

Lezione 4

Numeri Casuali

Generazione di Numeri Casuali

- ❑ Il calcolatore essendo una macchina deterministica non può generare una sequenza di veri numeri casuali. È possibile però che la macchina utilizzando particolari algoritmi sia in grado di produrre una sequenza di numeri che ai fini pratici per i quali sono generati si possano considerare casuali.
- ❑ Questi numeri sono detti **pseudo-casuali**. Vediamo un esempio di algoritmo di generazione $r_i = \text{mod}(a*r_{i-1} + c, m)$
- ❑ La funzione $\text{mod}(x,y) = x - \text{int}(x/y)*y$ ritorna il resto della divisione tra x e y
- ❑ Prendiamo per esempio (illustrativo) $c=1$, $a=4$ e $m=9$. Iniziamo con $r_1 = 3$ [questo valore iniziale è detto **seed (seme)**], allora $r_2 = 4$, $r_3 = 8$, $r_4 = 6$, ecc, $r_{10} = 3$, $r_{11} = 4$, ecc, ecc. La nostra sequenza è lunga 9 (come m che rappresenta il **periodo** dopo il quale la sequenza viene rigenerata).
- ❑ m deve essere il più grande intero possibile. In una macchina a 32 bit il massimo intero è $2^{32} - 1$

Generazione di Numeri Casuali

- ❑ Anche il parametro a va scelto in modo opportuno ($a = 16807$). Si noti che avendo fissato i parametri (a e c), l'algoritmo ripete la sequenza se gli viene dato lo stesso seed altrimenti produce un'altra sequenza.
- ❑ La riproducibilità della sequenza, il costo computazionale e la lunghezza del periodo sono aspetti molto importanti dell'algoritmo di generazione.
- ❑ Esistono molti generatori di numeri pseudo-casuali. Assumendo che `Random()` generi una distribuzione uniforme di numeri tra 0 e 1, se voglio generare una distribuzione uniforme di numeri ϕ tra 0 e 2π allora faccio $\phi = 2\pi \text{Random}()$
- ❑ Noi vogliamo ora vedere come sia possibile a partire da distribuzioni uniformi generare numeri, che chiamiamo d'ora in poi casuali, che siano distribuiti secondo un altro tipo di distribuzione.

Metodo della Trasformata

- ❑ Una variabile casuale r sia distribuita in modo uniforme. Noi vogliamo determinare una funzione $x(r)$ la quale segua una distribuzione $f(x)$. Sia $g(r)$ la pdf della variabile r .
- ❑ La probabilità di avere un valore tra r e $r + dr$ (cioè $g(r) dr$) deve essere uguale alla probabilità di avere un valore tra $x(r)$ e $x(r) + dx(r)$ (cioè $f(x) dx$) :

$$f(x) dx = g(r) dr$$

- ❑ Integrando passo alle cdf : $F(x(r)) = G(r)$

$$F(x(r)) = \int_{-\infty}^{x(r)} f(x') dx' = \int_{-\infty}^r g(r') dr' = r \quad [1]$$

- ❑ Se la distribuzione è uniforme, allora $G(r) = r$ ($0 \leq r \leq 1$).
- ❑ Se riesco a risolvere analiticamente la relazione [1], ottengo $x(r)$. Questo è possibile solo in qualche caso.

Distribuzione Esponenziale

- ❑ Consideriamo un esempio in cui si riesce a risolvere analiticamente la relazione appena vista. È questo il caso della distribuzione esponenziale:

$$\int_0^{x(r)} \frac{1}{\xi} e^{-x'/\xi} dx' = r$$

- ❑ Si sa integrare questa equazione ed la soluzione è: $x(r) = -\xi \log(1 - r)$
- ❑ Se una variabile r è uniforme in $[0,1]$, lo è anche la variabile $1 - r$. Quindi:

$$x(r) = -\xi \log r$$

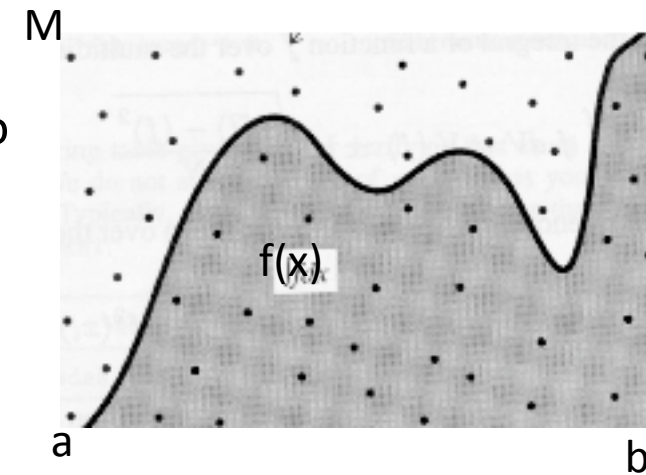
- ❑ Se genero uniformemente r tra 0 e 1, i valori $x(r)$ saranno distribuiti secondo una curva esponenziale (in queste distribuzioni ξ rappresenta il valore di aspettazione di x mentre ξ^2 è la varianza σ^2 di x).
- ❑ Nel caso di decadimenti di una risonanza studiata nel suo sistema di decadimento, ξ rappresenta la vita media della risonanza.

Metodo dell'Accettazione-Reiezione

❑ Quando non si può applicare il metodo della trasformata , allora si può utilizzare il metodo dell'accettazione-reiezione (hit and miss).

❑ Vogliamo generare numeri casuali compresi tra a e b distribuiti secondo una pdf $f(x)$.

❑ Indichiamo con M il massimo valore della funzione $f(x)$ in $[a, b]$.



1. Generiamo due numeri casuali r_1 e r_2 uniformemente distribuiti tra $[0, 1]$ e poniamo : $x_i = a + (b-a) \cdot r_1$ e per questo x_i prendiamo $y_i = M \cdot r_2$
2. Se $y_i > f(x_i)$ si ritorna al punto 1 (il punto y_i viene rigettato)
3. Se $y_i \leq f(x_i)$, allora x_i viene accettato. I punti accettati sono distribuiti come $f(x)$ per costruzione.

Metodo dell'Accettazione-Reiezione

- ❑ Questo metodo se la pdf $f(x)$ fosse stretta avrebbe bassa efficienza (e quindi alto costo computazionale) perchè la gran parte dei punti generati starebbe fuori della funzione $f(x)$ e quindi verrebbero rigettati.
- ❑ Il metodo può essere migliorato inscrivendo la funzione $f(x)$, invece che nel rettangolo visto prima, in una curva $g(x)$ che rappresenta una pdf di cui io so generare eventi. Per esempio potrei accorgermi che la funzione $f(x)$ è inscrivibile all'interno di una curva gaussiana.
- ❑ Allora genero punti in accordo alla gaussiana e li accetto o rigetto come fatto in precedenza. In questo caso l'efficienza può aumentare di molto!

Stima Monte Carlo di un Integrale

- ❑ Vogliamo calcolare l'integrale definito $I = \int_a^b f(x)dx$
- ❑ Consideriamo una sequenza di N punti casuali uniformemente distribuiti nell'intervallo [a,b] (**Uniform Sampling**). Il metodo Monte Carlo stima un valore approssimato dell'integrale utilizzando lo stimatore media (aritmetica):

$$\hat{I} = (b - a) \frac{1}{N} \sum_{i=1}^N f(x_i)$$

- ❑ Questo stimatore è non distorto e consistente. Al crescere di N il valore stimato tende al valore vero dell'integrale.

- ❑ La varianza dello stimatore della media è: $\sigma_{\bar{f}}^2 = \frac{\sigma_f^2}{N} = \frac{f^2 - \bar{f}^2}{N}$

$$\text{con } \bar{f} = \frac{1}{N} \sum_{i=1}^N f(x_i) \quad \text{e} \quad \bar{f}^2 = \frac{1}{N} \sum_{i=1}^N f^2(x_i)$$

Stima Monte Carlo di un Integrale

- $\frac{\sigma_f}{\sqrt{N}}$ è detto **errore standard della media**. Quindi tenendo conto dell'errore il valore dell'integrale stimato dal metodo Monte Carlo è:

$$I \approx (b-a)\bar{f} \pm (b-a)\frac{\sigma_f}{\sqrt{N}}$$

- Se volessimo calcolare un integrale di volume, $I \approx V\bar{f} \pm V\frac{\sigma_f}{\sqrt{N}}$ il valore approssimato sarebbe :

- Notiamo che indipendentemente dalle dimensioni dello spazio di integrazione l'errore scala come $\frac{1}{\sqrt{N}}$.

- Nell'integrazione numerica (con N passi) :
 - dimensione spazio = 1 $\rightarrow$ formula trapezi $\frac{1}{N^2}$ e Simpson $\frac{1}{N^4}$
 - dimensione spazio = d $\rightarrow$ formula trapezi $\frac{1}{N^{2/d}}$ e Simpson $\frac{1}{N^{4/d}}$

Stimatori Monte Carlo dell'Integrale

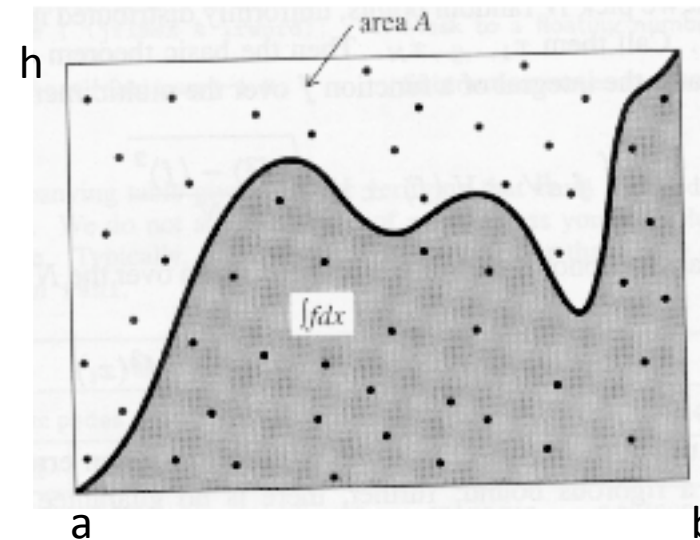
- ❑ Per avere una idea quantitativa: Se uno considera un numero di punti = 5 (un numero di soglia molto basso !!) in un integrale ad 1 dimensione, allora in un integrale in 10 dimensioni devo avere almeno 5^{10} (circa 10 milioni di punti nella formula dei trapezi !)
- ❑ Ma 5 punti sono in genere troppo pochi.
- ❑ Nelle stime MC degli integrali l'incertezza sulla misura dell'integrale scala come la radice quadrata del numero di punti N .
- ❑ Per integrazioni su spazi di dimensione ≤ 5 le integrazioni numeriche sono generalmente da preferire
- ❑ Le stime MC sono da preferire quando si è in spazi di dimensioni maggiori (≥ 6) ed anche in presenza di funzioni integrande particolarmente complesse!

Integrazione MC con Accett.-Reiezione.

- ❑ Area del rettangolo $A = (b-a) h$
- ❑ Scelgo un numero r_i casuale distribuito uniformemente tra 0 e 1 $\rightarrow x_i = a + (b-a) r_i$
- ❑ Per ogni x_i scelgo un secondo numero casuale u_i tra 0 e 1 e considero $y_i = u_i h$.
- ❑ Siano N il numero di punti considerati e N_s quelli tra questi N in cui y_i è minore o uguale a $f(x_i)$. Il valore approssimato dell'integrale è

$$I_n = A N_s / N$$

- ❑ Il metodo si generalizza ad integrali multidimensionali. L'efficienza può essere migliorata scegliendo opportunamente la figura in cui inscrivere la superficie (o il volume ecc) su cui si sta calcolando l'integrale!



Stima dell' Errore nell' Integrazione MC con Hit and Miss

□ L'errore standard su questa media è $\sigma_{\bar{f}} = \sqrt{\frac{\overline{f^2} - \bar{f}^2}{N}}$ con $\bar{f} = \overline{f^2} = \frac{N_s}{N}$

$$\sigma_{\bar{f}} = \sqrt{\frac{(N_s/N) - (N_s/N)^2}{N}} = \sqrt{\frac{N_s - N_s^2/N}{N^2}} = \frac{\sqrt{N_s - N_s^2/N}}{N}$$

$$I = (b-a)h \frac{N_s}{N} \pm (b-a)h \frac{\sqrt{N_s - N_s^2/N}}{N}$$

□ Se $N_s \ll N \rightarrow I = (b-a)h \frac{N_s}{N} \pm (b-a)h \frac{\sqrt{N_s}}{N}$ (N_s variabile poissoniana)

□ Tutto ciò vale indipendentemente dalle dimensioni dell'integrale (se per esempio è un integrale di volume, allora $(b-a)h \rightarrow V$ ed il resto non cambia)

Simulazione di Esperimenti

- ❑ Gli esperimenti devono essere progettati in modo tale da essere in grado di riuscire a fare la fisica per la quale sono stati progettati. Spesso devono migliorare risultati già presenti in letteratura.
- ❑ Per fare questo bisogna simulare l'apparato sperimentale e ottimizzarlo (scelta dei tipi di sottorivelatori, della geometria, ecc). Spesso le opzioni per lo stesso sottorivelatore sono diverse. Bisogna decidere quale opzione è la migliore.
- ❑ E non solo dal punto di vista della fisica ma anche dal punto di vista dei costi. Quindi un costo molto maggiore va giustificato con la possibilità di un risultato molto migliore, di meno rischi durante la presa dati, durata nel tempo, ecc .
- ❑ Per studiare il comportamento dell'apparato devo simulare eventi. Anche questi li simulo con tecniche Monte Carlo. Analizzo le possibili sorgenti di incertezze sistematiche, studio i bias, gli effetti di risoluzione finita dell'apparato, ecc

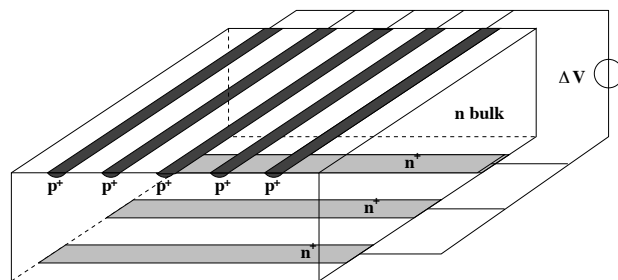
Simulazione di esperimenti

- ❑ **Full Monte Carlo Simulation.** Si simulano nel modo più dettagliato possibile gli eventi e si tracciano questi nell'apparato tenendo conto passo passo dei vari processi fisici che possono avvenire.
- ❑ **Fast Monte Carlo Simulation.** Nelle simulazioni degli eventi attraverso l'apparato sperimentale i processi fisici vengono simulati in modo approssimato usando valori medi delle varie grandezze in gioco. È il tipo di simulazione iniziale negli esperimenti.
- ❑ La Fast simulation è meno precisa della precedente ma molto più veloce (e molto più facile da realizzare nelle prime fasi di studio di un rivelatore). Permette di prendere le decisioni iniziali sulla struttura ottimale dell'apparato.
- ❑ Consideriamo ora il caso della simulazione del decadimento di una particella (per esempio un mesone B) nelle altre due particelle π e K all'interno di un certo rivelatore

Un Esperimento Semplificato

- ☐ Sia ξ la vita media della particella. In generale questo parametro è già misurato da altri in precedenza oppure che ha un valore suggerito dalla teoria. Se non so nulla, faccio qualche supposizione e simulo esperimenti con diversi valori di ξ .
- ☐ Estraggo numeri a caso da una distribuzione esponenziale col parametro uguale a ξ . Vogliamo simulare il decadimento del mesone B in $K\pi$
- ☐ Mi determino così un vertice di decadimento del B.
- ☐ Generato il vertice faccio decadere in questo punto il B nelle due particelle K e π sulla base dei possibili valori di quantità di moto che possono avere (nel rispetto dei principi di conservazione). Simulo così (p_{1x}, p_{1y}, p_{1z}) e (p_{2x}, p_{2y}, p_{2z}) delle due particelle K e π .

Un Esperimento Semplificato



❑ In figura lo schema di un rivelatore a doppio strato a piani ortogonali. Una particella carica passando attraverso una striscia produce coppie e^+ e^- e quindi lascia un segnale che viene letto. L'informazione ci dice che una particella carica è passata attraverso quella striscia

- ❑ Non sapendo dove è realmente passata la particella assumiamo una distribuzione uniforme e assegniamo alla particella come coordinata il valore centrale della striscia e come incertezza la larghezza della striscia (tipicamente $50\text{ }\mu\text{m}$) diviso la radice quadrata di 12.
- ❑ Per ogni rivelatore ho le coordinate della particella ricostruita dal rivelatore. L'insieme delle coordinate sui vari piani di microstrip li uso per ricostruire la traccia (pattern recognition) e quindi per calcolare le quantità di moto delle particelle $(p_{1x}', p_{1y}', p_{1z}')$ e $(p_{2x}', p_{2y}', p_{2z}')$ ricostruite dall'apparato.
- ❑ Con un fit delle tracce ricostruite mi determino il vertice di decadimento $V'(x,y,z)$

Simulazione di Eventi

- ❑ Le quantità di moto ricostruite ed i vertici ricostruiti hanno valori diversi da quelli simulati (che chiamiamo **MC veri**) . Questi effetti di distorsione sono dovuti alla risoluzione finita dell'apparato. Poichè gli eventi simulati seguono la stessa catena di ricostruzione degli eventi reali, questi subiscono effetti di distorsione simili a quelli degli eventi simulati.
- ❑ **Toy experiments**: Io rifaccio con simulazioni MC tanti esperimenti. Gli eventi sono simulati in modo completo (full simulation). Si simulano eventi di segnale ed eventi di tutti i tipi di fondo nelle proporzioni aspettate nei dati. Il numero di eventi MC simulati è pari (ma meglio 2-3 volte superiore) a quello dei dati.
- ❑ **Pure Toy Experiments**: Gli eventi sono simulati a partire dalle p.d.f. delle quantità rilevanti nell'analisi. Vedo che la massa di una particella è fittata da una gaussiana con determinati valori. Allora nei pure toy experiment simulo eventi in cui la massa della particella è estratta da una gaussiana con quei valori
- ❑ I pure toy sono meno precisi degli esperimenti (full) toy MC perchè non contengono alcun effetto di correlazione tra le variabili, non includono in modo preciso l'effetto di distorsione dell'apparato, ecc. Comunque è un primo metodo molto veloce e semplice per capire se le cose vanno nel verso giusto.^{1.7}

Uso di Esperimenti ed Eventi Simulati

- ❑ Quindi io simulo eventi MC, studio su questi eventi l'effetto dell'apparato. Da questo studio imparo come devo ottimizzare il rivelatore (numero di piani di microstrip per avere una buona efficienza di ricostruzione, la larghezza delle strisce, la distanza tra i piani, ecc.), eventualmente se devo cambiare tipo di rivelatore, ecc
- ❑ Questi eventi simulati poi mi sono molto utili nell'analisi dei dati per l'analisi delle correlazione tra le variabili in gioco, controllo e correzioni Dati/MC, calcolo di efficienze, per la validazione nei vari fit, per lo studio delle incertezze sistematiche, per lo studio della bontà del fit, ecc.
- ❑ Naturalmente nelle varie applicazioni si usano eventi simulati MC **quando non** si hanno a disposizione campioni di controllo di dati (veri).