

PROBABILITÀ - SCHEDA N. 3 VARIABILI ALEATORIE CONTINUE E SIMULAZIONE

(da un'idea di M. Impedovo "Variabili aleatorie continue e simulazione" Progetto Alice n. 15, 2004)

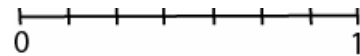
1. La simulazione

Nelle schede precedenti abbiamo usato il calcolatore per generare numeri casuali per simulare il lancio di un dado, di una moneta, eccetera. Molti software possiedono un comando per generare un numero "casuale" compreso tra 0 e 1. Si tratta, naturalmente, di numeri generati da un algoritmo e per questo sarebbe meglio chiamarli "pseudo-casuali".

Che cosa si intende per "numero casuale compreso tra 0 e 1"? Anche se dal punto di vista intuitivo l'espressione sembra chiara, non è facile precisarla. Siamo tentati di dire qualcosa del tipo "un numero appartenente a un insieme di numeri che hanno tutti la stessa probabilità di essere estratti". Le cose tuttavia sono un po' più complicate, perché la probabilità di estrarre un certo numero reale da un intervallo è sempre 0: non c'è modo di assegnare a ciascun numero reale dell'intervallo $[0, 1]$ una probabilità positiva in modo che la somma delle probabilità sia 1.

Un'immagine adeguata potrebbe essere la seguente: vogliamo scegliere, sul segmento $[0, 1]$, un numero reale (un punto) x in modo tale che, comunque si fissi un numero naturale M , x possa cadere con uguale probabilità in uno degli M sotto-intervalli individuati dai punti:

$$0, \frac{1}{M}, \frac{2}{M}, \dots, \frac{M-1}{M}, 1$$



In questo caso si dice che il numero casuale ha una **distribuzione uniforme**.

Un calcolatore non gestisce i numeri reali ma un opportuno sottoinsieme (finito) di numeri razionali: in linea di principio potremmo assegnare a ciascuno di essi una probabilità positiva, ma non è ciò che ci interessa. Noi vogliamo simulare la scelta di un punto a caso sul segmento $[0, 1]$; il generatore di numeri casuali può servire allo scopo.

Generiamo n numeri casuali da una distribuzione uniforme in $[0, 1]$, e consideriamoli la realizzazione di una variabile aleatoria A .

Abbiamo utilizzato il software Minitab per generare le realizzazioni campionarie di una variabile aleatoria uniforme per $n=100$, $n=500$, $n=1000$ e $n=10000$. In fondo alla scheda sono riportate le realizzazioni che abbiamo ottenuto per $n=1000$.

Di seguito sono riportati i grafici

- delle frequenze assolute delle realizzazioni campionarie (grafici 1); fate attenzione alla diversa scala;
- delle frequenze relative in percentuali (grafici 2); questi quattro grafici sono più facilmente confrontabili fra loro.

Tutti gli istogrammi sono stati costruiti con un ugual numero di classi di ampiezza 0.01.

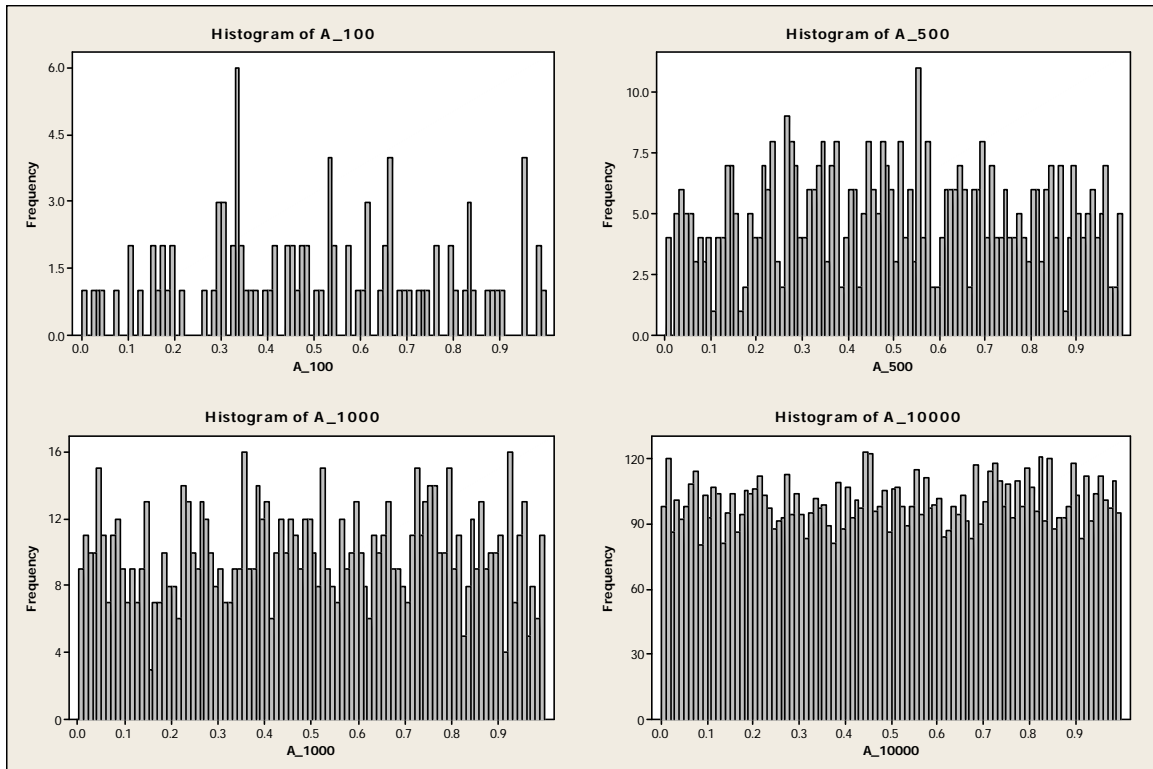


Figura 1 – Frequenze assolute

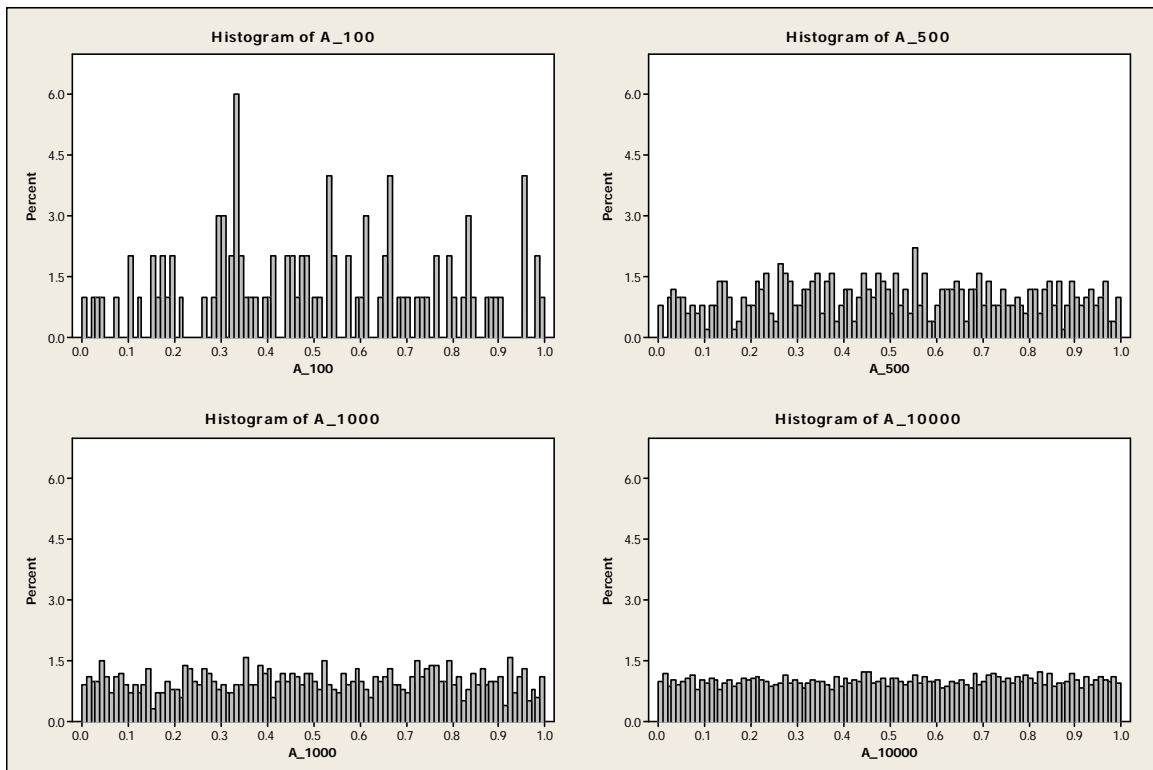


Figura 2 – Frequenze percentuali

Dagli istogrammi si vede che aumentando il numero n di generazioni le frequenze relative delle realizzazioni in ciascuna classe tendono a essere uguali.

Consideriamo ora i grafici delle frequenze cumulate empiriche per le 4 simulazioni. Anche in questo caso si osserva che aumentando il numero n di generazioni il grafico assomiglia sempre di più a una retta. Quale equazione ha la retta?

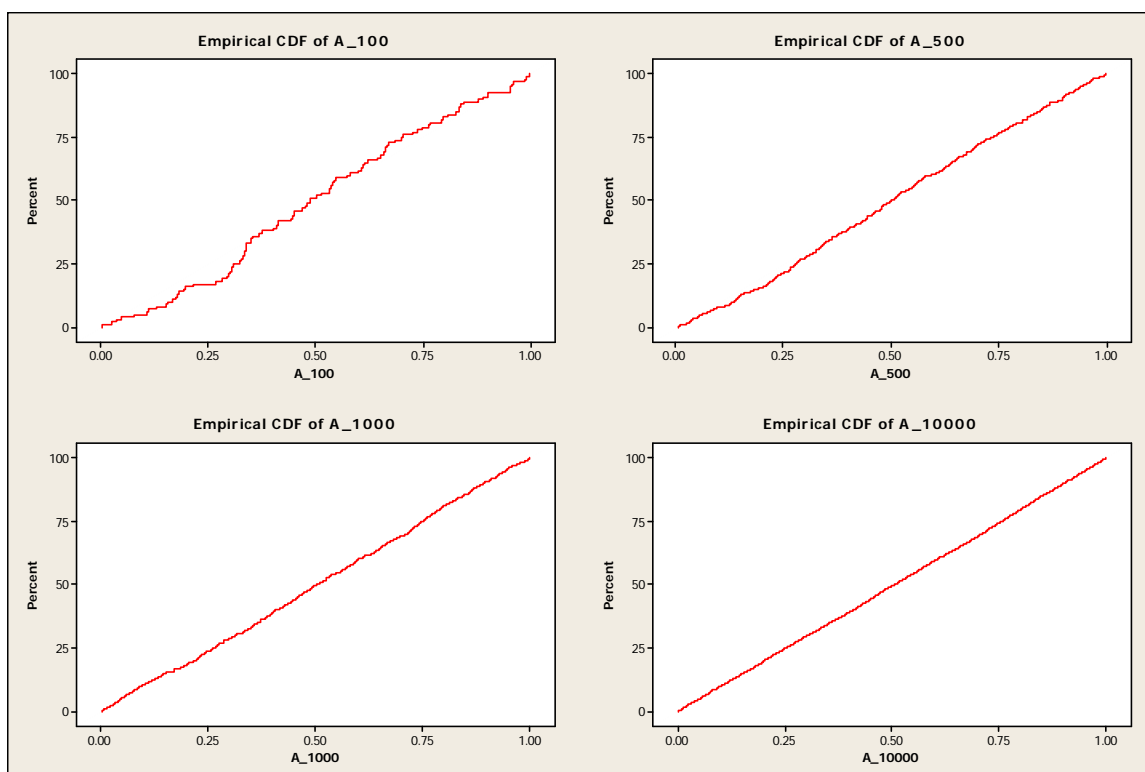


Figura 3 – Funzione di distribuzione cumulata (in percentuale)

Possibili esperienze (descritte in fondo alla scheda)

- E1. Simulazione di estrazione di numeri casuali con distribuzione uniforme in $[0, 1]$
- E2. Simulazione di estrazione di numeri casuali con distribuzione uniforme in $[4, 8]$

Trasformazioni di variabili aleatorie

Per quanto osservato nella seconda esperienza possiamo ipotizzare che se Y è una variabile aleatoria con distribuzione uniforme in $[0,1]$, anche

$$aY + b$$

ha distribuzione uniforme. Gli estremi dell'intervallo della nuova variabile sono $[b, a+b]$; si ottengono trasformando gli estremi dell'intervallo di Y con la stessa trasformazione:

$$\text{limite sinistro: } a \times 0 + b$$

$$\text{limite destro: } a \times 1 + b$$

Che cosa succede se la trasformazione non è lineare? Per esempio se X rappresenta l'area di un quadrato aleatorio costruita a partire da una variabile Y che modella il lato del quadrato. In tal caso $X = Y^2$. Vedremo un esempio simile nel prossimo paragrafo.

2. La derivata della funzione di distribuzione cumulata.

Utilizzando le realizzazioni della variabile Y con distribuzione uniforme in $[0, 1]$ possiamo anche simulare altre variabili aleatorie; per esempio:

$$X = 2Y^2$$

I valori assunti da X appartengono all'intervallo $[0, 2]$; perché? Ma quale distribuzione ha X ?

Consideriamo i 1000 numeri casuali già generati e da questi costruiamo una realizzazione di Y elevandoli al quadrato e moltiplicandoli per 2.

Qui a fianco sono riportati alcuni istogrammi delle realizzazioni.

Consideriamo una barra e indichiamo i suoi estremi sulle ascisse con x e $x+h$. L'altezza della barra indica la percentuale di realizzazioni campionarie che cadono tra x e $x+h$.

Osserviamo che la scelta dell'ampiezza h , e quindi del numero di classi, non è univoca.

Dagli istogrammi si vede chiaramente che X non ha distribuzione uniforme; infatti i valori vicino a 0 sono più frequenti dei valori vicino a 2.

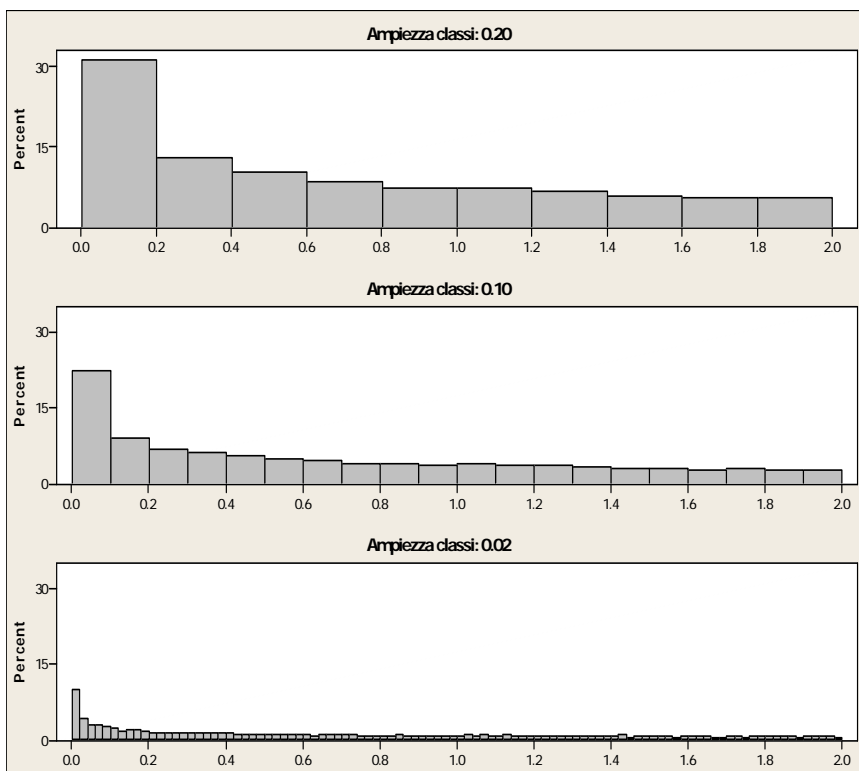


Figura 4 Istogrammi di X con classi di diversa ampiezza e $n = 1000$

A fianco è riportato il grafico della funzione di distribuzione cumulata, che indichiamo con \hat{F}_X moltiplicata per 100.

La percentuale di realizzazioni campionarie che cadono fra x e $x+h$ è data da:

$$(\hat{F}_X(x+h) - \hat{F}_X(x)) * 100$$

La funzione di distribuzione cumulata non cresce in modo costante, come nel caso della distribuzione uniforme: per valori vicini a 0 cresce più rapidamente che per valori vicino a 2.

Ti aspettavi questo andamento o pensavi che crescesse più rapidamente per valori vicini a 2?

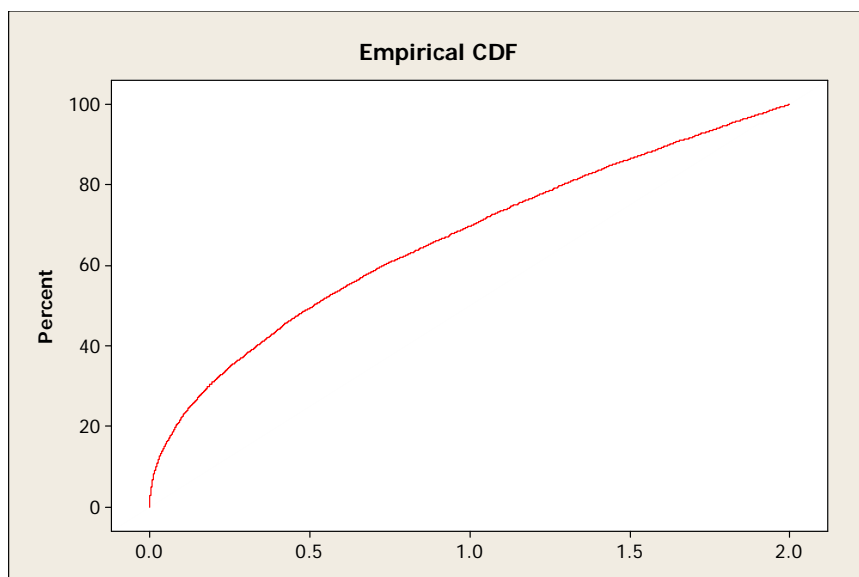


Figura 5 Funzione di distribuzione cumulata di X (in percentuale)

Cerchiamo di studiare in modo più approfondito come varia la pendenza locale di questa curva.

Iniziamo a studiare la pendenza della retta che passa per i due punti del grafico della funzione di distribuzione cumulata

$$(x, \hat{F}_X(x)) \quad \text{e} \quad (x+h, \hat{F}_X(x+h)).$$

La pendenza è data dal rapporto incrementale:

$$\frac{\hat{F}_X(x+h) - \hat{F}_X(x)}{h}$$

che, a meno del fattore $100/h$, è la percentuale di realizzazioni campionarie che cadono tra x e $x+h$. Questi valori sono le altezze delle barre degli istogrammi precedenti.

Scegliamo per esempio $x=0.4$ e controlliamo i valori delle altezze delle barre degli istogrammi in Minitab.

h = 0.20	h = 0.10	h = 0.02
$\widehat{F}_x(0.4 + 0.2) - \widehat{F}_x(0.4) = 10.14\%$	$\widehat{F}_x(0.4 + 0.1) - \widehat{F}_x(0.4) = 5.37\%$	$\widehat{F}_x(0.4 + 0.02) - \widehat{F}_x(0.4) = 1.32\%$

Calcoliamo quindi i rapporti incrementali al variare di h:

h = 0.20	h = 0.10	h = 0.02
$\frac{\widehat{F}_x(0.4 + 0.2) - \widehat{F}_x(0.4)}{0.2} = 0.507$	$\frac{\widehat{F}_x(0.4 + 0.1) - \widehat{F}_x(0.4)}{0.1} = 0.537$	$\frac{\widehat{F}_x(0.4 + 0.02) - \widehat{F}_x(0.4)}{0.02} = 0.660$

Qui a fianco sono riportati degli istogrammi in cui le barre hanno altezza uguale al rapporto incrementale nelle diverse classi.

Ciascuna barra ha *area* uguale alla percentuale di realizzazioni campionarie che cadono tra x e $x+h$.

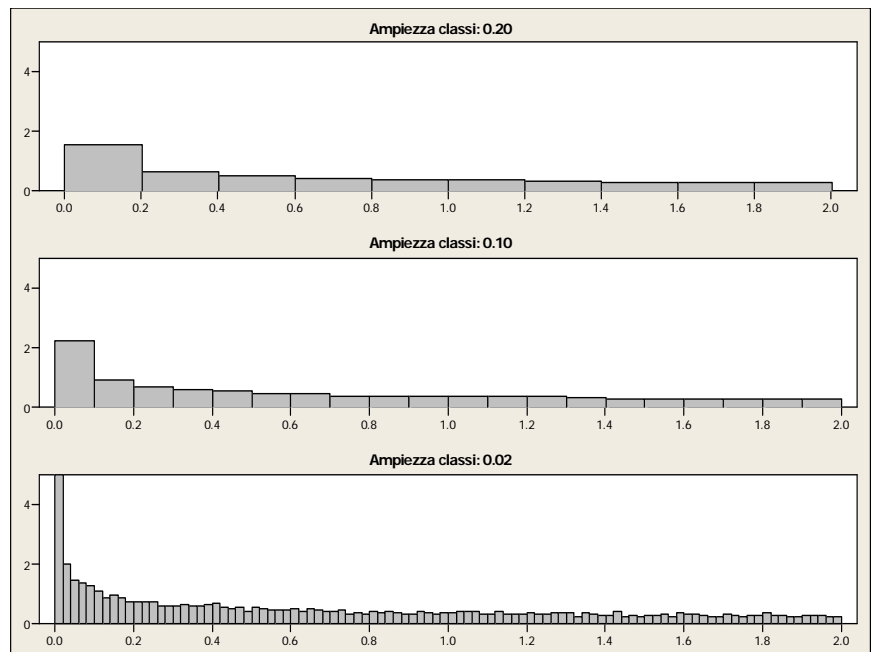


Figura 6 Istogrammi dei rapporti incrementali di X con classi di diversa ampiezza e $n = 1000$

Ricordiamo che, a proposito della Figura 3, relativa alla distribuzione uniforme, avevamo osservato che aumentando il numero di simulazioni n , la funzione di distribuzione cumulata empirica \widehat{F}_Y (cioè calcolata a partire dai dati simulati) "tende" alla funzione di distribuzione cumulata F_Y della variabile aleatoria Y . Questo vale in generale per ogni variabile aleatoria X .

Ossevando la Figura 6 possiamo notare che restringendo l'ampiezza delle classi, cioè il valore di h , l'istogramma dei rapporti incrementali "tende" ad individuare il grafico della derivata di \widehat{F}_x , o all'aumentare del numero di osservazioni, di F_x , per quanto detto in precedenza.

Per le variabili aleatorie continue (e con F_x continua o con altre condizioni che qui non precisiamo) diamo quindi una definizione di funzione di densità di probabilità f_x che è diversa da quella per variabili aleatorie discrete, ma mantiene proprietà analoghe, ossia:

- a) f_x è non negativa; b) $\int_{-\infty}^{+\infty} f_x(t) dt = 1$.

La **funzione di densità di probabilità** nel punto x è (dove esiste) la derivata di $F_x(x)$:

$$f_x(x) = \lim_{h \rightarrow 0} \frac{F_x(x+h) - F_x(x)}{h}$$

Qui sotto sono riportati un istogramma con ampiezza delle classi uguale a 0.005 e il grafico della derivata di F_x .

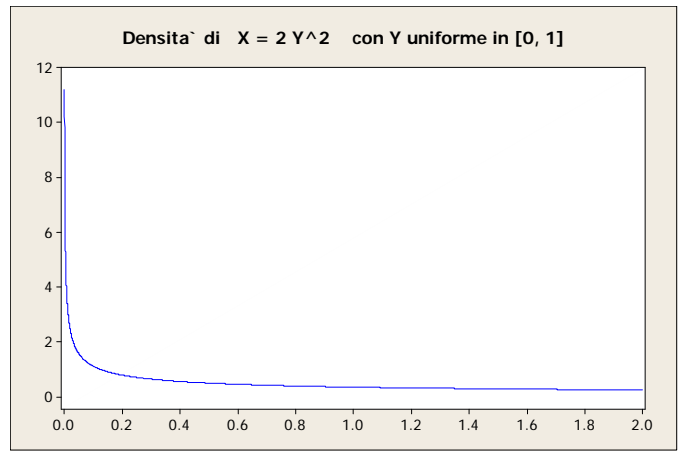
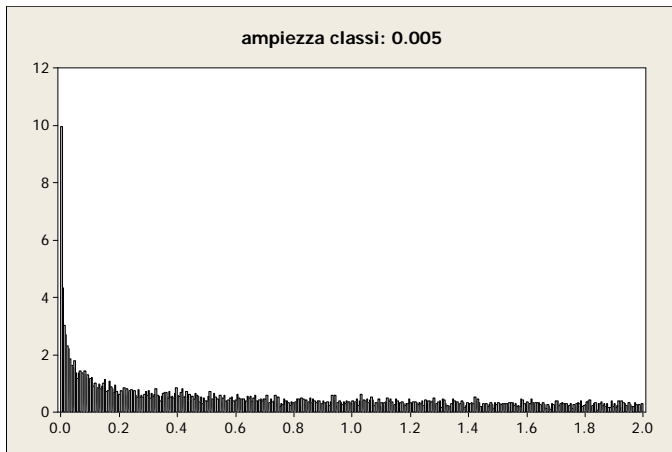


Figura 7 Istogramma dei rapporti incrementali con $h=0.005$ e funzione di densità di probabilità di X

3. Approfondimento: calcolo delle funzioni di distribuzione cumulata e di densità

3.1 Variabile aleatoria con distribuzione uniforme

La funzione di distribuzione cumulata di Y , variabile uniforme in $[0, 1]$, vale 0 prima di 0, vale 1 dopo 1 e fra 0 e 1 è la retta di equazione $F_Y(y) = y$. Quindi:

$$F_Y(y) = P(Y \leq y) = \begin{cases} 0 & \text{se } y < 0 \\ y & \text{se } 0 \leq y < 1 \\ 1 & \text{se } 1 \leq y \end{cases}$$

Esercizio. Controllate che questa funzione soddisfi le proprietà di una funzione di distribuzione cumulata.

La densità di probabilità di Y vale quindi:

$$f_Y(y) = F_Y'(y) = \begin{cases} 0 & \text{se } y < 0 \\ 1 & \text{se } 0 < y < 1 \\ 0 & \text{se } 1 \leq y \end{cases}$$

Esercizio. Controllate che questa funzione soddisfi le proprietà di una funzione di densità di probabilità.

La funzione di distribuzione cumulata di Y , variabile uniforme in $[a, b]$, vale 0 prima di a , vale 1 dopo b e fra 0 e 1 è la retta di equazione $F_Y(y) = \frac{1}{b-a}(y-a)$.

Esercizio. Verificate questo risultato.

Esercizio. Quanto vale la densità di probabilità di Y ? Calcolatela sia usando le proprietà della funzione di densità di probabilità, sia come derivata della funzione di distribuzione cumulata.

3.2 Variabile aleatoria $X = 2Y^2$ dove Y ha distribuzione uniforme $[0, 1]$

Come si calcola la funzione di distribuzione cumulata di X e poi la sua derivata?

$$F_X(x) = P(X \leq x) = P(2Y^2 \leq x) = P\left(Y^2 \leq \frac{x}{2}\right)$$

Essendo Y a valori positivi si ha:

$$F_x(x) = P\left(Y^2 \leq \frac{x}{2}\right) = P\left(Y \leq \sqrt{\frac{x}{2}}\right) = \begin{cases} 0 & \text{se } \sqrt{\frac{x}{2}} < 0 \text{ ovvero } x < 0 \\ \sqrt{\frac{x}{2}} & \text{se } 0 \leq \sqrt{\frac{x}{2}} < 1 \text{ ovvero } 0 \leq x < 2 \\ 1 & \text{se } 1 \leq \sqrt{\frac{x}{2}} \text{ ovvero } 2 \leq x \end{cases}$$

La funzione di densità di probabilità è la derivata di F_x , quindi:

$$f_x(x) = F'_x(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{\sqrt{8x}} & \text{se } 0 < x < 2 \\ 0 & \text{se } 2 < x \end{cases}$$

Inoltre $\lim_{x \rightarrow 0^+} f_x(x) = +\infty$

3.3 La variabile aleatoria esponenziale

Fenomeni naturali del tipo "tempo di vita di una lampadina", "tempo di emissione di una particella radioattiva", ... sono modellabili con una variabile aleatoria con distribuzione "esponenziale".

La funzione di distribuzione cumulata e la densità di probabilità dipendono da un parametro λ , che può assumere valori reali positivi.

$$F_x(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - \exp(-\lambda x) & \text{se } x \geq 0 \end{cases} \quad \text{e} \quad f_x(x) = \begin{cases} 0 & \text{se } x < 0 \\ \lambda \exp(-\lambda x) & \text{se } x > 0 \end{cases}$$

Qui sotto sono riportati i grafici delle funzioni di distribuzione cumulata di variabili aleatorie esponenziali per diversi valori del parametro.

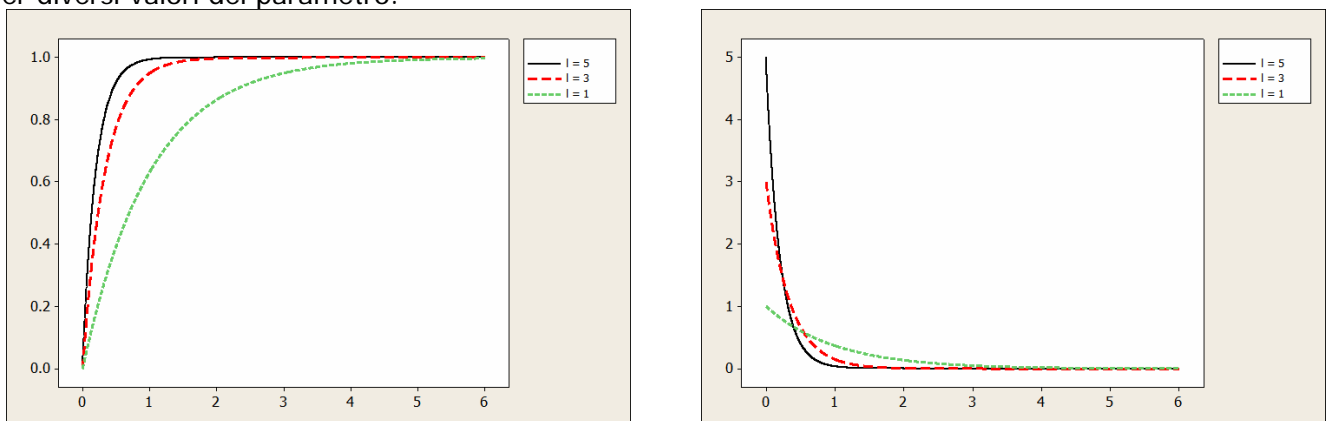


Figura 8 Funzione di distribuzione cumulata e funzione di densità di probabilità di v.a. esponenziali

Conoscere la forma analitica della funzione di distribuzione cumulata può essere utile per simulare delle realizzazioni della variabile aleatoria a partire dalle realizzazioni di una variabile aleatoria uniforme in $[0,1]$. Come?

4. La simulazione di variabili aleatorie discrete e continue a partire dall'Uniforme

Se si conosce la funzione di distribuzione cumulata di una variabile aleatoria X è possibile – almeno in linea di principio – simulare realizzazione campionarie di X a partire da simulazioni di una variabile aleatoria uniforme. Vediamo come.

Se la variabile aleatoria è **continua** e la sua funzione di distribuzione cumulata $F_X(x)$ è sempre **strettamente crescente** (nell'intervallo in cui X assume valori) basta simulare la realizzazione di una variabile uniforme U in $[0, 1]$ e invertire.

Più precisamente, indichiamo con u la realizzazione di U , allora la realizzazione di X sarà:

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

come indicato nella figura a fianco quando X ha distribuzione esponenziale con parametro $\lambda = 1$.

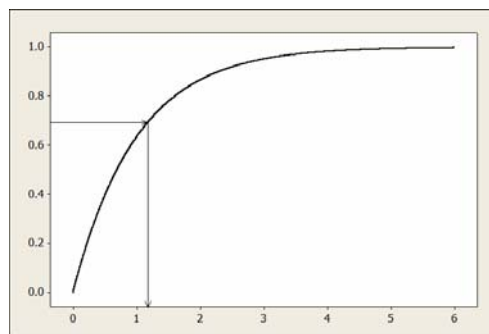


Figura 9 Schema di simulazione di una v.a. esponenziale

Se si può calcolare esplicitamente l'inversa della funzione di distribuzione cumulata, allora il calcolo della realizzazione campionaria di X è presto fatto.

Questo è il caso, per esempio della variabile aleatoria esponenziale dove $u = F_X(x) = 1 - \exp(-\lambda x)$ quindi invertendo, in modo da ottenere $x = F_X^{-1}(u)$, si ha:

$$\exp(-\lambda x) = 1 - u \quad -\lambda x = \log(1 - u) \quad x = -\log(1 - u) / \lambda$$

Quindi se, ad esempio $u=0.69$, allora $x = -\log(1 - 0.69) / 1 = 1.17$.

Quando l'inversa non è calcolabile esplicitamente si usano metodi numerici.

Quando invece la variabile aleatoria è **discreta** o la sua funzione di distribuzione cumulata $F_X(x)$ **non strettamente crescente**, si usa la cosiddetta pseudo-inversa, analogamente al caso del calcolo dei quantili.

$$\widetilde{F}_X^{-1}(u) = \min\{x \text{ tale che } F_X(x) \geq u\}$$

Quindi se, ad esempio, X è:

X	-2	1	5	7
$f_X(x)$	0.30	0.25	0.15	0.15
$F_X(x)$	0.30	0.55	0.80	1

e si ottenuto dalla simulazione di una variabile aleatoria uniforme in $[0, 1]$ il valore $u=0.72$, allora $x = 5$.

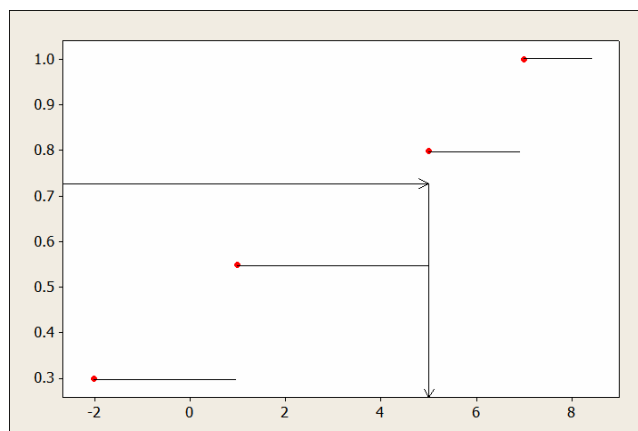


Figura 10 Schema di simulazione di una v.a. discreta

Possibile esperienza (descritta in fondo alla scheda)

E. Simulazione di realizzazioni campionarie di una variabile aleatoria con distribuzione esponenziale.

E1. Esperienza: estrazione di numeri casuali con distribuzione uniforme in [0, 1]

Simulate le realizzazioni campionarie di una variabile aleatorie uniforme in [0, 1] per $n=100$, $n=500$, $n=1000$ e $n=10000$ e ricostruite i grafici 1 e 2.

E2. Esperienza: estrazione di numeri casuali con distribuzione uniforme in [4, 6]

Considerate le realizzazioni precedenti per $n=1000$, indicate con A_{1000}

Trasformate la colonna A_{1000} moltiplicando per 2 e aggiungendo 4:

$$B_{1000} = 2 A_{1000} + 4$$

Costruite l'istogramma delle frequenze. Che distribuzione ha B_{1000} ?

E . Esperienza: Simulazione del lancio di una moneta e del lancio di un dado

Utilizzate le realizzazioni precedenti per $n=1000$ e simulate le realizzazioni di due variabili aleatorie discrete che descrivono il risultato del lancio di una moneta e il risultato del lancio di un dado.

E . Esperienza: Simulazione di realizzazioni di una variabile aleatoria esponenziale

Utilizzate le realizzazioni precedenti per $n=10000$ e simulate le realizzazioni di una variabile aleatoria con distribuzione esponenziale, avendo fissato il valore del parametro.

Costruite il grafico della funzione di distribuzione cumulata.

Costruite alcuni istogrammi con diverse ampiezze delle classi.

Confrontate i risultati ottenuti con la simulazione con quelli teorici.

