

---

# Generatori di numeri casuali e metodi Montecarlo (parte 1)

Laboratorio Trattamento Numerico dei Dati Sperimentali

Prof. L. Carminati  
Università degli Studi di Milano

La Statale contro la violenza sulle donne

---

## Sequenze di numeri casuali

- ❑ Cosa intendiamo per sequenze di numeri casuali ? (in generale se non c'è una specifica indicazione si intende numeri casuali equiprobabili)
  - ❑ Immaginiamo un sistema meccanico costituito da un'urna chiusa contenente palle identiche ma numerate da 1 a  $N$  (non è possibile vedere il numero dall'esterno)
  - ❑ Estraiamo (a mano o in modo automatico) una palla alla volta (con re-inserimento delle palle estratte) e segniamo la sequenza dei numeri stampati sulle palle estratte
- ❑ La sequenza di numeri che otteniamo è una sequenza di numeri veramente casuali equiprobabili (uniformi):
  - ❑ Il numero della palla estratta dopo  $n$  estrazioni non è predicibile in base alla conoscenza delle estrazioni precedenti
  - ❑ Tutti i numeri tra 1 e  $N$  sono equiprobabili
- ❑ In generale per produrre sequenze di veri numeri casuali si utilizzano processi fisici :
  - ❑ Estrazioni casuali di palline, lancio di monete, estrazioni di carte da un mazzo
  - ❑ Misurazione del tempo tra due raggi cosmici che raggiungono la superficie terrestre, misura del rumore di un dispositivo elettronico, tempo tra due decadimenti radioattivi, misurazioni di fenomeni atmosferici

---

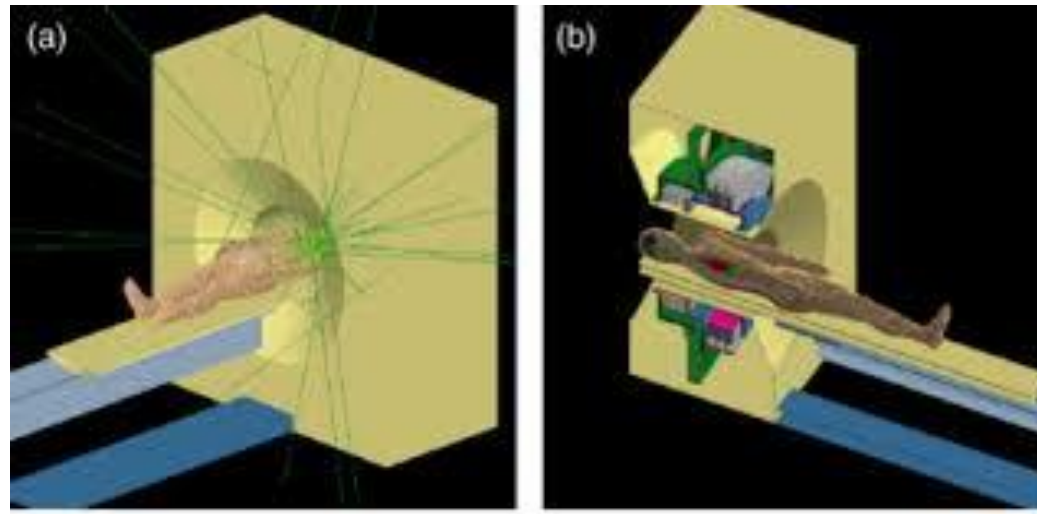
## So what ?

- ❑ Nonostante la cosa sembri piuttosto strana, avere a disposizione sequenze di numeri casuali (come vedremo bastano numeri casuali tra 0 e 1 distribuiti uniformemente) si presta ad un grande numero di applicazioni numeriche molto interessanti :
  - ❑ Metodi "MonteCarlo" per il calcolo di integrali
  - ❑ Simulazione di apparati sperimentali
  - ❑ Simulazione di sistemi fisici stocastici ( o finanziari )
  - ❑ Algoritmi probabilistici ( algoritmi genetici )
  - ❑ Videogames
  - ❑ Crittografia

## So what ?

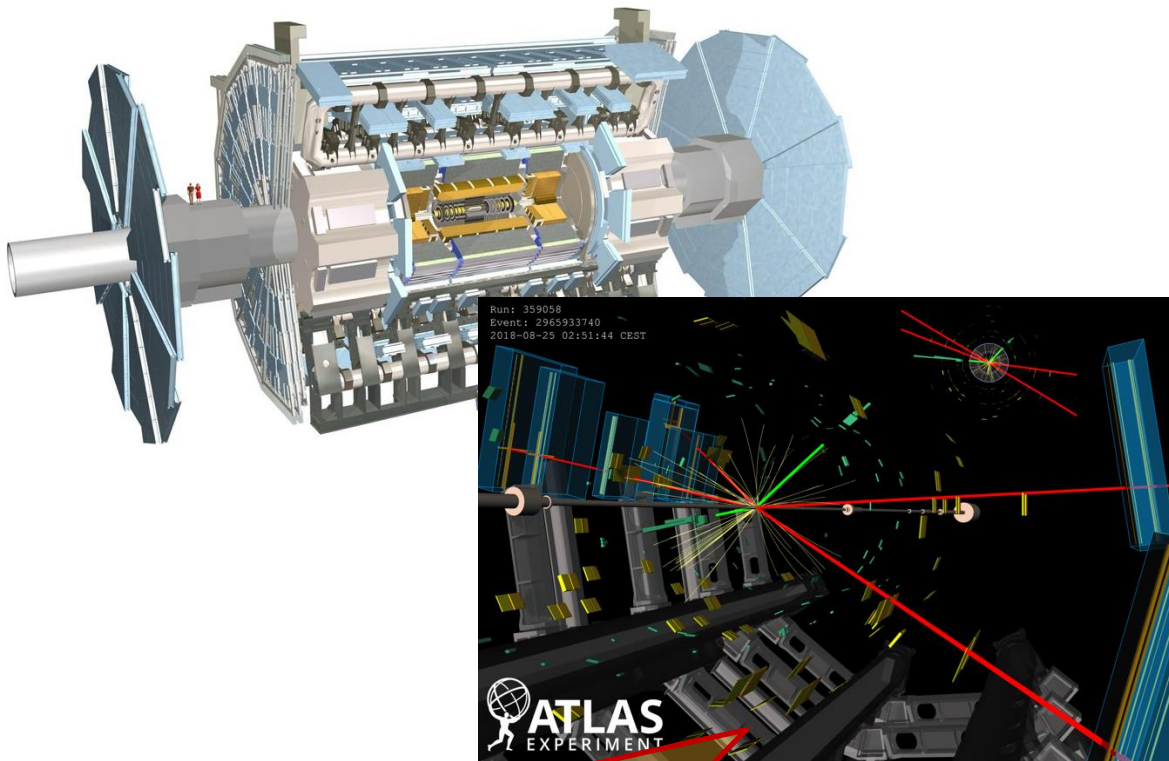
In questa lezione impareremo a scrivere algoritmi per generare numeri (pseudo) casuali che seguano densità di probabilità diverse (equiprobabili, gaussiane, esponenziali...)

- ❑ Curiosamente potremo usare la generazione di numeri casuali per calcolare integrali (con tecniche molto efficienti per integrazioni multidimensionali)
- ❑ Potremo studiare le incertezze relative a una misura simulando un gran numero di pseudo-misure
- ❑ Generare pseudo-dati complessi per studiare le caratteristiche di un fenomeno e ottimizzare apparati sperimentali



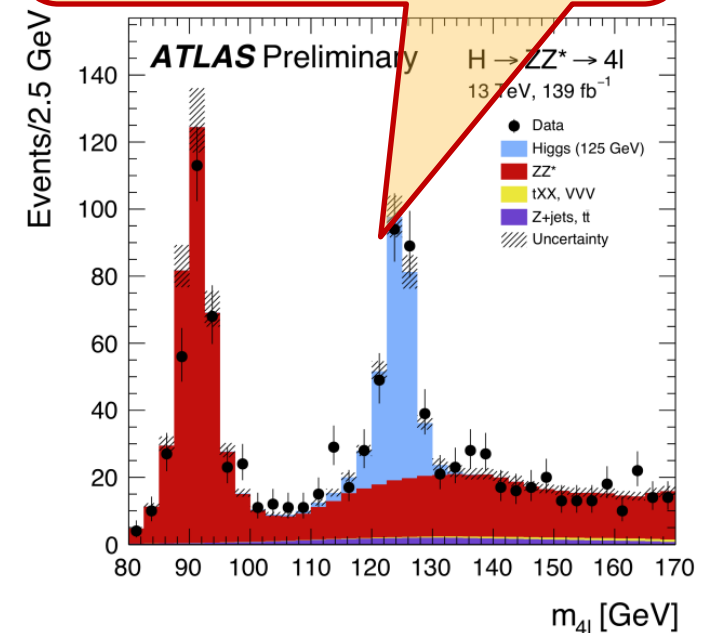
## So what ?

- ❑ Generare pseudo-dati complessi per studiare le caratteristiche di un fenomeno e ottimizzare apparati sperimentali



Un evento reale misurato nel rivelatore ATLAS di possibile decadimento del bosone di Higgs in  $2e2\mu$  in associazione ad un bosone Z che decade in due  $\mu$

In blu : una simulazione di cosa avresti dovuto vedere dall'analisi dei dati raccolti dal rivelatore ATLAS se esistesse un bosone di Higgs come predetto dalla teoria con una massa di 125 GeV.



---

## So what ?

- ❑ Nonostante la cosa sembri piuttosto strana, avere a disposizione sequenze di numeri casuali (come vedremo bastano numeri casuali tra 0 e 1 distribuiti uniformemente) si presta ad un grande numero di applicazioni numeriche molto importanti :
  - ❑ Metodi MC per il calcolo di integrali
  - ❑ Simulazione di apparati sperimentali
  - ❑ Simulazione di sistemi fisici stocastici
  - ❑ Algoritmi probabilistici (algoritmi genetici)
  - ❑ Videogames
  - ❑ Crittografia
  
- ❑ Come produrre un numero elevato (Milioni ? Miliardi? Milioni di Miliardi?) di numeri casuali da utilizzare nei nostri algoritmi numerici ?
  - ❑ Pescare da tabelle o database (non molto efficiente per velocità e uso memoria)
  - ❑ Interfacciare il calcolatore con sistemi fisici casuali (non molto pratico)
  - ❑ Farli generare dal calcolatore stesso mediante opportuni algoritmi

---

## So what ?

- ❑ Nonostante la cosa sembri piuttosto strana, avere a disposizione sequenze di numeri casuali (come vedremo bastano numeri casuali tra 0 e 1 distribuiti uniformemente) si presta ad un grande numero di applicazioni numeriche molto importanti :
  - ❑ Metodi MC per il calcolo di integrali
  - ❑ Simulazione di apparati sperimentali
  - ❑ Simulazione di sistemi fisici stocastici
  - ❑ Algoritmi probabilistici (algoritmi genetici)
  - ❑ Videogames
  - ❑ Crittografia
  
- ❑ Come produrre un numero elevato (Milioni ? Miliardi? Milioni di Miliardi?) di numeri casuali da utilizzare nei nostri algoritmi numerici ?
  - ❑ Pescare da tabelle o database (non molto efficiente per velocità e uso memoria)
  - ❑ Interfacciare il calcolatore con sistemi fisici casuali (non molto pratico)
  - ❑ **Farli generare dal calcolatore stesso mediante opportuni algoritmi**

---

## Sequenze di numeri casuali e numeri pseudo-casuali

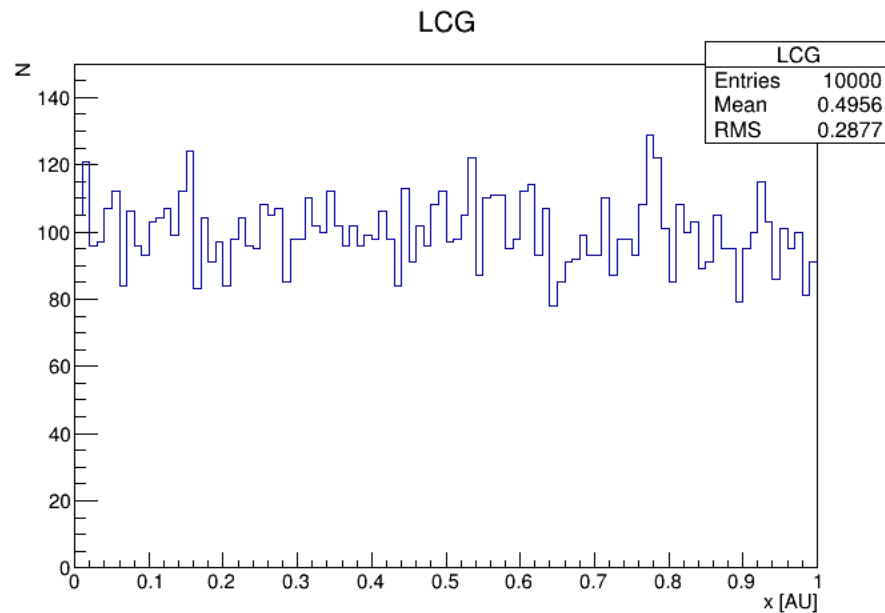
“Farli generare dal calcolatore stesso mediante opportuni algoritmi”

- ❑ La frase è chiaramente difficile da digerire: come è possibile far generare numeri casuali ad una macchina deterministica come il calcolatore ?
- ❑ Le sequenze di numeri prodotte dal calcolatore vengono denominate pseudo-casuali : essendo prodotte da algoritmi deterministici (al contrario dei numeri "effettivamente" casuali) non solo sono esattamente prevedibili ma possono anche essere riprodotte in maniera identica semplicemente ripetendo il procedimento di calcolo.
- ❑ Il loro utilizzo viene giustificato dal fatto che non sono necessari numeri che soddisfino tutti i criteri formali di casualità ma numeri che abbiano le stesse proprietà di quelli casuali (numeri che 'sembrino' casuali per gli scopi): non importa come siano stati generati, basta che, se sottoposti a verifica, risultino indistinguibili da quelli veramente casuali in termini di buon adattamento e indipendenza.
  - ❑ Li considero casuali se non sono capace di rendermi conto che non lo sono!
- ❑ Test di randomness: esistono in letteratura numerosi test per valutare la bontà di un algoritmo di generazioni di numeri casuali.
  - ❑ Se un algoritmo soddisfa i principali test di randomness possiamo considerare le sequenze generate come quasi-casuali

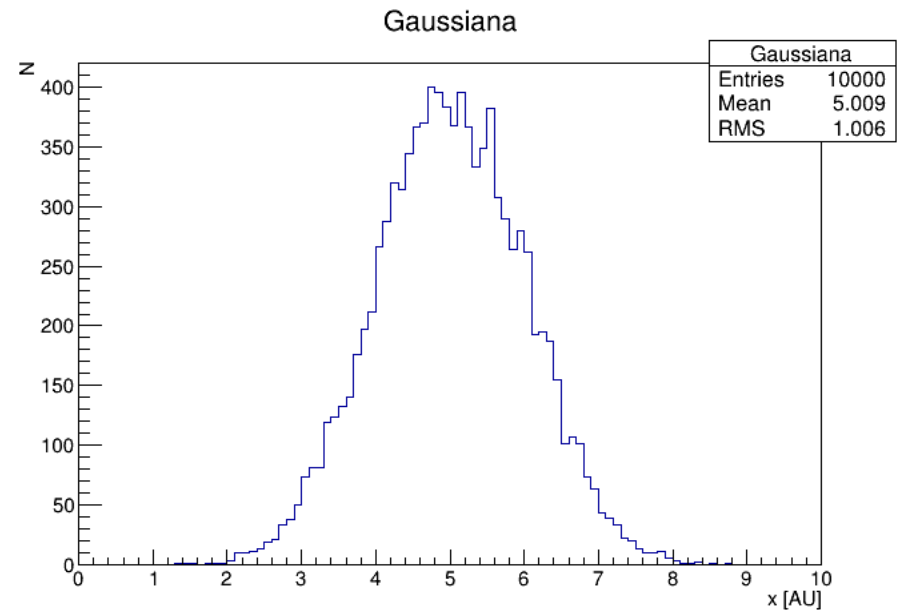


## Generatori di numeri casuali

In questa lezione impareremo a scrivere degli algoritmi per generare sequenze di numeri (pseudo) casuali che seguano densità di probabilità diverse (equiprobabili, gaussiane, esponenziali...)



Tutti i numeri tra 0 e 1 sono equiprobabili : su 10000 estrazioni in media circa 100 per bin



I numeri vicini a 5 sono più probabili ( i.e. vengono estratti più spesso )

---

## Outline:

- ❑ Tutta la statistica che ci serve in 6 passi: cos'è una variabile casuale (o variabile aleatoria) come al solito una rapida carrellata dei punti formali principali (corsi e testi dedicati di probabilità e statistica per gli interessati)
  - ❑ Usiamo come guida il classico lancio della moneta per introdurre gli elementi teorici di base
- ❑ Generatori di numeri casuali: teoria e implementazione (Lezione 10)
  - ❑ Generatore uniforme : LCG (Linear Congruential Generator)
  - ❑ Generazione di numeri casuali secondo una PDF generica : metodo della funzione inversa e metodo accept-reject
- ❑ Calcolo di integrali con metodi Montecarlo (Lezione 11) :
  - ❑ Metodo hit-or-miss
  - ❑ Metodo della media
- ❑ Simulazione di apparati sperimentali : simulare misure sperimentali utilizzando generatori di numeri casuali (Lezione 12)
  - ❑ Generiamo molte pseudo-misure per capire il comportamento di un apparato sperimentale (prisma di dispersione)

---

## Punto 1 : spazio campionario e spazio degli eventi

Consideriamo un certo fenomeno la cui descrizione non è di tipo deterministico, ad es. il lancio di un dado, il risultato di un sondaggio, l'istante in cui decade un nucleo radioattivo rispetto ad un  $t_0$ .

- ❑ Rappresenteremo matematicamente un esperimento tramite l'insieme  $\Omega$  dei suoi possibili risultati  $\Omega = \{\omega_1, \omega_2, \omega_3 \dots\}$  detto spazio degli eventi elementari (o spazio campionario)
  - ❑ Lancio di una moneta : eventi elementari sono  $\Omega = \{T, C\}$
  - ❑ Lancio di un dado : eventi elementari sono  $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - ❑ Tempo impiegato da un nucleo radioattivo a decadere  $\Omega = [0, +\infty)$
  
- ❑ Eventi: le collezioni di possibili risultati dell'esperimento che possiamo quindi rappresentare come un qualsiasi sottoinsieme di  $\Omega$ 
  - ❑ Ad esempio nell'esperimento del lancio di un dado :
    - ❑ l'evento "esce il numero 2" è rappresentato dal sottoinsieme  $\{2\}$ .
    - ❑ L'evento "esce un numero pari" è rappresentato dal sottoinsieme  $E = \{2, 4, 6\}$
    - ❑ l'evento "esce un numero  $\leq 3$ " è rappresentato dall'insieme  $F = \{1, 2, 3\}$ .
  - ❑ Nell'esperimento sul decadimento del nucleo radioattivo l'evento "il nucleo decade dopo l'istante  $T_1$  e prima dell'istante  $T_2$ " è rappresentato dall'insieme  $A = [T_1, T_2]$ .

## Punto 1 : spazio campionario e spazio degli eventi

□ A partire dallo spazio campionario  $\Omega$  costruiamo lo spazio degli eventi  $\mathbb{F}$  come l'insieme dei sottoinsiemi di  $\Omega$  (eventi) tale che sia chiuso rispetto alle operazioni di unione e complemento ovvero rispetti le seguenti proprietà:

1. L'insieme  $\Omega \in \mathbb{F}$
2. se  $A \in \mathbb{F} \Rightarrow \bar{A} \in \mathbb{F}$
3. se  $A, B \in \mathbb{F} \Rightarrow A \cup B \in \mathbb{F}$ 
  1.  $(\{E_n\} \subset \mathbb{F} \Rightarrow \bigcup_{n=1}^{\infty} E_n \in \mathbb{F})$

Una famiglia  $\mathbb{F} \subseteq \mathcal{P}(\Omega)$  di sottoinsiemi di  $\Omega$  che verifica le proprietà 1, 2 e 3.1 viene detta  $\sigma$ -algebra.

Insieme delle parti

□ Lancio di una moneta :  $\Omega = \{T, C\}$  e  $\mathbb{F} = \{T, C, \{T, C\}, \emptyset\}$

□ Se  $\Omega = \mathbb{R}$  allora  $\mathbb{F}$  contiene tutti i punti e gli intervalli della retta reale

□ Dato un fenomeno casuale, la coppia  $(\Omega, \mathbb{F})$ , dove  $\Omega$  è lo spazio campionario e  $\mathbb{F}$  è la  $\sigma$ -algebra generata da  $\Omega$ , è detta spazio misurabile.

## Punto 2 : definizione di probabilità di un evento

Definiamo la probabilità di un evento: ad ogni evento  $A$  in  $\mathbb{F}$  assegnamo un numero, compreso tra 0 e 1 che esprime intuitivamente quanto è verosimile che l'evento si verifichi. Ci sono vari modi di farlo :

□ Definizione assiomatica di Kolmogorov : dato uno spazio campionario  $\Omega$  e una  $\sigma$ -algebra  $\mathbb{F} \subseteq \mathcal{P}(\Omega)$ , una probabilità  $P$  su  $(\Omega, \mathbb{F})$  è un'applicazione  $P : \mathbb{F} \rightarrow [0, 1]$  che assegna ad ogni evento  $A \in \mathbb{F}$  un numero reale  $P(A)$ , con le seguenti proprietà :

1.  $0 \leq P(A) \leq 1$
2.  $P(\Omega) = 1$
3. data una successione  $\{A_n\}$  di eventi a due a due disgiunti (cioè tali che  $A_i \cap A_j = \emptyset \forall i \neq j$ ) si ha che  $P(\bigcup_n A_n) = \sum_n P(A_n)$ .

□ Definizione a priori : se tutti gli elementi sono equiprobabili

$$P(A) = \frac{\text{nr. eventi elementari in } A}{\text{nr. eventi elementari in } \Omega}$$

□ Definizione a posteriori (frequentista) : eseguo  $N$  esperimenti e conto quante volte ( $N_A$ ) esce un elemento di  $A$

$$P(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}$$

---

## Punto 3 : variabili aleatorie ( o variabili casuali )

Dato uno spazio probabilistico  $(\Omega, \mathbb{F}, P)$  una variabile aleatoria (o casuale) è una funzione che ad ogni evento elementare associa un numero reale, in simboli  $X: \Omega \rightarrow \mathbb{R} (\omega \rightarrow X(\omega))$

- ❑ In parole semplici una variabile casuale è un modo di trasformare i punti campionari in numeri. Ci sono molti modi di fare questo. Il motivo di trasformare i punti campionari in numeri è semplice: lavorare sui numeri è molto più semplice che lavorare sui punti campionari, anche perché questi ultimi possono essere di natura assai diversa fra un esperimento casuale ed un altro.
- ❑ Essendo una corrispondenza tra eventi campionari (ai quali è associata una probabilità) e numeri reali, si può definire una probabilità anche per la variabile aleatoria.
- ❑ Dobbiamo definire i valori che la variabile aleatoria può assumere e quali sono le probabilità associate. Possiamo farlo in vari modi: distinguiamo il caso in cui la variabile casuale è discreta o continua
  - ❑ Le variabili aleatorie discrete, ovvero variabili aleatorie  $X$  che possono assumere solo un numero discreto di valori
  - ❑ variabili aleatorie continue ovvero variabili aleatorie che assumono valori all'interno di un insieme continuo.

---

## Punto 4 : probabilità e variabili aleatorie : caso discreto

- ❑ Consideriamo un esperimento i cui risultati  $\Omega = \{\omega_i : i = 1, \dots, N\}$  siano discreti con  $N$  finito o infinito.
- ❑ Costruiamo lo spazio degli eventi  $\mathbb{F}$  e ad ogni evento in  $\mathbb{F}$  assegnamo una probabilità
- ❑ Indichiamo con  $X(\omega)$  la variabile aleatoria che assegna a ciascun risultato  $\omega_i$  il numero reale  $x_i = X(\omega_i)$ .
- ❑ La probabilità  $p_i$  che la variabile  $X(\omega)$  assuma il valore  $x_i$  possiamo definirla naturalmente come  $p_i = P(\{\omega_i | X(\omega_i) = x_i\})$ .
- ❑ Le probabilità così definite soddisfano le condizioni
  - ❑  $0 \leq p_i \leq 1 \quad \forall i=1, \dots, N$
  - ❑  $\sum_{i=1}^N p_i = 1$
- ❑ Definiremo probabilità cumulativa  $P_n$  (con  $n < N$ ) la probabilità che la variabile  $X$  assuma valori nell'insieme  $\{x_1, x_2, \dots, x_n\}$ , cioè'  $P_n = \sum_{i=1}^n p_i$

## Punto 4 : probabilità e variabili aleatorie : caso discreto

❑ Consideriamo il caso semplice del lancio di una moneta :

❑ Spazio campionario  $X = \{T, C\}$

❑ Spazio degli eventi :  $F = \{\{T\}, \{C\}, \{T, C\}, \emptyset\}$

❑ Assegniamo una probabilità :

	$\{T\}$	$\{C\}$	$\{T, C\}$	$\emptyset$
	↓	↓	↓	↓
	0.5	0.5	1	0

❑ Variabile aleatoria :  $X = \begin{cases} X(T) = +1 \\ X(C) = 0 \end{cases}$

❑ Assegniamo una probabilità alla variabile aleatoria

$$P(X) = \begin{cases} P(+1) = P(\omega_i | X(\omega_i) = +1) = 0.5 \\ P(0) = P(\omega_i | X(\omega_i) = 0) = 0.5 \end{cases}$$



---

## Punto 5 : probabilità e variabili aleatorie : caso continuo

Non possiamo banalmente assegnare ad ogni elemento di  $\Omega$  una probabilità, questa dovrebbe essere nulla (altrimenti  $P(\Omega)$  sarebbe infinita e non 1). Possiamo farlo in due modi diversi, attraverso la funzione di ripartizione o la funzione di densità

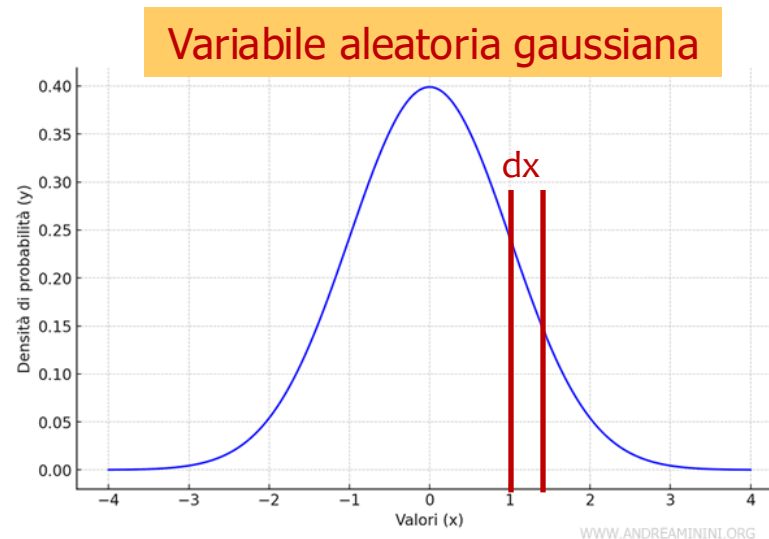
## Punto 5 : probabilità e variabili aleatorie : caso continuo

Non possiamo banalmente assegnare ad ogni elemento di  $\Omega$  una probabilità, questa dovrebbe essere nulla (altrimenti  $P(\Omega)$  sarebbe infinita e non 1 ). Possiamo farlo in due modi diversi, attraverso la funzione di ripartizione o la funzione di densità

1. Definizione di funzione di densità (di probabilità): sia  $X$  una variabile casuale continua che assume valori nell'intervallo  $(a, b)$ : La funzione di densità di  $X$  è la funzione

$$f(x) = \lim_{dx \rightarrow 0} \frac{P(x < X < x+dx)}{dx}$$

Il valore  $f(x)dx$  rappresenta la probabilità che la variabile  $X$  assuma valori tra  $x < X < x+dx$

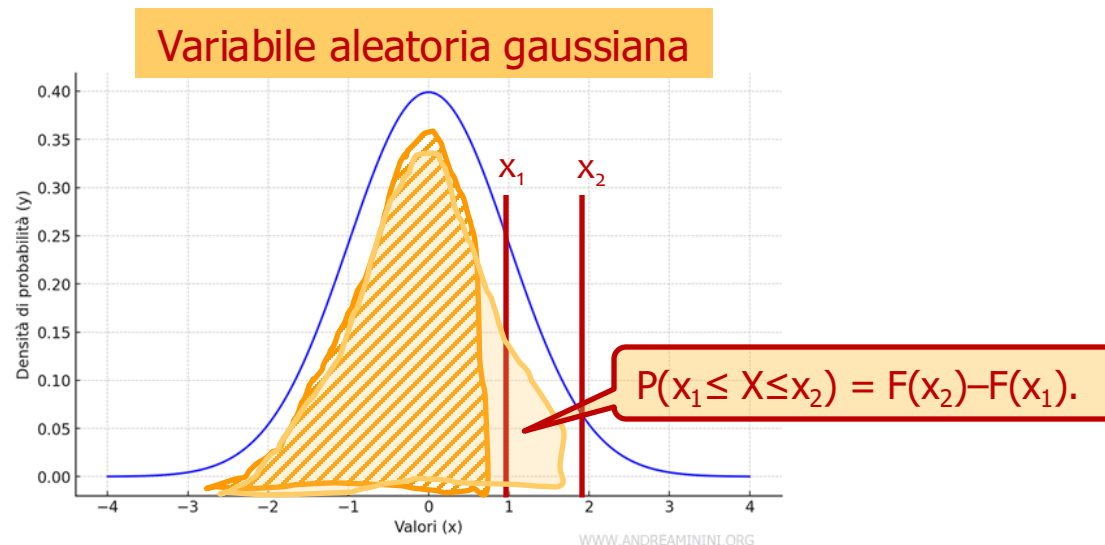


## Punto 5 : probabilità e variabili aleatorie : caso continuo

Non possiamo banalmente assegnare ad ogni elemento di  $\Omega$  una probabilità, questa dovrebbe essere nulla (altrimenti  $P(\Omega)$  sarebbe infinita e non 1 ). Possiamo farlo in due modi diversi, attraverso la funzione di ripartizione o la funzione di densità

2. Definizione di funzione di ripartizione (o funzione cumulativa): data una variabile casuale  $X$ , la funzione di ripartizione di  $X$  è la funzione  $F(\bar{x}) = P(X \leq \bar{x})$  intesa come  $P(\{\omega \in \Omega: X(\omega) \leq \bar{x}\})$ , dove  $\bar{x}$  è un qualsiasi numero reale.

- conoscendo la funzione di ripartizione di  $X$  è possibile ricavare la probabilità che  $X$  assuma valori in un qualsiasi intervallo  $[x_1, x_2]$ . Infatti  $P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$ .



---

## Punto 5 : probabilità e variabili aleatorie : caso continuo

Le due formulazioni sono chiaramente equivalenti : sia  $X$  una v.a. continua che assume valori nell'intervallo  $(a, b)$  (eventualmente  $a$  può essere  $-\infty$  e  $b$   $+\infty$ ). Allora per ricavare la funzione di densità dalla funzione di ripartizione e viceversa possiamo utilizzare

$$f(x) = \lim_{dx \rightarrow 0} \frac{P(x < X < x + dx)}{dx} = \lim_{dx \rightarrow 0} \frac{F(x + dx) - F(x)}{dx} = F'(x)$$

$$F(\bar{x}) = \int_{-\infty}^{\bar{x}} f(x) dx$$

---

## Punto 6 : valore di aspettazione e varianza di una variabile aleatoria

- ❑ In generale tenderemo poi a concentrarci sulla distribuzione della variabile aleatoria di fatto dimenticandoci dello spazio degli eventi da cui proviene
- ❑ Spesso non è necessario conoscere nel dettaglio la densità di probabilità di una variabile aleatoria ma può bastare conoscere alcuni suoi parametri fondamentali

1. Valore di aspettazione  $\mu(x) = \begin{cases} \sum_i x_i f(x_i) & (\text{caso vc discreta}) \\ \int_{-\infty}^{+\infty} x f(x) dx & (\text{caso vc continua}) \end{cases}$

2. Varianza  $\sigma^2 = \mu(x - \mu(x))^2 = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i) & (\text{caso vc discreto}) \\ \int_{-\infty}^{+\infty} (x - \mu(x))^2 f(x) dx & (\text{caso vc continua}) \end{cases}$

1. Deviazione Standard :  $\sqrt{\sigma^2}$

---

## Esempi di distribuzioni discrete (modelli per descrivere fenomeni casuali)

- Binomiale : consideriamo come esperimento  $n$  lanci di una moneta truccata dove  $p$  è la probabilità che esca testa e  $(1-p)$  la probabilità che esca croce. La probabilità di ottenere  $k$  successi in  $n$  lanci vale

$$f(X = k) = \frac{n!}{k! (n - k)!} p^k (1 - p)^{n-k}$$

$$\mu(X) = np$$

$$\sigma^2(X) = p(1 - p)n$$

- Poisson : essa descrive la probabilità che si verifichino  $n$  eventi di un tipo sapendo che in media se ne verificano  $\lambda$  in un intervallo (tempo, area o lunghezza o altro).

- La distribuzione di Poisson è anche conosciuta come distribuzione degli eventi rari in quanto approssima la distribuzione binomiale per probabilità molto basse di accadimento.

$$f(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

$$\mu(X) = \lambda$$

$$\sigma^2(X) = \lambda$$

---

## Esempi di distribuzioni continue

□ Distribuzione uniforme :  $f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases} \Rightarrow \begin{aligned} \mu(x) &= \frac{a+b}{2} \\ \sigma^2 &= \frac{(b-a)^2}{12} \end{aligned}$

□ Distribuzione esponenziale :  $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \Rightarrow \begin{aligned} \mu(x) &= \frac{1}{\lambda} \\ \sigma^2(x) &= \frac{1}{\lambda^2} \end{aligned}$

□ Distribuzione normale :  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow \begin{aligned} \mu(x) &= \mu \\ \sigma^2(x) &= \sigma^2 \end{aligned}$

---

## Generatori congruenti lineari (distribuzione uniforme)

Per generare numeri casuali al calcolatore possiamo usare dei semplice algoritmi (e quindi implicitamente accettando il fatto che i numeri generati sono pseudo-casuali ossia perfettamente predicibili )

$$x_{i+1} = (ax_i + c) \bmod(m)$$

- ❑ Linear Congruential Generator (LCG): sfruttiamo il resto di una divisione tra interi. Se il divisore è  $m$  allora il resto sarà distribuito tra  $0$  e  $m-1$
- ❑ È un algoritmo iterativo: occorre fornire un numero di partenza ( detto seme ). Stesso seme, stessa sequenza. ( dovremo poi verificare l'indipendenza dal seme )
- ❑ È un algorimo : il nuovo numero è correlato al precedente !
- ❑ Un generatore di numeri casuali : una classe con un metodo che restituisce un numero pseudo-casuale quando viene invocato



## Generatori congruenti lineari ( distribuzione uniforme )

Consideriamo a titolo di esempio un generatore LCG con  $a = 3$ ;  $c = 5$  ;  $m = 11$ ;

❑ Seme del generatore  $x_0 = 3$  :

❑  $x_1 = (3 \times 3 + 5) \bmod (11) = 3$

Non particolarmente interessante :  
generatore che restituisce sempre 3

❑ Seme del generatore  $x_0 = 1$  :

❑  $x_1 = (3 + 5) \bmod (11) = 8$

❑  $x_2 = (8 \times 3 + 5) \bmod (11) = 7$

❑  $x_3 = (7 \times 3 + 5) \bmod (11) = 4$

❑  $x_4 = (4 \times 3 + 5) \bmod (11) = 6$

❑  $x_5 = (6 \times 3 + 5) \bmod (11) = 1$

❑  $x_6 = (1 \times 3 + 5) \bmod (11) = 8$

❑  $x_7 = (8 \times 3 + 5) \bmod (11) = 7$

❑ .....

Meglio : generatore che restituisce 5  
numeri diversi poi la sequenza si  
ripete identica

Da questo punto la sequenza  
riparte sempre identica !

## Generatori congruenti lineari ( distribuzione uniforme ) : Goodman-Miller

- ❑ Il numero massimo di numeri diversi (periodo) che si possono ottenere da un generatore LCG è proprio  $m$  per una scelta opportuna dei parametri  $a$  e  $c$

$$\begin{cases} m = 2^{31} \\ a = 1664525 \\ c = 1013904223 \end{cases}$$

$$x_{i+1} = (ax_i + c) \bmod(m)$$

Escludiamo 1 per evitare problemi in future con la generazione di una distribuzione esponenziale

- ❑ Se poi vogliamo generare numeri uniformemente distribuiti in  $[0,1)$

$$r_i = \frac{x_i}{m}$$

- ❑ Per generare numeri distribuiti uniformemente in  $[a,b)$  :

$$y = a + (b - a) r$$

# RandomGenerator

```
#ifndef __RANDOMGEN_H__
#define __RANDOMGEN_H__

class RandomGen {
public:
    RandomGen(unsigned int);

    void SetA(unsigned int a) {m_a=a;}
    void SetC(unsigned int c) {m_c=c;}
    void SetM(unsigned int m) {m_m=m;}

    double Rand( ); // distribuzione uniforme tra 0 e 1
    double Unif(double xmin, double xmax); // distribuzione uniforme tra xmin e xmax
    double Exp(double lambda); // distribuzione esponenziale con costante lambda
    double Gaus(double mean, double sigma); // distribuzione gaussiana (Box-Muller)
    double GausAR(double mean, double sigma, ... ); // distribuzione gaussiana (Accept-Reject)

private:
    unsigned int m_a, m_c, m_m;
    unsigned int m_seed;
};

#endif
```

Costruttore : inizializza le costanti a,c,m

Possibilità di cambiare ( eventualmente ! ) i parametri del generatore

Un metodo specifico per ogni tipo di distribuzione che voglio simulare

```
#include "TApplication.h"
#include "TCanvas.h"
#include "TH1F.h"
#include "TAxis.h"

#include <iostream>
#include "RandomGen.h"

int main() {

    TApplication app("app",0,0);

    RandomGen myGen(1);

    int nmax = 10000;

    TH1F unif("Uniforme","Uniforme",100,4,11) ;

    for ( int k = 0 ; k < nmax ; k++ ) {

        unif.Fill( myGen.Unif(5,10) ) ;

    }

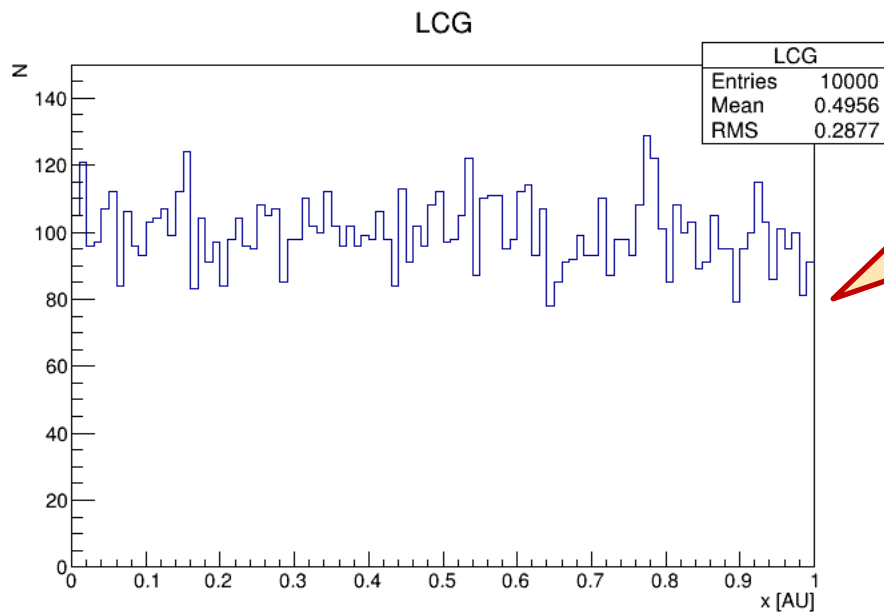
    TCanvas can2("Uniforme","Uniforme") ;
    can2.cd();
    unif.GetXaxis()->SetTitle("x [AU]");
    unif.GetYaxis()->SetTitle("N");
    unif.Draw();

    app.Run();

}
```

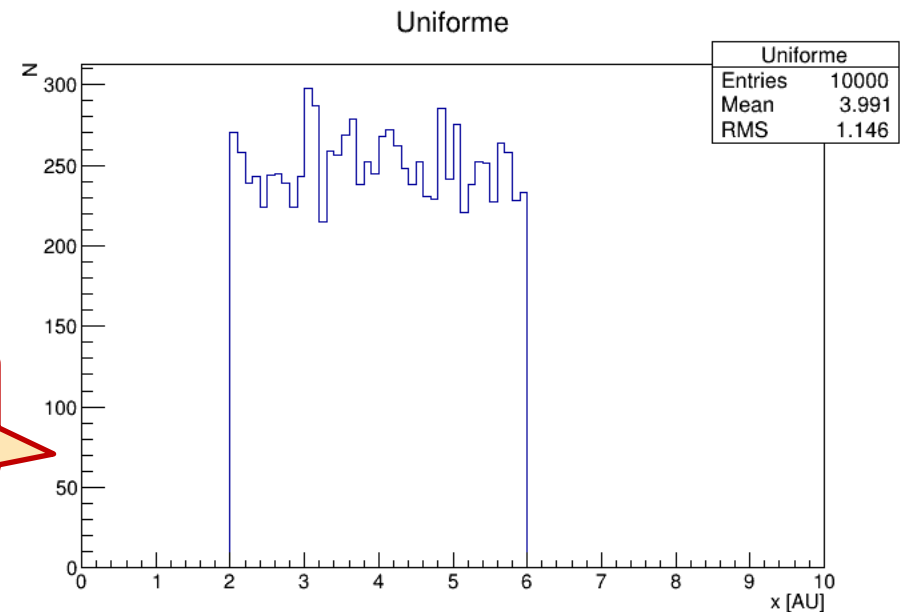


## Generatori congruenti lineari ( distribuzione uniforme ) :



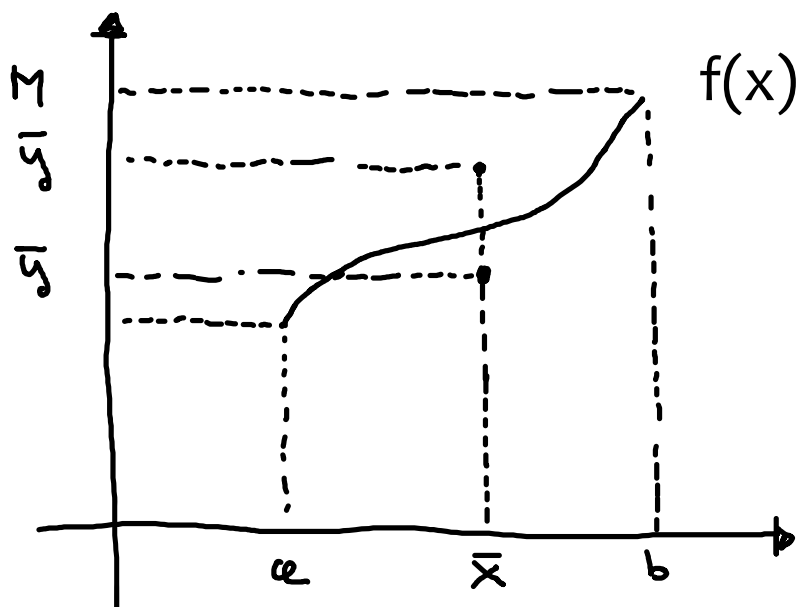
Generati 100000 valori di x : istogramma con 100 bins. Come atteso circa 100 entries per bin ( distribuzione uniforme tra 0 e 1)

Generati 100000 valori di x : istogramma con 100 bins tra 0 e 10. Come atteso circa 250 entries per bin ( distribuzione uniforme tra 2 e 6)



## Dalla distribuzione uniforme a una distribuzione qualunque, come fare ?

- ❑ Supponiamo di voler generare una sequenza di numeri pseudo-casuali in  $[a,b]$  che segua una densità di probabilità  $f(x)$  non necessariamente uniforme: possiamo usare un algoritmo Accept/Reject



- ❑ Estraggo un numero casuale  $\bar{x}$  uniformemente distribuito tra  $a$  e  $b$
- ❑ Estraggo un numero casuale  $\bar{y}$  uniformemente distribuito tra  $0$  e  $M$  ( massimo della  $f(x)$  )
- ❑ Se  $\bar{y} < f(\bar{x})$  accetto il punto  $\bar{x}$  e lo restituisco
- ❑ Se  $\bar{y} > f(\bar{x})$  ci riprovo

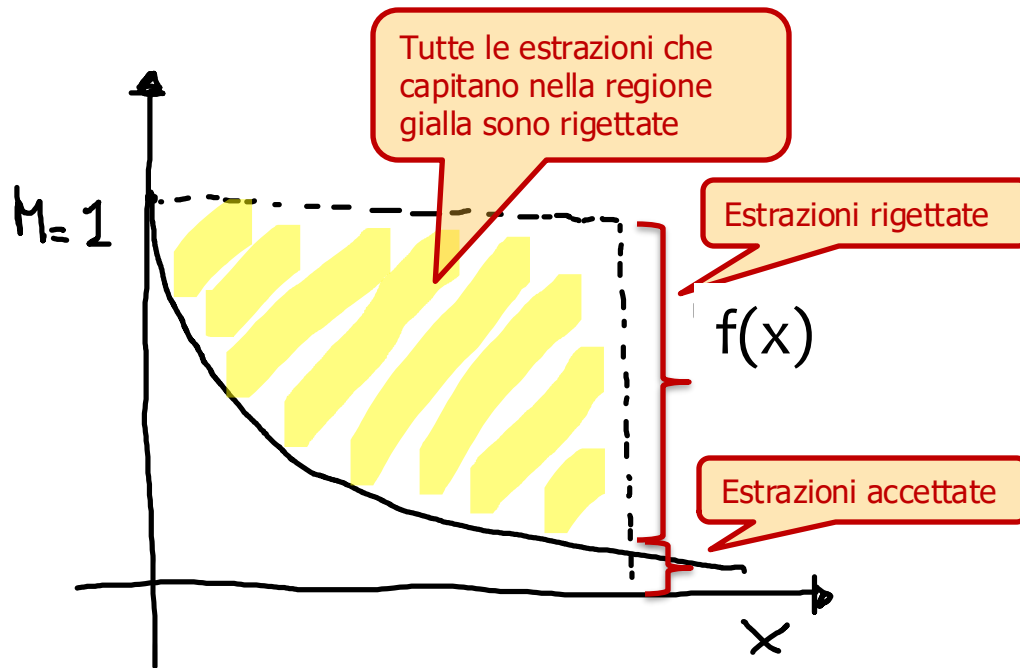
In pratica per ogni  $\bar{x}$  estratto (uniformemente in  $[a,b]$ ) decido se tenerlo o meno in base a quanto vale  $f(\bar{x})$  !  
Se  $f(\bar{x})$  è grande  $\bar{x}$  verrà statisticamente restituito più volte

Tecnica molto semplice e si applica con facilità ad ogni  $f(x)$  ma :

- ❑ Richiede la conoscenza a priori del massimo della  $f(x)$
- ❑ Richiede di impostare un intervallo di estrazione ( $a$  e  $b$ , se lo volessi fino a  $+\infty$  ? )
- ❑ può essere computazionalmente inefficiente in particolare nelle code di una distribuzione

## Dalla distribuzione uniforme a una distribuzione qualunque

Supponiamo di voler generare numeri casuali secondo una distribuzione esponenziale tra  $[0, +\infty)$  con il metodo Accept/Reject :



- ❑ richiede la conoscenza a priori del massimo della  $f(x)$  :
  - ❑ ok, facile il massimo è 1
- ❑ Richiede di impostare un intervallo di estrazione :
  - ❑ come rappresento  $+\infty$  ? Dovrò mettere un numero molto grande ( quanto ? )
- ❑ Può essere computazionalmente inefficiente:
  - ❑ per popolare bene le code ( specialmente quando  $x$  è "molto grande"

## Teorema della funzione inversa

Consideriamo una variabile aleatoria  $x$  con densità di probabilità  $f(x)$ . Se  $f(x)$  è integrabile e la sua cumulativa  $F(x)$  è continua e strettamente crescente (quindi invertibile) allora la variabile aleatoria  $y=F^{-1}(u)$  con  $u$  uniformemente distribuito in  $[0,1]$  è distribuita secondo  $f$

- ❑ Voglio generare  $x$  distribuita come una generica  $f(x)$
- ❑ Calcolo analiticamente ( se possibile )  $F$  e poi  $F^{-1}$
- ❑ Genero un numero casuale uniformemente distribuito in  $[0,1]$
- ❑  $y=F^{-1}(u)$  è distribuito secondo  $f$  !!

- ❑ Per dimostrare che  $y$  è distribuita come  $f$  dobbiamo mostrare che  $P(y < \bar{x}) = F(\bar{x})$

Se la  $F$  è monotona crescente posso applicarla ad entrambi

$$\begin{aligned} P(y < \bar{x}) &= P(F^{-1}(u) < \bar{x}) \\ &= P(F F^{-1}(u) < F(\bar{x})) \\ &= P(u < F(\bar{x})) = F(\bar{x}) \end{aligned}$$

Se  $u$  è distribuita uniformemente tra 0 e 1 la probabilità che sia minore di  $k$  è proprio  $k$

## Esempio : generatore esponenziale

❑ Consideriamo la densità di probabilità che vogliamo ottenere ( esponenziale )

$$f(y) = \begin{cases} \lambda e^{-\lambda y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

❑ Possiamo calcolare facilmente la funzione cumulativa :

$$F(\bar{y}) = \int_{-\infty}^{\bar{y}} \lambda e^{-\lambda y} dy = \int_0^{\bar{y}} \lambda e^{-\lambda y} dy = 1 - e^{-\lambda \bar{y}}$$

❑ Invertiamo la funzione cumulativa

$$U = 1 - e^{-\lambda \bar{y}} \Rightarrow U - 1 = -e^{-\lambda \bar{y}} \Rightarrow 1 - U = e^{-\lambda \bar{y}} \Rightarrow$$

$$y = -\frac{1}{\lambda} \ln(1 - U)$$



## Esempio : metodo di Box-Muller

- ❑ Consideriamo la densità di probabilità che vogliamo ottenere (gaussiana  $\mu=0$  e  $\sigma=1$ )

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- ❑ La gaussiana non è integrabile analiticamente proviamo ad aggirare il problema costruendo la probabilità combinata di due variabili aleatorie distribuite gaussianamente

$$f(x_1, x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} = \frac{1}{2\pi} e^{-\frac{(x_1^2 + x_2^2)}{2}}$$

$$F(\bar{x}_1, \bar{x}_2) = \int_{-\infty}^{\bar{x}_1} dx_1 \int_{-\infty}^{\bar{x}_2} dx_2 \frac{1}{2\pi} e^{-\frac{(x_1^2 + x_2^2)}{2}} \quad \begin{aligned} x'_1 &= r \cos(\theta) \\ x'_2 &= r \sin(\varphi) \end{aligned}$$

$$F(\bar{\vartheta}, \bar{r}) = \int_0^{\bar{\vartheta}} d\theta \int_{-\infty}^{\bar{r}} \frac{1}{2\pi} e^{-\frac{(r^2)}{2}} \boxed{r dr} = \frac{\bar{\vartheta}}{2\pi} \left(1 - e^{-\frac{\bar{r}^2}{2}}\right) = F(\bar{\vartheta})F(\bar{r})$$

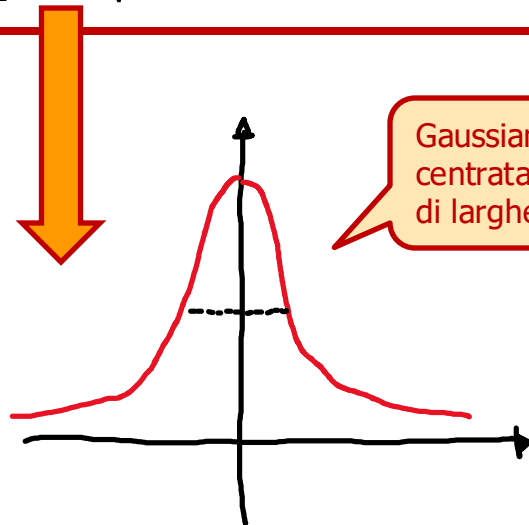
## Metodo di Box-Muller

$$t \in [0,1] \quad t = \frac{\theta}{2\pi} \Rightarrow \theta = 2\pi t$$

$$s \in [0,1] \quad s = \left(1 - e^{-\frac{r^2}{2}}\right) \Rightarrow r = \sqrt{-2 \ln(1 - s)}$$

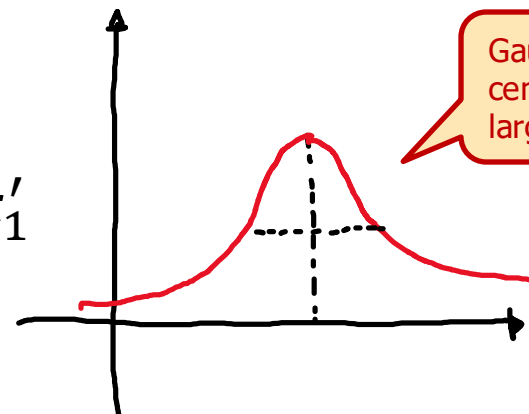
$$\begin{aligned} x'_1 &= \sqrt{-2 \ln(1 - s)} \cos(2\pi t) \\ x'_2 &= \sqrt{-2 \ln(1 - s)} \sin(2\pi t) \end{aligned}$$

Prendi due numeri casuali  $u$  ed  $s$  uniformemente distribuiti tra 0 e 1, allora  $x'_1, x'_2$  sono distribuiti come una gaussiana centrata in zero e di larghezza 1 !!



Gaussiana  
centrata in 0 e  
di larghezza 1

$$x'_{1t} = x_0 + \sigma_0 x'_1$$



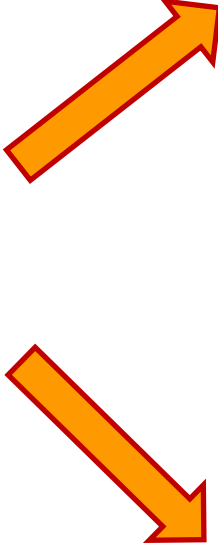
Gaussiana  
centrata in  $x_0$  e  
larghezza  $\sigma_0$

## Generare variabili aleatorie con PDF diverse

### Metodo accept/reject :

- ☐ Semplice implementazione, funziona per qualsiasi PDF vogliamo emulare
- ☐ Richiede impostazione di un intervallo di generazione e del massimo della funzione
- ☐ Può essere molto inefficiente

Generatore congruente lineare  
uniforme tra  $[0,1)$



### Metodo della trasformata ( o funzione inversa ):

- ☐ Richiede inversione della funzione cumulativa ( se possibile )
- ☐ Molto efficiente : per ogni numero estratto idealmente un numero fornito. Non richiede conoscenza di massimo o range di estrazione

# Generatori di numeri casuali

```
#ifndef __RANDOMGEN_H__
#define __RANDOMGEN_H__

class RandomGen {

public:

    RandomGen(unsigned int);

    void SetA(unsigned int a) {m_a=a;}
    void SetC(unsigned int c) {m_c=c;}
    void SetM(unsigned int m) {m_m=m;}

    double Rand( );
    double Unif(double xmin, double xmax);
    double Exp(double lambda);
    double Gaus(double mean, double sigma);
    double GausAR(double mean, double sigma, ... );

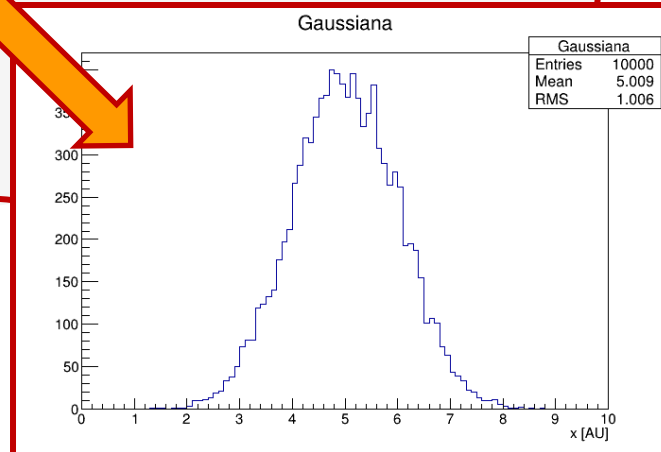
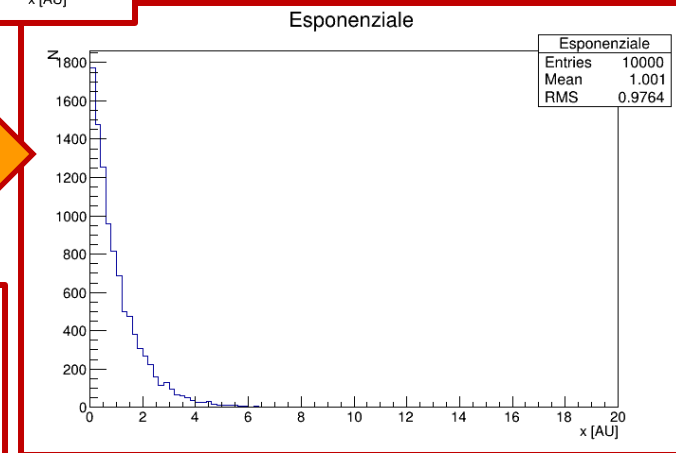
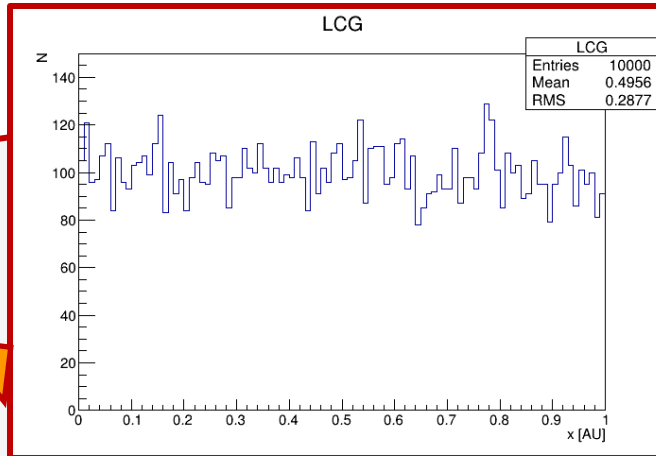
private:

    unsigned int m_a, m_c, m_m;
    unsigned int m_seed;

};

#endif
```

// distribuzione uniforme tra 0 e 1  
// distribuzione uniforme tra xmin e xmax  
// distribuzione esponenziale con cos  
// distribuzione gaussiana (Box-Muller)  
// distribuzione gaussiana (Accept-Reject)



## Legge dei grandi numeri e teorema del limite centrale

Sia  $x$  una variabile aleatoria distribuita secondo una densità di probabilità  $f(x)$  qualunque

$$\mu = \int_{-\infty}^{+\infty} x f(x) dx \quad \left( \sum x_i f_i \right) \quad \text{Caso discreto}$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \quad \left( \sum (x_i - \mu)^2 f_i \right) < +\infty$$

Caso discreto

Allora se costruiamo la variabile

$$Y_N = \frac{x_1 + x_2 + \dots + x_N}{N}$$

Si ha che ( se la varianza  $\sigma^2$  è finita ) :

- $Y_N \rightarrow \mu$  per  $N \rightarrow +\infty$
- $\sigma_{Y_N}^2 \rightarrow \frac{\sigma^2}{N}$  per  $N \rightarrow +\infty$

Il teorema del limite centrale inoltre afferma che la  $Y_N$  tende ad una distribuzione gaussiana

$$f(Y_N) \rightarrow \frac{1}{\sqrt{2\pi \left( \frac{\sigma^2}{N} \right)}} e^{-\frac{(Y_N - \mu)^2}{2 \left( \frac{\sigma^2}{N} \right)}} \quad \text{per } N \rightarrow +\infty$$

Montecarlo

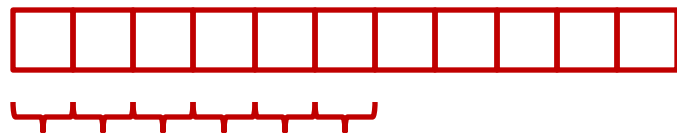
---

## Verifica del teorema del limite centrale

Generare una serie di numeri casuali uniformemente distribuiti in  $[0,1]$  e calcolare la somma eseguita su un numero  $n$  di elementi consecutivi della serie generata. Calcolare la varianza della serie di numeri generata e della serie delle somme. Verificare che questa scala con  $n$ .

- ❑ Passare da riga di comando sia il numero di elementi della serie di partenza (10000 può essere un buon numero) ed il numero di elementi su cui fare la somma. Creare due istogrammi che contengano la distribuzione dei numeri generata e la distribuzione delle somme. Verificare come cambia la distribuzione delle somme al variare di  $n$ .
- ❑ Scrivere un codice che faccia il ciclo in modo automatico : ciclo esterno fissa  $n$  crescente ( diciamo da 1 a 12 ). Ciclo interno genero 1000 somme di  $n$  numeri generati uniformemente. Fare un grafico della varianza della distribuzione delle 1000 somme in funzione di  $n$

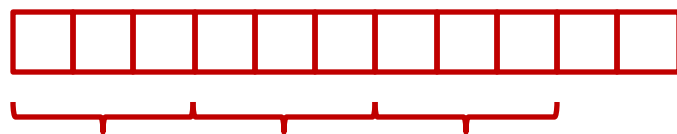
# Verifica del teorema del limite centrale



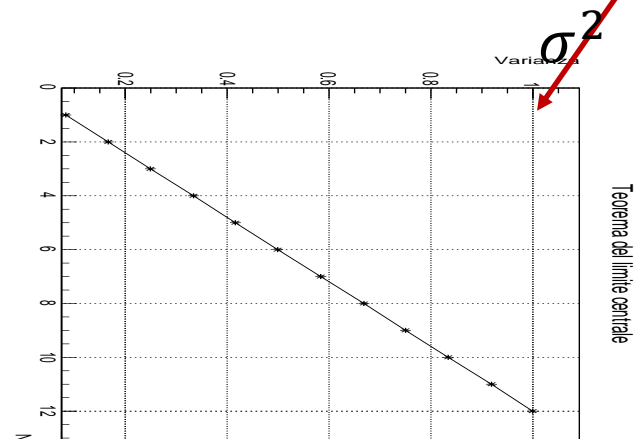
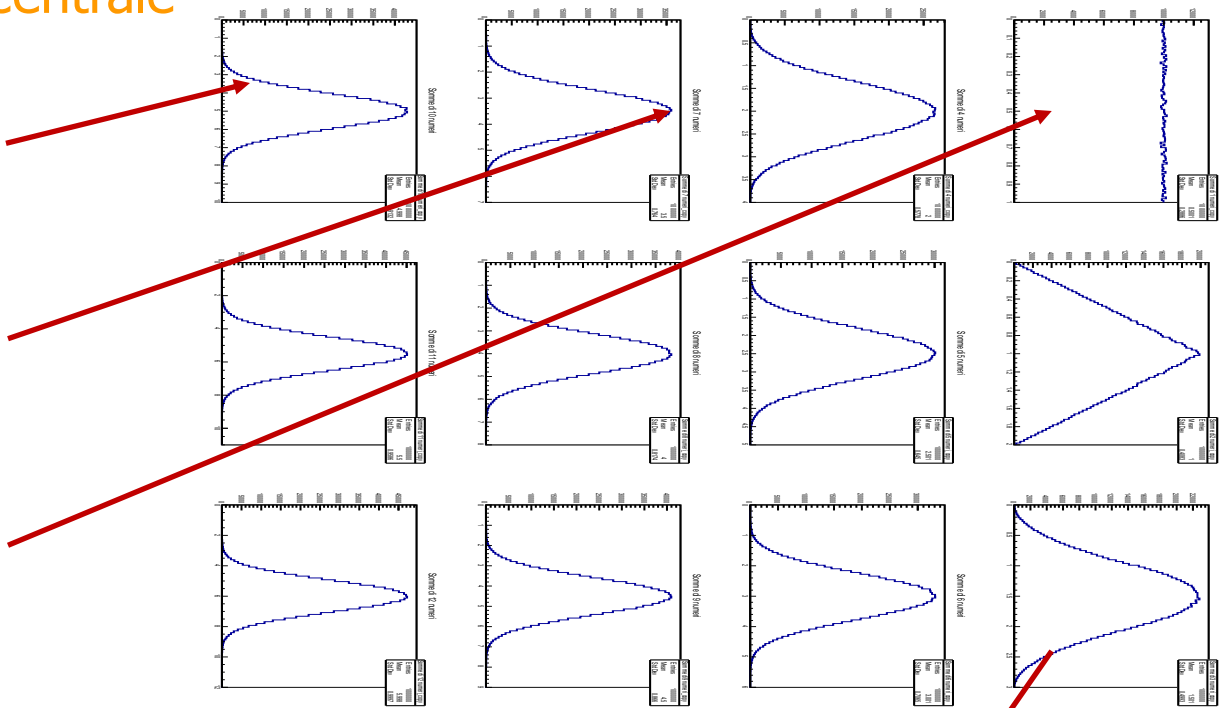
somme



somme



somme



## Test di randomness per generatori di numeri casuali

"The art of computer programming", D. E. Knut

- ❑ Quando viene ideato un generatore di numeri casuali, è necessario verificarne le proprietà. Come specificato precedentemente ci concentriamo sul generatore uniforme in  $[0,1)$  ( le altre distribuzioni nascono in modo semplice da questa). Le due proprietà che più ci interessano sono:
  - ❑ Uniformità della distribuzione dei numeri generati
  - ❑ Indipendenza del numero estratto dai precedenti
- ❑ Esistono tipicamente due tipologie di tests che possono essere utilizzati :
  - ❑ TEST EMPIRICI: esaminano un campione di uscita del generatore per identificare quanto il suo comportamento devia dalla casualità: test di frequenza, test seriale, test del gap e test del poker ...
  - ❑ TEST TEORICI: test 'a priori': risultati teorici che ci dicono in anticipo quanto bene verranno fuori quei test. Tali risultati teorici sono specifici per ogni tipo di generatore ma ci danno una comprensione più approfondita sui metodi di generazione rispetto ai risultati empirici.



## Test di randomness empirici : il test di ipotesi

Gli algoritmi per testare un generatore di numeri casuali si basano sul concetto di test di ipotesi: utilizziamo come esempio il test di uniformità.

- ❑ Partiamo due ipotesi, una dice che il generatore di numeri casuali è effettivamente distribuito uniformemente. La chiamiamo  $H_0$  (ipotesi nulla).
- ❑ L'altra ipotesi dice che il generatore di numeri casuali non è distribuito uniformemente. La chiamiamo  $H_1$  (ipotesi alternativa).
- ❑ Il test di ipotesi consiste nel cercare di rigettare  $H_0$  : provare che un generatore è davvero uniforme è complicato, meglio cercare almeno una evidenza di non-uniformità per concludere che il generatore non è uniforme («Uniforme fino a prova contraria»).

## Test di randomness empirici : il test di ipotesi

### ❑ Come si dichiara il risultato di un test?

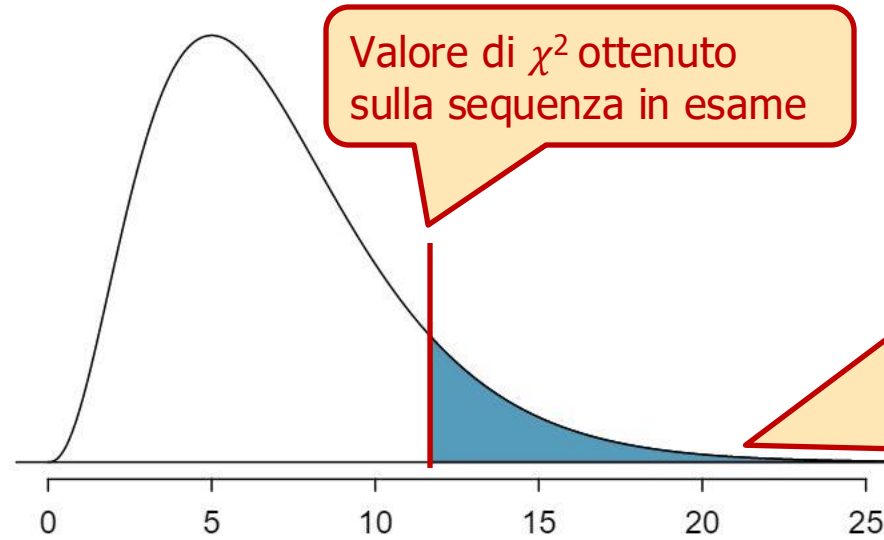
- ❑ Dobbiamo definire un test statistico ( per esempio uno dei più utilizzati è il  $\chi^2$ ) e soprattutto conoscere la distribuzione statistica del test che stiamo considerando
- ❑ Calcoliamo il p-valore  $p$ , l'integrale della distribuzione del test statistico  $-\infty$  al valore osservato, probabilità di ottenere un valore del test statistico maggiore di quello osservato
- ❑ Decidiamo un livello di significatività statistica  $\alpha$  (probabilità di rigettare  $H_0$  mentre  $H_0$  è vero che posso tollerare): se  $p < \alpha$  rigettiamo l'ipotesi nulla, il generatore NON è davvero uniforme. I valori tipici per  $\alpha$  sono 0.01 (1%) o 0.05 (5%).

### ❑ In generale si applicano si generano più sequenze di numeri casuali e si applica più di un test per rafforzare la confidenza nel risultato

### ❑ Esistono diverse suite di test codificati : [Dieharder](#) e [TestU01](#)

## Esempio di test empirico di uniformità : test del $\chi^2$

- ❑ Si chiede al generatore da mettere alla prova un set di numeri generati  $u_1 \dots u_n$  in  $[0,1)$  supposti essere distribuiti uniformemente
- ❑ Si divide l'intervallo  $[0,1)$  in  $k$  sotto-intervalli: la probabilità che un numero uniformemente estratto tra  $[0,1)$  entri nell'intervallo  $s$  è  $p_s = 1/k$
- ❑ Se si effettuano  $n$  estrazioni il valore atteso in ciascun intervallo è  $\bar{f}_s = n/k$
- ❑ Possiamo allora costruire la variabile aleatoria  $V = \sum_{s=1}^k \frac{(f_s - \bar{f}_s)^2}{\bar{f}_s}$  che è distribuita come un  $\chi^2$  con  $k-1$  gradi di libertà



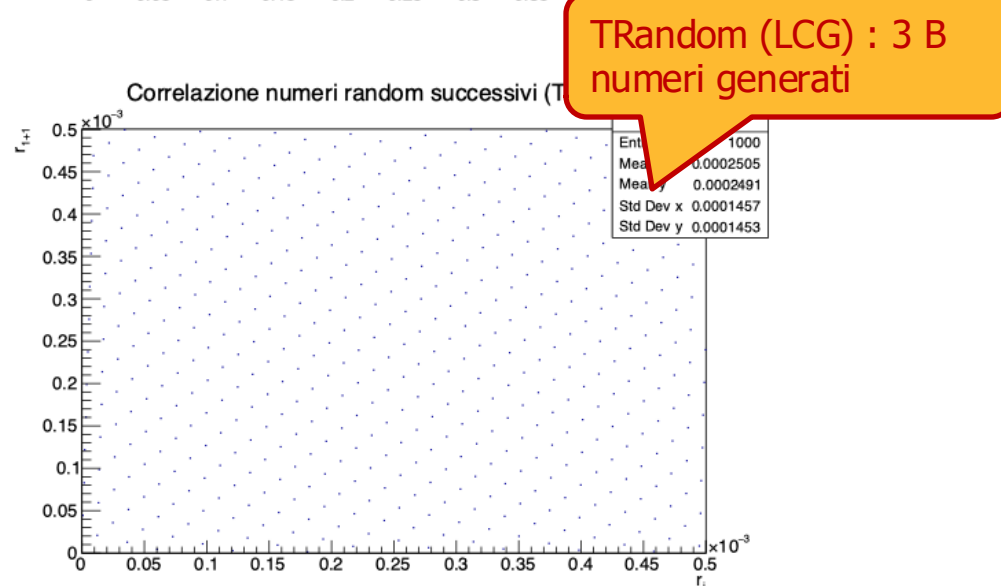
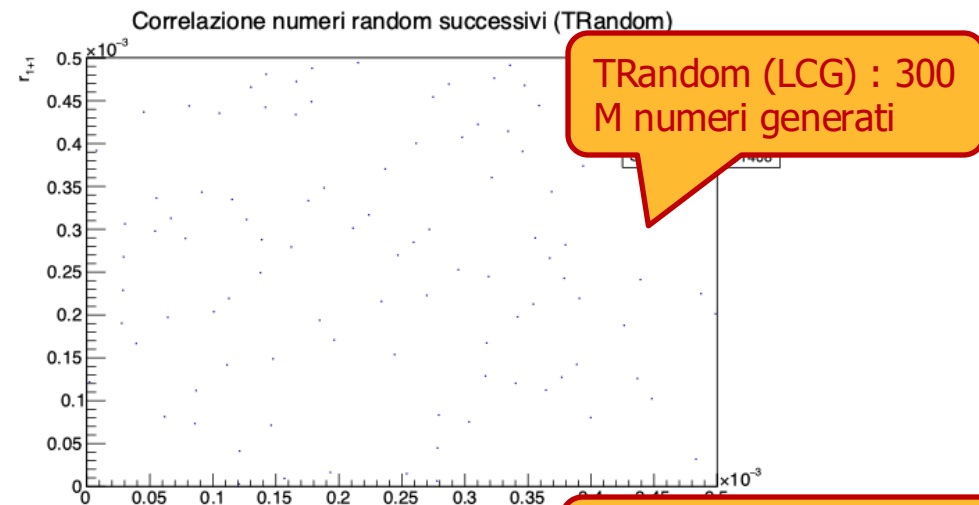
p-valore (area azzurra): probabilità di ottenere un valore di  $\chi^2$  maggiore di quello osservato sulla sequenza in esame.

- ❑ Se questo numero è  $< 0.05$  diciamo che il generatore NON è veramente uniforme (rigettiamo l'ipotesi nulla)
- ❑ Equivalentemente possiamo dire che è poco probabile che sia uniforme

## Esempi di test empirici di indipendenza

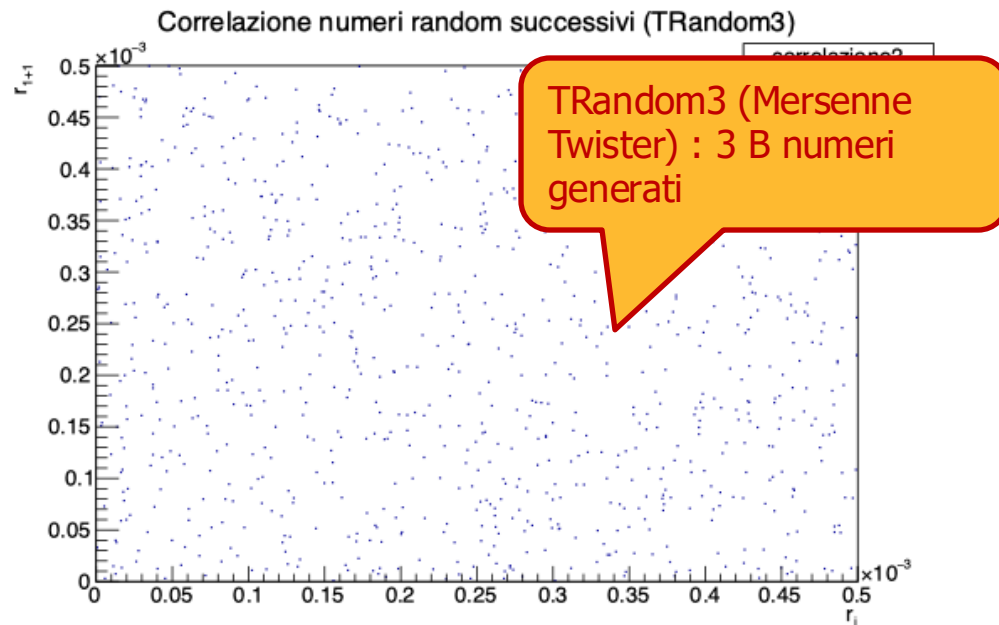
- ❑ Test del gap: questo test viene applicato a sottosequenze  $U_i, U_{i+1}, \dots, U_{i+r}$  e osserva la lunghezza del gap tra due successivi numeri pseudocasuali compresi in un determinato intervallo  $[a, \beta]$  con  $0 \leq a < \beta \leq 1$ .
  - ❑ Si genera una sequenza di numeri casuali  $U_0, \dots, U_N$ .
  - ❑ La sottosequenza  $U_{i-1}, U_i, \dots, U_{i+r-1}, U_{i+r}$  di  $r+2$  numeri, con  $U_{i-1}$  e  $U_{i+r} \in [a, \beta]$  mentre tutti gli  $U_i, \dots, U_{i+r-1}$  non appartengono a  $I$ , è chiamata gap di lunghezza  $r$ .
  - ❑ Dati  $a, \beta$  si conta il numero di gap di lunghezza  $0, 1, \dots, t-1$  e  $> t-1$  nella sequenza originale
  - ❑ Un test di  $\chi^2$  può essere definito sulla frequenza dei gaps di lunghezza  $i$  e di conseguenza può essere definito un livello di confidenza
- ❑ Test dei run: il test delle sequenze esamina la disposizione dei numeri in una sequenza per testare l'ipotesi di indipendenza.
  - ❑ Si calcola numero di run nella sequenza di numeri generati : un up-run è una sotto-sequenza di numeri ciascuno dei quali è seguito da un numero maggiore; un down-run è una sotto-sequenza di numeri ciascuno dei quali è seguito da un numero più piccolo
  - ❑ Se una sequenza di numeri ha troppe poche sequenze, è improbabile che si tratti di una vera sequenza casuale (es 0.08, 0.18, 0.23, 0.36, 0.42, 0.55, 0.63, 0.72, 0.89, 0.91)
  - ❑ Se una sequenza di numeri ha troppe sequenze, è improbabile che si tratti di una vera sequenza casuale (es 0,08, 0,93, 0,15, 0,96, 0,26, 0,84, 0,28, 0,79, 0,36, 0,57).
  - ❑ Un test statistico opportuno può essere definito sul numero di runs e di conseguenza può essere definito un livello di confidenza

## Esempio di test teorico : spectral test



- ❑ Se consideriamo coppie (n-uple) di punti successivi generati con un LCG notiamo che si dispongono su rette parallele ( iperpiani )
- ❑ Di per se non è un problema, granularità minima del nostro generatore (teoricamente indistinguibile da un set di numeri veramente random troncato con questa granularità )
- ❑ Si può calcolare la minima distanza tra due rette ( iperpiani ) e costruire un test che verifichi se questa è costante con il numero di dimensioni

## Esempio di test teorico : spectral test



- ❑ Se consideriamo coppie (n-uple) di punti successivi generati con un LCG notiamo che si dispongono su rette parallele ( iperpiani )
- ❑ Di per se non è un problema, granularità minima del nostro generatore (teoricamente indistinguibile da un set di numeri veramente random troncati con questa granularità )
- ❑ Si può calcolare la minima distanza tra due rette ( iperpiani ) e costruire un test che verifichi se questa è costante con il numero di dimensioni

## Esempi di distribuzioni discrete: binomiale

Consideriamo come esperimento  $n$  lanci di una moneta truccata dove  $p$  è la probabilità che esca testa e  $(1-p)$  la probabilità che esca croce

- ❑ Lo spazio degli eventi elementari è dato dalle  $n$ -uple  $\Omega = \{l_1, l_2, \dots, l_n\}$  con  $l_i$  che possono essere T ( successo ) o C ( insuccesso )
- ❑ Assegniamo ad ogni elemento di  $\Omega$  una probabilità : la probabilità di avere  $k$  successi (T) su  $n$  lanci è  $p^k(1-p)^{n-k}$ .
- ❑ Introduciamo la variabile aleatoria  $X = \{l_1, l_2, \dots, l_n\} = \#T$  (numero di successi)
- ❑ Assegniamo una distribuzione di probabilità alla variabile aleatoria  $X$ : la probabilità  $p^k(1-p)^{n-k}$  andrà moltiplicata per il numero di possibili  $n$ -ple  $(\omega_1, \omega_2, \dots, \omega_n)$  contenenti  $k$  volte T e  $n-k$  volte C, cioè  $\frac{n!}{k!(n-k)!}$ . Quindi la probabilità di ottenere  $k$  successi in  $n$  lanci vale

$$f(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$\mu(X) = np$$

$$\sigma^2(X) = p(1-p)n$$