# Importance Sampling

Importance sampling is essentially a form of Monte Carlo simulation in which sampling uses a new set of distributions for the random variables—for example, moving the mean value point to the MPP so that many more sample points fall into the failure region.

From: e-Design, 2015

Related terms:

Monte Carlo Method, Failure Probability, Gaussians, Limit State Function, Particle Filter, Random Variable □, Sample Point, Sampling Technique

View all Topics

# Monte Carlo Methods

Sergios Theodoridis, in Machine Learning, 2015

## 14.5 Importance Sampling

Importance sampling (IS) is a method for estimating expectations. Let $f(\mathbf{x})$ be a known function of a random vector variable, $\mathbf{x}$, which is distributed according to $p(x)$. If one could draw samples from $p(x)$, then the expectation in Eq. (14.1) could be approximated as in Eq. (14.2). We will now assume that we are not able to draw samples from $p(x)$, and to go one step further, assume that $p(x)$ is only known up to a normalizing constant, that is,

Let $q(x)$ be another distribution from which samples can be drawn. Then we can write

(14.25)

where $xi, i = 1,2,\ldots,N$, are samples drawn from $q(x)$ and

(14.26)

The normalizing constant can readily be obtained as

(14.27)

Combining Eqs. (14.25) and (14.27), we finally obtain

(14.28)

or

where  are the normalized weights. It is not difficult to show (Problem 14.6) that the estimate

(14.29)

corresponds to an unbiased estimator of the normalizing constant. This is very interesting, because computing the normalizing constant is particularly useful information in a number of tasks. Recall that the evidence function, discussed in Chapter 12, is a normalizing constant; see also [26] for related comments.

In contrast, the estimator associated with Eq. (14.28), being the result of a ratio, is unbiased only asymptotically and it is a *biased* one for finite values of $N$ (Problem 14.6). Hence, if one would have the luxury of a very large number $N$ of samples, Eq. (14.28) would be a good enough estimate. However, in practice, $N$ cannot be made arbitrarily large and the resulting estimate may not be satisfactory.

If $q(x) \approx p(x)$, or at least $q(x)$ is a fairly good approximation of $(x)$, then Eq. (14.28) would approximate Eq. (14.2). However, for most practical cases, this is not easy to obtain, especially in high-dimensional spaces. If $q(x)$ is not a good match to $(x)$, it is very likely that there will be regions where $(x)$ is large while $q(x)$ is much smaller. The corresponding weights will have large values, relative to those from other regions, and they will be the dominant ones in the summation (Eq. (14.28)).

The effect of it is equivalent to reducing the number, $N$, of samples. Moreover, it is also possible that $q(x)$ takes very small values in some regions, which makes it very likely that samples from such regions are completely absent in Eq. (14.28); see Figure 14.7. In such cases, not only may the resulting estimate be wrong, but we will not be aware of it, and the variance of the weights, $w(xn)$ and $w(xn)f(xn)$, may exhibit low values.

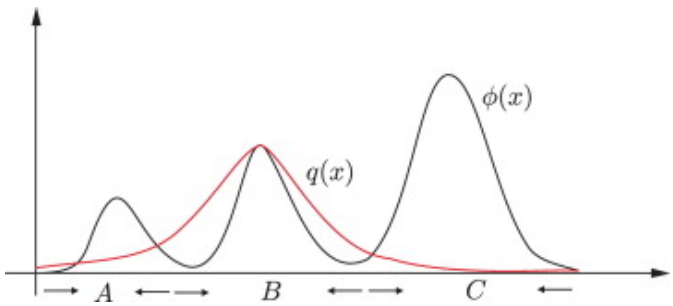


Figure 14.7. If $q(x)$ is not a good match of $(x)$, a number of undesired effects appear. Samples from region A will give rise to weights of much larger values compared to

these in region B. Due to the extremely low values of $q(x)$ in C, it is highly likely that given the finite size of the number of the samples, $N$, no samples will be drawn from this region, in spite of the fact that this is the most dominant region for $\square(x)(p(x))$.

*These phenomena are accentuated in high-dimensional spaces* (see, e.g., [26] and Problem 14.7).

To alleviate the previous shortcomings, a number of variants have been proposed to search for high-probability regions and make local approximations around the modes and use them in order to generate samples (see, e.g., [30] and the references therein).

> Read full chapter

# Estimation of collision probability between space debris and satellites

J. Morio, ... M. Balesdent, in Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems, 2016

## 11.4.2 Importance sampling

Only adaptive importance sampling techniques (i.e., CE and nonparametric adaptive importance sampling (NAIS)) can be applied to this case because the failure space is not known *a priori*.

### 11.4.2.1 Nonparametric adaptive importance sampling

Whatever the parameter tuning used for NAIS, this algorithm has not been able to converge in 100 iterations with a simulation budget of $10_5$ samples. The dimension of the input space is too large for a potentially efficient application of NAIS in this case.

### 11.4.2.2 Cross-entropy optimization

Cross entropy (CE) is applied to this debris-satellite case with the following parameter tuning:

- Number of samples per iteration $N$ : 1000.
- Value of the quantile parameter ($\square$) used to define the intermediary thresholds: 0.95.
-

Types of auxiliary pdf: Gaussian and Laplace. The center and bandwidth of the auxiliary pdf are optimized.

The probability estimates obtained with CE are indicated in Table 11.2. CE converges but estimates the target probability with a very significant bias. The strong non-connectivity of the failure space explains why CE is not adapted to this case and is not able to capture its different modes.

Table 11.2. Results obtained with CE using Gaussian and Laplace auxiliary pdfs of which the center and the bandwidth are optimized

| Aux. pdf | | RB () | RE () | Simulation budget | CE |
|---|---|---|---|---|---|
| Gaussian | 2.5 × 10–10 | – 100% | 203% | 4000 | n/a |
| Laplace | 6.2 × 10–9 | – 99% | 183% | 8000 | n/a |

> Read full chapter

# Selection of a simulation method

M. Elimelech, ... R.A. Williams, in Particle Deposition & Aggregation, 1995

## 7.3.2 Monte Carlo methods for various ensembles

Although the importance sampling established by Metropolis *et al.* (1953) is based on the canonical ensemble, it is a rather general sampling strategy and can be readily applied to other ensembles. On the other hand, the difficulty inherent in the Metropolis MC is providing a direct route to statistical properties, such as the entropy *S* and Helmholtz free energy A that require evaluation of the partition function *Q*, also raises the need to exploit other ensembles.

The extension of the method to the isothermal–isobaric (i.e. constant-NPT) ensemble was first made by Wood (1968, 1970), with the Markov chain of state having a limiting distribution proportional to exp [–(*PV*+ *E*)+*NIn V*]. Like the MMC, the isothermal-isobaric MC cannot calculate properties such as the Gibbs free energy since the method only samples importance regions in phase space.

To calculate the Gibbs free energy, the grand canonical MC method (GCMC) may be used. In this method the Markov chain of states is generated such that the limiting distribution is proportional to exp[–([*E*–*Nμ*)–In*N*!–3*N* In *B* + *N* In*V*], *B* being the thermal de Broglie wavelength defined as . A number of techniques for generating this chain have been proposed, e.g. Salsburg *et al.* (1959), Chesnut (1963), Norman

and Filinov (1969) and Rowley *et al.* (1975). Compared with the MMC, the grand canonical MC is much more complicated and difficult to handle and will not be presented here.

> Read full chapter

# Simulation techniques

M. Balesdent, ... R. Pastel, in Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems, 2016

## 5.3.1 Principle of importance sampling

The main idea of importance sampling (IS) (Engelund & Rackwitz, 1993; Kroese & Rubinstein, 2012; L'Ecuyer, Mandjes, & Tuffin, 2009) is to use an auxiliary distribution $h$ to generate more samples such that $\square(\mathbf{X}) > T$ than with the initial distribution $f$. A weight is then introduced in the probability estimate to take into account the change in the pdf which is used to generate the samples. IS takes advantage of the fact that

The IS probability estimate is then given by

(5.3)

where the rv $\mathbf{X}i, i = 1, \ldots, N$ are sampled iid from pdf $h$. The estimate of is unbiased. Its variance is given by the following equation:

(5.4)

with . The term $w(\mathbf{X})$ is often called the *likelihood function* in the importance sampling literature. The variance can be estimated using the classical Monte Carlo formula as follows:

(5.5)

The variance of strongly depends on the choice of $h$. If $h$ is well chosen, the IS estimate then has a much smaller variance than the Monte Carlo estimate and vice versa. The objective of IS is to decrease the estimate variance. We can thus define an optimal IS auxiliary density $h_{opt}$ as the density that minimizes the variance . Because variances are non-negative quantities, the optimal auxiliary density $h_{opt}$ is determined by canceling the variance in Equation (5.4). $h_{opt}$ is then defined with (Bucklew, 2004)

(5.6)

The optimal auxiliary density $h_{opt}$ depends unfortunately on the probability  that we estimate and consequently is unusable in practice. Nevertheless, $h_{opt}$ can be derived to determine an efficient sampling pdf. Indeed, a valuable auxiliary sampling pdf $h$ will be close to the density $h_{opt}$ with regard to a given criterion. An optimization of the auxiliary sampling pdf is then necessary.

This zero variance approach is difficult to implement directly for quantile estimation because the variance of quantile depends on the unknown quantity $f_2(q\square)$ (see Equation (5.2)). Nevertheless, using Equation (5.6), a simple and practically feasible alternative is to choose the following optimal IS auxiliary density to estimate the $\square$-quantile $q\square$:

This auxiliary density is often suggested in different articles (Cannamela et al., 2008; Egloff & Leippold, 2010) without further theoretical justification but based on the similarity with Equation (5.6).

Specific surveys on IS have been proposed as in Smith, Shafi, and Gao (1997) and Tokdar and Kass (2010). Some possible importance sampling heuristics are efficient only in very restrictive cases and thus are not analyzed for the sake of conciseness. In this book, we review in the next sections only the main IS algorithms that can be used for a general input–output function $\square(\cdot)$.

> Read full chapter

# TRANSPORT

Esam M.A. Hussein, in Radiation Mechanics, 2007

## Importance sampling

Some initial trials should be performed without importance sampling. If the variance of a quantity of interest is high, then these initial trials can provide insight into the problem and assist in assigning the spatial regions and energy regions within which importance sampling can be effective in reducing the variance. Importance sampling can be performed at the source, during the random walk and at scoring. The expected-value estimates of the next-event estimator can be viewed as a form of importance sampling. Source biasing should allow the production of more source particles, with suitably reduced weights, in the more important ranges of each variable: position, energy, and direction. For example, source particles directed to a region of interest should be sampled more often than those directed away from

it, provided of course that the weight of the source particle is accordingly adjusted. Variance reduction can also be achieved by the cut-off parameters, discussed below, where insignificant particles are not allowed to continue to score endlessly to the quantity of interest. However, the main methods of importance sampling in particle transport codes are splitting, Russian roulette, and exponential transformation.

Splitting should be applied in regions or energies that are expected to significantly contribute to the quantity of interest, but is unlikely to be reached; the opposite is true for Russian roulette. It is important, however, to control the amount of splitting to avoid the unnecessary creation of too many particles. Russian roulette takes a particle of weight $W$ and turns it into a particle of weight $W' > W$ with probability  and kills it with probability . In general, Russian roulette increases the history variance but decreases the time per history, while splitting achieves the opposite effect.

As shown in Section 4.1.7, the adjoint flux presents the importance of a particular point in the transport space to the response of a detector. Therefore, the adjoint flux is the best possible importance function for use in Monte Carlo sampling, as it can be shown that it leads to a zero variance [52]. However, knowing the adjoint flux entails solving the Boltzmann adjoint transport equation, Eq. (4.35), which is equal in effort to solving the direct transport problem to begin with. However, a crude solution of the adjoint transport equation can provide an approximation of the importance function, which in turn can reduce the variance of the original problem significantly.

The exponential transformation is a process that stretches or shrinks a particle's path-length between collisions. This is done by artificially reducing the macroscopic cross section in the preferred direction and increasing it in the opposite direction. A fictitious cross section, $\Sigma^*$, is related to the actual cross section, $\Sigma$, by $\Sigma^* = \Sigma t(1 - p\mu)$, where $\mu$ is the cosine of the angle between the preferred direction and the particle's direction and $p$ is a biasing parameter, $|p| < 1$; a constant or equal to . For and $\mu = 1$, $\Sigma^* = \Sigma s$ and the particle path is sampled from the distance to the next scattering, rather than from the mean-free-path  for all interactions. The weight is, consequently, adjusted by a factor of $\exp(-\Sigma a d)$, where $d$ is the distance of travel. Therefore, the exponential transformation works best in highly absorbing media and very poorly in highly scattering media. Exponential transformation is useful in deep penetration problems.

> Read full chapter

# Particle Filtering

Sergios Theodoridis, in Machine Learning, 2015

### 17.4.1 Degeneracy

Particle filtering is a special case of sequential importance sampling; hence, everything that has been said in Section 17.2 concerning the respective performance is also applied here.

A major problem is the degeneracy phenomenon. The variance of the importance weights increases in time, and after a few iterations only very few (or even only one) of the particles are assigned nonnegligible weights, and the discrete random measure degenerates quickly. There are two methods for reducing degeneracy: one is selecting a good proposal distribution and the other is resampling.

We know the optimal choice for the proposal distribution is

There are cases where this is available in analytic form. For example, this happens if the noise sources are Gaussian and the observation equation is linear (e.g., [11]). If analytic forms are not available and direct sampling is not possible, approximations of  are mobilized. Our familiar (from Chapter 12) Gaussian approximation via local linearization of  is a possibility [11]. The use of suboptimal filtering techniques such as the extended/unscented Kalman filter have also been advocated [37]. In general, it must be kept in mind that the choice of the proposal distribution plays a *crucial* role in the performance of particle filtering. Resampling is the other path that has been discussed in Section 17.2.2. The counterpart of Algorithm 17.1 can also be adopted for the case of particle filtering. However, we are going to give a slightly modified version of it.

> Read full chapter

# Estimation of launch vehicle stage fall-out zone

L. Brevault, ... J. Morio, in Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems, 2016

### 10.4.2 Importance sampling

Because the failure space is totally unknown, only adaptive importance sampling techniques (i.e., CE and nonparametric adaptive importance sampling (NAIS)) are applied to the launch vehicle fallout test case.

### 10.4.2.1 Nonparametric adaptive importance sampling

The following parameters are used as parameter tunings of NAIS:

- Number of samples per iteration $N$: 1000.
- Value of the quantile parameter ($\rho$) used to define the intermediary thresholds: 0.8.
- Type of kernels used as the auxiliary pdf: Gaussian and Laplace.

NAIS is applied to the launch vehicle fallout zone test case with success (see Table 10.2). Indeed, with a very restricted budget of 5000 samples, this algorithm allows to estimate the probability of interest with very low relative bias (0.5% with Gaussian kernel and 7% with Laplace kernel) and relative error (19% with Gaussian kernel and 27% with Laplace kernels, respectively), which results in a very high efficiency relative to CMC (5102 with Gaussian kernels and 2834 with Laplace kernels, respectively). The efficiency of the NAIS estimator can be explained by the fact that the dimension of the input space is not very important (the dimension is only 6) and is compatible with the use of kernel-based auxiliary pdf. Moreover, because the optimal auxiliary pdf is multimodal (see Figure 10.7), NAIS estimates the probability with accuracy whereas classical CE fails.

Table 10.2. Results obtained with NAIS for the launcher stage fallout zone test case

| **Kernel** | | **RB ()** | **RE ()** | **Simulation budget** | $\rho_{NAIS}$ |
|---|---|---|---|---|---|
| Gaussian | $9.95 \times 10^{-7}$ | 0.5% | 19% | 5000 | 5102 |
| Laplace | $1.06 \times 10^{-6}$ | 7% | 27% | 5000 | 2834 |

Figure 10.7. Histograms of NAIS-generated samples over the iterations (with Gaussian kernels). (a) $X_1$—iteration 1. (b) $X_1$—iteration 3. (c) $X_1$—iteration 5. (d) $X_2$—iteration 1. (e) $X_2$—iteration 3. (f) $X_2$—iteration 5. (g) $X_3$—iteration 1. (h) $X_3$—iteration 3. (i) $X_3$—iteration 5. (j) $X_4$—iteration 1. (k) $X_4$—iteration 3. (l) $X_4$—iteration 5. (m) $X_5$—iteration 1. (n) $X_5$—iteration 3. (o) $X_5$—iteration 5. (p) $X_6$—iteration 1. (q) $X_6$—iteration 3. (r) $X_6$—iteration 5.

### 10.4.2.2 Cross-entropy optimization

CE is applied to this test case with the following parameter tuning:

- Number of samples per iteration $N$ : 1000.
- Value of the quantile parameter ( ) used to define the intermediary thresholds : 0.95.
- Type of auxiliary pdf: Gaussian and Laplace. The center and bandwidth of the auxiliary pdf are optimized.

As illustrated in Figure 10.8, CE succeeds in catching only the most dominant mode of the optimal auxiliary pdf with Gaussian kernels. This is the result of the static parameterization of the auxiliary pdf with CE, which in this case is not compatible with multimodal distributions. This effect is less sensitive with Laplace distribution because its pdf tail is heavier than the tail of Gaussian pdf (Table 10.3). Thus, CE with Laplace pdf is able to successfully estimate a valuable probability, but its efficiency is still lower than NAIS efficiency.

Figure 10.8. Histograms of CE-generated samples at the last iteration (with Gaussian auxiliary pdf). (a) $X_1$. (b) $X_2$. (c) $X_3$. (d) $X_4$. (e) $X_5$. (f) $X_6$.

Table 10.3. Results obtained with CE using Gaussian and Laplace auxiliary pdf of which the center and the bandwidth are optimized

| Aux. pdf | | RB () | RE () | Simulation budget | CE |
|---|---|---|---|---|---|
| Gaussian | $6.0 \times 10^{-7}$ | – 39% | 54% | 8000 | 423 |
| Laplace | $9.1 \times 10^{-7}$ | – 8% | 59% | 8000 | 363 |

> Read full chapter

# Signal Processing for Hybridization

Mauricio A. Caceres Duran, ... Francesco Sottile, in Satellite and Terrestrial Radio Positioning Techniques, 2012

## 6.1.5.1 Sequential Importance Resampling

Sequential importance resampling (SIR) is an implementation of sequential importance sampling with a resampling step introduced to prevent the degeneration problem explained above. For the choice of the importance density function , it is possible to use the transition density , which can be defined by a motion equation. It has been shown that this importance function minimizes the variance of the importance weights conditioned upon the simulated trajectory and the observations up to time

$k$, $\mathbf{Z}_{1:k}$. Then, from (6.70), we see that the importance weights can be computed directly from the likelihood function . SIR, although requiring the ability to sample from  and to evaluate , which may not have an analytical solution in the general case, can be employed in the Gaussian state-space model with nonlinear transition equation. This is equivalent to taking Eqs (6.21) and (6.22) and considering the noise distributions  and  as mutually independent and Gaussian. For the resampling step, we make use of the effective sample size $N_{\text{eff}}$ introduced in Ref. [41]. This is defined as

(6.74)

which cannot be evaluated exactly but can be easily estimated by . The resampling step will be applied whenever  decreases below a given threshold $N_{\text{thres}}$, which is usually chosen to be $N_{\text{thres}} = 2N_s/3$. For the sake of clarity, Fig. 6.3 shows the general flow chart of the SIR procedure, whereas the method is outlined in Algorithm 6.3.

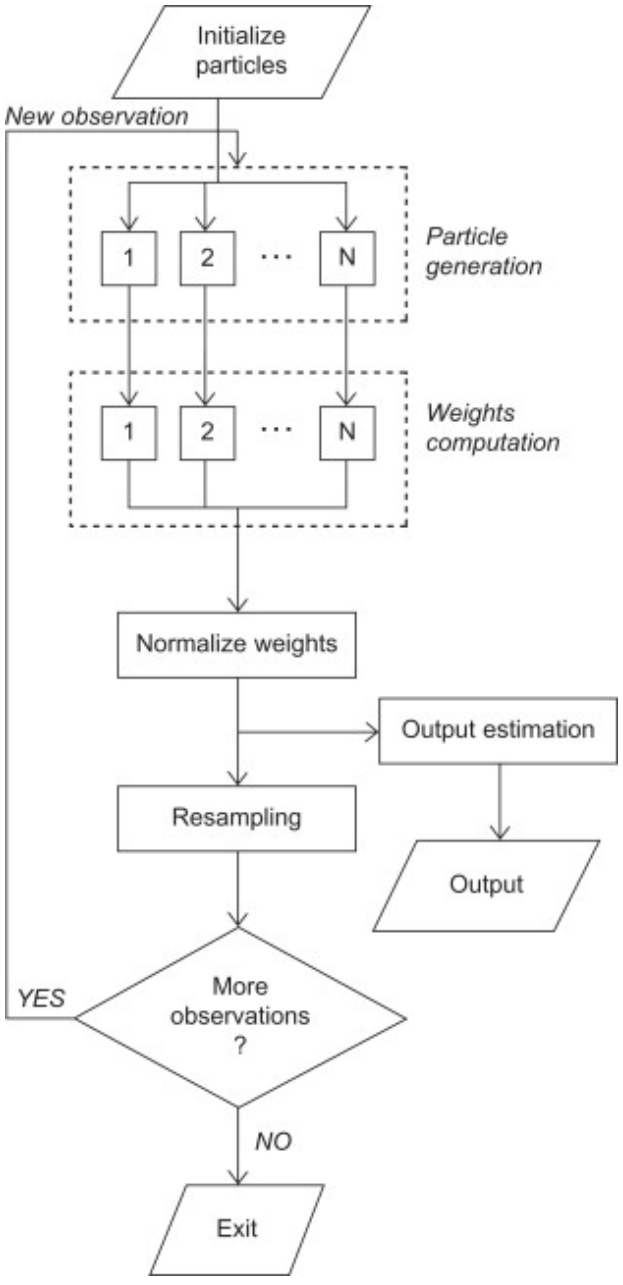

Figure 6.3. Flow chart of the SIR algorithm.

Algorithm 6.3

## Sequential Importance Resampling (SIR) Algorithm

**Require:** , observations and .

**Ensure:** approximates $\mathbf{x}k$ as $N_s \to \infty$.

1: Initialization: set $k = 0$, and generate $N_s$ samples from .

2: **For** $k = 1$ to **do**

3: **For** $i = 1$ to $N_s$ **do**

4: Sample Generate and simulate .

5: Evaluate importance weight .

6: **end for**

7: Normalize weights .

8: Compute

9: **if then**

10: Set .

11: **else**

12: Resample with replacement $N_s$ particles from the set according to the importance weights .

13: Reset weights to .

14: **end if**

15: Compute .

16: **end for**

> Read full chapter

# Special developments for time-variant systems

D. Jacquemart, ... M. Balesdent, in Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems, 2016

### 9.4.1 Principle

In the same way as described in Section 5.3 for static input–output function, the idea of importance sampling (IS) (Glynn & Iglehart, 1989; Juneja & Shahabuddin, 2001), also called *sequential importance sampling* when applied to Markov chains, is to replace the original probability distribution of the process by an auxiliary measure. Although the optimal change of density is not *a priori* Markovian, the IS auxiliary measure remains Markovian for convenience (Sandmann, 2005). An estimate of  is obtained by generating Markov chains  for $i = 1,\ldots,N$ with respect to new transition kernels  for $j = 1,\ldots,c$ and initial law . Then we compute the IS probability estimate with

with

and

The main difficulty of IS is to determine valuable sampling densities  that can reduce the CMC variance. For a general Markov process, there is, to our knowledge, no general method to determine an efficient IS auxiliary measure because of the very broad variety of instances that can be involved. Moreover, in practice, IS can lead to a degeneracy problem when only a few of the sample paths  will have significant weights $\Pi i$, and all the other sample paths will have very small weights. One common way to deal with degeneracy in particle filtering is resampling, but it is not adapted to rare event simulation. In the case of continuous state space when the value of $c$ in the relation $c\Delta t = S$ is not too large (e.g., say a few dozen), it is nevertheless possible (Johansen, 2006; Johansen, Del Moral, & Doucet, 2006) to determine potentially efficient sampling transition kernels using the sequential Monte Carlo sampler framework. Tuning these kernels requires the user to have a significant experience with the studied system. Two examples are provided in Johansen (2006). The IS algorithm is detailed in Algorithm 17.

**ALGORITHM 17 Importance sampling algorithm with Markov processes for probability estimation**

**Input:** The discrete time Markov process $(\mathbf{X}_j)_{0\le j\le c}$ with initial distribution $\pi_0(\mathbf{x}_0)$ and transition kernels $\pi_j(\mathbf{x}_j|\mathbf{x}_{j-1})$, the IS auxiliary initial transition distribution $\tilde{\pi}_0(\mathbf{x}_0)$ and kernels $\tilde{\pi}_j(\mathbf{x}_j|\mathbf{x}_{j-1})$, the number of samples $N$, the function $\varphi(\cdot)$, the threshold $T$, and the stopping time $c$.

**Output:** The probability estimate $\hat{\mathbb{P}}^{\mathrm{IS}}$.

1 Sample $(\mathbf{X}_j^{(1)})_{0\le j\le c}, (\mathbf{X}_j^{(2)})_{0\le j\le c}, \ldots, (\mathbf{X}_j^{(N)})_{0\le j\le c}$ with initial distribution $\tilde{\pi}_0(x_0)$ and transition kernels $\tilde{\pi}_j(\mathbf{x}_j|\mathbf{x}_{j-1})$.

2 Estimate $Z_i = \sup_{0\le j\le c} \varphi(\mathbf{X}_j^{(i)})$.

3 Estimate $\hat{\mathbb{P}}^{\mathrm{IS}} = \frac{\sum_{i=1}^N \omega_i \mathbf{1}_{Z_i>T}}{\sum_{i=1}^N \omega_i}$ with $\omega_i = \frac{\pi_0(\mathbf{x}_0^{(i)})\prod_{j=1}^c \pi_j(\mathbf{x}_j^{(i)}|\mathbf{x}_{j-1}^{(i)})}{\tilde{\pi}_0(\mathbf{x}_0^{(i)})\prod_{j=1}^c \tilde{\pi}_j(\mathbf{x}_j^{(i)}|\mathbf{x}_{j-1}^{(i)})}$.

4 **return** $\hat{\mathbb{P}}^{\mathrm{IS}}$.

> Read full chapter

# Distributed Kalman and Particle Filtering

Ali H. Sayed, ... Franz Hlawatsch, in Cooperative and Graph Signal Processing, 2018

## 6.3.3 Review of the Particle Filter

A PF performs an approximation of optimal sequential Bayesian state estimation that is based on Monte Carlo simulation and importance sampling [20, 47–49]. In a centralized scenario, the posterior pdf $f(x_n|z_{1:n})$ is represented in an approximative manner by $M$ randomly drawn samples or *particles* and corresponding weights , where $m = 1, 2, \ldots, M$. Specifically, at each time step $n$, propagation of the posterior pdf (i.e., $f(x_{n-1}|z_{1:n-1}) \to f(x_n|z_{1:n})$) is replaced by propagation of the particles and weights (i.e., ). As in the case of optimal sequential Bayesian estimation considered in Section 6.3.2, this propagation consists of a prediction step and an update or correction step.

In the prediction step at time $n$, for each preceding particle , a new particle is sampled from a suitably chosen proposal pdf (a.k.a. importance pdf) . In the simplest case, is chosen as , i.e., the state-transition pdf $f(x_n|x_{n-1})$ conditioned on ; the resulting PF algorithm is known as the sequential importance resampling (SIR) filter. However, more sophisticated "adapted" proposal pdfs that also involve the current measurement $z_n$ can result in improved estimation performance [49, 50] (see the section "Distributed proposal adaptation").

In the update step at time $n$, for each particle , a nonnormalized weight is calculated according to

(6.93)

For the SIR filter, this simplifies to

(6.94)

Subsequently, the weights are normalized according to . The set of particles and normalized weights  constitutes a Monte Carlo representation of the posterior pdf $f(x_n|z_{1:n})$. From , a corresponding approximation of the MMSE state estimate  in Eq. (6.92) can be computed as the weighted sample mean, i.e.,

(6.95)

Finally, if a suitable criterion is satisfied (as discussed in [20, 51]), the set  is resampled to avoid an effect known as particle degeneracy. In the simplest case [51], the resampled particles are obtained by sampling with replacement from the set , where is sampled with probability . This results in $M$ resampled particles . The weights are redefined as .

This recursive algorithm is initialized at time $n = 0$ by $M$ particles , $m = 1, 2, \ldots, M$, which are drawn from a suitable prior pdf $f(x_0)$. The initial weights are equal, i.e.,  for all $m$.

In a distributed implementation based on a decentralized network of agents, each agent runs a local instance of a PF, hereafter briefly referred to as "local PF." The local PF at agent $k$ observes only the local measurements $z_{n,k}$ directly; however, it receives from the neighbor agents indirect information about the measurements of the other agents in the network and also provides to the neighboring agents indirect information about its own measurements. The type of information that is exchanged between neighboring agents depends on the specific method used for distributed particle filtering. A DPF is based on the PF algorithm summarized above. However, it modifies that algorithm to account for the fact that each agent runs its own local PF, and it employs some networkwide distributed scheme (such as consensus, gossip, or diffusion) to disseminate and fuse local information provided by the agents. Several types of DPF methods have been proposed, including [13, 17, 52–64]. In what follows, we discuss two classes of DPF methods that rely on consensus and diffusion schemes for distributed information dissemination and fusion.

> Read full chapter