Sequential Importance Sampling

Sequential Importance Sampling (SIS)

- Let say we are interested to do IS for

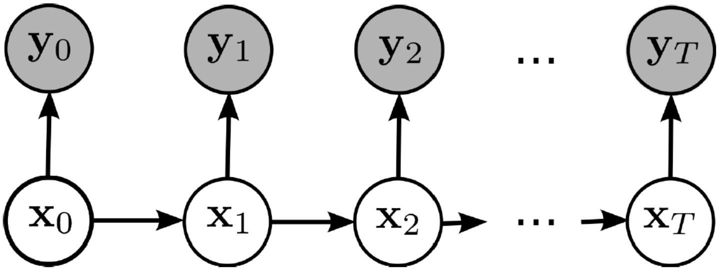
$$\pi(x_{0:T}) = \frac{\gamma(x_{0:T})}{Z},$$

- Here we are using notation: $x_{0:T} = (x_0, \ldots, x_T)$.
- Sequential Importance Sampling means applying IS sequentially/recursively
- Define a sequence of distributions

$$\{\pi_n(x_{0:n}) = \frac{\gamma_n(x_{0:n})}{\mathcal{Z}_n}\}_{n \leq T}, \quad \text{with } \pi = \pi_T.$$

and aim to reuse *samples/particles*.
- Note increasing size of state space (of π_n)

Example 1: stochastic filtering



Example 1: stochastic filtering

- Suppose $X_n \in \mathbb{R}^{d_x}$ being a hidden Markov process
  - $X_0 \sim \eta_\theta(\cdot)$, $X_n \sim f_\theta(\cdot|x_{n-1})$
- Let $Y_n \in \mathbb{R}^{d_y}$ form a sequence of conditionally i.i.d observations
  - $Y_n \sim g_\theta(\cdot|x_n)$
- What is the hidden signal $X_{0,}X_1, \ldots, X_T$?
- Can perform Bayesian inference using

$$\Pi_n(\cdot) = P\left[X_{0:n} \in \cdot |\, Y_{0:n}\right]$$

and use the marginal likelihood $Z_n = P_\theta(Y_1, \ldots, Y_n)$ for model selection.

Example 2: self avoiding random walk

- Have you played vintage snake game?

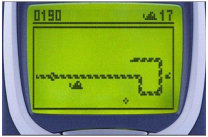


Figure: Good old days!

Example 2: self avoiding random walk

- Given X_0 let $X_n \in \mathbb{T}^2$ and consider a standard RW

$$p(X_n = x | X_{n-1} = y) = \begin{cases} \frac{1}{8}, & \text{if } x - y = 1 \\ 0 & \text{otherwise} \end{cases}$$

- Simulate from

$$\pi(X_1, \ldots, X_n) \propto 1_{x_n \neq x_{n-1} \neq \ldots \neq x_0}(X_1, \ldots, X_n) p(X_1, \ldots, X_n)$$

and compute $Z_n = P(x_n \neq x_{n-1} \neq \ldots \neq x_0)$

Sequential Importance Sampling (SIS)

1. Construct a product factorisation

$$\gamma_n(x_{0:n}) = \gamma_{n-1}(x_{0:n-1}) r_n(x_n | x_{0:n-1}),$$

   where r_n is simply a function of $x_{0:n}$

2. Construct proposal or instrumental density as

$$q_n(x_{0:n}) = q_{n-1}(x_{0:n-1}) q_n(x_n | x_{0:n-1}).$$

3. Then obtain a recursive expression for the IS weight

$$w_n(x_{0:n}) = w_{n-1}(x_{0:n-1}) \frac{r_n(x_n | x_{0:n-1})}{q_n(x_n | x_{0:n-1})}.$$

QUIZ: What is γ_n, r_n for Examples 1,2?

General SIS

At each $n \geq 0$ we have available $\{X^i_{0:n-1}, W^{\mathrm{i}}_{n-1}\}_{i=1}^N$.

1. Sampling
   - For $i = 1, \dots, N$,
     - sample particles as $X^i_n \sim q_n\left(\cdot | X^i_{0:n-1}\right)$,
     - Augment the path of the state as $X^i_{0:n} = [X^i_{0:n-1}, X^i_n]$.
2. Compute weight:
   - For $i = 1, \dots, N$, Compute weight

$$\widetilde{W}^i_n = W^i_{n-1} \frac{\gamma_n(X^i_{0:n})}{\gamma_{n-1}(X^i_{0:n-1}) q_n\left(X^i_n | X^i_{0:n-1}\right)} = W^i_{n-1} \frac{r_n(X^i_n | X^i_{0:n-1})}{q_n\left(X^i_n | X^i_{0:n-1}\right)},$$

   - Normalise weight $W^{\mathrm{i}}_n = \frac{\widetilde{W}^i_n}{\sum_{j=\mathbf{1}}^N \widetilde{W}^j_n}$.

SIS approximations

At time n, the approximations of π_n and Z_n after the sampling step are

$$\widehat{\pi}_n\left(dx_{0:n}\right) = \sum_{i=1}^{N} W_n^i \delta_{X_{0:n}^i}\left(dx_{0:n}\right), \tag{1}$$

$$\widetilde{Z}_n = \frac{1}{N}\sum_{i=1}^{N} w_n\left(X_{0:n}^i\right). \tag{2}$$

SIS approximations

- Let also
  - $\varphi : \mathcal{X}^n \to \mathbb{R}$ be a bounded measureable test function
  - the integral of interest be

$$I_n = \pi_n(\varphi) = \int \varphi(x_{0:n})\pi_n\left(x_{0:n}\right) dx_{0:n}$$

  - and its particle approximation

$$\begin{aligned}\hat{I}_n &= \hat{\pi}_n(\varphi) = \int \varphi(x_{0:n})\hat{\pi}_n\left(dx_{0:n}\right) \\ &= \sum_{i=1}^{N} W_n^i \varphi\left(X_{0:n}^i\right)\end{aligned}$$

Some Asymptotics with N

- Similar as self normalising IS:
  - basic difference is we are computing the weight recursively.
- Asymptotically consistent as $N \to \infty$. Asymptotic bias

$$\left(\widehat{I_n} - I_n\right) = -\frac{1}{N} \int_{\mathcal{X}^n} \frac{\left(\pi_n\left(x_{0:n}\right)\right)^2}{q_n(x_{0:n})} \left(\varphi(x_{0:n}) - I_n\right) dx_{0:n}$$

- Central Limit Theorem (CLT) holds:

$$\sqrt{N}\left(\widehat{I_n} - I_n\right) \Rightarrow \mathcal{N}\left(0, \sigma^2_{SIS}\right)$$

where

$$\sigma^2_{SIS} = \frac{1}{N}\left(\int_{\mathcal{X}^n} \frac{\left(\pi_n\left(x_{0:n}\right)\right)^2}{q_n(x_{0:n})} \left(\varphi(x_{0:n}) - I_n\right)^2 dx_{0:n}\right)$$

Variance of normalising constant estimator

- Normalising constant: use standard IS and weight final sample

$$\widetilde{Z}_n = \frac{1}{N} \sum_{i=1}^{N} w_n(X_{0:n}^i)$$

- Monte Carlo variance is as in standard IS:

$$\frac{\mathbb{V}ar\left[\widetilde{Z}_n\right]}{Z_n^2} = \frac{1}{N}\left(\int \frac{(\pi_n(x_{0:n}))^2}{q_n(x_{0:n})} dx_{0:n} - 1\right)$$

- So far it is not clear how we can exploit more sequential structure.

A sequential estimator for the normalising constant

- Lets write conditional distribution of X_n given the previous path:

$$\mathbb{P}\left(X_n \in A \middle| X_{0:n-1} = x_{0:n-1}\right) = \int_A \pi_n(x_n | x_{0:n-1}) dx_n$$

- Apply Bayes rule

$$\begin{aligned}\pi_n(x_n | x_{0:n-1}) &= \frac{\pi_n(x_{0:n})}{\pi_{n-1}(x_{0:n-1})} \\ &= \frac{\gamma_n(x_{0:n}) Z_{n-1}}{\gamma_{n-1}(x_{0:n-1}) Z_n} \\ &\propto r_n(x_n | x_{0:n-1})\end{aligned}$$

- Observe
  1. $\frac{Z_n}{Z_{n-1}}$ is normalising constant for r_n
  2. $\prod_{k=0}^{n} \frac{Z_k}{Z_{k-1}} = Z_n$, where we use for convenience $Z_{-1} = 1$.

A sequential estimator for the normalising constant

1. Construct estimator for each $\frac{Z_k}{Z_{k-1}}$ using standard IS, with proposal $q_k(x_k|x_{0:k-1})$, so

$$\widehat{\frac{Z_k}{Z_{k-1}}} = \sum_{i=1}^{N} W_{k-1}^i \omega_k(X_{0:k}^i),$$

where

$$\omega_k(x_{0:n}) = \frac{r_k(x_k|x_{0:k-1})}{q_k\left(x_k|x_{0:k-1}\right)}$$

2. Multiply estimators

$$\widehat{Z_n} = \prod_{k=0}^{n} \widehat{\frac{Z_k}{Z_{k-1}}} = \prod_{k=0}^{n} \sum_{i=1}^{N} W_{k-1}^i \omega_k\left(X_{0:k}^i\right)$$

Comparison of estimators

We have two estimators

1. Use standard IS and weight final sample

$$\widetilde{Z}_n = \frac{1}{N}\sum_{i=1}^{N} w_n(X^i_{0:n})$$

where recall $w_n(x_{0:n}) = \prod_{k=0}^{n} \omega_k\left(x_{0:k}\right)$

2. The sequential version

$$\widehat{Z_n} = \prod_{k=0}^{n} \frac{\widehat{Z_k}}{Z_{k-1}} = \prod_{k=0}^{n}\sum_{i=1}^{N} W^i_{k-1}\omega_k\left(X^i_{0:k}\right)$$

with $\omega_k\left(x^i_{0:k}\right) = \frac{r_k(x^i_k|x^i_{0:k-1})}{q_k\left(x^i_k|x^i_{0:k-1}\right)}$

QUIZ: which estimator is unbiased and why?

Choosing importance proposals

- Intuition for choosing q_n is same as in standard IS
  - can attempt minimise the rel. variance of the normalising constant $\hat{Z}_n$
  - or equivalently minimise the variance of the importance weights $\widetilde{W}_n^i$.
- This means q_n-s should be **very similar or close** to π_n-s
  - can use other approximations available, e.g. Laplace, Saddlepoint, etc.
- Can also monitor ESS as n progresses:

$$ESS_n^N = \frac{1}{\sum_{i=1}^N W_n^{i2}}$$

Discussion on SIS

- Approach can be useful for low/moderate n and low dimensional x_n-s
  - we will look into Example 1 above in more detail
- Eventually as n increases method will degenarate:
  - low weights will remain low for each particle
  - mass concentrates to few or one particle
  - weight variance eventually explodes

- Particle filtering and Sequential Monte Carlo addresses this by using **resampling** to stabilise the weights

Extensions

- Optimisation :
  - aim: find the mode of density $\pi(x)$
  - define a sequence of targets

$$\pi_n(x) \propto \pi(x)^{\gamma_n}$$

    where $\gamma_n > \gamma_{n-1}$.
  - As $\gamma_n \to \infty$ then π_n concentrates around the set of maximisers of $\widetilde{\pi}$

Extensions

- Rare Events:
  - compute probability of a small/rare tail $\pi(A)$.
  - define a sequence of targets

    $$\pi_n(x) \propto 1_{A_n} \pi(x)$$

    where $A = A_T \subset A_{T-1} \subset \ldots \subset A_0$.
  - normalising constant is $\pi(A)$

- Note: in the last two slides the sequence of densities is defined on *a static (non-increasing)* state space so slightly different than presentation so far
  - SMC samplers (see paper by Del Moral, Doucet and Jasra 06).