



Analisi dei dati di ingresso

Simulazione & Logistica, I modulo
Lezione n. 10

Corso di Laurea in Informatica Applicata
Università di Pisa, sede di La Spezia
A.a. 2008/09, I semestre



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

1



Contenuti

- Campioni e distribuzioni empiriche
- Processo di simulazione, dai dati alle distribuzioni
- Validazione dei dati
- Identificazione della distribuzione
- Valutazione della confidenza
- Esempio: l'impiegato tormentato, analisi dei dati



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

2



Variabili casuali e generatori

- Variabile casuale
 - Funzione $X : \Omega \rightarrow \mathbb{R}$
 - Rappresenta un fenomeno non deterministico
- Utilizzo pratico nella modellazione
 - Associazione di un fenomeno a una variabile casuale
 - Analisi del fenomeno (della variabile) tramite osservazioni
 - Riproduzione della variabile (del fenomeno) in simulazione
- Conosciamo e generiamo v.c. con distribuzioni note
- Problema di modellazione: come riconoscerle?



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

3



Variabili casuali in pratica

- Una v.c. è un contenitore di valori casuali
- Data X con distribuzione F , sono “equivalenti”
 - n valori successivi di X
 - i valori di n v.c. X_1, X_2, \dots, X_n tutte con distribuzione F
- Purché le n v.c. X_1, X_2, \dots, X_n siano indipendenti
- Osservazione di un fenomeno
 - Successione di osservazioni dello stesso sistema
 - Osservazione di sistemi simili
 - I dati sono trattati come tante v.c. indipendenti



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

4



Distribuzioni empiriche

- Strumento per trovare distribuzioni non note
 - X_1, X_2, \dots, X_n sono v.c. indipendenti
 - Non è nota la distribuzione F
 - Non sono note media o varianza (dipendono da F)
- Media del campione
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$
- Varianza del campione
$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$
- Cosa cambia rispetto a media e varianza di F ?



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

5



Parametri della v.c. e del campione

- Confronto
$$\mu_X = E[X] = \sum_i x_i f_X(x_i) \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$
$$\sigma_X^2 = \text{Var}[X] = \sum_i (x_i - \mu_X)^2 f_X(x_i) \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$
- Intuitivamente
 - Sommare tutti i valori pesandoli con la loro probabilità
 - Sommare i valori del campione (non tutti e forse replicati)
 - L'assunzione della probabilità uniforme è “affidata” alla frequenza di valori replicati

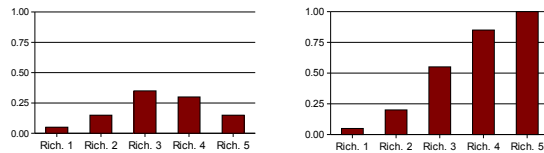


Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

6

Distribuzione campionaria

- Deriva dalla diretta osservazione dei dati
- Il fenomeno è associato a una v.c. discreta
 - Con associazione diretta ai valori finiti
 - Suddividendo il dominio continuo in n intervalli
 - I dati si ordinano e se ne contano le occorrenze
 - Si ricavano le funzioni di densità e di distribuzione



Dati in una simulazione

- Comunque devono esistere dei dati di partenza
 - Osservazioni ad hoc
 - Dati storici
 - Situazioni simili
- Uso di dati originali nella simulazione
 - Riproducono le osservazioni, con le stesse sequenze
- Uso di distribuzioni campionarie
 - Riproducono sequenze diverse, ma solo dei valori osservati
- Distribuzioni notevoli con parametri stimati
 - Ricavate osservando il campione
 - Riproducono sequenze e valori diversi

Dai dati alla distribuzione

- Recupero e validazione dei dati
 - Ricavare i dati
 - Verificare l'indipendenza delle osservazioni
- Identificazione del tipo di distribuzione
 - Costruzione della distribuzione empirica
 - Analisi dell'istogramma
 - Media e varianza campionarie come stima dei parametri
- Valutazione della confidenza
 - Intervalli di confidenza
 - Test di ipotesi
- Scelta e inizializzazione dei generatori



Indipendenza delle osservazioni

- Coefficiente di correlazione
 - Rapporto fra covarianza e prodotto delle varianze $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$
- Stima
 - Misura della correlazione
 - Osservazioni distanti j
 - Deve essere prossimo a zero
 - Analisi al variare di j
- Scatterplot
 - Connotazione qualitativa della correlazione
 - Visualizzazione grafica della correlazione
 - La disposizione deve essere irregolare



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

10



Analisi dell'istogramma

- Densità discreta della distribuzione empirica
- Confronto con
 - Funzioni di densità di distribuzioni notevoli
 - Istogrammi discreti vs curve continue
- Identificazione
 - Del tipo di distribuzione
 - Del dominio di definizione
 - Dei parametri
- Osservazione
 - Della forma
 - Dell'assenza di valori, "buchi" e "diradamenti" agli estremi
 - Dei punti notevoli della funzione



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

11



Stimatori dei parametri

- La media del campione è uno stimatore di μ
$$E[\bar{X}_n] = \mu$$
- La varianza del campione è uno stimatore di σ^2
$$E[S_n^2] = \sigma^2$$
- L'accuratezza della stima cresce al crescere di n

$$Var[\bar{X}_n] = \frac{1}{n} \sigma^2$$



Giovanni A. Cignoni - SLo1: Simulazione - www.di.unipi.it/~giovanni/

12

Intervallo di confidenza

- Stimare la media di X
 - Con fiducia che l'errore sia entro un certo intervallo
 - Ovvero, con probabilità α che la differenza fra la media campionaria e la media stimata della distribuzione sia $> d$
 - Ovvero, con probabilità $1 - \alpha$ che $\mu = \bar{X}_n \pm d$
- Quanti valori di X occorrono?
 - Z variabile casuale normale standard
 - c tale che $P(Z < c) = 1 - \alpha/2$
 - occorrono k valori, con k tale che $c \frac{S_k}{\sqrt{k}} < d$



Test di ipotesi

- Problema: determinare la correttezza di un'ipotesi
 - Per accettarla con ragionevole fiducia
 - Probabilità α di sbagliare accettando l'ipotesi
 - Probabilità $1 - \alpha$ di azzeccare accettando l'ipotesi
- Test Chi-Quadro (χ^2)
 - "Distanza" fra le densità del campione e della distribuzione
 - Calcolabile come funzione primitiva di molti strumenti di calcolo
- Valori canonici di α : 0.05, 0.01, 0.001
- Il test si fa dare una valutazione a un'ipotesi, non si fa il test per cercare l'ipotesi che "funziona meglio"



Riferimenti

- A.M. Law, W.D. Keaton, *Simulation Modeling & Analysis*, Cap. 4, McGraw-Hill, 1991
- G. Gallo, *Note di Simulazione*, capp. 3.2, 3.3, 4.1 - 4.3

