Simulare scenari probabilisitici elementari

- Andrea Marin - Università Ca' Foscari – Corso di laboratorio di Probabilità e Statistica

Dal generatore di numeri casuali in [0, 1) a...

- La maggior parte dei linguaggi fornisce un generatore di numeri casuali in grado di estrarre sequenze di numeri nell'intervallo [0,1) con probabilità uniforme
- Come possiamo estrarre numeri casuali...
  - In intervelli diversi?
  - Con distribuzioni non uniformi?

Trasformazioni di v.c.

- Possiamo vedere un'estrazione di un numero pseudo-casuale come una variabile aleatoria
- L'obiettivo è quello di trasformare una variabile aleatoria uniforme in [0,1) in una variabile di nostro gradimento
- Questo argomento richiederà maggiori approfondimenti, oggi tratteremo alcuni aspetti fondamentali

Dal continuo al discreto

- Abbiamo a disposizione un generatore di numeri pseudo-casuali nell'intervallo reale [0,1)
- Vogliamo ottenere un numero intero nell'intervallo [0, T-1]
- La trasformazione richiesta prevede di passare da v.c. Continua a v.c. Discreta
- Soluzione semplice: se r è il numero estratto basta calcolare: floor(r*T)

Connessioni con la teoria...

- Consideriamo il caso in cui tutti gli eventi elementari dello spazio campionario siano equiprobabili
- Assumiamo che lo spazio contenga T eventi e_i, $i=1,...,T$
- Vogliamo determinare la probabilità associata ad un evento E definito come un sottoinsieme proprio di questi eventi elementari

Nella simulazione...

- Generiamo numeri ”reali” r nell'intervallo [0,1)
- Calcoliamo s=floor(r*T)
- Nel caso in cui r*T stia in [i-1,i) allora l'intero estratto sarà i-1
- L'evento elementare estratto è l'evento e_i
- Noto l'evento estratto e_i possiamo testare la condizione ”e_i appartiene ad E”

Algoritmo fondamentale

- Inizializza il contatore di eventi: n_E=0
- Per un totale di ”n” prove
  - * Effettua un esperimento
  - * Se è accaduto l'evento E allora incrementa n_E
- Stima la probabilità di E come n_E/n

- Lo svolgimento dell'esperimento significa estrarre un intero casuale e decidere quale evento elementare dello spazio campionario è accaduto.

Un primo semplice esempio

- Stimare la probabilità di ottenere un punteggio inferiore o uguale a 3 lanciando due dadi bilanciati
- Abbiamo 36 eventi elementari equiprobabili:
  - (1,1)->0 (2,1)->1 (1,2)->2.... (6,6)->35
- L'evento E cercato corrisponde all'intervallo [0,2]
- Nella simulazione estraiamo un numero casuale in [0,1)

Conteggio

- Se r è il numero estratto, calcoliamo s = floor(36*r)
- Se 0 <= s <= 2 allora incrememementa il contatore di eventi
- Al termine del numero di prove prefissate n calcola n_E/n
- Come determinare n?

Decidere il numero di esperimenti

- Si desidera determinare per simulazione la probabilità ti ottenere esattamente 3 teste in 5 lanci consecutivi di una moneta
- Mapping degli eventi elementari C-> 0, T ->1
- Estratto r, calcoliamo s = foor(r*2)

Pseudocodice del simulatore

- n_E=0
- Ripeti n volte
  - Heads = 0
  - Ripeti 5 volte
    - S = floor (rand * 2)
    - Heads = s + Heads
  - Se Heads == 3 allora n_E = n_E + 1;
- Fine for
- Prob = n_E/n

Risultati

- Simulazioni:
- Valore di n?

n	50	100	1.000	10.000	50.000
n_E	18	28	321	3.165	15.631
n_E/n	0.3600	0.2800	0.3210	0.3165	0.31262

Simulazione di probabilità condizionate

- $Pr\{A|B\}$ è la probabilità dell'evento A noto che si è verificato B
- Molto utile in pratica (specie se abbinato alla legge delle probabilità totali)
- Se B è un sottoinsieme di Ω e $B^c=\Omega \setminus B$
- Allora $Pr\{A\}=Pr\{A \cap B\} + Pr\{A \cap B^c\}$
- Ma $Pr\{A \cap B\} = Pr\{A|B\}Pr\{B\}$ quindi:
- $Pr\{A\}=Pr\{A|B\}Pr\{B\} + Pr\{A|B^c\}Pr\{B^c\}$

Esempio trivial

- Dato un mazzo di carte. Si pesca una carta e non la si guarda. Quindi si pesca la seconda carta. Qual è la probabilità che questa sia un asso?
- Ovviamente la risposta è 1/10
- Dal punto di vista delle probabilità totali: A evento di pescare un asso alla seconda pescata. B evento di aver pescato un asso dalla cima del mazzo. B^c evento di aver pescato una carta diversa da un asso dalla cima del mazzo.

Esempio trivial (continuazione)

- $Pr\{B\}=1/10$
- $Pr\{B^c\}=9/10$
- $pr\{A|B\} = 3/39$
- $pr\{A|B^c\} = 4/39$
- Quindi: $Pr\{A\}=3/39*1/10+4/39*9/10=39/390=1/10$

Algoritmo di simulazione per la probabilità condizionata Pr{A|B}

- n_A=n_B=0
- Ripeti ”n” esperimenti
  - Esegui l'eperimento
  - Se l'evento B accade
    - Incremementa n_B
    - Se l'evento A accade
      - Incrementa n_A
- Pr{A|B}=n_A/n_B

Esempio

- Abbiamo una scatola con 2 palline bianche, 2 rosse e una nera
- Si estraggono due palline senza ripetizione
- Si vuole stimare la probabilità che la seconda pallina sia bianca dato che la prima è bianca
- A={seconda pallina bianca}
- B={prima pallina bianca}
- Stima di Pr{A|B}

Pseudo-codice

- n_A=n_B=0
- Ripeti n volte
  - W=2; R=2; B=1;
  - s=floor(rand * (W+R+B))
  - If (s < W) then
    - Incrementa n_B
    - s=floor(rand * (W+R+B-1))
    - If (s < W-1) then incrementa n_A
- Pr{A|B} = n_A/n_B

Tabella delle simulazioni

- Come al solito il problema consiste nel determinare n:
- Tabella

n	50	100	1.000	10.000	100.000	500.000	1E6
n_A	1	13	94	1.028	9.921	50.173	99.803
n_B	20	37	403	3.972	39.923	199.752	399.812
n_A/n_B	0.0500	0.35135	0.23325	0.25881	0.25026	0.25118	0.24962

Simulazioni per la stima della media

- In questo caso dobbiamo adottare un approccio differente da quello usato per la stima delle probabilità
- Ripetiamo l'esperimento un grande numero di volte
- Annotiamo il valore della v.c. che ci interessa
- Per ottenere il valore atteso della v.c. calcoliamo la media aritmetica di questi valori

Esempio

- Un'urna contiene b palline bianche e n nere
- Le palline sono estratte dall'urna con reinserimento fino a che non compare una pallina bianca
- Stimare la probabilità di di dover eseguire un numero di pescaggi pari a x
- Stimare la media del numero di pescaggi

Soluzione

- Il numero di pescaggi X è una v.c. con distribuzione geometrica
- La probabilità di dover fare x pescaggi Pr{X=x} è data dalla probabilità di pecare x-1 palline nere consecutive e quindi una bianca.
- Sia p=b/(n+b)
- Gli esperimenti sono indipendenti
- $Pr\{X=x\}=(1-p)^{x-1}*p$
- La media dei pescaggi è 1/p

Altro esempio con la geometrica

- In una strada passa un'auto ogni dieci minuti
- Supponiamo di osservare la strada in un tempo ”discretizzato”, ogni minuto.
- Il numero di minuti tra un'auto e la successiva è distribuito geometricamente con media 1/p=10 minuti
- Gli automobilisti sono molto gentili e si fermano sempre a raccogliere gli autostoppisti
- Un autostoppista sa che in quella strada passa mediamente una macchina ogni dieci minuti, ed arriva in un certo istante.
- Quanto dovrà mediamente attendere?

Un esempio di performance evaluation

- Due stazioni comunicano su una linea soggetta ad errori indipendenti. La probabiltà di trasmettere un bit errato è p
- Suddividiamo i bit da trasmettere in frame ciascuno dei quali comporta un certo overhead nella trasmissione dovuto alla presenza dell’header (destinatario, mittente ecc...) e dei bit necessari per il controllo dell’errore (CRC).

Come si decide la grandezza di un frame?

- Se scelgo un frame di grandi dimensioni
  - Riduco l'overhead
  - E' più probabile che il frame subisca un errore e che quindi debba essere ritrasmesso
- Se scelgo un frame di piccole dimensioni
  - Aumento l'overhead
  - La ritrasmissione prevede l'invio di pochi bit
- Vogliamo studiare l'efficienza, cioè il numero medio di bit significativi inviati, rispetto al numero totale inviato

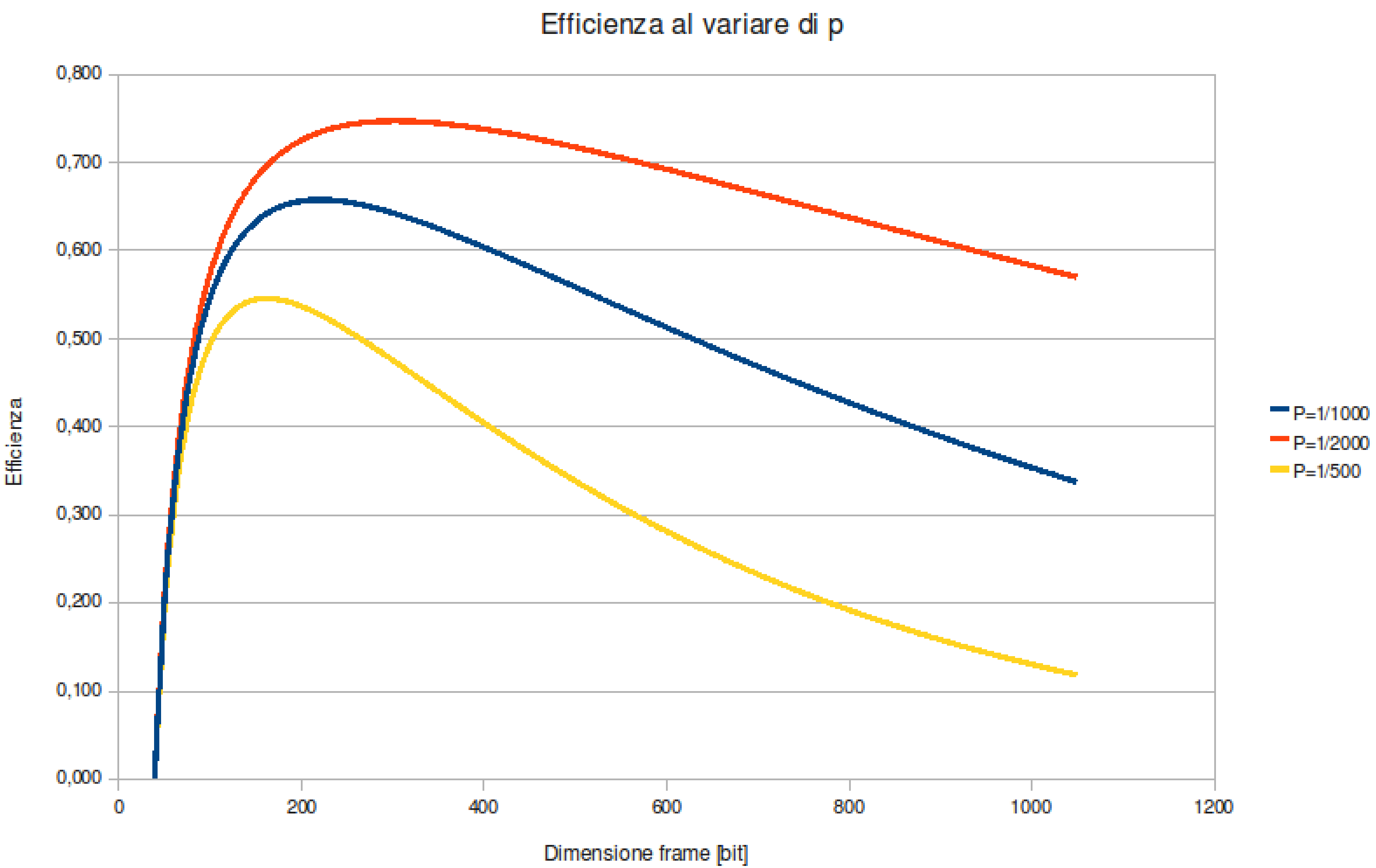
Definizione del problema

- p probabilità di un bit errato
- F dimensione di un frame in bit
- H dimensione dell'overhead di un frame in bit
- E[N]numero medio di trasmissioni di un frame perchè il destinatario lo riceva correttamente
- Eff = (F-H)/(NF)

Calcolo di N

- Probabilità di trasmissione di un frame corretto è $Pc=(1-p)^F$
- Quante trasmissioni sono necessarie per far arrivare un frame corretto? N è una v.c.:
  - $N \sim (1-Pc)^{n-1} Pc$
- La distribuzione è geometrica!
- Ricaviamo $E[N]=1/Pc$
- Da cui $eff=(F-H)(1-p)^F/F$

Dimensione ottimale di un frame?



Ma allora la simulazione a che serve?

- I risultati analitici in pratica sono piuttosto rari
- Nell'esempio precedente è difficile che gli errori siano indipendenti (correlazione degli eventi)
- La simulazione è in grado di studiare un ampio spettro di casi
- Problemi nell'interpretazione dei risultati...

Nuovo esercizio

- In un casinò si svolge il seguente gioco con carte. Il giocatore estrae dal mazzo una carta dopo l'altra (si assuma reinserimento e rimescolamento del mazzo) fino a quando pesca una carta di seme rosso (cuori o quadri). A questo punto il punteggio del giocatore è dato dal valore più alto delle carte pescate (fante=11, donna=12, re=13). Il giocatore vince se il suo punteggio è maggiore di N.
- Si determini per simulazione il più piccolo valore intero di N che consenta al casinò un margine di guadagno
- Si ripeta la simulazione assumendo che il giocatore peschi sempre 3 carte (con reinserimento), confrontando il valore di N stimato dalla simulazione con quello calcolato teoricamente.
- Consegna con le solite modalità fra 15 gg.