

RICHIAMI SU SIMULAZIONE E METODO MONTE CARLO

Agenda

- introduzione alla simulazione e al metodo Monte Carlo
- tecniche di simulazione
 - ▷ generatore congruenziale lineare, variabili uniformi
 - ▷ metodo dell'inversa generalizzata
 - ▷ metodo di accettazione-rifiuto
 - ▷ metodi ad hoc
 - ▷ simulazione di v.a. multidimensionali
- analisi dei risultati

Simulazione e Monte Carlo

- variabile aleatoria (v.a) X , o un evento E , **difficili** da osservare / non ripetibili
- simulare X o $E \rightsquigarrow$ considerare un'altra v.a., X' , o un E' , **facili** da osservare, osservarli e **far finta** (simulare) che le loro osservazioni siano quelle di X (o di E)
- il **procedimento** è **corretto** se le nuove variabili riproducono il **medesimo stato di incertezza** originario $\rightsquigarrow X'$ deve avere la stessa distribuzione di probabilità di $X \rightsquigarrow E$ ad E' devono avere la stessa probabilità

Simulazione e Monte Carlo

- la simulazione di variabili aleatorie è estremamente efficace per
 - ▷ generare scenari
 - ▷ calcolare speranze matematiche di v.a. complesse tramite il metodo **Monte Carlo**
- applicazioni:
 - ▷ generare traiettorie future di variabili rilevanti per un assicuratore (rendimenti degli investimenti, esborsi e numeri di sinistri, ...) \rightsquigarrow **economic scenario generators**
 - ▷ probabilità di rovina \rightsquigarrow **solvency**
 - ▷ calcolare il prezzo di strumenti finanziari derivati
 - ▷ ...
- simulazione Monte Carlo
 - ▷ funziona bene in ambito **multidimensionale**
 - ▷ **elevato tempo di calcolo**

Simulazione e Monte Carlo

- metodo Monte Carlo (**legge dei grandi numeri**): X v.a. con media finita, X_1, \dots, X_n, \dots i.i.d. con la stessa distribuzione di X , allora, con prob. 1,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X]$$

- **Monte Carlo**:
 - ▷ simula $X_1 = x_1, \dots, X_n = x_n$ (n grande!)
 - ▷ approssima $\theta = E[X]$ con

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

Simulazione e Monte Carlo

- primissime applicazioni del metodo Monte Carlo: **calcolo di integrali multipli** di elevata dimensione:

$$\int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} f(y_1, \dots, y_m) dy_1 \cdots dy_m = [\text{sostituzione di variabili}]$$
$$= \int_0^1 \cdots \int_0^1 g(u_1, \dots, u_m) du_1 \cdots du_m = E[g(U_1, \dots, U_m)]$$

con U_1, \dots, U_m i.i.d. con distribuzione uniforme in $[0, 1]$

Simulazione e Monte Carlo

- le v.a. X_i simulate devono avere la stessa distribuzione di X
↔ simula prima $U \sim U(0, 1)$, poi **trasforma** U in modo da ottenere una v.a. con distribuzione come quella di X
- due alternative per generare $U(0, 1)$:
 - ▷ procedimento **fisico** ↔ **veri** numeri casuali
 - ▷ computer ↔ sequenza **deterministica** che **sembra** casuale ↔ numeri **pseudocasuali**
- qualunque sia il metodo prescelto, deve essere statisticamente testato per
 - ▷ l'indipendenza
 - ▷ la distribuzione $U(0, 1)$
- numeri pseudocasuali (computer) preferibili:
 - ▷ velocità
 - ▷ riproducibilità ↔ analisi di **sensibilità**, **statica comparata**
 - ▷ portabilità

Numeri Pseudocasuali

- **numeri pseudocasuali** generati attraverso algoritmi deterministici
- la gran parte di questi algoritmi produce dapprima una sequenza ricorsiva s_1, \dots, s_n, \dots tramite
 - ▷ $s_0 = s \rightsquigarrow$ **seme** (assegnato)
 - ▷ $s_{n+1} = f_1(s_n), n = 0, 1, 2, \dots$

poi ricava u_1, \dots, u_n, \dots di numeri pseudocasuali (i.i.d. $U(0, 1)$) tramite

$$u_n = f_2(s_n), n \geq 1$$

dove f_1 e f_2 sono funzioni reali, in particolare f_2 prende valori nell'intervallo $[0, 1]$

Numeri Pseudocasuali

- periodo ℓ :

minimo ℓ per cui esiste d tale che $s_d = s_{d+\ell}$

\rightsquigarrow la sequenza si ripete dopo ℓ iterazioni

- noti (s, f_1, f_2) \rightsquigarrow si può riprodurre l'intera sequenza (u_n)
- esempio tipico, **generatore congruenziale lineare**:

$$s_{n+1} = f_1(s_n) = (A s_n + C) \bmod M, \quad u_n = f_2(s_n) = \frac{s_n}{M}$$

con A, C, M interi tali che $A > 0, C \geq 0, M > 0$ \rightsquigarrow il periodo non può superare M

- i moderni calcolatori usano algoritmi più complessi ma basati su simili idee (es. Mersenne Twister, periodo $2^{19937} - 1$)
- come simulare variabili non uniformi?

Metodo dell'Inversa

- si vuole simulare X con F_X funzione di ripartizione e inversa generalizzata (o psudoinversa, o quantile, \dots):

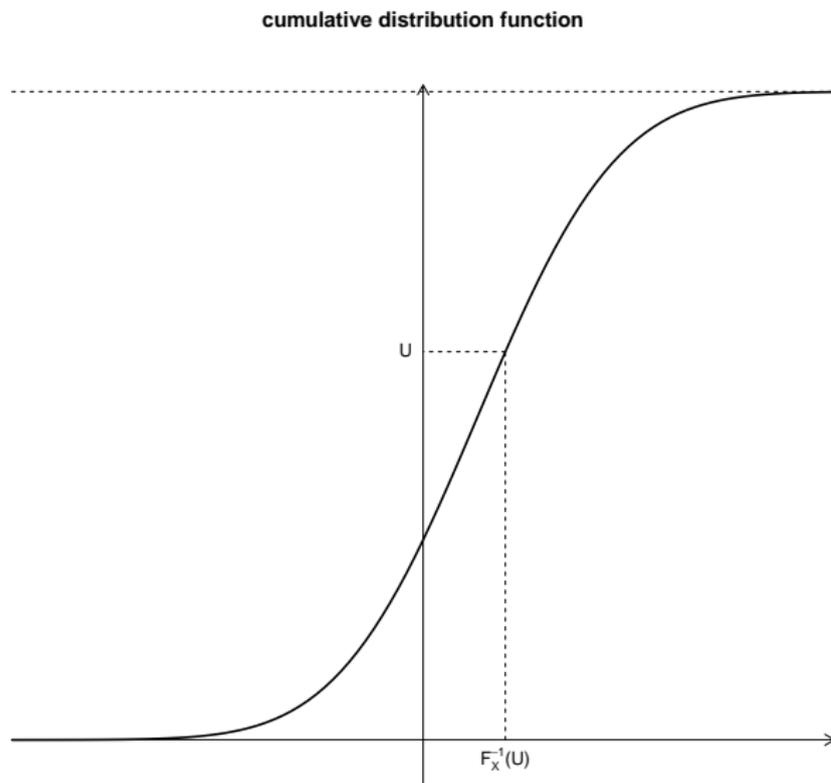
$$F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\} \quad \text{per } 0 < p < 1$$

- F_X è invertibile $\rightsquigarrow F_X^{-1}$ usuale inversa
- Teorema:

se $U \sim U(0, 1)$ allora $F_X^{-1}(U)$ ha funzione di ripartizione F_X

- Algoritmo:
 1. simula $U \sim U(0, 1)$
 2. restituisci $X' = F_X^{-1}(U)$

Metodo dell'Inversa



Metodo dell'Inversa

- ESEMPIO. Simula v.a. con distribuzione Weibull(c, γ)

▷ funzione di ripartizione della Weibull(c, γ):

$$\rightarrow F_X(x) = 1 - \exp(-cx^\gamma), \quad x > 0, \quad \text{dove } c > 0, \gamma > 0$$

▷ inversa: risolvere l'equazione $F_X(x) = p$ con $0 < p < 1$, \rightsquigarrow si ottiene

$$x = F_X^{-1}(p) = \left(-\frac{1}{c} \log(1-p) \right)^{1/\gamma}$$

\uparrow

▷ simulare Weibull(c, γ) \rightsquigarrow simula $U \sim U(0, 1)$ poi calcola

$$X' = \left(-\frac{1}{c} \log U \right)^{1/\gamma}$$

▷ $\gamma = 1 \rightsquigarrow \text{Exp}(c)$

Metodo dell'Inversa

- ESEMPIO. Sia T_x durata residua di vita di un individuo di età x
 - ▷ la sua funzione di ripartizione è (μ = intensità di mortalità)

$$F_{T_x}(t) = \Pr[T_x \leq t] = {}_tq_x = 1 - {}_tp_x = 1 - \exp\left(-\int_0^t \mu(x+v)dv\right),$$

- ▷ se μ è una Gompertz, $\mu(z) = \alpha \exp(\beta z)$ con $\alpha, \beta > 0$

$$F_{T_x}^{-1}(p) = \frac{1}{\beta} \log \left[1 - \frac{\beta}{\alpha} e^{-\beta x} \log(1-p) \right],$$

- ▷ simula T_x con

$$T'_x = \frac{1}{\beta} \log \left[1 - \frac{\beta}{\alpha} e^{-\beta x} \log(U) \right]$$

$$F_{T_x}(t) = 1 - e^{-\frac{\alpha}{\beta} B (e^{\beta t} - 1)} \quad t > 0$$

$$B = e^{\beta x}$$

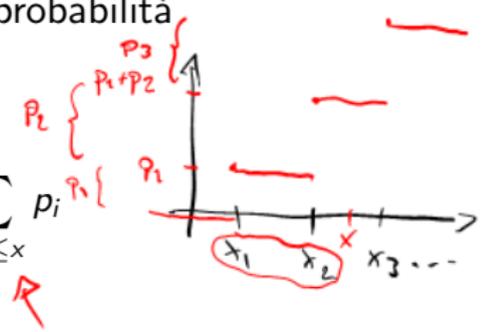
$$f_{T_x}(t) = \alpha \cdot B \cdot e^{\beta t} - \frac{\alpha}{\beta} B (e^{\beta t} - 1) \quad t > 0$$

Metodo dell'Inversa

- v.a. discrete, funzione di ripartizione **costante a tratti** (non invertibile)
- se X ha valori $x_1, x_2, \dots, x_n, \dots$ con probabilità $\Pr[X = x_i] = p_i$,

▷ la funzione di ripartizione di X è

$$F_X(x) = \sum_{i: x_i \leq x} p_i$$



p_i = il salto di tale funzione in x_i

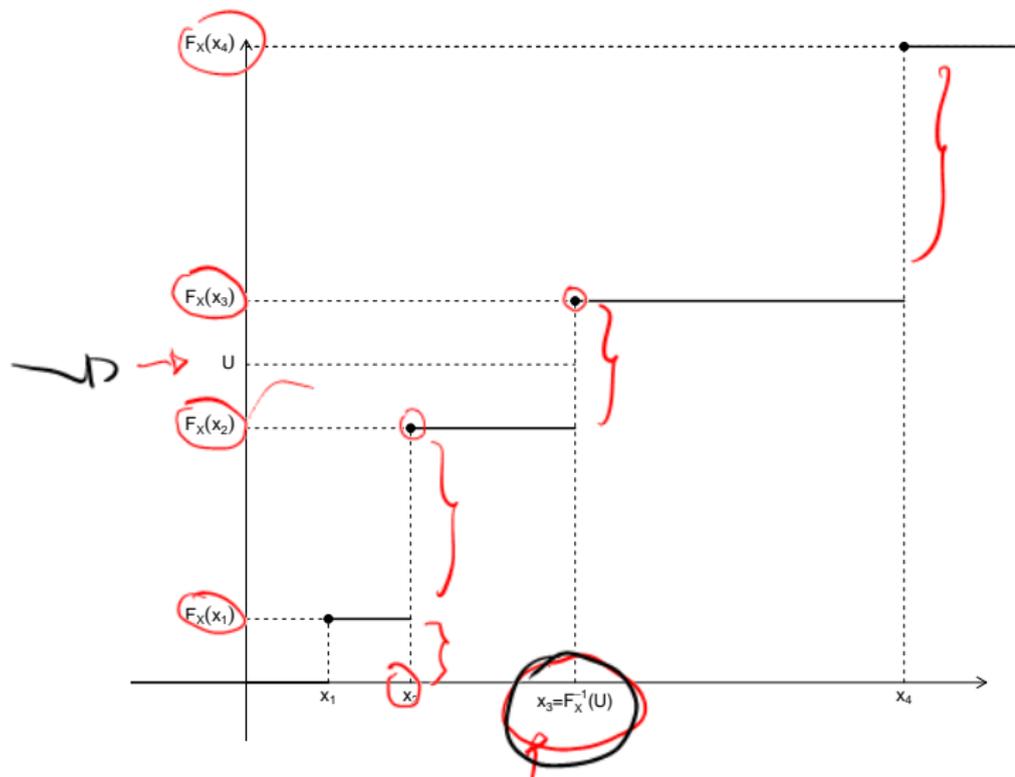
▷ inversa generalizzata:

$$F_X^{-1}(p) = \min_i \left\{ x_i : F_X(x_i) = \sum_{j=1}^i p_j \geq p \right\}$$

Metodo dell'Inversa

$$U \quad F_X^{-1}(U) = \inf \{x, F_X(x) \geq U\}$$

cumulative distribution function



Metodo dell'Inversa

- Algoritmo nel caso discreto:
 1. simula $U \sim U(0, 1)$
 2. trova i tale che $F_X(x_{i-1}) < U \leq F_X(x_i)$
 3. restituisci $X' = x_i$
- per calcolare $X' = x_i$:
 - ▷ una ricerca sequenziale
 - ▷ una ricerca binaria
 - ▷ approccio specifico alla distribuzione data (Poisson, binomiale, ...)

Metodo dell'Inversa

- ESEMPIO. Sia $K_x = \lfloor T_x \rfloor$ durata residua di vita troncata di un individuo di età x
($K_x = i \Leftrightarrow i \leq T_x < i + 1, \quad i = 0, 1, 2, \dots$)
 - ▷ data una tavola di mortalità (q_y)
 - ▷ $\Pr[K_x = i] = {}_i|1q_x = {}_ip_x q_{x+i} \rightsquigarrow F_{K_x}(i) = {}_{i+1}q_x$

Algoritmo:

1. simula $U \sim U(0, 1)$
2. trova i tale che ${}_i q_x < U \leq {}_{i+1} q_x$
3. restituisci $K'_x = i$

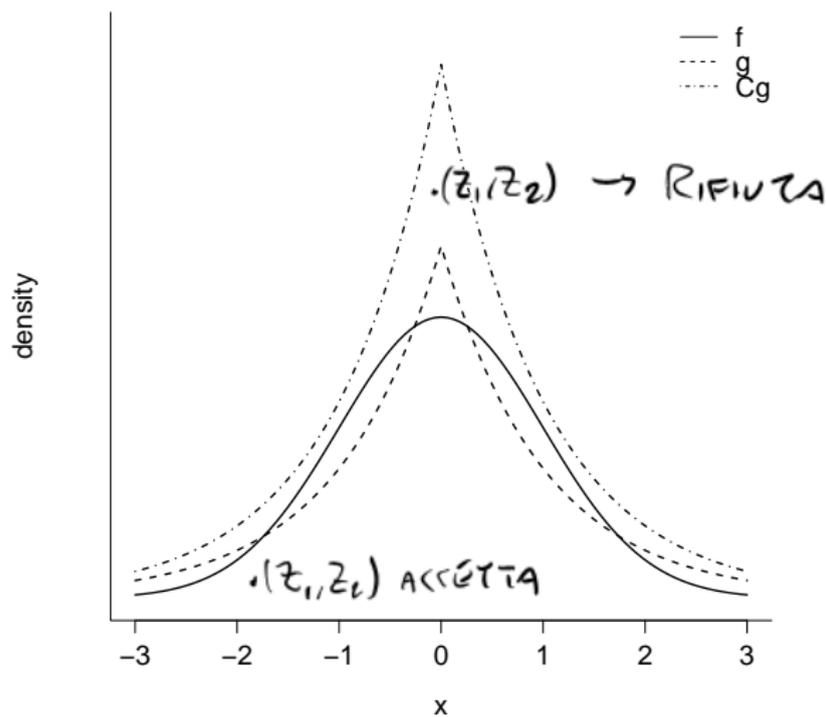
Metodo di Accettazione-Rifiuto

- si vuole simulare una v.a. X con densità f , e si sa simulare una v.a. con densità g per cui esiste $C > 1$ tale che

$$f(x) \leq C \cdot g(x) \text{ per ogni } x \in \mathbb{R}$$

- ▷ il grafico di g può essere 'gonfiato' in modo da dominare quello di f
- ▷ la scelta di g è vincolata alla forma di f
- si sceglie a caso un punto (Z_1, Z_2) sotto la curva $C \cdot g$; se il punto cade anche sotto la curva f allora l'ascissa Z_1 viene accettata come valore simulato di X , altrimenti viene rifiutata (e si ritenta con un nuovo punto)
- tale procedimento genera una v.a. X' con densità f

Metodo di Accettazione-Rifiuto



Metodo di Accettazione-Rifiuto

- Algoritmo:

1. simula $Z_1 \sim g$
2. simula $Z_2 \sim U(0, C \cdot g(Z_1)) \sim C \cdot g(Z_1) \cdot U$ con $U \sim U(0, 1)$
3. se $Z_2 \leq f(Z_1)$ allora **accetta** $X' = Z_1$, altrimenti **rifiuta** e torna al passo 1.

- probabilità di accettazione

$$\Pr \left[U \leq \frac{f(Z_1)}{C \cdot g(Z_1)} \right] = \int_{-\infty}^{+\infty} \frac{f(y)}{C \cdot g(y)} g(y) dy = \frac{1}{C},$$

\rightsquigarrow numero atteso di tentativi prima di ottenere un'accettazione = $C \rightsquigarrow$ valore **ottimo** di C :

$$C^* = \max_{x \in \mathbb{R}} \frac{f(x)}{g(x)}$$

($C \cdot g$ più **vicino** possibile a f)

Metodo di Accettazione-Rifiuto

- ESEMPIO. T_x durata residua di vita di un individuo di età x con intensità di mortalità Makeham, $\mu(z) = \gamma + \alpha \exp(\beta z)$. Simulare T_x usando come densità dominante quella di una Gompertz ($\gamma = 0$)
 - ▷ densità di T_x : (posto $B = e^{\beta x}$)

$$f(t) = (\gamma + \alpha B e^{\beta t}) \exp \left\{ -\gamma t - \frac{\alpha B}{\beta} (e^{\beta t} - 1) \right\}$$

- ▷ costante ottima:

$$C^* = 1 + \frac{\gamma}{\alpha B}$$

- Algoritmo:
 1. simula $Z_1 \sim \text{Gompertz}$
 2. simula $U \sim U(0, 1)$
 3. se

$$U \leq \frac{\gamma e^{-\beta Z_1} + \alpha B}{\gamma + \alpha B} e^{-\gamma Z_1},$$

accetta $T'_x = Z_1$, altrimenti **rifiuta** e ritorna a 1.

SIMULARE $T_x \sim$ MAXIMAM CON A/R

$$h(t) = \frac{f(t)}{g(t)} = \frac{(\gamma + \alpha B e^{\beta t}) e^{-\gamma t - \frac{\alpha B}{\beta} (e^{\beta t} - 1)}}{\alpha B e^{\beta t} - \frac{\alpha B}{\beta} (e^{\beta t} - 1)}$$

$f =$ MAXIMAM
 $g =$ COMPENZAZIONE
($\gamma = 0$)

$$= \left(\frac{\gamma e^{-\beta t}}{\alpha B} + 1 \right) e^{-\gamma t} \quad \text{DECRESCENTE}$$

MAX $h(t)$
 $t > 0$

$$\sim f^* = 0$$

$$h(0) = C^* = \text{costante ottima} \\ = \frac{\gamma + \alpha B}{\alpha B}$$

CONDIZIONE D) ACCETTAZIONE

$$Z_1 \sim \text{GOMPERTZ} \quad , \quad U \sim U(0,1)$$

$$Z_2 \sim C^* \cdot g(Z_1) \cdot U$$

ACCETTA SE $Z_2 \leq f(Z_1) \Leftrightarrow$

$$\Leftrightarrow U \leq \frac{f(Z_1)}{C^* g(Z_1)} = \dots = \left(\frac{\gamma e^{-\beta Z_1} + \alpha B}{\gamma + \alpha B} \right) e^{-\delta Z_1}$$

Metodi ad hoc - Distribuzione Normale

- alcuni metodi ad hoc possono essere usati per simulare v.a. in situazioni particolari, sfruttando la loro specifica struttura; alcuni semplici esempi:

- ▷ simulare v.a. $Y \sim N(\mu, \sigma^2)$: simula normale standard X e poi calcola $Y' = \mu + \sigma X$
- ▷ simulare v.a. Y lognormale: simula normale X e poi calcola $Y' = \exp(X)$
- ▷ simulare una v.a. Y binomiale di parametri (n, p) : simula n eventi indipendenti $A_i, i = 1, \dots, n$ con $\Pr[A_i] = p$ e poi calcola

$$Y' = I_{A_1} + \dots + I_{A_n}$$

- ▷ simulare una v.a. Y chi quadrato con n gradi di libertà (n intero): simula n normali standard X_1, \dots, X_n e poi calcola $Y' = X_1^2 + \dots + X_n^2$, oppure (n pari) simula $n/2$ esponenziali standard $Z_1, \dots, Z_{n/2}$ e poi calcola $Y' = 2(Z_1^2 + \dots + Z_{n/2}^2)$

Metodi ad hoc - Distribuzione Normale

- per simulare v.a. **normali standard** esistono vari **procedimenti ad hoc**, oltre a quelli generali di inversione della cdf (lo standard in \mathbf{R}) e di A/R
- **Teorema del limite centrale**:
 - ▷ X_1, \dots, X_n v.a. i.i.d. con media θ e varianza τ^2 finite, allora

$$\frac{\frac{1}{n} \sum_{i=1}^n X_i - \theta}{\tau/\sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\theta}{\tau\sqrt{n}} \rightarrow N(0, 1) \text{ in distribuzione}$$

- ▷ scelta più semplice: $X \sim U(0, 1) \rightsquigarrow \theta = 1/2$ e $\tau^2 = 1/12$
- ▷ numero **"magico"**: $n = 12$. Algoritmo:
 1. simula U_1, \dots, U_{12} da $U(0, 1)$
 2. calcola $X' = \sum_{i=1}^{12} U_i - 6$
- ▷ malgrado n sia molto piccolo, il procedimento supera bene i test di normalità ed è molto rapido da utilizzare

Metodi ad hoc - Distribuzione Normale

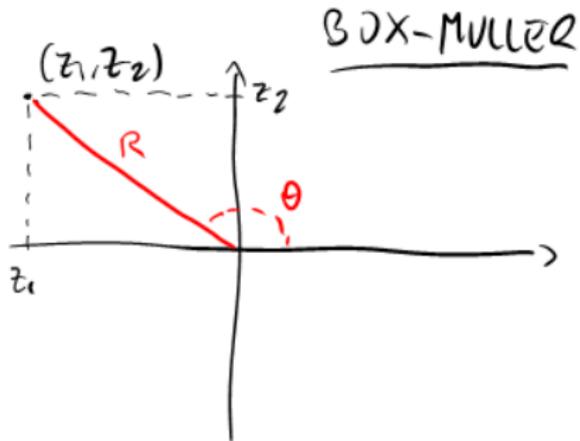
- Algoritmo di **Box-Muller**:
 1. simula U_1, U_2 da $U(0, 1)$
 2. calcola

$$Z_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2)$$

$$Z_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

(Z_1, Z_2) sono una coppia di **normali standard indipendenti**

- idea: se il punto del piano (Z_1, Z_2) ha coordinate normali standard indipendenti, la sua distanza dall'origine $R = \sqrt{Z_1^2 + Z_2^2}$ e l'angolo $\beta = \arctan \frac{Z_2}{Z_1}$ sono indipendenti con distribuzione $\text{Exp}(1/2)$ e $U(0, 2\pi)$
- svantaggio: bisogna calcolare funzioni 'speciali' come sin, cos



z_1, z_2 INDIPENDENTI,
 $N(0, 1)$

\Leftrightarrow

R, θ INDIPENDENTI

$$R \sim \text{EXP}\left(\frac{1}{2}\right)$$

$$\theta \sim U(0, 2\pi)$$

Metodi ad hoc - Distribuzione Normale

- Algoritmo di **Polar-Marsaglia** (variante di Box-Muller):
 1. simula U_1, U_2 da $U(0, 1)$
 2. calcola $V_1 = 2U_1 - 1, V_2 = 2U_2 - 1$ e $S = V_1^2 + V_2^2$
 3. se $S > 1$ **rifiuta** e torna al passo 1., altrimenti **accetta** restituendo

$$Z_1 = \sqrt{\frac{-2 \log S}{S}} V_1$$

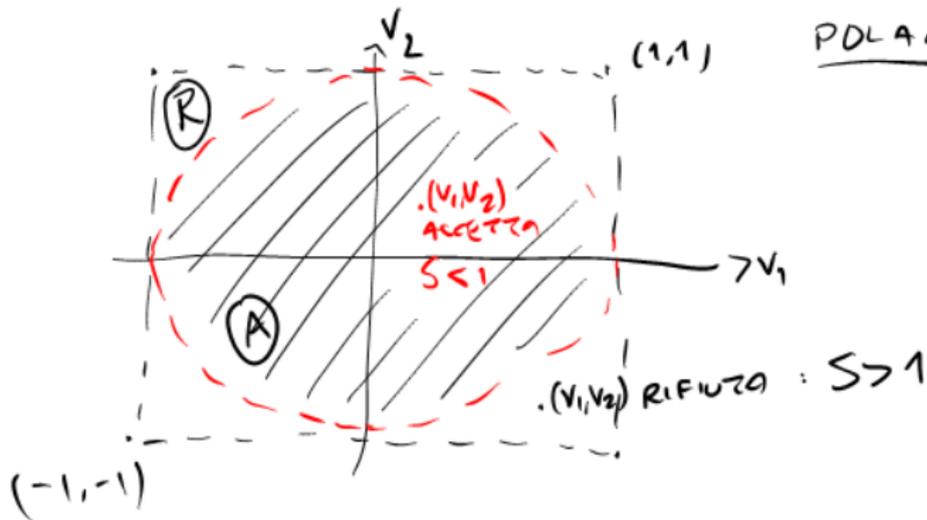
$$Z_2 = \sqrt{\frac{-2 \log S}{S}} V_2$$

(Z_1, Z_2) sono una coppia di normali standard indipendenti

- idea: il punto (V_1, V_2) appartiene al quadrato di vertici $(\pm 1, \pm 1)$ ed è rigettato se non appartiene al cerchio unitario
 \rightsquigarrow probabilità di accettazione è $\frac{\pi}{4} \approx 0.78$

POLAR - MARSAGLIA

$$S = V_1^2 + V_2^2$$



Simulazione di V.A. Multidimensionali

- simulare un vettore di v.a. (X_1, X_2, \dots, X_m) (e.g. che potrebbe rappresentare, ad es., **la traiettoria di un processo stocastico**) si procede come segue:
 - 1. simula $X'_1 = x_1$ dalla distribuzione marginale F_{X_1} ,
 - 2. simula $X'_2 = x_2$ dalla distribuzione condizionata $F_{X_2|X_1=x_1}$
 - \vdots
 - m. simula X'_m dalla distribuzione condizionata $F_{X_m|X_1=x_1, X_2=x_2, \dots, X_{m-1}=x_{m-1}}$
- Metodi ad-hoc in cui ci si riconduce a variabili indipendenti (e.g. moto Browniano, processo di Poisson, ...)

PASSAGGIATA ALGEBRA CON ORIT
(DERIVA)

$$K_{t+1} = K_t + \delta + \varepsilon_t, \quad (\varepsilon_t) \text{ i.i.d.}$$
$$K_0 = \bar{K} \in \mathbb{R} \quad \varepsilon_t \sim N(0, \sigma^2)$$

$$\Leftrightarrow K_t = \bar{K} + t \cdot \delta + \sum_{u=1}^t \varepsilon_u$$

$$\left[\text{SE } \varepsilon_t = \begin{cases} -1 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases}, \delta = 0 \Rightarrow K_t = \begin{array}{cc} \text{GUADAGNO} & \text{GNO} \\ \text{IN UN} & \text{GIOCO} \\ \text{BUONO} & \text{RIPETUTO} \end{array} \right]$$

Metodo Monte Carlo: Analisi dei Risultati

- X_1, X_2, \dots, X_n i.i.d. con media finita θ , allora

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i \approx \theta$$

- se inoltre la varianza delle X_i è τ^2 finita, allora $(\hat{\theta}_n - \theta)/(\tau/\sqrt{n}) \approx N(0, 1)$ (in distribuzione) \rightsquigarrow accuratezza della stima:

$$\hat{\theta}_n \approx \theta + \frac{\tau}{\sqrt{n}} N(0, 1)$$

\rightsquigarrow tasso di convergenza è dell'ordine di \sqrt{n} \rightsquigarrow per aumentare la precisione di una cifra decimale bisogna aumentare n di un fattore pari a 100

Analisi dei Risultati

- si può ottenere un intervallo di confidenza per θ nel modo usuale:

$$\Pr \left[\hat{\theta}_n - z_{1-\alpha/2} \frac{\tau}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n + z_{1-\alpha/2} \frac{\tau}{\sqrt{n}} \right] = 1 - \alpha$$

dove $z_p : \Phi(z_p) = p$ (Φ f. di ripartizione della normale standard), α piccolo, e.g.. 1%, o 5%

- stimatore di τ^2 :

$$\hat{\tau}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\theta}_n)^2$$

- intervallo di confidenza:

$$\left(\hat{\theta}_n - z_{1-\alpha/2} \text{SE}, \hat{\theta}_n + z_{1-\alpha/2} \text{SE} \right)$$

standard error: $\text{SE} = \frac{\hat{\tau}_n}{\sqrt{n}}$

CALCOLO DI INTEGRALI VIA QUADRATURA

$$\int_a^b f(x) dx \approx \sum_{i=1}^n w_i f(x_i)$$

$x_i = \text{NODI}$

ZERO DI
UN POLINOMIO
APPROPRIATO

$w_i = \text{PESI}$

ERRORE (STIMA)

$$\int_a^b f(x) dx - \sum_{i=1}^n w_i f(x_i) = \bar{E} \approx \hat{E}$$

FORMULA
ESATTA
SE f
POLINOMIO
DI GRADO
 $2n-1$

ESEMPIO: INTEGRAZIONE in \mathbb{R}^n

$$\int_{[0,1]^n} e^{-\|x\|^2} dx_1 \dots dx_n =$$

$$= \int_0^1 e^{-x_1^2} dx_1 \dots \int_0^1 e^{-x_n^2} dx_n = \left(\int_0^1 e^{-x^2} dx \right)^n$$

$$= \left(\frac{1}{\sqrt{2}} \int_0^1 e^{-\frac{y^2}{2}} dy \right)^n = \left(\frac{\sqrt{11}}{\sqrt{211}} \int_0^1 e^{-\frac{y^2}{2}} dy \right)^n$$

$$= \left(\sqrt{11} \left(\Phi(1) - \frac{1}{2} \right) \right)^n$$

$$\begin{aligned} &= \Phi(1) - \Phi(0) \\ &= \Phi(1) - \frac{1}{2} \end{aligned}$$

CAMBIO DI VARIABILI

$$\int_{-\infty}^{+\infty} f(x) dx$$

$$y = \frac{1}{e^x + 1} \quad x \in \mathbb{R} \rightarrow y \in [0, 1]$$

$$\Leftrightarrow e^x + 1 = \frac{1}{y} \Leftrightarrow e^x = \frac{1}{y} - 1 = \frac{1-y}{y}$$

$$\Leftrightarrow x = \log\left(\frac{1-y}{y}\right) \quad dx = -\frac{1}{y(1-y)} dy$$

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^1 f\left(\log\left(\frac{1-y}{y}\right)\right) \frac{1}{y(1-y)} dy$$