STATE ESTIMATION OF A SUPPLY CHAIN USING IMPROVED RESAMPLING RULES FOR PARTICLE FILTERING

Nurcin Celik
Young-Jun Son

Systems and Industrial Engineering
The University of Arizona
Tucson, AZ, 85721, USA

ABSTRACT

Resampling rules for importance sampling play a critical role in achieving good performance of the particle filters by preventing the sampling procedure from generating degenerated weights for particles, where a single particle abruptly possesses significant amount of normalized weights, and from wasting computational resources by replicating particles proportional to these weights. In this work, we propose two new resampling rules concerning minimized variance and minimized bias, respectively. Then, we revisit a half-with based resampling rule for benchmarking purposes. The proposed rules are derived theoretically and their performances are compared with that of the minimized variance and half width-based resampling rules existing in the literature using a supply chain simulation in terms of their resampling qualities (mean and variance of root mean square errors) and computational efficiencies, where we identify the circumstances that the proposed resampling rules become particularly useful.

1 INTRODUCTION

Successful supply chain management in globally competitive markets necessitates coherent planning and control across as well as within each strategic, tactical, and operational units (Dreyer et al., 2009; Rupp and Ristic, 2000). In the decision making process of coherent planning and control, effective monitoring and hence obtaining the latest information reflecting current supply chain status and capabilities becomes critically important, while not disrupting its ongoing operations. However, supply chains generate massive datasets at each measurement point due to their large-scale, dynamic and complex nature, where both the collection and the analysis become quite challenging and computationally intensive. Even though this situation holds true for the strategic and tactical levels, it becomes even more obvious at the operational level where the number of parameters as well as the frequency of update for each parameter grow significantly. In order to enable timely monitoring, simulation-based analysis, and control of these supply chains at the operational level in an economical and effective way, Celik and Son (2010) proposed a data-driven adaptive simulation scheme incorporating Bayesian inferencing by means of particle filters. Particle filtering (also known as Sequential Monte Carlo or sequential importance sampling with resampling) defines a class of simulation-based estimation techniques, which have been used to solve various types of sequential Bayesian inference problems that are encountered in wide range of areas such as econometrics (Casarin and Sartore, 2008; Flury and Shephard, 2008), signal and image processing (Brasnett et al., 2007; Xu and Li, 2007), robotics (Schulz and Burgard, 2001), and recently supply chain management (Celik and Son, 2010). In this study, the particle filtering algorithm intends to estimate the actual status of the supply chain members in terms of "mean time between failures" for each machine, which is further used in preventive maintenance and part routing scheduling problems in the supply chain control.

The idea behind the particle filtering resides in effective sampling from a sequence of probability distributions, and the algorithm is structurally composed of two major steps including importance sampling and resampling. Resampling, by definition is drawing repeated samples from subsets of available data based on a given criteria. It plays a critical role in the performance of the particle filter as it may resolve the potential issues of weight degeneration where after a few iterations, all but one particle will have negligible weight (Ristic et al., 2004) and waste of computational resources by replicating particles in proportion to their weights. In this work, we enhance the efficiency of the generic particle filtering algorithms by presenting improvements in their resampling techniques, namely variance-based resampling and bias-based resampling efficiency rules, respectively. Then, we revisit a half width-based resampling rule for benchmarking purposes. Here, all these three resampling rules arise from three distinct statistical standpoints. The proposed rules are first derived theoretically and their performances are benchmarked against the performances of the resampling rule developed by Kong et al. (1994) and half width-based resampling rule revisited in this work in terms of their resampling qualities and computational efficiencies using a simulation study.

2   PROPOSED RESAMPLING RULES FOR PARTICLE FILTERING ALGORITHMS

In order to provide a consistent notation throughout the article, we briefly describe the sequence of operations of generic particle filter in Figure 1, where a random vector $\mathbf{x}_k = [x_k^1, x_k^2, x_k^3, \dots, x_k^i]$, $i = 1, \dots, N_s$ is assumed to be with distribution $p(\mathbf{x})$, the proposal distribution $q(\mathbf{x}|\mathbf{z})$ is assumed to approximate the posterior $p(\mathbf{x}|\mathbf{z})$ and $q(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})p(\mathbf{x})$, k stands for the iteration number, N_s denotes the sample size, x_0^1 denotes each individual sample drawn at iteration 0, $q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k)$ is the proposal density, $\mathbf{z}_{1:k}$ are the measurements from times 1 to k, and δ is the Dirac peak. In this work, our focus is to develop efficient resampling rules to be used as part of Step 2 of the algorithm given in Figure 1.

Algorithm 1: Generic Particle Filter Algorithm

[Step 0: Initialization]

[Step 1: Importance Sampling]

   Set $k = k + 1$

   For $i = 1, \dots, N_s$, sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k)$ and set $\tilde{\mathbf{x}}_{0:k}^i = (\mathbf{x}_{0:k-1}^i, \mathbf{x}_k^i)$

   For $i = 1, \dots, N_s$, evaluate the importance weights via $\tilde{w}_k^i = \tilde{w}_{k-1}^i \left[\frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{0:k-1}^i, \mathbf{z}_{1:k})}\right]$

   Normalize the importance weights via $w_k^i = \frac{\tilde{w}_k^i}{\sum_{j=1}^{N_s} \tilde{w}_k^j}$

[Step 2: Resampling]

   Resample N_s particles randomly with replacement from the current particle set $\tilde{\mathbf{x}}_{0:k}^i; i = 1, \dots, N_s$ with probabilities proportional to w_k^i

   For each particle $\mathbf{x}_{0:k}^i; i = 1, \dots, N_s$, set $w_k^i = \frac{1}{N_s}$

   Set $k = k+1$

   Go to Step 1 as the next measurement arrives

[Step 3: Estimation]

   Estimate functions (i.e., $h(\mathbf{x})$) of the true posterior via $p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^i)$

Figure 1: Operations of generic particle filter (sequential importance sampling with resampling)

2.1 Importance Sampling

Following the same notation that is provided before in generic particle filtering algorithm (see Figure 1), we consider a random vector $\mathbf{x}_k = [x_k^1, x_k^2, x_k^3, ..., x_k^i]$, $i = 1, ..., N_s$ with new samples of size N_s for each iteration k. However, throughout this section, by hiding the index k, we use only the random vector $\mathbf{x} = [x^1, x^2, x^3, ..., x^i]$ to represent the sample of any iteration k to facilitate explanation. Here, random variable $\mathbf{x}$ is assumed to have the probability density function $p(\mathbf{x})$ (or probability mass function) and a function $h(\mathbf{x})$ (which is described below). In Step 3 of the particle filtering algorithm, our goal is to estimate the expected value of a function h of $\mathbf{x}$ (e.g., μ), where $\mu = E_p[h(\mathbf{x})] = \int h(\mathbf{x})p(\mathbf{x})d\mathbf{x}$, using the Monte Carlo approximation provided in (1). Here, $\hat{\mu}_1$ is an unbiased estimate of μ.

$$\hat{\mu}_1 = \frac{1}{N_s} \sum_{i=1}^{N_s} h(\mathbf{x}^i) \tag{1}$$

where $x^1, x^2, x^3, ..., x^{N_s}$ are independent samples drawn from $p(\mathbf{x})$.

Importance sampling (Kong, 1992; Ristic 2004) suggests estimating properties of a desired distribution (e.g., the expected value in our case), while having samples generated from a different alternative distribution rather than the original distribution of interest when the original distribution is not known or very hard to generate samples from. In this work, we represent that alternative distribution as $f(\mathbf{x})$. Using this alternative distribution, we can denote μ as given in (2), and its unbiased Monte Carlo estimate, $\hat{\mu}_2$, as given in (3).

$$\mu = \int \frac{h(\mathbf{x})p(\mathbf{x})}{f(\mathbf{x})} f(\mathbf{x})d\mathbf{x} = \int h(\mathbf{x})w(\mathbf{x})f(\mathbf{x})d\mathbf{x} = E_f[h(\mathbf{x})w(\mathbf{x})] \tag{2}$$

$$\hat{\mu}_2 = \frac{\sum_i^{N_s} h(\mathbf{x}^i)w(\mathbf{x}^i)}{N_s} \tag{3}$$

$$\bar{w} = \frac{1}{N_s} \sum_1^{N_s} w(\mathbf{x}^i) \quad \text{and} \quad w'(\mathbf{x}^i) = \frac{w(\mathbf{x}^i)}{\bar{w}} \tag{4}$$

where $w(\mathbf{x}) = p(\mathbf{x})/f(\mathbf{x})$, and $w(\mathbf{x}^i)$ and $w'(\mathbf{x}^i)$ are the importance sampling weight of sample i and its normalized version, respectively.

The reason behind the introduction of term $w'(\mathbf{x}^i)$ is the fact that while $E_f[w(\mathbf{x})] = 1$, the sample mean $\bar{w}$ might not necessarily be equal to $E_f[w(\mathbf{x})]$. Therefore, the term $w'(\mathbf{x}^i)$ is introduced by having a sample mean of one by its normalized definition, to resolve this bias issue, and force $\bar{w}$ to be an unbiased estimate of $E_f[w(\mathbf{x})]$. Our estimate of μ can be represented as in (5).

$$\hat{\mu}_3 = \frac{\sum_i^{N_s} h(\mathbf{x}^i)w'(\mathbf{x}^i)}{N_s} = \frac{\frac{1}{N_s}\sum_i^{N_s} h(\mathbf{x}^i)w(\mathbf{x}^i)}{\frac{1}{N_s}\sum_i^{N_s} w(\mathbf{x}^i)} = \frac{\bar{z}}{\bar{w}} \quad \text{where} \quad z = h(\mathbf{x})w(\mathbf{x}) \tag{5}$$

In this work, we have developed an improved stratified resampling rule based on the efficiency of $\hat{\mu}_3$ using the ratio of $\text{Var}_f[\hat{\mu}_3]/\text{Var}_p[\hat{\mu}_1]$ as a measure of relative efficiency between sampling from $p(\mathbf{x})$ and $f(\mathbf{x})$. In the prior literature (Kitagawa, 1996; Kong, 1992), in several cases, the estimation of the ratio $\text{Var}_f[\hat{\mu}_3]/\text{Var}_p[\hat{\mu}_1]$ is made via the Delta method which is a generalized combination form of central limit theorem and Slutsky's theorem. In some other particular cases, the approximation of the aforementioned ratio highly relies on the first order Taylor series expansion assuming that the higher order terms are negligible. In this study, we relax this quite strong assumption given the fact that the current sampling accuracy has a significant impact on determining the future sampling steps. In particular, we use the second

order Taylor series expansion of ratio of variables in order to obtain a closed form solution that can be practically useful and efficient when employed in the particle filtering algorithms.

2.2   Closed Form Representation of the Variance of Ratio of Variables using Theorem of Taylor Series Expansion

In this section, we provide the proof to obtain a closed form formula of the variance of ratio of variables using the theorem of Taylor series expansion. The widely known Taylor series expansion of any function g of two variables (i.e., $g(x,y)$) can be found in earlier derivations shown by Guterman and Nitecki (2006), and Randall (2006). In this work, however, we are interested in the application of the Taylor expansion in a variance function of ratio of variables where we make use of the second order expansion in particular while the derivation of closed form formula is provided for any order n.

Let us consider a function of two variables $g(x,y)$, whose total change of the function can be due to changes either in variable x or variable y. Therefore, if dx and dy are assumed to be constant, the direction of the mentioned change in the (x, y) plane can be shown by (6).

$$dg = dx\frac{\partial g}{\partial x} + dy\frac{\partial g}{\partial y} \tag{6}$$

Next, the chain of equations which derives the second total differential of the function $g(x,y)$ by taking the differential of the first total differential of $g(x,y)$, is shown in (7)-(10).

$$d^2g = dx\frac{\partial(dg)}{\partial x} + dy\frac{\partial(dg)}{\partial y}; \quad d^2g = dx\frac{\partial\left(dx\frac{\partial g}{\partial x}+dy\frac{\partial g}{\partial y}\right)}{\partial x} + dy\frac{\partial\left(dx\frac{\partial g}{\partial x}+dy\frac{\partial g}{\partial y}\right)}{\partial y} \tag{7}$$

$$d^2g = dx\frac{\partial\left(dx\frac{\partial g}{\partial x}+dy\frac{\partial g}{\partial y}\right)}{\partial x} + dy\frac{\partial\left(dx\frac{\partial g}{\partial x}+dy\frac{\partial g}{\partial y}\right)}{\partial y}; \quad d^2g = (dx)^2\frac{\partial^2 g}{\partial x^2} + 2dxdy\frac{\partial^2 g}{\partial x\partial y} + (dy)^2\frac{\partial^2 g}{\partial y^2}$$

$$d^2g = \left(dx\frac{\partial g}{\partial x} + dy\frac{\partial g}{\partial y}\right)^2 g(x,y)$$

Here, the generalization of the derivation given in (7) to the total differential of order n yields to (8).

$$d^n g = \left(dx\frac{\partial}{\partial x} + dy\frac{\partial}{\partial y}\right)^n g(x,y) \tag{8}$$

By using the closed form shown in (8), the Taylor series expansion of $g(x,y)$ for the neighborhood of the point (a, b) can be written as in (9).

$$g(x,y) = g(a,b) + \left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]g + \frac{1}{2!}\left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]^2 g + \quad ... \tag{9}$$

At this point, by using the closed form formula shown in (9) for the second order Taylor series expansion, our function of ratio of variables $g(x,y) = y/x$ can be approximated via the series of (10)-(13) around the neighborhood of the point (a,b).

$$g(x,y) = y/x \tag{10}$$

$$dg = dx\frac{\partial g}{\partial x} + dy\frac{\partial g}{\partial y} = [-\frac{y}{x^2}(x-a) + \frac{1}{x}(y-b)] \tag{11}$$

where $dx\frac{\partial g}{\partial x} = -\frac{y}{x^2}(x-a)$ and $dy\frac{\partial g}{\partial y} = \frac{1}{x}(y-b)$

$$d^2g = dx\frac{\partial(dg)}{\partial x} + dy\frac{\partial(dg)}{\partial y} = [(x-a)(\frac{-2ay}{x^3} + \frac{b}{x^2}) + (y-b)(\frac{a}{x^2})] \tag{12}$$

where $\frac{\partial(dg)}{\partial x} = (-\frac{2ay}{x^3} + \frac{b}{x^2})$ and $\frac{\partial(dg)}{\partial y} = (\frac{a}{x^2})$

$$g(x,y) = \frac{b}{a} + [-\frac{y}{x^2}(x-a) + \frac{1}{x}(y-b)] + \frac{1}{2!}[(x-a)\left(\frac{-2ay}{x^3} + \frac{b}{x^2}\right) + (y-b)\left(\frac{a}{x^2}\right)] \tag{13}$$

Next, using the closed form formula derived in (13), we first estimate the variance of $g(x,y) = y/x$ using the second order Taylor approximation as shown in (14)-(15) considering $a = E[x]$ and $b = E[y]$.

$$g(x,y) \approx \frac{b}{a} + [-\frac{b}{a^2}(x-a) + \frac{1}{a}(y-b)] + \frac{1}{2!}[(x-a)\left(\frac{-2ab}{a^3} + \frac{b}{a^2}\right) + (y-b)\left(\frac{a}{a^2}\right)] \tag{14}$$

$$= \frac{b}{a} + [-\frac{3b}{2a^2}(x-a) + \frac{3}{2a}(y-b)]$$

$$Var(g(x,y)) \approx Var\left(\frac{b}{a} + \left[-\frac{3b}{2a^2}(x-a) + \frac{3}{2a}(y-b)\right]\right) \tag{15}$$

$$= Var\left(\frac{b}{a}\right) + Var(-\frac{3b}{2a^2}(x-a) + \frac{3}{2a}(y-b))$$

$$= 0 + \frac{9b^2}{4a^4}Var(x) + \frac{9}{4a^2}Var(y) + 2\left(-\frac{3b}{2a^2}\right)\left(\frac{3}{2a}\right)Cov(x,y)$$

$$= \frac{9b^2}{4a^4}Var(x) + \frac{9}{4a^2}Var(y) + \frac{9b}{2a^3}Cov(x,y)$$

Utilizing the estimate of the variance of the function $g(x,y) = y/x$ shown in (15), we then derive the variance estimate of $g(x,y) = y/x$ as shown in (16) where N_s is the sample size.

$$Var(g(x,y)) \approx \frac{9b^2}{4a^4}\frac{Var(x)}{N_s} + \frac{9}{4a^2}\frac{Var(y)}{N_s} + -\frac{9b}{2a^3}\frac{Cov(xy)}{N_s} \tag{16}$$

$$= \frac{9}{4a^2N_s}(\frac{b^2}{a^2}Var(x) + Var(y) - \frac{2b}{a}Cov(x,y))$$

2.3   Resampling Rule 1: Variance-based Relative Sampling Efficiency Rule

In the resampling step, the goal is to eliminate the particles with low weights while distributing more particles in more probable regions. The decision of resampling can be made based on different reasoning (rules). In our first contribution of this study, we develop a stratified resampling rule, which is optimum in terms of minimized variance, by following the footsteps of a well-known stratified resampling rule, which was first proposed by Kong et al., (1994). In this section, by using the Taylor series approximation

for variance of the ratio estimates, we develop an improved stratified resampling rule which focuses on the ratio of $Var_f[\hat{\mu}_3]/Var_p[\hat{\mu}_1]$ as a measure of relative efficiency between sampling from $p(\mathbf{x})$ and $f(\mathbf{x})$. To this end, following the explanation given in (1)-(6), we obtain a closed formula relationship between $Var_f[\hat{\mu}_3]$ and $Var_p[\hat{\mu}_1]$ as derived in (17)-(22).

$$Var_f[\hat{\mu}_3] = Var_f\left(\frac{z}{w}\right) \approx \frac{9}{4a^2 N_S}\left(\mu^2 Var_f(w) + Var_f(z) - 2\mu Cov_f(w,z)\right) \tag{17}$$

$$= \frac{9}{4N_S}\left(\mu^2 Var_f(w) + Var_f(z) - 2\mu Cov_f(w,z)\right)$$

where $a = E_f[w] = 1$.

$$Cov_f(w,z) = E_f[wz] - E_f[w]E_f[z] = E_f[hw^2] - \mu = E_p[hw] - \mu \tag{18}$$

$$= Cov_p(w,h) + \mu E_p[w] - \mu$$

where $h = h(x)$.

$$Var_f(z) = E_f[w^2 h^2] - E_f^2[wh] = E_p[wh^2] - \mu^2 \tag{19}$$

$$E_p[wh^2] \approx E_p[w]E_p^2[h] + \frac{1}{2}Var_p(h)(2E_p[w]) + Cov_p(w,h)(2E_p[h]) \tag{20}$$

$$= \mu^2 E_p[w] + Var_p(h)E_p[w] + 2\mu Cov_p(w,h)$$

Then, in our estimate of the variance $Var_f[\hat{\mu}_3]$, the resultant equation becomes as in (21).

$$Var_f[\hat{\mu}_3] \approx \frac{9}{4N_S}\left(Var_p(h)E_p[w] + \mu^2\left(1 + Var_f(w) - E_p[w]\right)\right) \tag{21}$$

$$= \frac{9}{4}\left(Var_p(\hat{\mu}_3)\left(1 + Var_f(w)\right)\right)$$

where $\frac{Var_p(h)}{N_S} = Var_p(\hat{\mu}_1)]$, and $E_p[w] = E_f[w^2] = Var_f(w) + 1$.

Finally, the proposed estimate of the ratio of variances $Var_f[\hat{\mu}_3]$, and $Var_p[\hat{\mu}_1]$ reduces to the following.

$$\frac{Var_f[\hat{\mu}_3]}{Var_p[\hat{\mu}_1]} \approx \frac{9}{4}\left(1 + Var_f(w)\right) \tag{22}$$

Similar to the stratified resampling rule developed by Kong et al., (1994), the proposed improved stratified resampling rule does not involve $h(x)$. This makes the proposed estimation particularly useful as a measure of the relative efficiency of importance sampling. During importance sampling (e.g., within Sequential Monte Carlo algorithms), the effective sample size in the resampling stage is defined as below with $Var_f(w)$ being estimated by the sample variance of the standardized weights.

$$\frac{N_S}{\frac{Var_f[\hat{\mu}_3]}{Var_p[\hat{\mu}_1]}} \approx \frac{N_S}{\frac{9}{4}\left(1 + Var_f(w)\right)} = \frac{4N_S}{9\left(1 + Var_f(w)\right)} \tag{23}$$

2.4 Resampling Rule 2: Bias-based Relative Sampling Efficiency Rule

From the second order Taylor series expansion of $g(x, y)$ for a point in the neighborhood of the point (a, b), we obtain the approximation given in (24).

$$g(x,y) = g(a,b) + \left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]g + \frac{1}{2!}\left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]^2 g + \cdots \quad (24)$$

Given our function, $g(z, w) = z.w$, we determine the first and second order derivates as shown below.

$$\frac{\partial g(z,w)}{\partial z} = w, \quad \frac{\partial g(z,w)}{\partial w} = z, \quad \frac{\partial^2 g(z,w)}{\partial z^2} = 0, \quad \frac{\partial^2 g(z,w)}{\partial w^2} = 0, \quad \frac{\partial^2 g(z,w)}{\partial w \partial z} = 1 \quad (25)$$

Furthermore, the approximation of our function, $g(z, w) = z.w$, can be denoted as follows.

$$g(z,w) \approx g(a, b) + b(z - a) + a(w - b) + \frac{1}{2}[2(z - a)(w - b)] \quad (26)$$
$$= a.b + b(z - a) + a(w - b) + (z - a)(w - b)$$

Now, if we take the expectation of this function knowing that $a = E_g[z]$ and $b = E_g[w]$, we obtain the general open form for the covariance.

$$\begin{aligned}
E_g[zw] &= E_g[ab + b(z - a) + a(w - b) + (z - a)(w - b)] \quad (27)\\
&= ab + bE_g[z - a] + aE_g[w - b] + E_g[(z - a)(w - b)]\\
&= ab + bE_g[z] - ab + aE_g[w] - ab + E_g[(z - a)(w - b)]\\
&= E_g[z]E_g[w] + E_g[w]E_g[z] - E_g[z]E_g[w] + E_g[z]E_g[w] - E_g[z]E_g[w] + E_g[(z - a)(w - b)]\\
&= E_g[z]E_g[w] + E_g[(z - a)(w - b)]\\
&= E_g[z]E_g[w] + Cov_g(z, w)
\end{aligned}$$

Now, we switch to the notation used to determine our resampling rule as part of importance sampling. As mentioned earlier in (5), and repeated below is our variable of interest as the ratio of mean estimators.

$$\mu_3 = \frac{\sum_i^{N_s} h(x^i)w(x^i)}{N_s} = \frac{\frac{1}{N_s}\sum_1^{N_s} h(x^i)w(x^i)}{\frac{1}{N_s}\sum_1^{N_s} w(x^i)} = \frac{\bar{z}}{\bar{w}} \quad \text{where} \quad z = h(\mathbf{x})w(\mathbf{x}) \quad (28)$$

The bias of the ratio estimator above can be determined as follows using the proofs from Koop (1951) and Koop (1976), respectively. It should be noted here that we used the letter f for our particular function of interest rather than g, which was used earlier to represent the more general derivations.

$$Bias(z, w) = E_f\left[z\left(\frac{1}{w} - \frac{1}{E_f[w]}\right)\right] \quad (29)$$

$$E_f\left[z\left(\frac{1}{w} - \frac{1}{E_f[w]}\right)\right] = -\frac{1}{E_f[w]}Cov_f\left(\frac{z}{w}, w\right) \quad (30)$$

Now, since $E_f[w] = 1$, the equation becomes as given in (31).

$$E_f\left[z\left(\frac{1}{w} - \frac{1}{E_f[w]}\right)\right] = -Cov_f\left(\frac{z}{w}, w\right) \quad (31)$$

Provided $z = hw$, and $E_f[w] = 1$, our approximation of bias becomes as follows.

$$-Cov_f\left(\frac{z}{w}, w\right) = -E_f\left[\frac{z}{w}w\right] + E_f\left[\frac{z}{w}\right]E_f[w] = -E_f[hw] + E_f\left[\frac{hw}{w}\right] = -E_f[hw] + E_f[h]E_f[w] \quad (32)$$
$$= -Cov_f(h, w)$$

Therefore, the square of the bias, which is usually considered as part of the mean square error becomes as the following.

$$[Bias(z, w)]^2 = [Cov_f(h, w)]^2 \quad (33)$$

Finally, given a sample where w can be evaluated up to a constant, and $Cov_f(h, w)$ can be estimated by the sample covariance between the standardized weights and the distribution h, the bias based rule of effective sample size to be considered in the resampling stage of importance sampling becomes as the following.

$$\frac{N_s}{[Cov_f(h, w)]^2} \quad (34)$$

2.5 Resampling Rule 3: Half Width-based Sampling Efficiency Rule

Assuming that the initial sample size during the importance sampling is N_s', $(1 - \alpha)$ is the confidence level on μ where $0 < \alpha < 1$, σ_h is the desired half width, $t_{n-1,1-\alpha/2}$ is the value of a t-score in a two-tailed test with α-value, $n - 1$ degrees of freedom using students t-distribution, and s is our current sample standard deviation, the probability of obtained values being in between the upper and lower control limits is given in (35).

$$P\left(-t_{n-1,1-\alpha/2}\frac{s}{\sqrt{N_s}} \leq \mu - \bar{x} \leq t_{n-1,1-\alpha/2}\frac{s}{\sqrt{N_s}}\right) = 1 - \alpha \quad . (35)$$

The half width σ_h, is then defined as $\sigma_h = t_{n-1,1-\alpha/2}s/\sqrt{N_s}$ when we replace $z_{1-\alpha/2}$ with $t_{n-1,1-\alpha/2}$ based on the fact that a t-distribution with infinitely-many degrees of freedom is a normal distribution. Now if we solve the half width formula given above for N_s,

$$N_s = t_{n-1,1-\alpha/2}^2 \frac{s^2}{\sigma_h^2} \quad (36)$$

To this end, if $N_s < N_s'$, then no further action is required since we can estimate the mean μ with a $(1 - \alpha)$ confidence level using the current sample size. Otherwise if $N_s > N_s'$, then we will need to increase the sample size by $N_s - N_s'$ samples in order to reach the effective sample size. By the definition of half width, $N_s \to \infty$ with probability 1 as $\sigma_h \to 0$. Assuming that the measurement errors are independent and identically distributed, if $\sigma_h > 0$, then $P(N_s < \infty) = 1$ and $\lim_{\sigma_h \to 0} P(\mu + \sigma_h \leq \bar{x} \leq \mu + \sigma_h) \geq 1 - \alpha$. Hence, the confidence interval formed is asymptotically valid as σ_h reaches to zero.

3 EMPIRICAL RESULTS VIA SIMULATIONS

In the previous sections, three different resampling rules have been discussed from three different statistical perspectives, where two new resampling rules (variance-based and bias-based relative sampling efficiency rules) were proposed, and one was revisited (half-width based sampling efficiency rule). Theoretical derivations have been also provided for the proposed rules as well as the revisited one. In this section, these three algorithms are compared against each other as well as against a resampling rule pro-

posed by Kong et al., (1994) which we refer as "Kong rule" in terms of their resampling qualities (mean and variance of their root mean square errors) and computational effort (computation time). For this benchmarking, large scale supply chain simulations which have been built to represent a semiconductor manufacturing supply chain that is comprised of three echelons (wafer manufacturing fab, semiconductor manufacturing fab, and assembly and packaging fab) has been used. The status parameters sought in these simulations are "mean time between failures (MTBF)" for each machine in the considered shop floor (member) of this supply chain, where each shop floor includes at least hundred machines. These machines perform various operations such as diffusion, etch, photo, metals, and probe processes. By knowing the current system status (mean time between failures in this case), the final goal of the analysis is then to evaluate various preventive maintenance and part routing scheduling policies under various scenarios. All of the results shown in this section are obtained out of 30 replications of the same simulation with different seed values, where MTBF is the state parameter.

Figures 2 and 3 show the results obtained for the resampling qualities in terms of mean and variance root mean square errors (RMSE) of their estimates. The proposed bias-based relative efficiency sampling rule is shown to be the least effective among all of the compared algorithms since the mean RMSE values recorded for this resampling rule is greater than that of all the others. Half-width based sampling efficiency rule, while performing slightly better than the proposed bias-based relative sampling efficiency rule, is outperformed by both the proposed variance-based relative sampling efficiency rule as well as the Kong rule. Even though, mean RMSE decreases as the sample size increases, regardless of the used resampling rule, the proposed variance-based relative sampling efficiency rule produces the smallest RMSE on average. Regarding variance of the recorded RMSE values, Kong rule produces the largest variance in its results up to a sample size of 1000, and after this sample size, recorded variances of RMSE values decrease dramatically, and even becomes smaller than that of both half-width and bias-based resampling rules. The proposed variance-based resampling rule on the other hand, depicts the smallest amounts of variances in RMSE in most of the cases. However, when the sample size is 200, half-width based resampling rule generates slightly lesser amount of variance in RMSE than that of variance-based resampling rule; and when the sample size is 1400, proposed bias-based resampling rule produces slightly lesser variance in RMSE when compared to the results of the same.
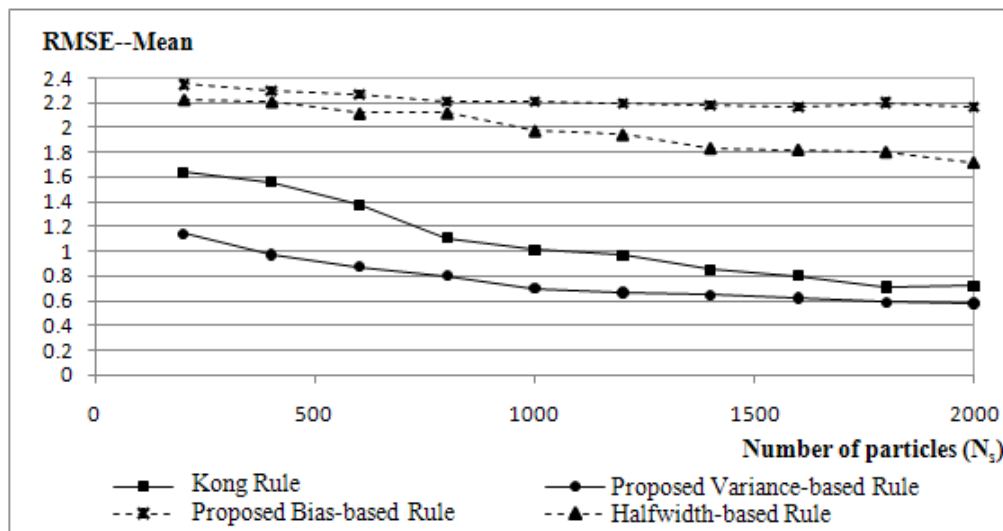

Figure 2: Mean of RMSE values for state estimates as a function of the number of particles for the supply chain simulation model
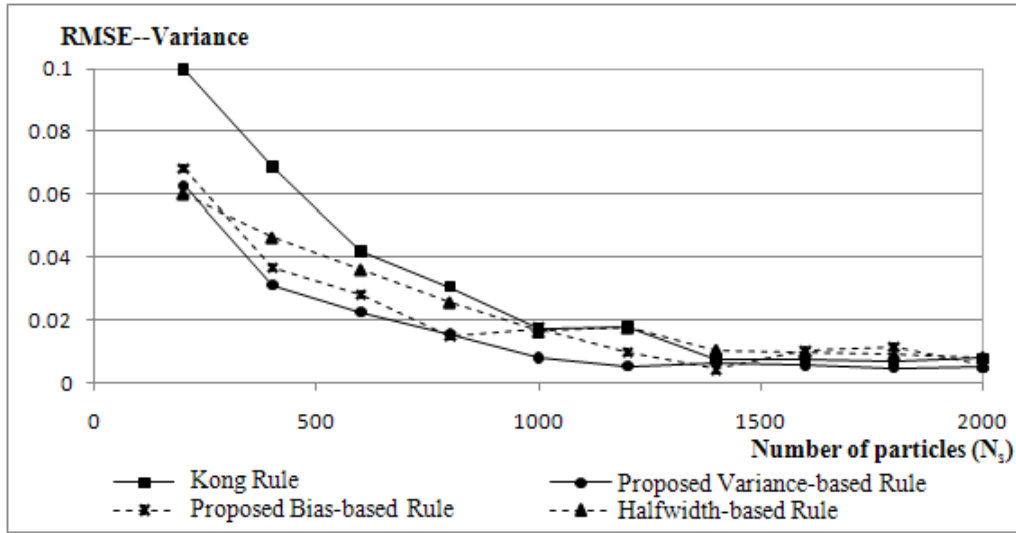
Figure 3: Variance of RMSE values for state estimates as a function of the number of particles for the supply chain simulation model

Based on the above results, the proposed bias-based relative sampling efficiency rule is not recommended for the cases where reasonable accuracy is necessary with a particle set size of less than 1000 due to computational resource unavailability. When the particle set size can be ranged between 1000 and 1500, the proposed bias based relative sampling efficiency rule can be selected over half-width based resampling rule as the former one results in smaller variances against very close mean values. On the other hand, the use of Kong rule might be risky as well when the sample set size is limited with 1000 particles, since the variance of the estimates are the highest among all the methods compared in this study. Overall, based on the estimation results plotted in this work, the proposed variance-based relative sampling rule is recommended to be employed in particle filtering algorithms, when the sample set size is upper bounded with 1000 particles, and the best possible accuracies are targeted with limited computational resources. When this limitation is relaxed up to 2000 particles, the difference between the Kong rule and the proposed variance-based relative sampling rule becomes minimum, hence both algorithms are equally recommended. It should be noted here that, the results obtained in this work is based on the simulation considered in this work, and the performance of the proposed variance-based relative sampling efficiency rule may involve even more significant computational savings while preserving accuracy in the system estimates when it is applied to the real-world supply chain control situation, that is greater in both scale and complexity. In addition, while the performance of the bias-based relative sampling rule is shown to be weak when compared to other methods, the performance of the methods combining these proposed rules may still outperform the individual rules. Future venues of this study will focus on the performance of these combined methods.
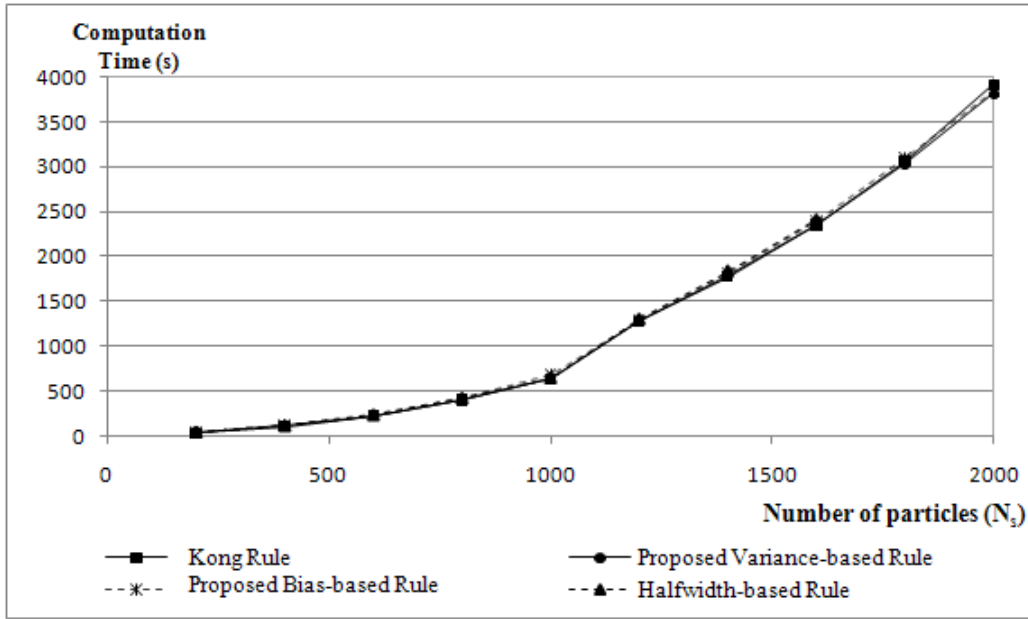
Figure 4: Computational effort required for different resampling rules as a function of the number of particles

As part of the experiment in this work, the computational effort required for the algorithms are investigated by measuring the time required to perform resampling of weight sequences. Figure 4 shows the measured times each resampling algorithm requires. As can be seen from the same figure, the computational effort necessary for the implementation of these algorithms are directly related to the particle set size and not related to the selection of the resampling method since all of the rules discussed in this study have computational complexities of $O(N_s^2)$, where N_s is the particle set size as mentioned earlier.

4   CONCLUSION AND FUTURE WORK

In this study, three resampling rules have been discussed in order to be used within particle filtering algorithms from three different perspectives. Among them, two rules, the variance and bias-based relative efficiency rules, are the improved version of the rules existing in the literature, whereas one of them (half-width sampling efficiency rule)is revisited one existing in the literature. The proposed rules have been first derived theoretically and then benchmarked against the revisited half-width based resampling rule as well as widely known Kong rule using a supply chain simulation in terms of their resampling qualities and computational efficiencies. Results obtained from our simulation experiment show that the proposed variance based resampling rule outperforms all the other three algorithms including the proposed bias-based resampling rule in terms of the recorded mean RMSE values. When estimating the states of the large scale, dynamic and complex systems such as supply chains and target tracking, the utilization of this rule is recommended compared to the other three rules discussed in this work as part of particle filtering algorithms. This selection will enable better accuracy results via less number of particle set sizes, which in turn prevents the simulations from wasting computational efforts. The proposed variance-based relative sampling rule may result in even greater savings in terms of computational resources when applied to the real-world complex scenarios, where both the state estimation and the measurement are more time-consuming and computationally challenging. Appropriate selection of the resampling rule is also critical in achieving desired accuracy levels in estimations when the number of samples those can be drawn is limited with the nature of the process. On the other hand, the performance of the proposed bias-based relative sampling rule was proven to be weak when compared to the other methods. However, the perfor-

mance of the methods combining two or more of the aforementioned rules (including the bias-based relative sampling rule) may still outperform the individual rules. Future venues of this research concerns itself with focusing on the performance of these blended resampling schemes. As all four of the discussed algorithms have complexities of $O(N_s^2)$, they are equivalent when selection of the resampling is based solely on the computational effort.

REFERENCES

Anscombe, F.J. 1952. Large sample theory of sequential estimation. *Proceedings of the Cambridge Philosophical Society*, 48, 600-607.

Arnaud, E. and Le Gland, F. 2009. SMC with adaptive resampling : large sample asymptotics, *Proceedings of the 2009 IEEE Workshop on Statistical Signal Processing*, Cardiff, August 31-September 3, 2009, 481-48.

Bayraksan, G. 2005. Monte Carlo Sampling-Based Methods in Stochastic Programming, Dissertation Presented to the Faculty of the Graduate School of The University of Texas at Austin.

Brasnett. P. L. Mihaylova, D. Bull, and N. Canagarajah. 2007. Sequential Monte Carlo Tracking by Fusing Multiple Cues in Video Sequences. Image and Vision Computing, Elsevier Science, 28 (1), 1217-1227.

Casarin, R. and D. Sartore. 2008. Matrix-State Particle Filter for Wishart Stochastic Volatility Processes, Discussion Paper (No: 0816), Dipartimento di Scienze Economiche, Università degli Studi di Brescia, Italy.

Celik, N., and Y-J. Son. 2010. Sequential Monte Carlo-based Fidelity Selection in Dynamic-Data-Driven Adaptive Multi-Scale Simulations, *submitted to International Journal of Production Research*.

Douc, R., and O. Cappe. 2005. Comparison of resampling schemes for particle filtering, *Proceedings of the 4th International Symposium on In Image and Signal Processing and Analysis*, 64-69.

Doucet, A., N. De Freitas, and N. Gordon. 2001. Eds. Sequential Monte Carlo Methods in Practice, Springer, New York, 2001.

Flury, T. and N. Shephard. 2008. Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models, Report (No: 413), Department of Economics, University of Oxford, UK.

Goodman, L. A., 1960, On the exact variance of products, *Journal of the American Statistical Association*, 55 (292), 708- 713.

Guterman, M. M., and Z.H. Nitecki. 2006. Differential Equations-A First Course, Delhi:Indo American Books.

Liu J.S., and R. Chen. 1998. Sequential Monte Carlo Methods for Dynamic Systems, Journal of the American Statistical Association, 93 (443), 1032-1044.

Kitagawa, G. 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics, 5, 1-25.

Kong, A. 1992. A note on importance sampling using standardized weights, Technical Report, 348.

Kong, A., J.S. Liu, and W.H. Wong. 1994. Sequential imputations and Bayesian missing data problems, Journal of American Statistical Association, 89, 278-288.

Koop, J. C., 1951. A note on the bias of the ratio estimate, *Bulletin of the* International *Statistical Institute*., 21(2), 141-146.

Koop, J. 1962. On upper limits to the difference in bias between two ratio estimates, *Metrika*, (5) 1, 145-149.

Meng, X.L. 1993. On the absolute bias ratio of ratio estimators, *Statistics and Probability Letters*, 18, 345-348.

Patterson, D. 2008. The Delta method and applications, Report.

Raj, D. 1964. A note on the variance of the ratio estimate, *Journal of the American Statistical Association*, 59 (307), 895- 898.

Randall, D. A. 2006. Taylor Series, Technical Report.

Ristic, B. M. Arulampalam, and A. Gordon. 2004. Beyond Kalman Filters: Particle Filters for Target Tracking. Artech House.

Schulz, D., and W. Burgard. 2001. Probabilistic state estimation of dynamic objects with a moving mobile robot. Robotics and Autonomous Systems, 34 (2-3).

Tin, M. 1965, Comparison of some ratio estimators, *Journal of the American Statistical Association*, 60(309), 294-307.

Xu, X. and B. Li. 2007. Adaptive Rao-Blackwellised Particle Filter and its Evaluation for Tracking in Surveillance. IEEE Transactions on Image Processing, 16 (3), 838-849.

AUTHOR BIOGRAPHIES

NURCIN CELIK is a Ph.D. Candidate in the Department of Systems and Industrial Engineering at the University of Arizona. She received her B.S. degree in Industrial Engineering from Bogazici University in Turkey and M.S. degree in Systems and Industrial Engineering at the University of Arizona. Her research interests lie in the areas of integrated modeling and decision making for large-scale, complex and dynamic systems such as supply chains and distributed electricity grids. She has received several awards including IIE Outstanding Graduate Research Award (2009), the Best Ph.D. scientific poster award in the Ph.D. Colloquium at IERC (2009), and the University of Arizona Outstanding Graduate Student Award (2008, 2010). She is a founding member and the treasurer of INFORMS Student Chapter at the University of Arizona. She is also a student member of IIE, INFORMS and SME. She can be reached at <nurcinkoyuncu@gmail.com>.

YOUNG-JUN SON is a Professor of Systems and Industrial Engineering and the Director of Advanced Integration of Manufacturing Systems and Technologies Center at University of Arizona. His research focuses on the coordination of a data-driven, multi-scale, networked-federated simulation and decision model needed for design and control in various applications, including extended manufacturing enterprise, homeland security, renewable energy network, and social network. He is an associate editor of the International Journal of Modeling and Simulation and the International Journal of Simulation and Process Modeling. He has received several research awards such as the SME 2004 Outstanding Young Manufacturing Engineer Award, IIE 2005 Outstanding Young Industrial Engineer Award, IERC Best Paper Awards (2005 in modeling and simulation; 2008 in homeland security; 2009 in modeling and simulation), and the Best Paper of the Year Award in 2007 from International Journal of Industrial Engineering. He can be reached by email at <son@sie.arizona.edu>.