# Campionamento di variabili aleatorie

- Andrea Marin

  Università Ca' Foscari Venezia
- Corso di Probabilità e Statistica a.a. 2009/2010

# Premessa

- Soluzione della prima esercitazione: l'analisi teorica
- Richiami: un esperimento consiste nel lancio di un dato equilibrato. Se il valore ottenuto è compreso tra 4 e 6, l'esito dell'esperimento è quel valore. Altrimenti si rilancia il dado è il valore ottenuto nell'esperimento è dato dalla somma del primo lancio e del secondo, in caso il totale sia minore o uguale di 6. Altrimenti il lancio non è valido

# Impostazione

Evento A: lancio primo dado con esito $X_1 = 4,5,6$

Evento B: lancio secondo dado con esito tale che $X_1 + X_2 \leq 6$ e $X_1 \in \{1,2,3\}$

I due eventi sono disgiunti, per cui vale la relazione:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

# Determinazione della probabilità di successo

- Chiaramente Pr(A)=1/2
- Pr(B)=Pr($X_1+X_2 \leq 6$ | $X_1 \leq 3$)Pr($X_1 \leq 3$)
- Se $X_1 = 1$, $X_2$ può assumere valori 1,2,3,4,5
- Se $X_1 = 2$, $X_2$ può assumere valori 1,2,3,4
- Se $X_1 = 3$, $X_2$ può assumere valori 1,2,3
- Pr($X_1 = 1$|$X_1 \leq 3$ )=1/6*2=1/3
- Pr(B)=(5+4+3)*1/18*1/2=1/3
- Pr(A $\cup$ B) = 1/3+1/2 = 5/6

# Determinazione della media degli eventi validi

- Definiamo la v.c. Y come segue:

$$Y = \begin{cases} X_1 & \text{se } 4 \leq X_1 \leq 6 \\ X_1 + X_2 & \text{se } 1 \leq X_1 \leq 3 \end{cases}$$

- La probabilità di y_i = 1/36, 2/36, 9/36, 9/36, 9/36, 3/36, 2/36, 1/36 per i=2,...,9
- Pr(Y$=$i|Y$\leq 6$ )=Pr(Y$\leq 6$ | Y$=$i)Pr(Y=i)/Pr(Y$\leq 6$)
- Questa espressione per i=2,6 vale Pr(i)*6/5.
- Da cui calcoliamo la media: E(Y|Y≤ 6)=(2*1+3*2+4*9+5*9+6*9)/30=143/30

# Definizione del problema

- Variabili casuali distribuite diversamente
- Il generatore di numeri casuali fornito con la maggior parte delle librerie/software è assimilabile ad un v.c. continua Uniforme nell'intervallo [0,1)
- Problema: come possiamo generare v.c. (discrete o continue) con distribuzioni diverse da quella uniforme?

# Esempi già visti

- Problema già affrontato intuitivamente per casi semplici
- Generare un numero naturale casuale Y tra [2,6]:

$$X \sim \mathcal{U}[0, 1)$$

$$Y = \lfloor X * 5 \rfloor + 2$$

$$\Rightarrow Y \sim U[2, 6]$$

# La variabile di Bernoulli

- Sia p la probabilità che un evento si verifichi
- La v.c. di Bernoulli Y è definita come:

$$\Pr\{Y = 1\} = p \qquad \Pr\{Y = 0\} = 1 - p$$

- Sia X una v.c. Aleatoria uniforme in [0,1)
- Possiamo ottenere Y con la seguente trasformazione:
- $$Y = \mathbf{1}_{X > 1 - p}$$
- Dove 1 rappresenta la funzione indicatrice

# Generazione di v.c. continue

- Ipotesi di lavoro:

$$X \text{ v.c. continua a valori in } \mathbf{R}$$

$$F_X(x) = \Pr\{X \leq x\}$$

- Com'è possibile generare campionamenti della v.c. X a partire dai campionamenti di una uniforme in [0,1)?

# Un primo passo

- Data la v.c. X e la sua funzione di ripartizione $F_X$ dimostriamo che la v.c. U definita come:

$$U = F_X(X)$$

- Ha distribuzione uniforme nell'intervallo [0,1]

# Dimostrazione caso semplice

- Supponiamo $F_X$ *strettamente* monotona crescente
- Dimostriamo $F_U(u)=u$ assumendo $0 \leq u \leq 1$

$$F_U(u) = \Pr\{U \leq u\} = \Pr\{F_X(X) \leq u\}$$

$$= \Pr\{X \leq F^{-1}(u)\} = F(F^{-1}(u)) = u$$

I casi in cui $u \notin [0,1]$ sono banali

# Caso generale

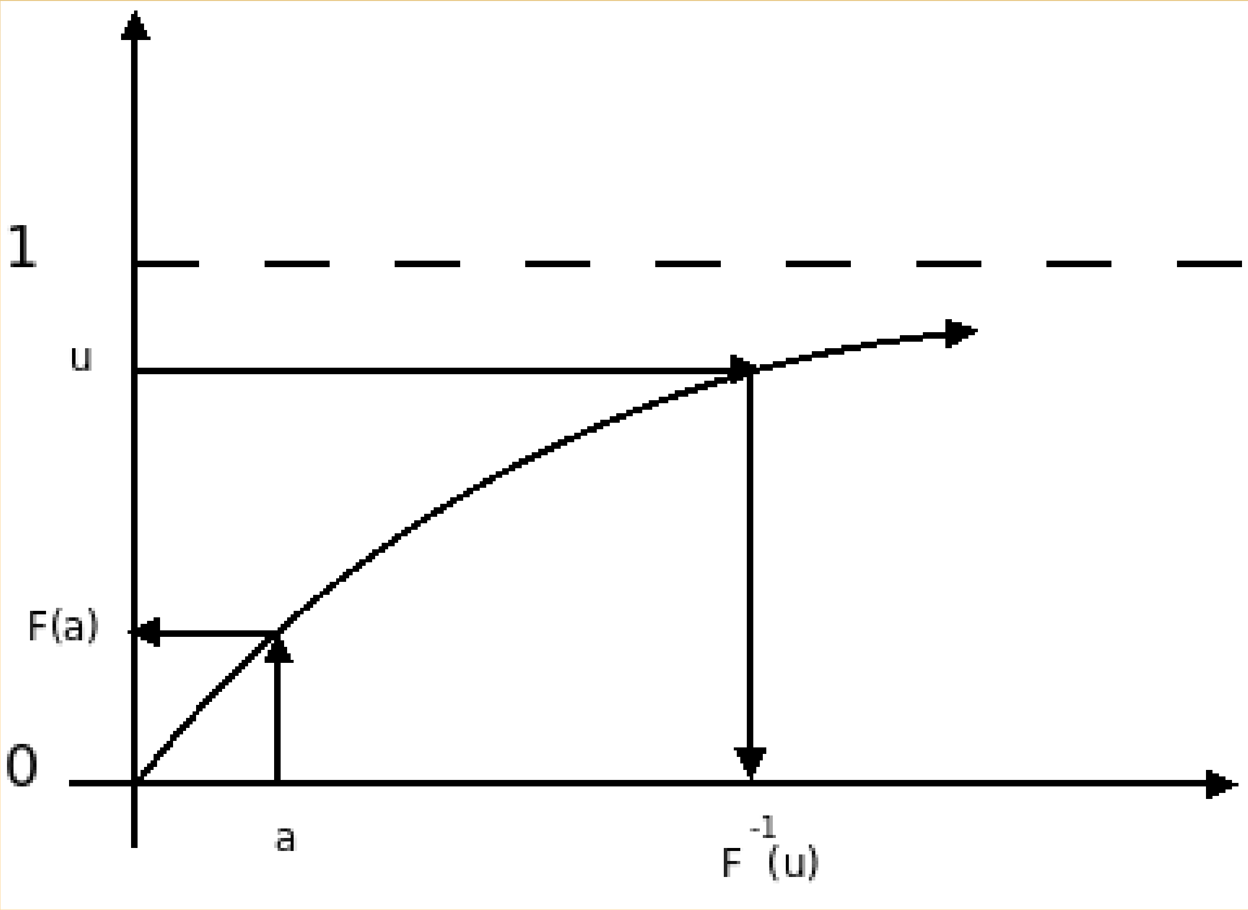
- $F_X(x)$ monotona crescente (non necessariamente strettamente)
- Definiamo:

$$F^{-1}(U) = \inf\{x : F(x) = u, 0 < u < 1\}$$

- Quindi possiamo procedere in modo analogo al precedente

# Metodo di inversione

- Campioniamo una v.c. U in [0,1)
- Quindi poniamo:
  $$X = F^{-1}(U)$$
- X è una v.c. continua la cui funzione di ripartizione è F



# Esempio 1

- Generare una v.c. Uniforme nell'intervallo [3, 5]
- In questo caso $f_X(x) = 1/(5-3) = 1/2$ per $0 \leq x \leq 5$ (0 altrove)

La funzione di ripartizione per $0 \leq x \leq 5$ è:

$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz = \left[\frac{1}{2}u\right]_3^x = \frac{1}{2}x - \frac{3}{2}$$

- Applicando l'inversione:

$$F_X^{-1}(x) = 2u + 3$$

# Esempio 2

- Vogliamo campionare da una v.c. X distribuita in modo esponenziale con parametro λ
- Funzione di ripartizione:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

- Invertiamo per x>0, ottenendo:

$$1 - e^{-\lambda x} = u \quad x = -\frac{1}{\lambda}\ln(1-u)$$

# Esempio 2 (parte 2)

- Osserviamo che se U è una v.c. Uniforme in [0,1] così lo è anche 1-U. Per cui possiamo scrivere:

$$X = -\frac{1}{\lambda}\log(U)$$

# Curiosità sull'esponenziale (Esempio del professore pigro)

- Un lampadario possiede due bulbi
- La durata dei bulbi è indipendente ed entrambe sono distribuite in modo esponenziale con parametro $\lambda$
- Quando una delle due lampadine di brucia il professore pigro non la sostituisce ma le cambia entrambe quando sono entrambe rotte
- Supponiamo di voler determinare, per simulazione, per quanto tempo il professore sta con una sola lampadina

# Impostazione della simulazione

- Campioniamo la durata della prima lampadina $X_1$
- Campioniamo la durata della seconda $X_2$
- Il tempo T in cui il professore sta con con una sola lampadina è:

$$T = \max(X_1, X_2) - \min(X_1, X_2)$$

# Ricampionamento?

- Supponiamo di voler calcolare l'istante di tempo in cui si guasta la prima lampadina:

$$T_G = \min(X_1, X_2)$$

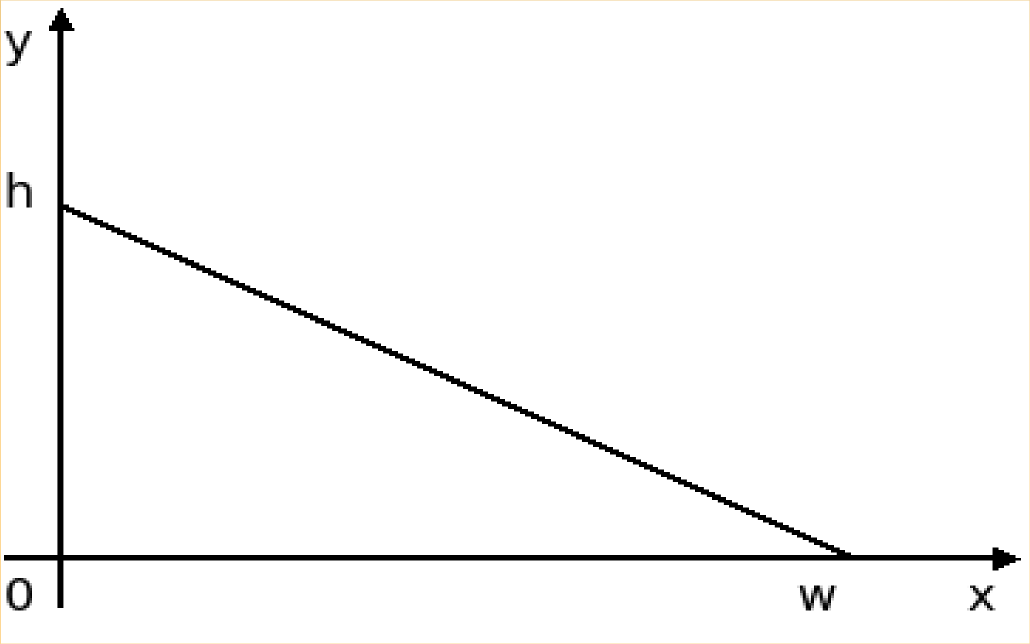
- A questo punto la vita residua della seconda lampadina è la v.c. Y la cui funz. di ripartizione è:

$$\Pr\{Y \leq y\} = \Pr\{X_i \leq y | X_i > T_G\}$$

- Com'è distribuita Y? Possiamo ricampionare?

# Task (parte A) /1

- Si consideri una v.c. Continua la cui funzione di densità è illustrata in figura:




- Con il vincolo $hw=2$

# Task (parte A) /2

- Formalmente la funzione di densità è:

$$f_X(x) = \begin{cases} (1 - x/w)h & 0 \leq x \leq w \\ 0 & \text{altrove} \end{cases}$$

- Scrivere in Matlab/Octave un generatore di numeri pseudo-casuali discributi secondo questa funzione di densità.
- Confrontare la funzione di ripartizione stimata con il campionamento con quella teorica.

# Campionamento di variabili casuali discrete

- Il metodo di inversione si applica anche a v.c. Discrete con un po' di attenzione...
- Consideriamo una v.c. Discreta X su m valori dove:

$$\Pr\{X = x_j) = p_j, \qquad \sum_{j=1}^{m} p_j = 1$$

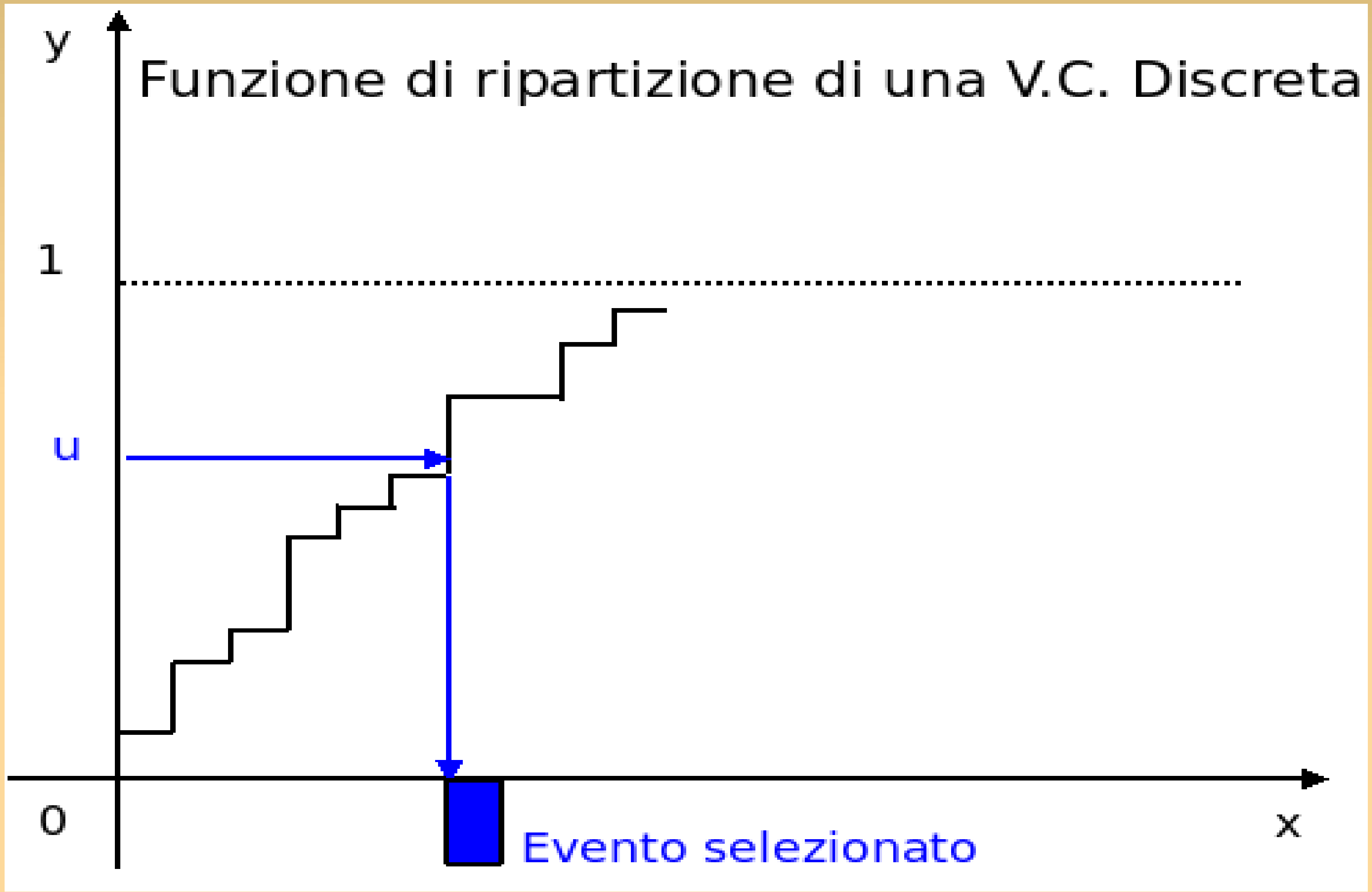
$$F_k = \Pr(X \leq x_k) = \sum_{j=1}^{k} p_j$$

# Metodo di inversione

- Sia U una v.c. discreta in [0,1], allora la v.c. X può essere definita come:

$$X = \begin{cases} x_1 & \text{se } U \leq F_1 \\ x_2 & \text{se } F_1 < U \leq F_2 \\ \vdots & \\ x_m & \text{se } F_{m-1} < U \leq F_m \end{cases}$$

# Intuizione grafica



# Campionamento della v.c. di Poisson per ricorsione

- Citazione: *If distributions were graded on a scale of one to ten, the Poisson clearly merits a ten.* [Taylor, Karlin]
- Perchè tanta importanza?
  - Legge degli eventi rari
  - Alta trattabilità matematica

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, \ldots$$

# Generazione di v.c. Di Poisson per inversione

- Osserviamo la seguente relazione ricorsiva tra le probabilità:

$$p_0 = e^{-\lambda} \quad p_{i+1} = \frac{\lambda}{i+1} p_i$$

- Sia u il valore estratto dal generatore pseudo-causale
- Se u$\leq$F$_i$ allora sarà estratto il valore i
- Altrimenti calcoliamo p$_{i+1}$ e F$_{i+1}$=F$_i$+p$_{i+1}$ e ripetiamo il test del punto precedente

# Problemi del precedente approccio

- Funziona bene quando il parametro è grande (rispetto alla precisione numerica)
- In caso di parametri bassi possono esserci problemi di stabilità numerica (nel calcolo della funzione di ripartizione $F_i$ )

# Richiami sulla v.c. geometrica

- Se p è la prob. di successo di un evento e (1-p) quella di insuccesso, la probabilità di k esperimenti indipendenti per ottenere il primo successo è:

$$p_X(k) = p(1-p)^{k-1} \qquad k = 1, 2, \ldots$$

- La funzione  di ripartizione è data da:

$$F_X(k) = \Pr\{X \leq k\} = \sum_{i=1}^{k} p(1-p)^{i-1} = 1 - (1-p)^k$$

# Generazione per definizione

```
K = 0;
do
   K = K + 1;
   U = rand;
Until (U>(1-p))

K è il numero estratto
```

# Generazione per inversione

- Invertiamo la funzione di ripartizione, sia q=1-p

$$1 - q^k = u \;\Rightarrow q^k = 1 - u$$

$$k = \log_q(1-u) = \frac{\log(1-u)}{\log q}$$

$$k = \lfloor \frac{\log(u)}{\log(1-p)} \rfloor + 1$$

# Task (parte B)

- Si vuole campionare la v.c. Binomiale in cui:

$$p_i = \binom{m}{i} p^i (1-p)^{m-i} \quad i = 1, 2, \ldots, m$$

- Dopo aver stabilito una relazione ricorsiva tra $p_{i+1}$ e $p_i$ si implementi il campionamento per inversione
- Implementare il campionamento in base alla definizione di binomiale, ovvero contando il numero di successi su m esperimenti indipendenti di Bernoulli in cui un successo ha probabilità p

# Task (Parte B)

- Confrontare le distribuzioni dei campionamenti (a parità di parmetri devono dare la stessa distribuzione)
- Effettuare 20000 campionamenti di una variabile binomiale con parametri (8, 0.4) e stabilire quale delle due implementazioni è più efficiente (hint: usare il comando Octave etime, clock)