

# PANORAMICA SUGLI ALGORITMI MULTICANONICI E NUOVO ALGORITMO MIGLIORATO, CON APPLICAZIONE ALLA PMD

A. Bononi, N. Rossi, A. Orlandini

Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma, v.le G. Usberti 181/A, 43100 Parma, Italy. Email: bononi@tlc.unipr.it

*A tutorial introduction to the standard Multicanonical Monte Carlo (MMC) algorithm is provided, and a novel fast MMC algorithm is introduced and compared to both standard MMC and Wang-Landau for group delay estimation in fibers with polarization mode dispersion.*

## 1. Introduzione

Obiettivo di questo lavoro è una introduzione didattica al recente metodo multi-canonico Monte Carlo (MMC) [1] per la stima della densità di probabilità (PDF) di una variabile aleatoria (VA), ed al suo legame con la più nota tecnica di importance sampling (IS) [2]. Sarà inoltre introdotto un nuovo metodo MMC migliorato, e ne sarà mostrata l'efficacia nella stima della PDF del ritardo differenziale di gruppo (DGD) di fibre con dispersione modale di polarizzazione (PMD). Il metodo MMC è stato introdotto nell'ambito delle comunicazioni ottiche principalmente da D. Yevick e dal gruppo di C. R. Menyuk [3,4].

## 2. Teoria

In comunicazioni ottiche è spesso di interesse conoscere la PDF di una VA  $Y = g(X)$ , funzione computazionalmente onerosa di un vettore aleatorio  $X \in \Gamma$ , dove  $\Gamma$  è lo spazio degli stati, ed è nota la PDF  $f_X(x)$  di  $X$  definita su  $\Gamma$ . Per esempio, se  $Y$  è il DGD di una fibra,  $X$  rappresenta l'insieme dei DGD e delle orientazioni aleatorie dei vari segmenti del random waveplate model col quale si emula la fibra. Se  $Y$  è la corrente al campionario in un sistema digitale di comunicazione,  $X$  rappresenta l'insieme di tutti i campioni di rumore lungo la linea di trasmissione e dei simboli aleatori adiacenti al simbolo di interesse che contribuiscono a fissare il valore di  $Y$ .

Per stimare per simulazione la PDF  $f_Y(y)$  di una VA continua  $Y$  sul range desiderato  $\mathcal{R}_Y$ , si discretizza  $\mathcal{R}_Y$  in  $M$  intervalli (*bin*) di ampiezza  $\Delta y$  centrati sui valori  $\{y_1, \dots, y_M\}$ . Se la funzione massa di probabilità (PMF) della  $Y$  discretizzata sul bin  $i$ -mo è  $P_i \equiv P(y_i) \triangleq P\{Y \approx y_i\}$ , allora si valuta la PDF dalla PMF come  $f_Y(y) \simeq P_i/\Delta y$  quando  $\Delta y$  è piccolo. Se  $D_i = \{x \in \Gamma : g(x) \approx y_i\}$  è il dominio in  $\Gamma$  che si mappa tramite  $g(\cdot)$  nel bin  $i$ -mo, e  $I_i(Y)$  è l'*indicatore* dell'evento  $\{Y \approx y_i\}$ , allora la PMF desiderata si può esprimere come valore atteso dell'indicatore:

$$P_i = \int_{D_i} f_X(x) dx = \int_{\Gamma} I_i(g(x)) f_X(x) dx = E[I_i(g(X))]. \quad (1)$$

Per questo nel metodo Monte Carlo (MC) classico si estraggono  $N$  campioni indipendenti  $\{X_1, \dots, X_N\}$  dalla PDF  $f_X(x)$ , e si stima il valore atteso dell'indicatore come conteggio relativo:  $\hat{P}_i^{MC} \triangleq \frac{1}{N} \sum_{j=1}^N I_i(g(X_j)) = \frac{N_i}{N}$ , essendo  $N_i$  il numero di campioni che cadono nel bin  $i$ -mo.

L'errore quadratico relativo nella stima del bin  $i$ -mo, definito come  $\varepsilon_i \triangleq \text{Var}[\hat{P}_i]/P_i^2$ , è per il MC:  $\varepsilon_i^{MC} = \frac{1-P_i}{NP_i}$ , che è circa l'inverso del valor atteso  $E[N_i] = NP_i$ . Nel MC la maggior parte dei campioni cade nei bin della zona modale, e nulla o quasi cade nei bin nelle code della PMF.

Ecco perchè per rendere possibile la stima della PMF nei bin delle code (eventi rari) è conveniente usare il metodo IS, che si basa sulla seguente scrittura alternativa di (1) [2]:

$$P_i = \int_{\Gamma} I_i(g(x)) \left[ \frac{f_X(x)}{f_X^*(x)} \right] f_X^*(x) dx = E^*[I_i(g(X))w(X)] \quad (2)$$

dove  $f_X^*(x)$  è una PDF di  $X$  *deformata* in modo da enfatizzare il conteggio nelle code (cioè  $f_X^*(x) > f_X(x)$  per  $x \in D_i$  nei bin  $i$  delle code),  $w(x) \triangleq f_X(x)/f_X^*(x)$  è detto *peso* della IS, ed  $E^*$

indica valore atteso rispetto alla PDF  $f_X^*(x)$ . Dalla (2) deriva la stima IS della PMF del bin i-mo:

$$\hat{P}_i^{IS} \triangleq \frac{1}{N} \sum_{j=1}^N I_i(g(X_j))w(X_j) = \left(\frac{N_i^*}{N}\right) \left[\frac{1}{N_i^*} \sum_{n=1}^{N_i^*} w(X_n)\right] \quad (3)$$

ottenuta estraendo  $N$  campioni indipendenti dalla PDF deformata e pesando ogni campione nel bin i-mo (ce ne sono  $N_i^*$ ) con il peso IS, che è funzione nota. La seconda uguaglianza mostra un metodo equivalente [5] di esprimere la stima IS come prodotto del conteggio relativo  $\hat{H}_i \triangleq N_i^*/N$  (che diremo *istogramma delle visite*) e del peso medio stimato dei campioni che cadono nel bin i-mo  $\bar{w}_i \triangleq \frac{1}{N_i^*} \sum_{n=1}^{N_i^*} w(X_n)$ . Si dimostra che lo stimatore IS è sempre centrato ( $E^*[\hat{P}_i^{IS}] = P_i$ ) e che l'errore quadratico relativo nella stima IS del bin i-mo è [5]:

$$\varepsilon_i^{IS} = \frac{1}{N} \left\{ \frac{1}{H_i} \left( \frac{\text{Var}^*[w(X)|X \in D_i]}{(E^*[w(X)|X \in D_i])^2} + 1 \right) - 1 \right\} \quad (4)$$

dove  $H_i = E^*[\hat{H}_i]$  è la probabilità che un campione generato con  $f_X^*$  cada nel bin i-mo, e  $E^*[w(X)|X \in D_i] = E^*[\bar{w}_i]$ . È evidente dalla (4) che una IS che ha peso  $w_i$  costante per ogni  $x \in D_i$  (uniform weight IS, UWIS) annulla  $\text{Var}^*[w(X)|X \in D_i]$ , e dunque riduce l'errore al valore:  $\varepsilon_i^{UWIS} = \frac{1}{N} \left\{ \frac{1}{H_i} - 1 \right\}$ , che è circa l'inverso del valor atteso  $NH_i$  del conteggio  $N_i^*$ , con evidente riduzione rispetto a  $\varepsilon_i^{MC}$  sui bin delle code in cui  $P_i \ll H_i \ll 1$ . La UWIS ottima per la stima *del solo bin i-mo* è inoltre quella che rende  $H_i = 1$ , ed è nota come *zero-variance* IS. Essa non è realizzabile in quanto necessita della conoscenza del dominio  $D_i \in \Gamma$  e della sua probabilità  $P_i$ , che è invece l'incognita che si vuole stimare.

Esiste anche una UWIS "ottima" per la stima dell'intera PMF su *tutti* i bin, che diremo flat-histogram IS (FHIS) [5]. Essa è ottenuta con la PDF deformata  $f_X^*(x) = \frac{f_X(x)}{MP(g(x))}$ , dove  $P(g(x)) \triangleq P\{Y \approx g(x)\}$  è la probabilità del bin in cui cade  $y = g(x)$ . Si noti che per costruzione la FHIS appartiene alla famiglia delle UWIS, in quanto  $w(x) = MP_i$  per ogni  $x \in D_i$ . Anche la FHIS presuppone nota la PMF incognita che si vuole stimare, e dunque non è realizzabile, ma è *approssimabile* come mostreremo nel seguito. La FHIS gode delle due seguenti proprietà [5]: i)  $H_i = E^*[\hat{H}_i] = 1/M$  per ogni bin (istogramma delle visite piatto in media, da cui il nome FHIS); ii)  $\varepsilon_i^{FHIS} = \frac{M-1}{N}$  per ogni bin: tutti i bin sono stimati con la stessa precisione relativa, in quanto in media ogni bin riceve lo stesso numero di campioni. Si dimostra che, tra tutte le UWIS, la FHIS è quella che *minimizza il peggior valore dell'errore relativo  $\varepsilon_i$  tra tutti i bin* [5]. Si capisce dunque da tali proprietà che FHIS è il meglio che si possa fare avendo a disposizione  $N$  campioni per stimare la probabilità di  $M$  eventi (la PMF di  $Y$ ) con identica precisione.

**Metodi FH e MC.** I metodi flat-histogram (FH) sono una famiglia di algoritmi, tra cui MMC, Wang-Landau (WL) [6] ed altri [7], che, partendo dalla PDF  $f_X(x)$ , costruiscono una sequenza di PDF deformate  $f_X^{(n+1)}(x) = \frac{f_X(x)}{c_n \Theta_n(g(x))}$ ,  $n = 0, 1, 2, \dots$  di tipo UW (dove  $\Theta_n \triangleq \{\Theta_n(y_i)\}_{i=1}^M$ , con elementi tutti non nulli, è una stima provvisoria della PMF di  $Y$  al passo  $n$ -mo e  $c_n$  la sua costante di normalizzazione) da cui si traggono campioni  $\{X_j\}$  che vengono utilizzati per formare una nuova stima  $\Theta_{n+1}$  della PMF di  $Y$ . La generazione dei campioni avviene con tecniche simili al metodo reject/accept di Von Neumann [5], nel quale l'idea è di sintetizzare  $f_X^*(x) = f_X(x|\mathcal{E})$ , cioè generando il campione  $X_j$  con  $f_X(x)$  ed accettandolo come campione di  $f_X^*$  se  $X_j$  verifica una specifica condizione  $\mathcal{E}$ . Ciò permette di generare campioni da  $\frac{f_X(x)}{c_n \Theta_n(g(x))}$  senza conoscere a priori i domini  $D_i$  in cui è partizionato lo spazio degli stati dalla funzione  $g(\cdot)$ , in quanto per ogni  $X_j$  generato si calcola  $Y_j = g(X_j)$ , il che permette di valutare  $\Theta_n(Y_j)$  che entra nella valutazione della condizione di accettazione  $\mathcal{E}$ .

Ciò che principalmente distingue gli algoritmi FH tra loro è la specifica *legge di aggiornamento*

$\Theta_n \rightarrow \Theta_{n+1}$  dalla quale dipende la velocità di convergenza. A convergenza, per confronto con la  $f_X^*$  FHIS, si ha  $c_n \rightarrow M$  e  $\Theta_n \rightarrow \underline{P} \triangleq \{P_i\}_{i=1}^M$ . La convergenza alla FHIS è empiricamente rivelata da un istogramma delle visite piatto in media. Il primo dei metodi FH fu l'MMC, introdotto da Berg *et al.* nel 1992 [1].

Nel MMC la legge di aggiornamento è basata sulla stima UWIS. Al passo (o *iterazione*)  $n + 1$ ,  $N$  campioni sono tratti da  $f_X^{(n+1)}$ , per ciascun campione si calcola  $Y_j = g(X_j)$ , da questi di valuta l'istogramma delle visite  $\hat{H}_i^{(n+1)} \triangleq \hat{H}_{n+1}(y_i) = N_i^{(n+1)}/N$ , ed una stima IS aggiornata della PMF di  $Y$  è data dalla (3):

$$\Theta_{n+1}(y_i) = \left( \frac{N_i^{(n+1)}}{N} \right) \left[ \frac{1}{N_i^{(n+1)}} \sum_{n=1}^{N_i^{(n+1)}} w(X_n) \right] = \hat{H}_{n+1}(y_i) c_n \Theta_n(y_i) \quad (5)$$

dove si è usato il peso costante  $w_i = c_n \Theta_n(y_i)$  della  $f_X^{(n+1)}$ . Come per ogni UWIS, si ha che [5]  $H_{n+1}(y_i) = E^{(n+1)}[\hat{H}_{n+1}(y_i)] = \frac{P_i}{c_n \Theta_n(y_i)}$ , da cui si ricava che  $\forall n E[\Theta_{n+1}(y_i)] = P_i$ , cioè lo stimatore MMC è centrato ad ogni passo. Ciò è evidente anche dal fatto che (5) è uno stimatore IS, dunque centrato per costruzione. Questa è una primaria differenza col metodo WL [6], che fornisce invece stime non centrate (*biased*), che diventano unbiased solo asintoticamente. Usando la (5), si faccia ora il rapporto tra PMF in bin adiacenti (il che elimina la costante  $c_n$ ) e se ne prenda il logaritmo, ottenendo la legge di aggiornamento

$$\beta_{n+1}(i) \triangleq \log \left( \frac{\Theta_{n+1}(y_i)}{\Theta_{n+1}(y_{i-1})} \right) = \beta_n(i) + \delta_{n+1}(i)$$

dove  $\delta_n(i) \triangleq \log(\hat{H}_n(y_i)/\hat{H}_n(y_{i-1}))$  è la stima rumorosa al passo  $n$  del *log-ratio* dell'istogramma  $H_n$ . Iterando, si ottiene  $\beta_{n+1}(i) = \sum_{t=1}^{n+1} \delta_t(i)$ , cioè la stima del log-ratio della PMF viene calcolata come somma degli stimatori  $\delta_t(i)$  ad ogni passo  $t = 1, \dots, n + 1$ . Questa non è la miglior stima che si possa ottenere combinando  $n + 1$  stimatori unbiased. Infatti la stima unbiased a minima varianza (MVUE) è  $\beta_{n+1}(i) = \sum_{k=1}^{n+1} \hat{g}_k(i) \delta_k(i)$ , dove  $\hat{g}_k(i) = \frac{g_k(i)}{\sum_{t=1}^{n+1} g_t(i)}$ , con  $g_t(i) \triangleq \frac{1}{\text{Var}[\delta_t(i)]}$ . Questa si traduce in una legge di aggiornamento  $\beta_{n+1}(i) = (1 - \hat{g}_{n+1}(i)) \beta_n(i) + \hat{g}_{n+1}(i) \delta_{n+1}(i)$ . Berg [1] suggerisce di approssimare  $\text{Var}[\log(\hat{H}_n(y_i))] \cong \frac{c}{\hat{H}_n(y_i)}$  usando i campioni ( $c$  è una costante che si elide in  $\hat{g}_n(i)$ ) e dunque approssima  $g_t(i) \cong \frac{1}{c} \frac{\hat{H}_t(y_{i-1}) \hat{H}_t(y_i)}{\hat{H}_t(y_{i-1}) + \hat{H}_t(y_i)}$ . Inoltre Berg usa l'aggiornamento  $\beta_{n+1}(i) = \beta_n(i) + \hat{g}_{n+1}(i) \delta_{n+1}(i)$  (ignorando il termine  $(1 - \hat{g}_{n+1}(i))$  della MVUE) che esplicitamente si traduce nelle legge di aggiornamento dell'*MMC standard* [1]:

$$\frac{\Theta_{n+1}(y_i)}{\Theta_{n+1}(y_{i-1})} = \frac{\Theta_n(y_i)}{\Theta_n(y_{i-1})} \left[ \frac{\hat{H}_{n+1}(y_i)}{\hat{H}_{n+1}(y_{i-1})} \right]^{\hat{g}_{n+1}(i)} \quad (6)$$

La legge di Berg funziona meglio della MVUE, e la ragione principale è che gli stimatori  $\delta_t(i)$  non sono unbiased.

**Il motore dei FH: Metropolis-Hastings.** Tutta la teoria è stata sinora esposta volutamente glissando sul meccanismo usato per generare la VA di stato  $X$  con distribuzione modificata  $f_X^*$  necessaria in tutti gli algoritmi FH. L'algoritmo più usato per tale generazione è il Metropolis-Hastings (MH) [8] e fa parte della famiglia degli algoritmi Markov-chain Monte Carlo (MCMC), che consistono nello scegliere i campioni  $X_1, \dots, X_N$  come realizzazioni di una catena di Markov con distribuzione a regime  $f_X^*$ . Tale sequenza di campioni è visualizzabile come un random-walk nello spazio degli stati. Pertanto i campioni sono correlati, e la taglia  $N$  di ogni iterazione MMC deve essere sufficientemente grande da garantire la convergenza delle statistiche del campione  $X_1, \dots, X_N$  ad  $f_X^*$ . L'effetto delle correlazioni è, a pari  $N$ , un aumento dell'errore relativo valutato

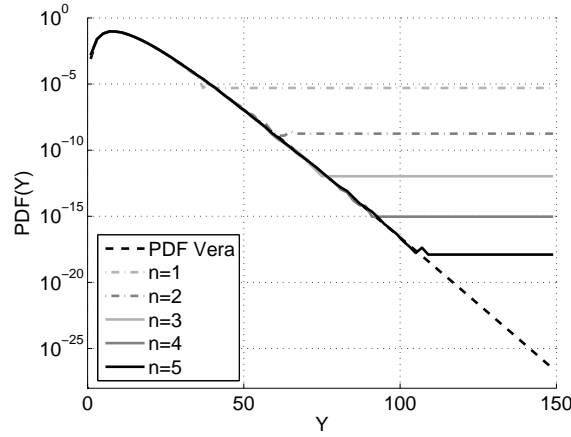


Fig. 1: PDF di un chi-quadro teorica a confronto con MMC con  $N = 10^5$  campioni per ciclo con  $n = 1, \dots, 5$  cicli.

in precedenza per campioni indipendenti. Con un approccio simile al metodo reject-accept, il MH ad ogni passo  $t$  dell'iterazione  $n + 1$  del MMC genera un proposal  $X_t = X_{t-1} + \Delta_t$  dove tipicamente  $\Delta_t$  è una VA indipendente dallo stato precedente  $X_{t-1}$ , a media nulla ed uniforme su un limitato range di "esplorazione", e lo accetta con probabilità  $\min[\frac{f_X^{(n+1)}(X_t)}{f_X^{(n+1)}(X_{t-1})}, 1]$ ; altrimenti lo rifiuta e mantiene  $X_t = X_{t-1}$ . Dunque il MH permette di estrarre (a regime) campioni da  $f_X^{(n+1)}(x) = \frac{f_X(x)}{c_n \Theta_n(g(x))}$  pur ignorando i domini  $D_i \in \Gamma$ , basandosi solo sull'*odds ratio*  $\frac{f_X^{(n+1)}(X_t)}{f_X^{(n+1)}(X_{t-1})} = \frac{f_X(X_t)}{\Theta_n(g(X_t))} \frac{\Theta_n(g(X_{t-1}))}{f_X(X_{t-1})}$  per valutare il quale non serve conoscere la costante  $c_n$ .

**Esempio MMC.** Per capire con un esempio elementare il funzionamento del MMC standard, vediamo la stima della PDF della VA  $Y = \sum_{i=1}^{10} X_i^2$ , dove le  $X_i$  sono gaussiane standard indipendenti e lo spazio degli stati è  $\Gamma = \mathbb{R}^{10}$ . La soluzione è qui nota essere un Chi-quadro, e potremo confrontarla con la stima MMC. L'intervallo di  $Y$  esplorato è  $\mathcal{R}_Y = [0, 150]$ , diviso in  $M = 75$  bin uniformi di taglia  $\Delta y = 2$ .

La Fig. 1 mostra le stime MMC della PDF  $\hat{f}_Y(y_i) = \Theta_n(y_i)/\Delta y$ ,  $i = 1, \dots, M$ , per i cicli  $n = 1, \dots, 5$ , con numero di campioni per ciclo  $N = 10^5$ . Il primo ciclo MMC è in sostanza un MC puro, in quanto si assume  $\Theta_0$  pari alla distribuzione uniforme. Quando un bin ed il successivo non sono esplorati, il valore di sinistra della PMF è propagato dalla (6) al bin corrente, e ciò crea i "floor" visibili in figura. La caratteristica saliente del MMC è quella di definire subito in modo preciso la zona modale, e lentamente scendere ciclo per ciclo allargando l'esplorazione rispetto alla zona modale. L' MMC standard è inefficiente in quanto ad ogni ciclo successivo al primo buona parte dei campioni per ciclo  $N$  viene sprecata nella reiterata esplorazione della zona modale.

### 3. MMC veloce

Dopo l'esposizione *tutorial* precedente, in questa sezione viene proposta una originale versione più efficiente dell'MMC (*MMC veloce*) che eviti per quanto possibile la reiterata esplorazione della zona modale. L' MMC veloce pre-definisce una partizione del range da esplorare  $\mathcal{R}_Y = A_1 \cup \dots \cup A_K$  in  $K$  sotto-intervalli, e fa uso della stessa legge di aggiornamento (6) dell'MMC standard ad ogni iterazione, con il presupposto però di interdire il random walk (e dunque l'aggiornamento dell'istogramma  $\hat{H}$  delle visite) in un particolare sottointervallo  $A_k$  a partire dall'iterazione  $n + 1$  successiva a quella in cui la PMF stimata ha, per ogni bin  $y_i \in A_k$ , un'accuratezza  $\delta_i$  desiderata:

$$\delta_i = \left| \frac{\Theta_n(y_i) - \Theta_{n-1}(y_i)}{\Theta_n(y_i)} \right| < \varepsilon \quad (7)$$

dove  $\varepsilon$  è un livello di accuratezza prefissato. Le iterazioni dell'MMC veloce giungono al termine una volta ottenuta l'accuratezza desiderata sull'intero intervallo  $\mathcal{R}_Y$ . Per concretezza e per fornire maggiori dettagli implementativi, si supponga  $\mathcal{R}_Y$  suddiviso in tre sottointervalli  $A_1, A_2, A_3$ , dove ad  $A_1$  corrisponde la zona modale, mentre ad  $A_2$  e  $A_3$  corrispondono le code. La nuova strategia prevede di effettuare un primo ciclo multicanonico inizializzato con PMF uniforme  $\Theta_0$  sull'intero  $\mathcal{R}_Y$ , che coincide con una simulazione MC pura. Si effettua poi un secondo ciclo MMC ancora sull'intero  $\mathcal{R}_Y$  per ottenere un primo valore di  $\delta_i$  dalla (7) per ogni bin in  $\mathcal{R}_Y$ , ottenendo la stima  $\Theta_2$  dalla (6). Ipotizzando che per ogni bin  $y_i$  di  $A_1$  sia  $\delta_i < \varepsilon$ , la PMF valutata su  $A_1$  è considerata definitiva. Nel ciclo  $n = 3$ , le visite vengono concentrate in  $\mathcal{R}_Y \setminus A_1$  e proibite in  $A_1$ . Calcolata nuovamente  $\delta_i$  per ogni bin in  $\mathcal{R}_Y \setminus A_1$  per  $\Theta_2$ , se per ogni  $y_i \in A_2$  è  $\delta_i < \varepsilon$ , la stima della PMF in  $A_2$  viene considerata definitiva e l'intervallo di esplorazione del random walk si restringe successivamente ad  $A_3$ . Un passo fondamentale di questo algoritmo è quello di concentrare le visite di  $Y$  solo nei sottointervalli desiderati, creando dunque una sorta di barriera tra tali intervalli e quelli proibiti. Nell'esempio, supponendo che all'iterazione attuale  $n$  la zona modale  $A_1$  sia proibita, si effettua inizialmente un random walk sull'intero  $\mathcal{R}_Y$  col solo scopo di "entrare" nelle code  $\mathcal{R}_Y \setminus A_1$ , senza tuttavia aggiornare l'istogramma delle visite  $\hat{H}_n(y_i)$ , finché il campione  $Y$  proposto non cade in  $\mathcal{R}_Y \setminus A_1$  e la transizione viene accettata tramite l'algoritmo MH. Dopo questa fase preliminare, inizia il random walk in  $\mathcal{R}_Y \setminus A_1$  "utile" in cui si aggiorna  $\hat{H}_n$ . Se durante questa fase viene proposto un campione in  $A_1$ , la transizione viene sempre rigettata ed in  $\hat{H}_n$  viene ri-contato il campione precedente. Si opera analogamente nell'iterazione in cui  $A_1 \cup A_2$  sono proibiti dal random walk utile in  $A_3$ . Per avere continuità della PMF stimata sugli  $\{A_k\}$ , ogni sottointervallo ha un bin di sovrapposizione con quelli adiacenti. In questo modo la legge di aggiornamento (6) propagata dal primo all'ultimo bin produce una funzione senza discontinuità che, dopo ri-normalizzazione, fornisce la stima aggiornata.

**Wang-Landau.** Anche la tecnica WL adotta l'algoritmo Metropolis per generare campioni nello spazio degli stati con distribuzione limite  $f_X^{(n)}(x) = \frac{f_X(x)}{c_{n-1} \Theta_{n-1}(g(x))}$ , ma, differentemente da MMC, la PMF stimata viene aggiornata *ad ogni passo*  $n$  del random walk (anziché dopo  $N$  passi come in MMC) con la seguente legge [6,7]:

$$\Theta_n(y_i) = \begin{cases} f_m \cdot \Theta_{n-1}(y_i) & \text{se } g(X_n) \approx y_i \\ \Theta_{n-1}(y_i) & \text{altrimenti} \end{cases} \quad (8)$$

dove  $f_m > 1$  è un parametro tenuto costante per tutto il ciclo  $m$ -mo del WL. A parole: se il campione  $g(X_n)$  estratto al tempo  $n$  cade nel bin  $i$ , si incrementa la stima su quel bin, altrimenti la si lascia invariata. La (8) incrementa il denominatore di  $f_X^{(n+1)}(x)$  e permette dunque di rendere meno probabile ai passi successivi l'esplorazione dei bin in cui si sono già collezionate visite, spingendo il random walk ad esplorare nuovi bin. Tipicamente le prime visite si concentrano nella regione modale della PDF, e dunque WL subito inibisce tali bin e tende a spostare le visite velocemente verso le code, benché il profilo della PMF stimata nella zona modale sia ancora lontano dal valore corretto. Al ciclo  $m$ -esimo questo procedimento continua finché non si ottiene un istogramma delle visite "piatto" su tutti i bin di  $\mathcal{R}_Y$  secondo una tolleranza prefissata (attorno al 20% [6]). All'inizio del ciclo successivo  $m + 1$  si riazzerà l'istogramma delle visite e si riduce il parametro  $f_{m+1}$  che determina il raffinamento della stima, con la legge  $f_{m+1} = \sqrt{f_m}$  [6]. Nel WL standard si continua ad iterare finché  $f_m < \exp(10^{-8})$  [6], mentre, per un confronto equo col nostro MMC, nella nostra applicazione del WL si è scelto di ultimare la simulazione una volta soddisfatta la condizione  $\delta_i < \varepsilon$  su ogni bin di  $\mathcal{R}_Y$ .

**Confronto MMC, WL e MMC veloce.** Per evidenziare il notevole guadagno computazionale del nostro metodo MMC veloce rispetto alle tecniche MMC standard e WL, si è scelto il classico problema della stima della PDF  $f_y(y)$  del DGD totale  $Y = \Delta\tau$  di una fibra modellizzata come concatenazione di 25 sezioni di fibra a mantenimento di polarizzazione (PMF) con DGD locale  $\Delta\tau_{loc}=1$  ps e stati principali di polarizzazione (PSP) uniformemente distribuiti sulla sfera

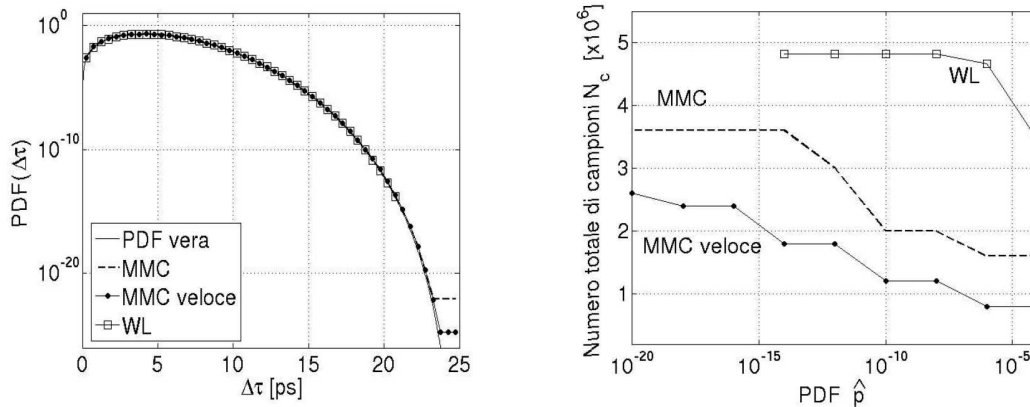


Fig. 2: (Sinistra) PDF del DGD teorica a confronto con MMC e MMC veloce (taglia ciclo  $N=2 \cdot 10^5$ ) e WL; (destra) numero di campioni necessario nei tre casi per ottenere i valori della PDF in ascissa.  $\varepsilon=5\%$ .

di Poincaré, in cui l'espressione esatta di  $f_Y(y)$  è nota [9]. Qui lo stato  $X$  è rappresentato da azimuth ed ellitticità dei PSP di ogni PMF. Ci si è posti come primo obiettivo la stima della PDF fino a  $10^{-20}$  con un'accuratezza (7) del 5% ( $\varepsilon = 0.05$ ) sull'intervallo di osservazione  $\mathcal{R}_Y = [0, \dots, 25]$  ps avente prefissata suddivisione uniforme in  $M = 50$  bin per tutti gli algoritmi. In Fig. 2 (sinistra), sono riportate le stime ottenute dalle diverse tecniche a confronto con la curva teorica (linea continua). Si nota che il metodo WL (quadrati) consente una stima accurata della PDF in tempi computazionali ragionevoli non oltre  $10^{-14}$ , data la tipologia di binning utilizzata. Entrambe le tecniche MMC, che utilizzano cicli da  $N = 2 \cdot 10^5$  campioni l'uno, permettono di raggiungere l'accuratezza imposta, mostrando un buon accordo con la curva teorica. In Fig. 2 (destra) si mostra il numero di campioni  $N_c$  totale necessario per raggiungere i valori della PDF in ascissa (che corrispondono a quelli in ordinata in Fig. 2 (sinistra)) con  $\varepsilon=5\%$ , confrontando la modalità e la velocità di convergenza degli algoritmi esaminati. Con entrambe le tecniche MMC la convergenza nella zona modale ( $f_Y > 10^{-6}$ ) della PDF è molto rapida e si estende alla coda con una velocità che è funzione del numero di campioni per iterazione  $N$  e del numero di bin considerati. Si noti che l'MMC veloce presenta un guadagno computazionale evidente rispetto all'MMC, risparmiando circa il 50% dei campioni per valori della PDF fino a  $10^{-14}$  e il 30% intorno a  $10^{-20}$ . La tecnica WL ha una convergenza diversa: infatti, l'accuratezza nella zona modale viene ottenuta con un numero più elevato di campioni (ad esempio, WL richiede  $4 \cdot 10^6$  campioni contro i  $1.5 \cdot 10^6$  dell'MMC e gli  $8 \cdot 10^5$  dell'MMC veloce), poi  $N_c$  satura rapidamente a partire da  $10^{-6}$ . Tale comportamento deriva dalle caratteristiche proprie del WL sopra descritte.

#### 4. Conclusioni

Dopo una introduzione tutorial sul legame che intercorre tra IS e MMC, e sulle caratteristiche dell'MMC, si è proposto un algoritmo MMC migliorato che consente un buon guadagno computazionale rispetto sia all'MMC standard che al WL, come dimostrato sul classico problema della stima del DGD di una fibra con PMD.

#### Bibliografia

- [1] B. A. Berg et al., *Fields Instr. Commun.*, vol. 26, pp. 1-24, 2000.
- [2] M. Jeruchim, *IEEE J. Sel. Areas. Commun.*, vol. SAC-2, pp. 153-170, Jan. 1994.
- [3] D. Yevick, *IEEE Photon. Technol. Lett.*, vol. 15, pp. 224-226, Feb. 2003.
- [4] G. Biondini et al., *IEEE Photon. Technol. Lett.*, vol. 14, pp. 310-312, Mar. 2002.
- [5] A. Bononi, [www.tlc.unipr.it/bononi/ricerca/seminars/MMCcourse.pdf](http://www.tlc.unipr.it/bononi/ricerca/seminars/MMCcourse.pdf), Jan. 2007.
- [6] F. Wang et al., *Phys. Rev. Lett.*, vol. 86, pp. 2050-2053, Mar. 2002.
- [7] F. Liang, *J. Stat. Phys.*, vol. 122, pp. 511-529, Feb. 2006.
- [8] W. K. Hastings, *Biometrika*, vol. 57, pp. 97-109, Apr. 1970.
- [9] M. Karlsson, *J. Lightw. Technol.*, vol.19, pp. 324-331, Mar. 2001.