osservati nel sistema reale.

Per condurre un test di validazione usando dati di ingresso storici è importante che tutti i dati di ingresso ($T_{1,i}$, $T_{2,i}$, ..., $T_{7,i}$) e tutti i dati di risposta del sistema (Z_i), siano stati raccolti nello stesso periodo di tempo altrimenti il confronto tra le risposte del modello e quelle del sistema potrebbe essere ingannevole. Le risposte (Y_i e Z_i , $i = 1, 2, 3$) dipendono dagli ingressi ($T_{h,i}$, $h = 1, \dots, 7$) non meno di quanto possono dipendere dalla struttura del sistema o del modello. Con un solo insieme di dati storici di ingresso e uscita può essere ottenuto un solo insieme di dati di uscita simulati e non è possibile applicare nessun test statistico basato su misurazioni riassuntive. Se invece sono disponibili k insiemi di dati di ingresso storici e k osservazioni delle variabili di uscita del sistema $Z_{i,1}$, $Z_{i,2}$, ..., $Z_{i,k}$, dove il valore $Z_{i,j}$ corrisponde al j -esimo insieme di ingresso, diventa possibile fare un test statistico oggettivo. Come nella sezione precedente illustriamo il procedimento considerando il tempo medio di permanenza $Z_{3,j}$. Per gli altri parametri si è proceduto in modo analogo. Sono stati usati, per la

validazione con dati storici, dati di ingresso e associate risposte del sistema relativi a una settimana, quindi nel nostro caso $k = 6$. Per ogni insieme di dati di ingresso è stato mandato in esecuzione il modello osservando i risultati simulati del tempo medio di permanenza $Y_{3,j}$, $j = 1, \dots, 6$ corrispondenti alle risposte del sistema $Z_{3,j}$. Se i k insiemi di dati di ingresso sono abbastanza omogenei è ragionevole assumere che le k differenze osservate $d_j = Z_{3,j} - Y_{3,j}$, $j = 1, \dots, 6$ siano identicamente distribuite. È stato fatto il test di Kruskal-Wallis per verificare l'omogeneità dei tempi di arrivo con risultati positivi. La raccolta dei dati di ingresso è stata fatta in giorni diversi, per cui è ragionevole assumere che le sei differenze d_1, \dots, d_6 siano statisticamente indipendenti costituendo un campione casuale. In questo caso ogni $Z_{3,j}$ e $Y_{3,j}$ è una media campionaria per cui, per il teorema del limite centrale, le variabili $d_j = Z_{3,j} - Y_{3,j}$ sono approssimativamente normalmente distribuite con una qualche media μ_d e varianza σ_d^2 .

Il test statistico appropriato è quindi il t -test con ipotesi nulla che non ci sia differenza tra le medie

$$H_0 : \mu_d = 0$$

contro l'alternativa di differenze significative

$$H_1 : \mu_d \neq 0$$

Per prima cosa calcoliamo la media campionaria delle differenze

$$\bar{d} = \frac{1}{k} \sum_{j=1}^k d_j$$

e la loro varianza campionaria

$$S_d^2 = \frac{1}{k-1} \sum_{j=1}^k (d_j - \bar{d})^2$$

I valori delle risposte del sistema, delle risposte del modello e dei calcoli appena discussi sono illustrati nella Tabella 4.7.

<i>Insieme dati di ingresso</i> j	<i>Risposta del sistema</i> $Z_{3,j}$	<i>Risposta del modello</i> $Y_{3,j}$	<i>Differenze osservate</i> $d_j = Z_{3,j} - Y_{3,j}$	<i>Quadrati differenze dalla media</i> $(d_j - \bar{d})^2$
1	21.4691	21.106	0.3631	0.938
2	12.6532	13.445	-0.7918	0.0348
3	34.6462	37.2378	-2.5916	3.9447
4	26.8293	24.6214	2.2079	7.9149
5	18.1957	19.4991	-1.3034	0.4872
6	20.8199	22.3366	-1.5167	0.8304
			$\bar{d} = -0.6054$	$S_d^2 = 2.83$

Tabella 4.7: Validazione con il tempo di permanenza

A questo punto calcoliamo la statistica t con la formula:

$$t_0 = \frac{\bar{d} - \mu_d}{S_d/\sqrt{k}} = \frac{-0.6054 - 0}{1.6823/\sqrt{6}} = -0.8815$$

e dalla tavola dei punti percentuale della distribuzione t di Student, per un livello di significatività $\alpha = 0.05$, prendiamo il valore critico $t_{\alpha/2, k-1} = 2.57$, con $k - 1 = 5$ gradi di libertà. Siccome

$$|t_0| = 0.8815 \leq t_{\alpha/2, k-1} = 2.57$$

non rigettiamo H_0 e concludiamo che questo test non fornisce l'evidenza di una inadeguatezza del modello.

<i>Insieme dati di ingresso</i> j	<i>Risposta del sistema</i> $Z_{1,j}$	<i>Risposta del modello</i> $Y_{1,j}$	<i>Differenze osservate</i> $d_j = Z_{1,j} - Y_{1,j}$	<i>Quadrati differenze dalla media</i> $(d_j - \bar{d})^2$
1	7.4613	7.3793	0.082	0.0006
2	2.4309	2.8239	-0.3929	0.2027
3	12.6179	10.57	2.0479	3.9623
4	12.5379	12.0688	0.4691	0.1696
5	5.0499	7.4799	-2.4301	6.1869
6	10.3547	9.7868	0.5679	0.2607
			$\bar{d} = 0.0573$	$S_d^2 = 2.1566$

Tabella 4.8: Validazione con il tempo di attesa accettazione

<i>Insieme dati di ingresso</i> j	<i>Risposta del sistema</i> $Z_{2,j}$	<i>Risposta del modello</i> $Y_{2,j}$	<i>Differenze osservate</i> $d_j = Z_{2,j} - Y_{2,j}$	<i>Quadrati differenze dalla media</i> $(d_j - \bar{d})^2$
1	8.6189	8.5023	0.1167	0.9091
2	4.4611	4.8809	-0.4198	0.1739
3	16.6081	21.4183	-4.8102	15.7881
4	8.4525	7.1775	1.2751	4.4599
5	7.3817	6.2169	1.1648	4.0063
6	4.2779	6.6251	-2.3472	2.2814
			$\bar{d} = -0.8368$	$S_d^2 = 5.5237$

Tabella 4.9: Validazione con il tempo di attesa box

Le due tabelle precedenti riportano i valori delle variabili di risposta del sistema e del modello e i risultati dei calcoli relativi ai tempi medi di attesa all'accettazione ed ai box. Il calcolo delle statistiche ha restituito $t_0 = 0.0956$ per l'accettazione e $t_0 = -0.8721$ per i box prelievi. Il valore critico è lo stesso usato per il tempo di permanenza, cioè $t_{0.025,5} = 2.57$ e quindi accettiamo il modello anche in relazione a questi due parametri.

Possiamo così concludere che, alla luce dei risultati forniti dai tre test, il modello del Centro Prelievi replica il comportamento del sistema reale con un buon grado di precisione.

4.3 Qualche prova con il modello

Sono state fatte alcune prove che possono servire, come modesto esempio, per illustrare le possibili applicazioni e personalizzazioni del modello di simulazione.

Due di queste sono state fatte sul modello del sistema attuale, agendo semplicemente sui bottoni dell'interfaccia utente, impostando diverse configurazioni del numero delle risorse attive, cioè degli sportelli di accettazione, dei box prelievi e delle macchinette per la riscossione del ticket.

Per la prima prova abbiamo impostato i seguenti valori:

- Giorno = Mercoledì;
- Numero di sportelli accettazione aperti = 3;
- Numero di macchinette accese = 2;
- Numero di box prelievi aperti = 5.

Abbiamo eseguito cento replicazioni della simulazione facendo in modo che i parametri impostati rimanessero costanti per tutta la durata dei run, contrariamente a ciò che succede nel sistema reale (si vedano le Tabelle 4.1 e 4.2 relative alle aperture).

Attraverso la sezione *Risultati della simulazione* dell'interfaccia, abbiamo potuto osservare che i tempi medi di attesa, sia all'accettazione che ai box, avevano valori vicini a sei minuti e il tempo medio di permanenza era all'incirca 17 minuti. Quest'ultimo risulta notevolmente più basso rispetto al valore rilevato dal sistema reale (circa 34 minuti).

Come ulteriore esempio abbiamo condotto una prova analoga alla precedente, ma con valori dei parametri differenti:

- Giorno = Mercoledì;
- Numero di sportelli accettazione aperti = 4;
- Numero di macchinette accese = 2;
- Numero di box prelievi aperti = 8.

Ossia abbiamo supposto che il personale fosse sufficiente per poter sfruttare al massimo le risorse, cioè tutti gli sportelli, le macchine e i box. Dopo le solite cento replicazioni abbiamo osservato che i tempi medi di attesa all'accettazione e ai box erano, anche in questo caso, quasi uguali e assumevano valori prossimi ai due minuti, mentre il tempo totale di permanenza si aggirava intorno ai 10 minuti.

I due esempi appena visti rappresentano un uso molto banale del modello del Centro Prelievi e non si potrebbe avere la pretesa di definirli “esperimenti” in quanto questi ultimi dovrebbero essere condotti dopo averli rigorosamente pianificati in stretta collaborazione con gli esperti del sistema reale. Si dovrebbero simulare diversi scenari non solo modificano le variabili controllabili del modello, ma anche modificando la sua

struttura, magari implementando diverse politiche di gestione e confrontando i diversi risultati con specifici metodi statistici. Tutti gli scenari da sperimentare hanno una stretta dipendenza dagli obiettivi che si vogliono raggiungere e dai vincoli strutturali propri del sistema.

Parte II

Reparto di Gastroenterologia

Capitolo 5

Formulazioni e modelli concettuali

5.1 Centro antidiabetico

Il centro antidiabetico è in funzione tutti i giorni dalle ore 7:00 alle ore 13:00 esclusi sabato (solo urgenze) e domenica.

Nel sistema possono essere individuate tre classi di **entità**: *pazienti*, *medici* e *infermieri*. Si può descrivere il comportamento di ogni entità di ciascuna classe, attraverso le sue attività. La Figura 5.1 descrive il ciclo delle attività dei pazienti.

Nel centro vengono svolte due attività principali: i prelievi e le visite. Gli addetti all'ambulatorio sono un medico e un' infermiera.

I pazienti possono recarsi al Centro per diversi motivi, ossia per prenotare un appuntamento¹, per fare un prelievo con relativa visita , oppure per ritirare i risultati di eventuali esami. Hanno anche la possibilità di fissare gli appuntamenti attraverso il telefono. Le prenotazioni possono essere fatte durante tutto l'orario d'apertura del Centro. Sia quando telefona che quando si reca personalmente, il paziente aspetta solamente che l'infermiera termini l'attività in corso in quel momento. La prenotazione di

¹Oppure per spostare la data di un appuntamento fissato in precedenza

un appuntamento o un eventuale variazione, viene fatta tenendo conto delle preferenze del paziente riguardo al medico, oppure nel primo giorno disponibile.

I prelievi sono svolti nel periodo di tempo che va dalle ore 7:00 alle ore 9:30. In questa fase sono coinvolti l'infermiera e il paziente. Si fissa un appuntamento per un determinato giorno (dietro richiesta di un medico esterno oppure interno) ed il numero medio di prenotazioni giornaliere è di 12 unità. Le analisi dei campioni prelevati sono fatte da un laboratorio (esterno al centro antidiabetico) che restituisce i risultati entro le ore 11:30, ora in cui inizia la fase delle visite.

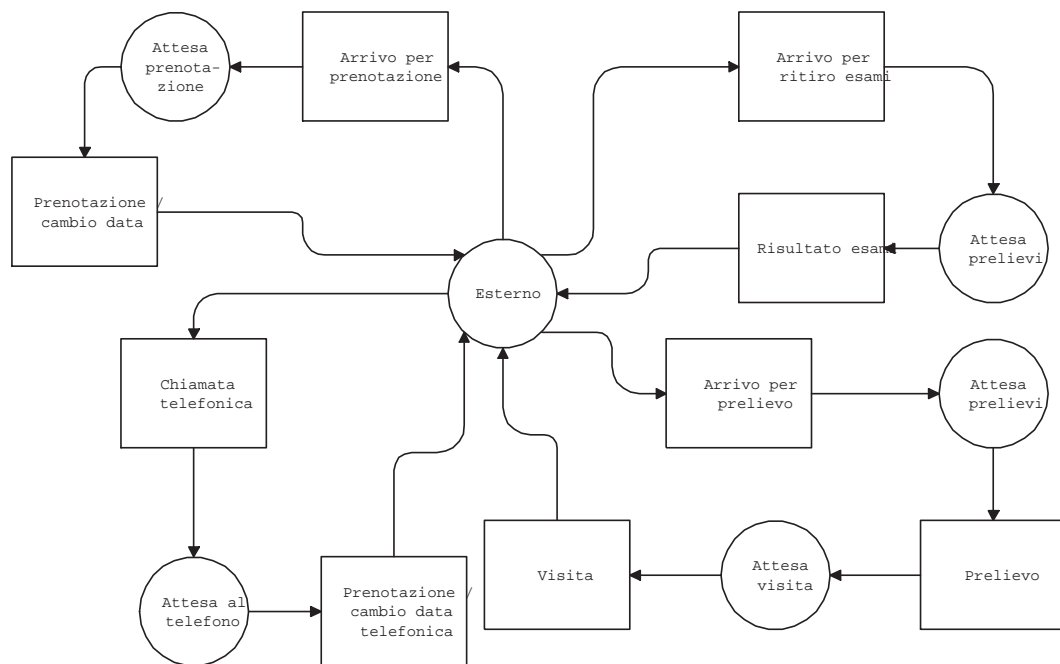


Figura 5.1: Diagramma dei cicli di attività pazienti

Il paziente può arrivare per un prelievo dalle ore 7:00 alle 9:30 e si mette in una fila di attesa ordinata in base al tempo di arrivo (FIFO). Nel Centro vengono fatti mediamente 12 prelievi al giorno. Prima di ciascun prelievo l'infermiera annota i dati del paziente in una scheda e al termine dei prelievi (intorno alle ore 9:30) deposita le

provette su un carrellino dal quale saranno prelevate da un addetto che passa ogni ora e le consegna al laboratorio di analisi. Finita la fase del prelievo i pazienti aspettano l'inizio delle visite mentre l'infermiera compila i moduli dei pazienti ed eventualmente può fare le consegne dei risultati degli esami supplementari. Tutto questo fino alle ore 11:00, ora in cui l'infermiera si reca al laboratorio di analisi per ritirare i risultati. Alle 11:30 arriva il medico ed iniziano le visite. I pazienti vengono visitati con lo stesso ordine con cui hanno fatto il prelievo. Il medico, con l'aiuto dell'infermiera, visita il paziente e prescrive la terapia e gli eventuali esami aggiuntivi. Alla fine della visita si consegna il modulo con i risultati e si fissa l'appuntamento per il controllo successivo in base ai dati laboratoristici attuali e storici (sono rilevanti HbA1c e Fruttosamina). Se il controllo del diabete non è ottimale, il paziente viene rivisto a breve scadenza oppure si propone il ricovero in Day Hospital. La visita successiva viene stabilita in modo che il paziente venga seguito sempre dallo stesso medico.

Le due figure che seguono descrivono il comportamento del medico e dell'infermiera.

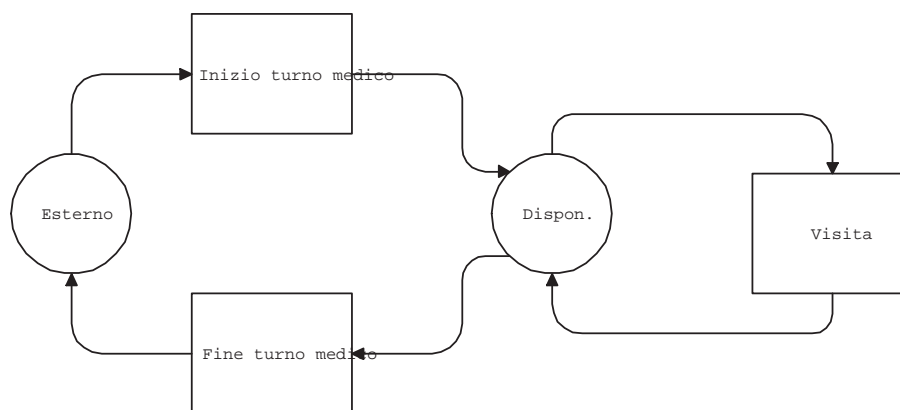


Figura 5.2: Diagramma dei cicli di attività medico

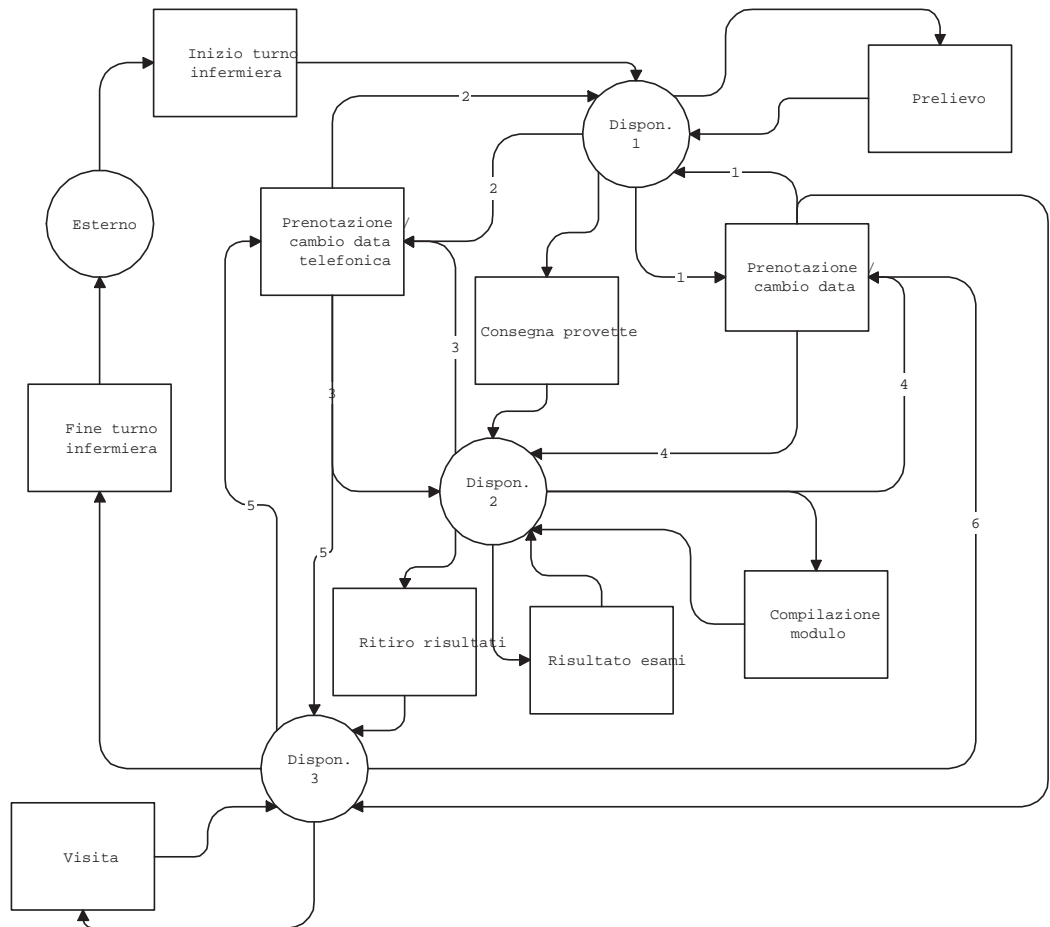


Figura 5.3: Diagramma dei cicli di attività infermiere

Riassumendo:

Entità

- Pazienti;
- Medico;
- Infermiera.

Attività

1. arrivo per prenotazione (7:00 - 13:00):
il paziente arriva per prendere un appuntamento e si mette in fila;
2. arrivo per prelievo (7:00 - 9:30);
3. arrivo per ritiro esami;
4. chiamata telefonica (7:00 - 13:00):
il paziente telefona per prendere un appuntamento e attende l' inizio della comunicazione;
5. inizio turno medico (11:30):
arriva il medico e aspetta il primo paziente;
6. fine turno medico (13:00):
quando non ci sono più pazienti da visitare il medico lascia l'ambulatorio;
7. inizio turno infermiera (7:00);
8. consegna provette:
quando sono stati fatti tutti i prelievi, le provette sono messe dall'infermiera su un carrellino dal quale vengono prelevate da un addetto che passa ogni ora e le consegna al laboratorio di analisi;
9. ritiro risultati (~11:00):
l'infermiera si reca al laboratorio di analisi per ritirare i risultati e ritorna al Centro. Il tempo medio per questa attività è di 20 minuti;
10. fine turno infermiera (13:00):
se sono finite le visite, l'infermiera lascia il Centro;

11. prelievo (7:00 - 9:30):

i prelievi vengono svolti dall'infermiera. Prima di ogni prelievo l'infermiera annota i dati personali del paziente. Tutta l'operazione richiede mediamente 5 minuti. Dopo che è stato fatto il prelievo il paziente aspetta di essere visitato. La fila di attesa per le visite è la stessa di quella per i prelievi;

12. visita (11:30 - 13:00):

il medico, con l'aiuto dell'infermiera, visita il paziente e prescrive la terapia e gli eventuali esami aggiuntivi. Alla fine della visita si consegna il modulo con i risultati e si fissa l'appuntamento per il controllo successivo, in base ai dati laboratoristici attuali e storici (sono rilevanti HbA1c e Fruttosamina) ed in base al diario clinico. La data viene stabilita in modo che il paziente venga seguito sempre dallo stesso medico;

13. risultato esami (9:30 - 11:30):

vengono consegnati al paziente i risultati degli esami supplementari (il tempo della durata della consegna è incognito);

14. prenotazione/cambio data (7:00 - 13:00):

la prenotazione di un appuntamento o un eventuale variazione, vengono fatte tenendo conto delle preferenze del paziente riguardo al medico, oppure nel primo giorno disponibile;

15. prenotazione/cambio data telefonica (7:00 - 13:00):

le telefonate vengono ricevute solo nel caso in cui l'infermiera non sia impegnata in un'altra attività. In quest'ultimo caso, chi telefona sta in attesa finché l'infermiera non è disponibile. La prenotazione di un appuntamento o un eventuale variazione,

vengono fatte tenendo conto delle preferenze del paziente riguardo al medico, oppure nel primo giorno disponibile;

16. compilazione modulo (9:30 - 11:30):

compilazione scheda paziente e/o cartella clinica;

17. attività educativa (autocontrollo).

5.2 Ambulatorio di Gastroenterologia e Day Hospital

L'ambulatorio di Gastroenterologia ed il Day Hospital sono in funzione tutti i giorni dalle ore 7:00 alle ore 13:00 escluso il sabato (solo urgenze) e la domenica. Questi due ambulatori vengono analizzati assieme perché il personale impiegato è comune e consiste in un medico e un infermiere. L'infermiere è sempre lo stesso, mentre i medici seguono dei turni giornalieri. Siccome l'attività del Day Hospital è assimilabile a quella della corsia ed il nostro interesse è rivolto ad attività ambulatoriali, non si considererà in dettaglio il suo funzionamento interno ma verrà visto come una semplice attività con certi tempi di servizio.

Nell'ambulatorio si svolgono mediamente 12 visite al giorno. Al momento dell'appuntamento vengono stabiliti giorno e ora della visita in modo tale che l'arrivo dei pazienti avvenga a distanza di 20 minuti l'uno dall'altro impedendo così di congestionare l'attività ambulatoriale e la sala di attesa. L'intervallo tra due appuntamenti è giustificato dal fatto che sebbene le visite abbiano una durata molto variabile dipendente dalla complessità del caso mediamente durano 20 minuti.

L'attività in Day Hospital abbiamo detto essere assimilabile a quella della corsia con la differenza che le prestazioni si esauriscono nell'arco di una singola giornata, evitando così il ricovero del paziente in ospedale.

5.2.1 Ambulatorio di Gastroenterologia

Nella maggioranza dei casi i pazienti richiedono un appuntamento telefonicamente, ma in alcuni casi si recano di persona al reparto. In ogni caso, la prassi seguita è la stessa, cioè gli appuntamenti per le visite ambulatoriali vengono fissati in modo che la data della visita venga scelta in base alle preferenze del paziente riguardo al medico, altrimenti si fissa l'appuntamento per il primo giorno disponibile tenendo conto che il

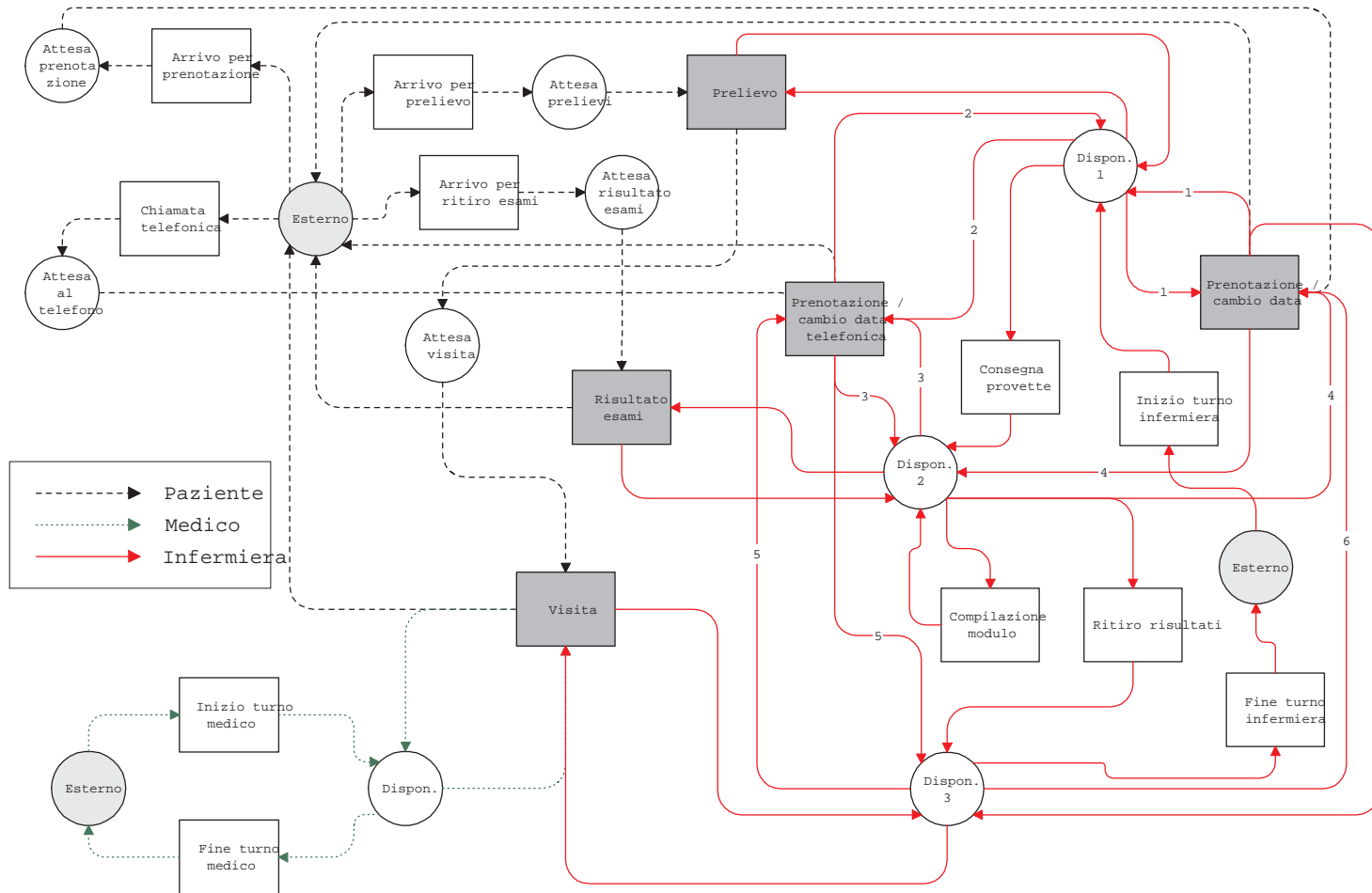


Figura 5.4: Diagramma dei cicli di attività Centro Antidiabetico

numero di visite giornaliere è circa 12. L'ora della visita, come già detto, è fissata in modo che i pazienti arrivino a distanza di circa venti minuti l'uno dall'altro, evitando così attese inutili. Oltre che per fissare gli appuntamenti delle visite, i pazienti possono telefonare o presentarsi direttamente nel reparto anche per modificare la data di un appuntamento già fissato in precedenza e anche in questo caso, la nuova data viene fissata con le suddette modalità.

La fase appena descritta coinvolge le entità infermiere e paziente che, come si può notare dalla Figura 5.8, sono viste come entità cooperanti.

Un'altra attività è la visita gastroenterologica in ambulatorio.

Il paziente arriva all'ora stabilita e generalmente attende la disponibilità del medico che potrebbe essere impegnato in un'altra visita oppure nel Day Hospital. Nell'attesa consegna all'infermiere la ricevuta dell'avvenuto pagamento del ticket. Il medico che fa le visite in ambulatorio in alcuni casi è costretto ad interromperle per fare degli interventi urgenti su un paziente ricoverato nel Day Hospital.

L'infermiere alla fine di ogni visita, chiama il paziente successivo nella lista di attesa. La visita coinvolge direttamente il medico e il paziente .

Dopo aver visitato il paziente, il medico prescrive la terapia e/o richiede eventualmente alcuni esami a completamento della diagnosi da effettuarsi all'interno del Reparto di Gastroenterologia oppure in strutture esterne ed in quest'ultimo caso il paziente viene rivisto in una visita successiva. Nel caso di esami interni, l'infermiere provvede direttamente a fissare gli appuntamenti. Nel caso di malati cronici al termine della visita viene fissato il successivo controllo. Oltre alle visite programmate possono essere richieste delle visite urgenti e in questi casi il paziente quando arriva, aspetta solo la disponibilità del medico ed ha generalmente precedenza rispetto alle visite programmate. I malati che hanno necessità di una visita urgente provengono dal Pronto Soccorso oppure da altri reparti.

Nella fase appena descritta ci sono diverse interazioni tra le entità:

- il medico con il paziente durante la visita;
- l'infermiere con il paziente per fissare gli appuntamenti degli esami interni e della visita successiva;
- inoltre l'infermiere interagisce anche con altre strutture del reparto per ciò che riguarda gli appuntamenti degli esami interni.

Dopo aver effettuato tutte le visite, il medico finisce il turno in ambulatorio e se non ci sono pazienti nel Day Hospital va via. La stessa cosa vale anche per l'infermiere se sono già passate le 13:00, altrimenti aspetta la fine del turno.

Le Figure 5.5, 5.6 e 5.7, mostrano il ciclo di attività per ciascuna entità : *paziente*, *medico* e *infermiere*.

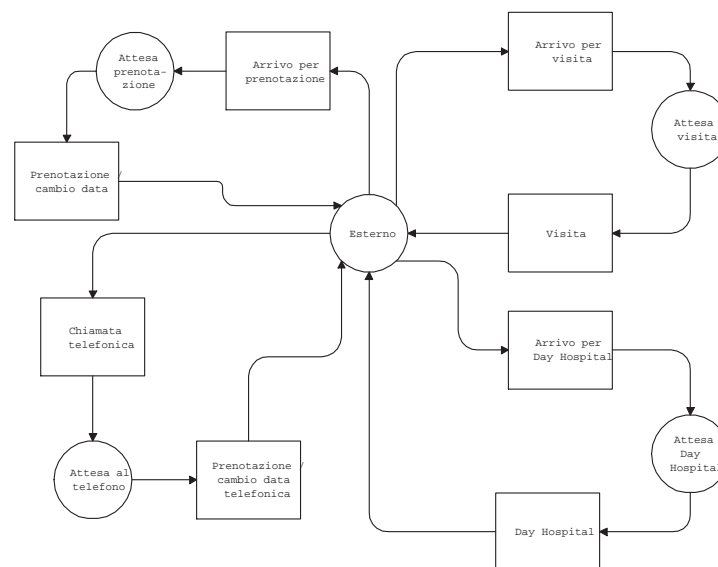


Figura 5.5: Diagramma dei cicli di attività pazienti

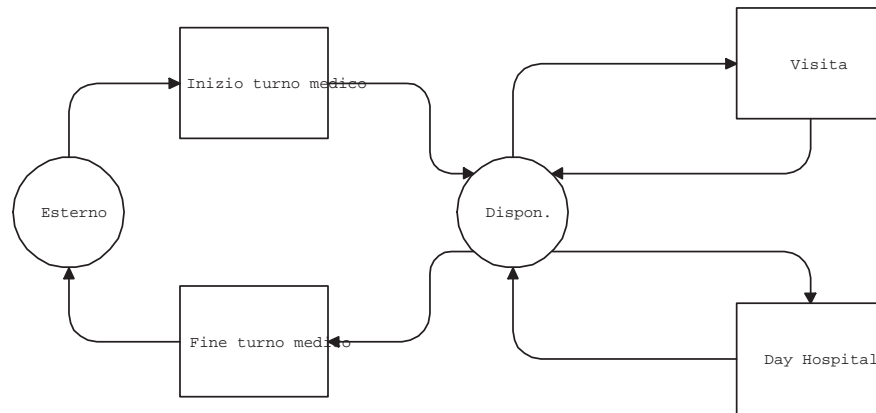


Figura 5.6: Diagramma dei cicli di attività medico

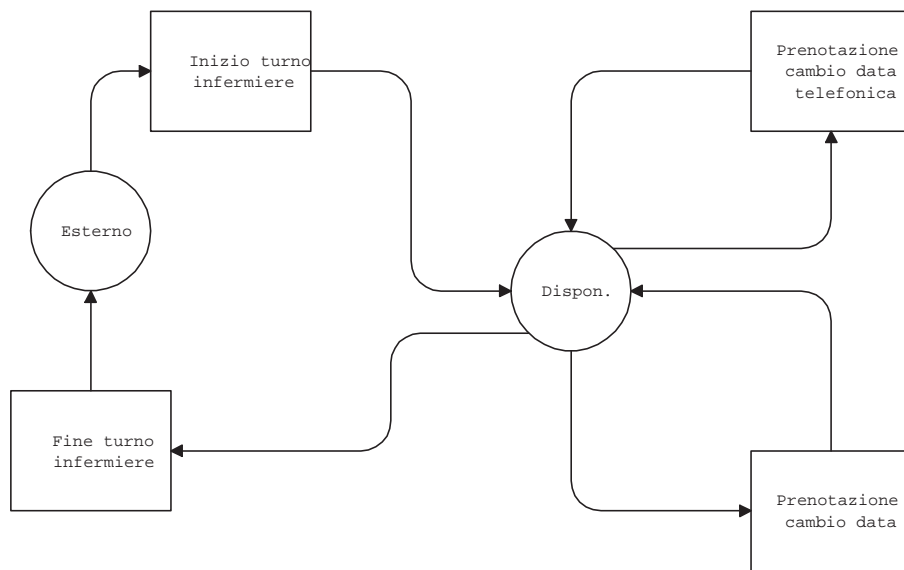


Figura 5.7: Diagramma dei cicli di attività infermiere

Si noti che nella Figura 5.8, è stata introdotta un'attività diversa dalle solite che è quella che riguarda il Day Hospital e la cui descrizione, a grandi linee, verrà data in seguito.

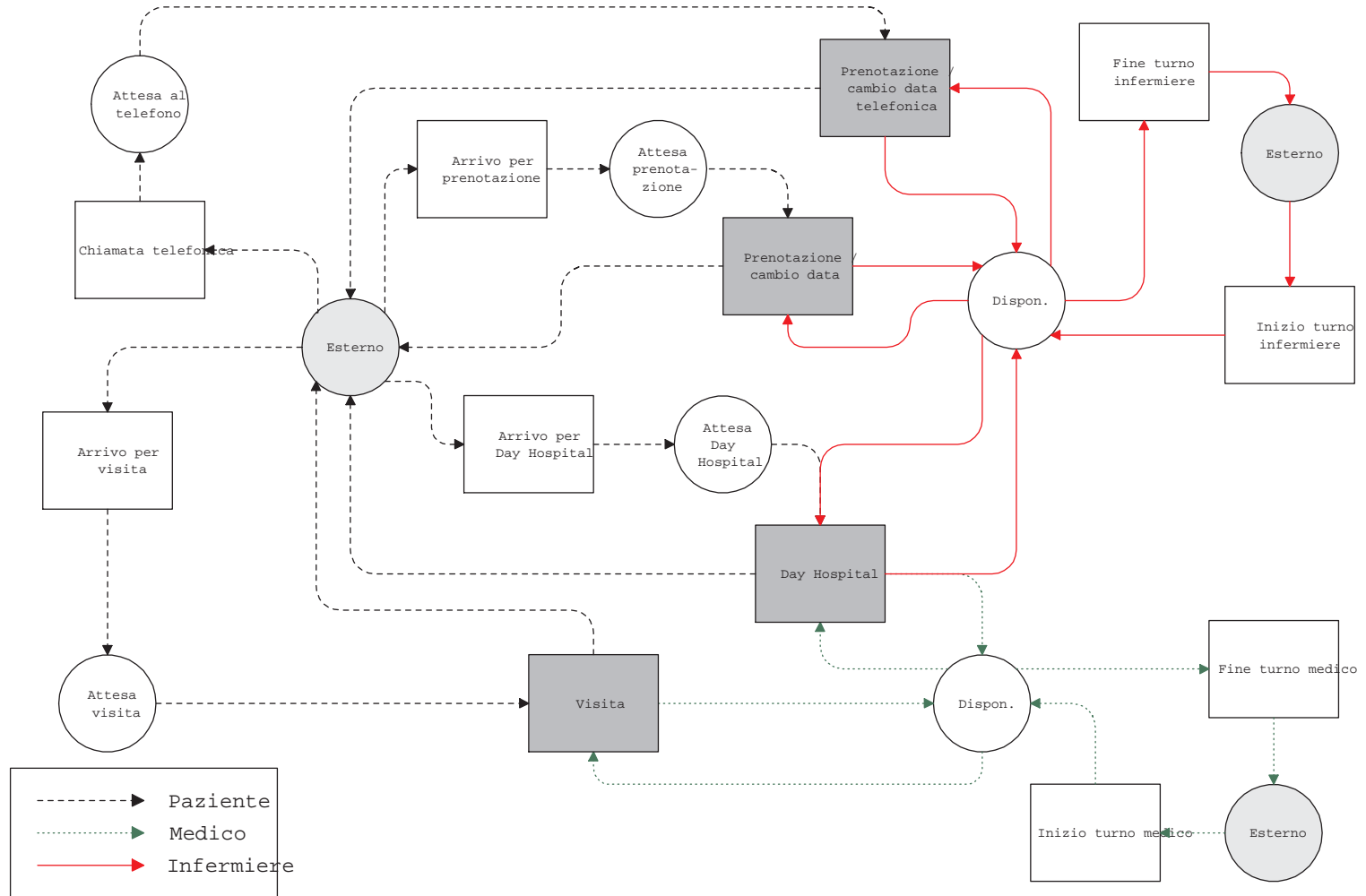


Figura 5.8: Diagramma dei cicli di attività Ambulatorio e Day Hospital

5.2.2 Day Hospital

In precedenza si accennava al fatto che il Day Hospital è da considerarsi come un piccolo reparto, dove si eseguono interventi relativamente impegnativi, che non possono essere eseguiti ambulatorialmente o a domicilio e che altrimenti richiederebbero il ricovero in ospedale. L'esistenza di questa struttura è giustificata dal fatto che in questo modo si risparmiano risorse come posti letto, medici e infermieri delle corsie, rendendole utilizzabili per interventi più impegnativi e prolungati. Generalmente il ricovero nel Day Hospital si risolve in una singola giornata ma in alcuni casi occorrono più giorni ed allora il paziente ritornerà più volte.

Gli interventi che si eseguono in questa struttura possono essere classificati come segue:

- clinico:
 - diagnostico;
 - terapeutico;
- educativo:
 - individuale;
 - di gruppo.

Il Day Hospital educativo viene eventualmente svolto in cooperazione con associazioni *no profit* e di volontariato (per esempio Associazione Diabetici, IBD, ecc. . .), altri specialisti e familiari.

Le prenotazioni per il Day Hospital si fanno, di solito uno o due giorni prima. Il numero di interventi terapeutici che possono essere eseguiti contemporaneamente è pari

a 5 (essendo disponibili 3 letti e due poltrone), mentre per gli interventi diagnostici non ci sono limiti salvo che la somma di tutti gli interventi, sia diagnostici che terapeutici, non sia superiore a sei. Le prenotazioni, nella stragrande maggioranza dei casi, vengono fatte telefonicamente da malati cronici oppure dal medico stesso nei casi in cui se ne ravvisi la necessità.

Gli interventi diagnostici coinvolgono:

- medico, paziente e infermiere:

quando il paziente arriva il medico svolge una visita preliminare, apre la cartella clinica e dopo la visita chiede all'infermiere la preparazione del piano degli esami da eseguire.

- paziente e infermiere:

l'infermiere dopo aver compilato il piano degli esami da eseguire, prepara le richieste degli esami esterni e fissa gli appuntamenti telefonicamente, possibilmente da effettuarsi in giornata o eventualmente nel giro di qualche giorno. Nel caso di esami da eseguirsi nel Day Hospital o in strutture interne al reparto l'infermiere provvede a fissare gli appuntamenti telefonicamente o personalmente e di solito gli esami vengono eseguiti in giornata.

Gli interventi terapeutici coinvolgono il paziente, l'infermiere ed il medico.

L'attività in questa struttura inizia alle 7:00, ed il primo ad arrivare è l'infermiere che provvede a preparare i letti e gli strumenti necessari per gli eventuali interventi.

Dalle 8:00 in poi arrivano gli eventuali pazienti che hanno un appuntamento.

Riassumendo:

Entità

- Paziente;

- Medico;
- Infermiere.

Attività

1. Arrivo per prenotazione (7:00 - 13:00):
il paziente arriva per prendere un appuntamento e si mette in attesa della disponibilità dell'infermiere.
2. Arrivo per visita (7:30 - 12:30):
il paziente arriva all'ora stabilita per la visita, consegna all'infermiere la ricevuta dell'avvenuto pagamento del ticket e attende il suo turno.
3. Chiamata telefonica (7:00 - 13:00):
il paziente telefona per prendere o spostare un appuntamento e attende l' inizio della comunicazione.
4. Inizio turno medico (7:00 - 8:00):
arriva il medico e aspetta il primo paziente del Day Hospital o dell'ambulatorio.
5. Fine turno medico:
quando non ci sono più pazienti da visitare, il medico lascia l'ambulatorio e se anche il Day Hospital è libero, va via.
6. Inizio turno infermiere (7:00):
l'infermiere arriva nel Day Hospital e le sue attività principali saranno l'occuparsi dei pazienti che arriveranno in questa struttura e il fissare gli appuntamenti delle visite.

7. Fine turno infermiere (13:00):

se le visite sono finite e allo stesso tempo nel Day Hospital non ci sono pazienti, l'infermiere va via.

8. Visita (7:00 - 13:00):

il medico visita il paziente e prescrive la terapia e gli eventuali esami aggiuntivi. Nel caso in cui gli esami siano da farsi nel reparto, l'infermiere prende i relativi appuntamenti ed eventualmente fissa l'appuntamento per il controllo successivo. La data viene stabilita in modo che il paziente venga seguito sempre dallo stesso medico.

9. Prenotazione/cambio data (7:00 - 13:00):

per le visite in ambulatorio, la prenotazione di un appuntamento o un eventuale variazione, vengono fatte tenendo conto delle preferenze del paziente riguardo al medico, oppure nel primo giorno disponibile. Si fissa l'appuntamento per un determinato giorno e ad una determinata ora in modo che i pazienti arrivino ad intervalli di venti minuti. Tale intervallo coincide con la durata media di una visita.

Le prenotazioni per il Day Hospital si fanno, di solito uno o due giorni prima. Il numero di interventi terapeutici che possono essere eseguiti contemporaneamente è pari a tre, mentre per gli interventi diagnostici non ci sono limiti salvo che la somma di tutti gli interventi, sia diagnostici che terapeutici, non sia superiore a sei. Le prenotazioni, nella stragrande maggioranza dei casi, vengono fatte telefonicamente.

In questa attività interagiscono il paziente e l'infermiere.

10. Prenotazione/cambio data telefonica (7:00 - 13:00):

le telefonate vengono ricevute solo nel caso in cui l'infermiere non sia impegnato in

un'altra attività. In quest'ultimo caso, chi telefona sta in attesa finché l'infermiere non è disponibile. La prenotazione di un appuntamento o un eventuale variazione, vengono fatte con le stesse modalità del punto precedente.

11. Arrivo per Day Hospital.

12. Day Hospital:

questa attività viene svolta dallo stesso medico che svolge le visite in ambulatorio e quindi le due attività (visite ambulatoriali e Day Hospital) procedono parallelamente. Se ci sono dei pazienti nel Day Hospital, il medico fa la spola tra l'ambulatorio e questa struttura, generalmente tra una visita e l'altra. In situazioni critiche può capitare che si interrompa una visita in ambulatorio e il medico intervenga nel Day Hospital. L'infermiere è sempre lo stesso mentre i medici ruotano a turno.

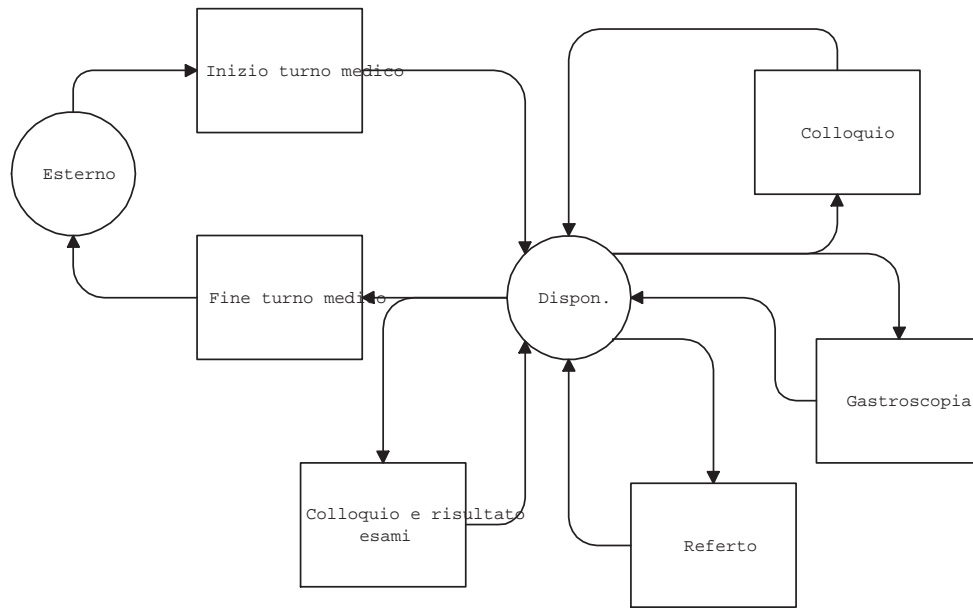


Figura 5.10: Diagramma dei cicli di attività medico

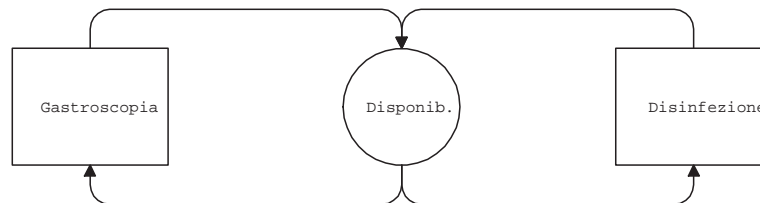


Figura 5.11: Diagramma dei cicli di attività macchina

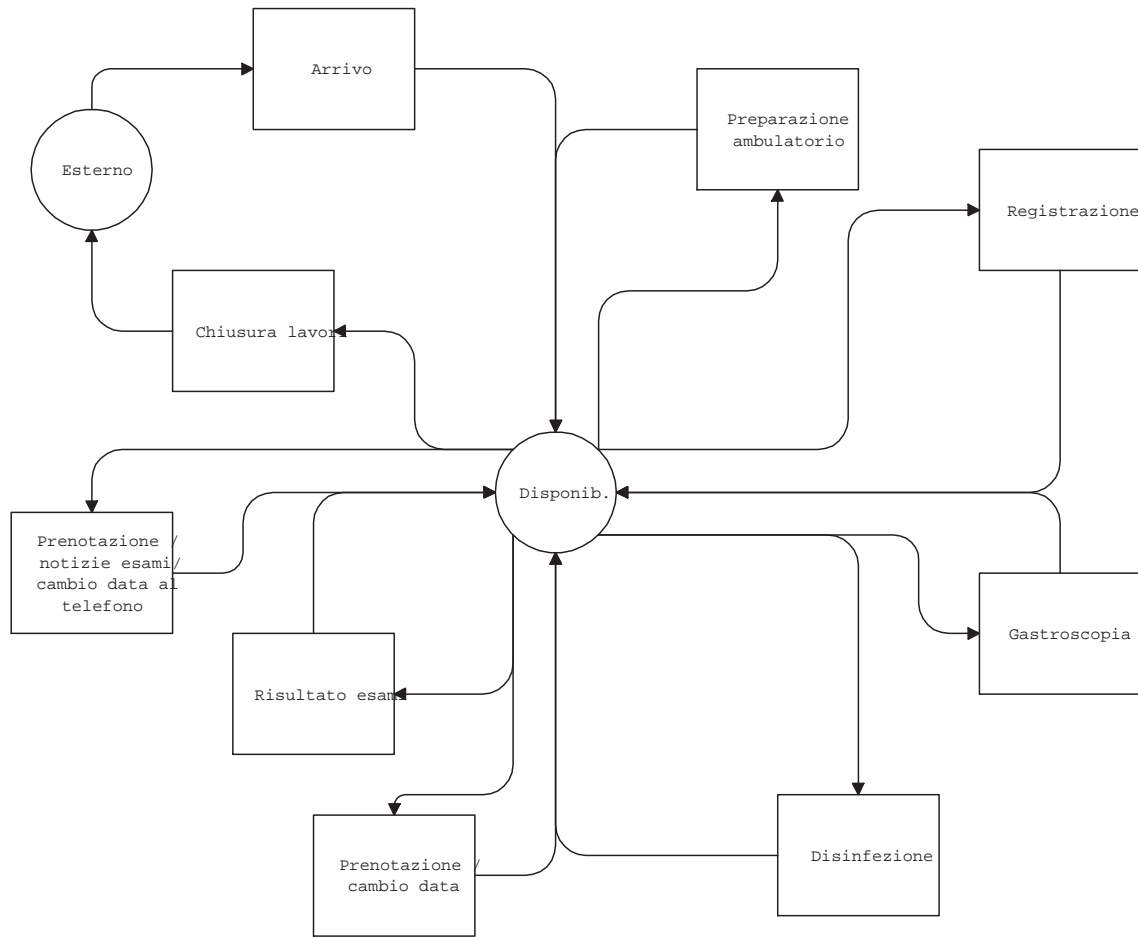


Figura 5.12: Diagramma dei cicli di attività infermiere

Conclusioni

Durante lo svolgimento di questa tesi si è potuto notare come la simulazione sia uno strumento utile e flessibile che permette di comprendere e riprodurre il funzionamento di un sistema reale, rendendo possibile l'analisi delle ripercussioni derivanti da modifiche delle politiche di gestione o strutturali del sistema.

Nella prima Parte della tesi abbiamo descritto in dettaglio le fasi principali dello sviluppo del modello di simulazione del Centro Prelievi.

Il lavoro è iniziato con la raccolta di informazioni attraverso ripetuti incontri con il personale del Centro e diverse osservazioni dirette del sistema. In questa fase è stato di particolare aiuto l'utilizzo dei diagrammi dei cicli di attività che ci hanno permesso di rappresentare graficamente il comportamento delle entità e le caratteristiche principali del sistema. Inoltre, abbiamo potuto appurare che questi diagrammi sono risultati un valido strumento di dialogo, tra noi e gli esperti del sistema, nei momenti di revisione del lavoro svolto. Le informazioni raccolte ci hanno permesso di capire il problema e di definire i confini del sistema stabilendo quali fossero gli elementi che caratterizzavano la sua struttura. Tutto questo ha costituito la base per la costruzione del modello concettuale.

Una volta terminata la costruzione del modello concettuale siamo stati in grado di capire quali fossero le variabili di ingresso che regolavano il comportamento del sistema e abbiamo quindi pianificato la fase successiva riguardante la raccolta dei dati. Uno

dei punti critici è stato quello del reperimento e del coordinamento di un gruppo di dodici persone con il compito di rilevare i tempi durante tutto il tempo di apertura del Centro Prelievi per un periodo di tre settimane.

L'analisi dei dati raccolti ha richiesto notevoli sforzi perché la teoria non fornisce dei metodi canonici sul modo di procedere ed è solo con l'esperienza che si riesce ad ottenere dei buoni risultati. D'altra parte, era necessario avere un modello dei dati di ingresso valido perché da questo dipende in larga misura il successo di uno studio di simulazione. Esistono in commercio dei pacchetti software specifici per l'analisi statistica dei dati ma, non avendone la disponibilità, il lavoro è stato svolto utilizzando il foglio elettronico Microsoft Excel. Perciò, la determinazione di un modello dei dati che superasse la fase di validazione, fatta sfruttando i test statistici di *Goodness-of-fit*, ha richiesto lunghi tempi e diversi tentativi.

La scelta di usare Extend, per l'implementazione del modello, si è rivelata felice perché, rispetto a un linguaggio *general-purpose*, ci ha permesso di contenere i tempi di sviluppo oltre ad essere stato di grande aiuto, con le animazioni e le tracce, sia in fase di *debugging* che in fase di verifica del modello grazie alla possibilità di poter osservare a video l'andamento della simulazione che ha consentito di individuare facilmente le anomalie strutturali. Durante la validazione si sono inoltre rivelate utili le sue capacità statistiche, la possibilità di interagire con altri programmi (con Excel) e la presentazione degli output.

Il risultato di questa parte della tesi è stato quello di aver fornito alla Direzione Sanitaria uno strumento di semplice uso anche da parte di persone poco esperte nel campo della simulazione che consente non solo di monitorare il sistema attuale, ma anche di avere indicazioni su eventuali modifiche da apportare. Il simulatore permette di condurre esperimenti modificando i valori delle variabili controllabili mediante la sua interfaccia utente. Sempre attraverso l'interfaccia è possibile controllare a video i

risultati della simulazione oppure memorizzarli in file con l'opportunità di fare dei controlli incrociati tra diversi possibili scenari. È inoltre possibile, mediante le finestre di dialogo dei generatori casuali (per esempio quelli che regolano il processo di arrivo degli utenti), modificare parametri relativi a fattori esterni (per esempio fattori demografici) e osservare le conseguenze.

Non sarebbe particolarmente difficile, per l'utente finale, intervenire anche al livello della struttura del modello per simulare modifiche strutturali del Centro Prelievi.

La seconda Parte della dissertazione, ha riguardato lo studio degli ambulatori del reparto di Gastroenterologia. Sebbene sia stata un'analisi parziale, causata dalla prematura interruzione, la raccolta delle informazioni e la loro traduzione in modelli concettuali ha richiesto un notevole impegno data la relativa complessità del sistema. Un possibile sviluppo potrebbe essere il suo completamento, considerando il sistema ambulatoriale del reparto nella sua globalità, studiando i percorsi dei diversi tipi di pazienti attraverso gli ambulatori.

Elenco delle figure

1.1	Simboli per i diagrammi dei cicli di attività	19
1.2	Diagramma dei cicli di attività Pazienti	23
1.3	Diagramma dei cicli di attività Distributore Numeri	24
1.4	Diagramma dei cicli di attività Sportello Accettazione	24
1.5	Diagramma dei cicli di attività Macchina Automatica Pagamento Ticket	24
1.6	Diagramma dei cicli di attività Servizi Igienici	25
1.7	Diagramma dei cicli di attività Box Prelievi	25
1.8	Diagramma dei cicli di attività Centro Prelievi	26
2.1	Istogramma dei tempi di interarrivo	48
2.2	Istogramma dei tempi di interarrivo e grafico di \hat{f}	49
2.3	Grafico Quantile-Quantile per i tempi di interarrivo	49
3.1	Connettori di Valori e Item	65
3.2	Un modello Singolo Servente Singola Coda	66
3.3	Modello con Tempo di Servizio Casuale	67
3.4	Discret Event Plotter Aggiunto al Modello	67
3.5	Assegnazione dell'Attributo "Tipo"	68
3.6	Tabella dei Costi del Blocco Activity Delay	69
3.7	Grafico della Lunghezza della Coda	71

3.8	Rappresentazione grafica del modello	75
3.9	Diagramma a blocchi del modello di simulazione	76
3.10	Generazione utenti	77
3.11	Processo di arrivo	78
3.12	Tempi di servizio e coda sala d'attesa	79
3.13	Tempo di servizio accettazione	80
3.14	File sportelli accettazione	81
3.15	Sportelli accettazione	82
3.16	Raccolta tempi di inizio accettazione	83
3.17	Pagamento ticket	84
3.18	Tempi di servizio box	85
3.19	Zona prelievi	86
3.20	Fila box	87
3.21	Box prelievi	88
3.22	Calcolo tempi di attesa accettazione e permanenza	89
3.23	Calcolo tempi di attesa box e macchinette	89
3.24	Calcolo delle statistiche e uscita dal sistema	90
3.25	Dati in ingresso	91
3.26	Simulation setup	92
3.27	Risultati della simulazione	93
3.28	Grafici della lunghezza delle code di attesa	94
3.29	Grafico del tempo di permanenza	95
4.1	Curve caratteristiche operative il t test bilaterale per diversi valori della dimensione del campione n	112
5.1	Diagramma dei cicli di attività pazienti	125

5.2	Diagramma dei cicli di attività medico	126
5.3	Diagramma dei cicli di attività infermiere	127
5.4	Diagramma dei cicli di attività Centro Antidiabetico	132
5.5	Diagramma dei cicli di attività pazienti	134
5.6	Diagramma dei cicli di attività medico	135
5.7	Diagramma dei cicli di attività infermiere	135
5.8	Diagramma dei cicli di attività Ambulatorio e Day Hospital	136
5.9	Diagramma dei cicli di attività paziente	142
5.10	Diagramma dei cicli di attività medico	143
5.11	Diagramma dei cicli di attività macchina	143
5.12	Diagramma dei cicli di attività infermiere	144
5.13	Diagramma dei cicli di attività Gastroscofia	145

Elenco delle tabelle

1.1	Priorità all'accettazione	16
1.2	Priorità ai box	18
2.1	Tabella dei valori critici $c_{1-\alpha}$, $c'_{1-\alpha}$, $c''_{1-\alpha}$	45
2.2	Tabella tempi (minuti) ordinati in ordine crescente	46
2.3	Valori per la costruzione dell'istogramma	47
2.4	Test χ^2 per i tempi di interarrivo	50
2.5	Risultati delle analisi dei tempi di servizio	54
2.6	Risultati delle analisi dei tempi di interarrivo	55
2.7	Probabilità di arrivo dei diversi tipi di utenti	57
4.1	Aperture sportelli accettazione	104
4.2	Aperture box prelievi	105
4.3	Risposte del sistema	106
4.4	Risposte del modello finale	108
4.5	Statistiche per il t -test	113
4.6	Valori $\hat{\delta}$	114
4.7	Validazione con il tempo di permanenza	118
4.8	Validazione con il tempo di attesa accettazione	119
4.9	Validazione con il tempo di attesa box	119

Bibliografia

- [Bau98] Marc Baudoin. *Impara L^AT_EX (... e mettilo da parte)*. École Nationale Supérieure de Techniques Avancées, 32, boulevard Victor, 75739 PARIS cedex 15, 1998.
- [BCN96] Jerry Banks, John S. Carson, and Barry L. Nelson. *Discrete-Event System Simulation*. Prentice Hall, second edition, 1996.
- [Dav85] Ruth Davies. An assessment of models of a health system. *J. Opl Res. Soc.*, 36(8), 1985.
- [DD87] Huw Davies and Ruth Davies. A simulation model for planning services for renal patients in europe. *J. Opl Res. Soc.*, 38(8), 1987.
- [DD95] Huw T.O. Davies and Ruth Davies. Simulating health systems: Modelling problems and software solutions. *European Journal of Operational Research*, (87), 1995.
- [HM90] W.W. Hines and D.C. Montgomery. *Probability and Statistics in Engineering and Management Scienze*. Wiley, New York, third edition, 1990.
- [Iaz78] Giuseppe Iazeolla. *Introduzione alla simulazione discreta*. Boringhieri, 1978.
- [Ima97a] Imagine That, Inc., San Jose, CA. *Extend + BPR User's Manual*, 1997.

- [Ima97b] Imagine That, Inc., San Jose, CA. *Extend + Manufacturing User's Manual*, 1997.
- [Ima97c] Imagine That, Inc., San Jose, CA. *Extend User's Manual*, 1997.
- [Lam94] Leslie Lamport. *TEX: a document preparation system*. Addison-Wesley, second edition, 1994.
- [Lee98] Lawrence Leemis. Input modeling. In *Proceedings of the 1998 Winter Simulation Conference*, volume 1. ed. D.J. Medeiros, Edward F. Watson, John S. Carson and Mani S. Manivannan, 1998.
- [LK91] Averill M. Law and W. David Kelton. *Simulation modeling and analysis*. McGraw-Hill Book Company, second edition, 1991.
- [NF67] T.H. Naylor and J.M. Finger. Verification of computer simulation models. *Management Sci.*, (14), 1967.
- [NHP98] Jalal Nikoukaran, Vlatka Hlupic, and Ray J. Paul. Criteria for simulation software evaluation. In *Proceedings of the 1998 Winter Simulation Conference*, volume 1. ed. D.J. Medeiros, Edward F. Watson, John S. Carson and Mani S. Manivannan, 1998.
- [Pid84] Michael Pidd. *Computer Simulation in Management Science*. John Wiley and Sons, first edition, 1984.
- [Ros97] Sheldon M. Ross. *Simulation*. Academic Press, second edition, 1997.
- [Sar94] Robert G. Sargent. Verification and validation of simulation models. In *Proceedings of the 1994 Winter Simulation Conference*. ed. J.D. Tew, S. Manivannan, D.A. Sadowski and A.F. Seila, 1994.

-
- [Shi97] Jingsheng Shi. A conceptual activity cycle-based simulation modeling method. In *Proceedings of the 1997 Winter Simulation Conference*. ed. S. Andradóttir, K.J. Healy, D.H. Withers and B.L. Nelson, 1997.
- [Ste74] M.A. Stephens. Edf statistics for goodness of fit and some comparison. *J. Am. Statist. Ass.*, (69), 1974.