

27.1 Sequential Importance Sampling (SIS)

Data una $p(x)$, con $x \in \mathbb{R}^n$, riscriviamo $q(x)$ come:

$$q(x) = \prod_{k=2}^n q(x_k | x_{k-1}, \dots, x_1) q(x_1)$$

Dati gli N campioni $x^i \sim q(x)$ (per $i = 1, \dots, N$) con $x^i \in \mathbb{R}^n$ si ha

$$\pi(x^i) = \frac{p(x^i)}{q(x^i)} = \prod_{k=2}^n \underbrace{\frac{p(x_k^i | x_{k-1}^i, \dots, x_1^i)}{q(x_k^i | x_{k-1}^i, \dots, x_1^i)}}_{\alpha_k^i} \underbrace{\frac{p(x_1^i)}{q(x_1^i)}}_{\alpha_1^i}$$

$$\pi(x^i) = \prod_{k=1}^n \alpha_k^i$$

$$\pi_h(x^i) = \prod_{k=1}^h \alpha_k^i$$

$$\pi_h(x^i) = \pi_h^i = \alpha_h^i \pi_{h-1}^i$$

Una volta calcolati tutti i π per ciascuna componente, questi vanno rinormalizzati.

Ad ogni passo, durante il quale viene generato un nuovo campione, vengono eseguite le seguenti operazioni:

- Campionamento
- Aggiornamento del peso
- Rinormalizzazione

27.2 Filtri Particellari

Dati N campioni, ad ogni istante di tempo t si ha un vettore degli stati $x_t \in \mathbb{R}^m$.
Vogliamo calcolare

$$p(\underbrace{x_t, x_{t-1}, \dots, x_1}_{X_t} | \underbrace{y_t, y_{t-1}, \dots, y_1}_{Y^t})$$

cioè vogliamo essere in grado di generare campioni da questa densità.

Useremo poi una q diversa, $q(x_t, x_{t-1}, \dots, x_1 | y_t, y_{t-1}, \dots, y_1)$, da cui campionare. Riscriviamo la p come:

$$\begin{aligned} p(x_t, x_{t-1}, \dots, x_1 | y_t, y_{t-1}, \dots, y_1) &= \frac{p(X_t, y_t | y_{t-1}, \dots, y_1)}{p(y_t | y_{t-1}, \dots, y_1)} \\ &= \frac{p(x_t, y_t | X_{t-1}, Y^{t-1}) p(X_{t-1} | Y^{t-1})}{p(y_t | Y^{t-1})} \\ &= \frac{p(y_t | x_t, X_{t-1}, Y^{t-1}) p(x_t | X_{t-1}, Y^{t-1}) p(X_{t-1} | Y^{t-1})}{p(y_t | Y^{t-1})} \\ &= \frac{\overbrace{p(y_t | x_t)}^{\text{modello misura}} \overbrace{p(x_t | x_{t-1})}^{\text{modello dinamico}} p(X_{t-1} | Y^{t-1})}{p(y_t | Y^{t-1})} \\ &\propto p(y_t | x_t) p(x_t | x_{t-1}) p(X_{t-1} | Y^{t-1}) \end{aligned}$$

dove sono state sfruttate la formula di Bayes e la markovianità del processo.

Supponendo che però da p non si possa campionare, benché lo si vorrebbe, ciò che si può fare è proporre una q siffatta:

$$q(X_t | Y^t) = q(x_t | X_{t-1}, Y^t) q(X_{t-1} | Y^t) := \underbrace{q(x_t | x_{t-1}, y_t)}_{\text{parte presente}} \underbrace{q(X_{t-1} | Y^{t-1})}_{\text{parte passato}}$$

dove la parte presente è ottenuta ipotizzando un processo markoviano di dinamica, mentre nella parte passata è stata eliminata la misura y_t in quanto si vuole appunto una densità che dipende solo dalle misure passate. Questo non significa che la parte a destra dell'equazione sia uguale a quella sinistra, ma essendo $q(x)$ definibile a “piacere”, viene scelta in modo tale da ottenere modalità di campionamento facilmente implementabili.

Il peso α_t (peso corrente), secondo le equazioni derivate qui sopra, sarà dato in questo caso da:

$$\alpha_t = \frac{p(y_t | x_t) p(x_t | x_{t-1})}{q(x_t | x_{t-1}, y_t)} .$$

Il più semplice approccio è quello di scegliere la q in base al modello *a-priori* della dinamica:

$$\begin{aligned} q(x_t | x_{t-1}, y_t) &:= p(x_t | x_{t-1}) \\ &\Downarrow \\ \alpha_t &= p(y_t | x_t) \end{aligned}$$

Ovviamente, per ogni scelta di q si ottiene un diverso filtro particellare. La scelta di q infatti risulta il parametro di progettazione più importante dei filtri particellari e molta della ricerca si concentra appunto nella progettazione di q facili da implementare ma che allo stesso tempo garantiscano una buona approssimazione di p .

Per la particolare scelta di proposal density qui sopra, il filtro particellare si riduce a:

- All'istante $t = 1$ (per $i = 1, \dots, N$) si ha $x_1^i \sim q(x_1)$, dove $i = 1, \dots, N$ sono le particelle, aventi peso $\alpha_1^i = p(y_1|x_1^i)$. Come si vede il peso è tanto più grande quanto migliore è l'evidenza di y_1 data la particella x_1^i .
- All'istante $t = 2$, in base al modello, si genera la traiettoria successiva $x_2^i \sim p(x_2|x_1^i)$ ed il peso corrispondente è dato da $\alpha_2^i = p(y_2|x_2^i)$.

Dopo alcuni passi si osserva che i pesi relativi variano molto tra di loro; ciò implica che $Var_q(\pi)$ cresce. Per tale ragione si ricampiona allo scopo di renderli "più uguali possibili" e quindi abbattere la varianza. Questo però si ottiene a scapito di una perdita di indipendenza tra i campioni e quindi una peggiore prestazione.

27.3 Resampling

Partendo dall'aspettazione $E_q[\pi] = 1$ e campionando secondo la densità q , ottengo la media campionaria:

$$\begin{aligned}\bar{\pi} &= \frac{1}{N} \sum_{i=1}^N \pi(x_i) \\ &= \frac{1}{N} \sum_{i=1}^N \pi^i \simeq 1, \quad N \text{ grande.}\end{aligned}$$

Questa approssimazione viene sfruttata nel calcolo della varianza campionaria, che diventa quindi:

$$\begin{aligned}\widehat{Var}_q \pi &= \frac{1}{N} \sum_{i=1}^N (\pi^i - 1)^2 = \frac{1}{N} \sum_{i=1}^N [(\pi^i)^2 - 2\pi^i + 1] \\ &= \frac{1}{N} \sum_{i=1}^N [(\pi^i)^2 - 1] = \frac{1}{N} \sum_{i=1}^N (\pi^i)^2 - 1\end{aligned}$$

Dalla precedente espressione risulta evidente che il peggioramento introdotto dal termine $1 + Var_q \pi$ derivata dall'aver campionato usando q anzichè p e che questo si riflette nella variabilità dei pesi π^i . È quindi importante identificare quindi un parametro che mi aiuti a decidere quando sia necessario ricampionare. Dall'espressione

$$Var_q(\hat{\mu}_q(f)) \simeq \frac{1}{N} Var_p(f)[1 + Var_q \pi]$$

è possibile estrarre il numero “effettivo” di particelle, espresso come

$$N_{EFF} = \frac{1 + Var_q(\pi)}{N},$$

riportando dunque la varianza alla forma più consueta di $Var_q(\hat{\mu}_q(f)) \simeq \frac{1}{N_{EFF}} Var_p(f)$.

Sostituendo la varianza vera con quella campionaria il parametro N_{EFF} si può stimare come

$$\hat{N}_{EFF} = \frac{N}{1 + \widehat{Var}_q(f)} = \frac{N}{\frac{1}{N} \sum_{i=1}^N (\pi^i)^2} = \frac{N^2}{\sum_{i=1}^N (\pi^i)^2} = \frac{N^2}{\sum_{i=1}^N \left(\pi^i \cdot \frac{\sum_{i=1}^N \pi^i}{\sum_{i=1}^N \pi^i} \right)^2}$$

Ricordando che $\frac{\pi^i}{\sum \pi^i}$ è l'espressione del peso normalizzato w_i , e che $\sum_{i=1}^N \pi^i \simeq N$, si ha:

$$\hat{N}_{EFF} = \frac{N^2}{\left(\sum_{i=1}^N \pi^i \right)^2 \cdot \sum_{i=1}^N w_i^2} \simeq \frac{1}{\sum_{i=1}^N w_i^2}$$

E' facile dedurre i valori entro cui N_{EFF} è definito: nel caso migliore, tutte le particelle hanno egual peso $\frac{1}{N}$, e si ha

$$w_i = \frac{1}{N} \Rightarrow N_{EFF} = \frac{1}{\sum_{i=1}^N \frac{1}{N^2}} = N.$$

Nel peggior caso, tutto il peso si concentra su un'unica k -esima particella, cioè si ha $w_k = 1$ e $w_j = 0, j \neq k$. Intuitivamente, $N_{EFF} = 1$. Fissando quindi una soglia minima di particelle effettive \bar{N} , si esegue il ricampionamento quando $N_{EFF} < \bar{N}$. In genere \bar{N} viene scelto con un valore tra 1/2 e 1/3.

27.3.1 Resampling Multinomiale

Una prima forma di resampling è chiamata multinomiale: concettualmente, consiste nell'eguire un normale ricampionamento da una densità multinomiale dove $\mathbb{P}[x = x_i] = w_i$, in modo che le particelle con peso maggiore vengano ricampionate più spesso, quelle con peso minore meno spesso o addirittura affatto (vengono perciò “uccise”).

27.3.2 Residual Resampling

Un'altra forma di ricampionamento valuta i pesi $0 \leq w_i \leq 1$: moltiplicandoli per il numero totale delle particelle N si ottengono dei valori $0 \leq Nw_i \leq N$ che indicano il numero di nuove particelle in cui “scomporre” la i -esima. Poichè di solito questo numero non è mai intero, si esegue un'operazione di *floor*:

$$N_i = \lfloor Nw_i \rfloor$$

Tale operazione induce la definizione del corrispettivo residuo $\beta_i < 1$:

$$\beta_i = Nw_i - N_i$$

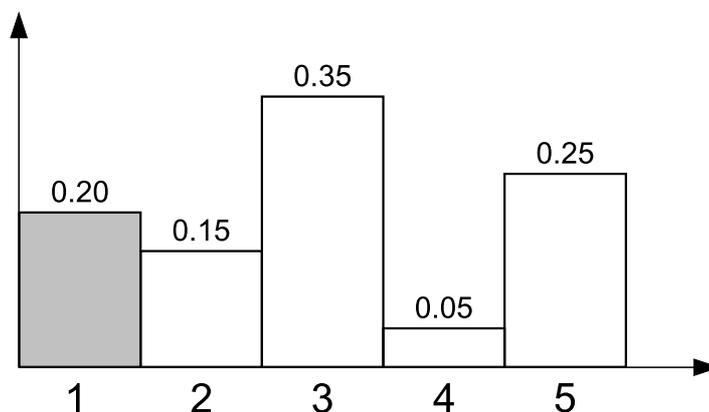
Non intendendo uccidere alcuna particella, è desiderabile mantenere costante la quantità N : tuttavia, essendo spesso $\sum N_i < N$, è necessario un secondo ricampionamento che sfrutti i residui (Residual Resampling). I vari β_i vengono rinormalizzati a

$$\tilde{\beta}_i = \frac{\beta_i}{\sum_{i=1}^N \beta_i}$$

ottenendo così una densità di probabilità multinomiale, da cui verranno campionati un numero di particelle $N - \sum_{i=1}^N N_i$ necessario per portare il totale a N .

Esempio

Si supponga di avere un campionamento a $N = 5$ particelle di cui effettuare il resampling, i cui pesi w_i sono distribuiti come indicato dal seguente grafico.



Eseguendo prima un campionamento sulla parte intera, si ha:

$$\begin{aligned} N \cdot w_i &= 5 \cdot w_{1..5} = 1, 0.75, 1.75, 0.25, 1.25 \\ N_i &= \lfloor N \cdot w_i \rfloor = 1, 0, 1, 0, 1 \\ \beta_i &= N \cdot w_i - N_i = 0, 0.75, 0.75, 0.25, 0.25 \end{aligned}$$

La struttura di N_i ci assicura che, dopo la prima valutazione sulle parti intere, nelle posizioni 1, 3 e 5 ci sia almeno una particella.

Normalizzando e campionando anche i residui, si ha:

$$\begin{aligned} \sum_{i=1}^N \beta_i &= 2 \\ \bar{\beta}_i &= \frac{\beta_i}{\sum_{i=1}^N \beta_i} = 0, 0.375, 0.375, 0.125, 0.125 \end{aligned}$$

Da questa densità di probabilità campiono \bar{N} altre particelle, con

$$\bar{N} = N - \sum_{i=1}^N N_i = 2$$

Una possibile realizzazione è riportata in figura: si noti come la particella numero 4, che in origine aveva un peso molto piccolo, sia stata eliminata. I due grafici non sono molto simili, ma ciò è dovuto al ridotto numero di particelle: con $N \rightarrow \infty$, ogni realizzazione del resampling tende verso l'originale.

