
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Author(s): Juho Kokkala and Simo Sarkka
Title: On the (Non-)Convergence of Particle Filters with Gaussian Importance Distributions
Year: 2015
Version: Final Published version

Please cite the original version:

On the (Non-)Convergence of Particle Filters with Gaussian Importance Distributions. In Proceedings of the 17th IFAC Symposium on System Identification (SYSID), (IFAC PapersOnline, Volume 48, Issue 28), Pages 793–798. Beijing, China, October 2015. DOI: 10.1016/j.ifacol.2015.12.226

Rights: © 2015 International Federation of Automatic Control IFAC. Reprinted with permission.

This publication is included in the electronic version of the article dissertation:
Kokkala, Juho. Particle and Sigma-Point Methods for State and Parameter Estimation in Nonlinear Dynamic Systems. Aalto University publication series DOCTORAL DISSERTATIONS, 30/2016.

All material supplied via Aaltodoc is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

On the (non-)convergence of particle filters with Gaussian importance distributions^{*}

Juho Kokkala^{*} Simo Särkkä^{**}

^{*} *Aalto University, FI-02150 ESPOO, Finland.
(e-mail: juho.kokkala@aalto.fi)*

^{**} *Aalto University, FI-02150 ESPOO, Finland.
(e-mail: simo.sarkka@aalto.fi)*

Abstract: We consider convergence properties of particle filters with Gaussian importance distributions for certain time-varying Poisson regression models. We analyze both the classical bounded-importance-weight condition and a more recent moment condition. We show that Gaussian importance distributions based on Laplace approximations or non-linear Kalman filters lead to particle filters that are not guaranteed to converge. We also suggest avoiding the problem by a certain split-Gaussian modification that naturally arises from ensuring bounded weights. Although in this paper we concentrate on the time-varying Poisson regression model, we argue that our findings have implications in more general particle filtering problems.

© 2015, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Convergence of numerical methods; Monte Carlo method; stochastic systems; nonlinear systems.

1. INTRODUCTION

This paper is concerned with the use of Gaussian approximation based important distributions in sequential Monte Carlo based particle filtering. As is well known (see, e.g., Doucet et al., 2000; van der Merwe et al., 2000; Doucet et al., 2001; Cappé et al., 2005; Särkkä, 2013), the computational efficiency of sequential importance sampling based particle filtering is heavily affected by the choice of the importance distribution. For general particle filtering, the use of Gaussian approximation based importance distributions is often suggested. Typically, these approximations are formed by using a single step of a non-linear Kalman filter (as in, e.g., unscented particle filter, UPF; van der Merwe et al., 2000) or by using a local Laplace approximation.

However, it is also known, but not too well documented, that the use of Gaussian approximation based importance distributions may lead to failure of convergence of the particle filter. In this paper, we study this kind of convergence challenge arising from using Laplace approximation based importance distributions in time-varying Poisson regressions of the form

$$y_k \sim \text{Poisson}(\exp(\beta_k^T x_k)) \quad (1)$$

$$x_k \sim \mathcal{N}(A_k x_{k-1}, Q_k), \quad (2)$$

where $y_k \in \mathbb{N}$ are the observed counts (measurements), β_k is a vector of covariates and $x_k \in \mathbb{R}^n$ are the unknown regression coefficients that vary according to a Gauss–Markov model characterized by the transition matrices A_k and noise covariances Q_k . Such models may arise, for example, in mortgage default risk modeling (Aktekin et al., 2013). Our analysis can also be generalized to other models, but for concreteness, in this paper we concentrate on the above class of models.

^{*} This work was supported by the Academy of Finland projects 266940 and 273475.

The above model has the form of a probabilistic state-space model such that the unknown regression coefficients x_k are the states, and therefore particle filtering (e.g., Doucet et al., 2001; Cappé et al., 2005; Särkkä, 2013) can be used to recursively solve the regression problem. In particle filtering, a discrete approximation to the filtering distribution is represented by a finite set of weighted points, called particles. The particle filter algorithm proceeds through the measurements by propagating the particles according to an importance distribution and updating the weights according to the measurement and dynamic model densities. The bootstrap filter (Gordon et al., 1993) used the dynamic model as the importance distribution. However, the performance of the particle filter may be improved by adjusting the importance distribution based on the new measurement. Various importance distributions have been proposed (see, e.g., Doucet et al., 2000; van der Merwe et al., 2000; Doucet et al., 2001; Cappé et al., 2005; Särkkä, 2013).

The convergence of a particle filter means that when the number of particles goes to infinity, the approximation converges to the exact Bayesian filtering solution. The convergence has been proven only under certain conditions in the model and in the importance distribution. The convergence of particle filters has been considered in a large number of papers and books (see, e.g., Del Moral and Guionnet, 1999; Crisan and Doucet, 2002; Doucet et al., 2001; Del Moral, 2004; Hu et al., 2008, 2011; Del Moral, 2013; Mbalawata and Särkkä, 2014a,b; Mbalawata and Särkkä, 2015, and references therein).

The main contribution of this paper is to observe that using a Gaussian importance distribution (based on, e.g., the Laplace approximation) in our example class of models does not lead to a particle filter which would converge according to the present convergence theory for particle

filters. Although in this article we concentrate on models of this specific kind, we think that the same problem arises in a much larger class of models as well and can occur quite commonly in particle filtering. We also show that in this particular case we can circumvent this problem by using a split-Gaussian importance distribution instead. The key observation is that with the Poisson regression measurement model, the Gaussian approximation weights are unbounded only in the negative direction of the predicted observation, and therefore a conventional Gaussian approximation may be directly used in the positive direction. Ensuring bounded weights in the negative direction naturally leads to the use of the split-Gaussian distribution.

2. PARTICLE FILTERS AND THEIR CONVERGENCE

2.1 Particle filtering

Particle filtering is concerned with approximate Bayesian filtering in probabilistic state-space models of the form

$$\begin{aligned} x_k &\sim p(x_k | x_{k-1}), \\ y_k &\sim p(y_k | x_k). \end{aligned} \quad (3)$$

In particle filtering, the filtering distribution $p(x_k | y_{1:k})$ is approximated by a weighted set of particles $\{(w_k^{(i)}, x_k^{(i)}) : i = 1, \dots, N\}$ such that for any test function (typically bounded Borel test functions are assumed) ϕ we have

$$\int \phi(x_k) p(x_k | y_{1:k}) dx_k \approx \sum_{i=1}^N w_k^{(i)} \phi(x_k^{(i)}). \quad (4)$$

At each measurement y_k , the particle set is updated sequentially as follows:

- (1) For each $i = 1, \dots, N$, draw a new sample $x_k^{(i)}$ from the importance distribution

$$x_k^{(i)} \sim q(\cdot | x_{k-1}^{(i)}, y_{1:k}). \quad (5)$$

- (2) For each $i = 1, \dots, N$, update the weights via the formula:

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(y_k | x_k^{(i)}) p(x_k^{(i)} | x_{k-1}^{(i)})}{q(x_k^{(i)} | x_{k-1}^{(i)}, y_{1:k})}, \quad (6)$$

and normalize them to sum to unity.

- (3) If the particle set is too degenerate, do resampling.

The computational efficiency of the particle filter depends on the importance distribution q . The optimal importance distribution, which minimizes the variance of the weights is $q(x_k | x_{k-1}^{(i)}, y_{1:k}) = p(x_k | x_{k-1}^{(i)}, y_k)$ (Doucet et al., 2000) and therefore it is often preferred. However, sampling from it is often infeasible and the analytical form of the density might not be known. Therefore, one usually selects the importance distribution based on some approximation of the optimal importance distribution. Alternatives suggested in the literature typically use approximations based on the extended Kalman filter (Doucet et al., 2000) or the unscented Kalman filter (van der Merwe et al., 2000), more general non-linear Kalman filters (Särkkä, 2013), or the Laplace approximation (Cappé et al., 2005). In the so-called bootstrap filter (Gordon et al., 1993), the dynamic model $p(x_k | x_{k-1})$ is directly used as the importance

distribution. See also the survey by Šimandl and Straka (2007).

2.2 Convergence results for particle filters

A particle filter is said to converge if the approximation (4) becomes exact in the limit of infinite number of particles, $N \rightarrow \infty$. This "becoming exact" may be defined in multiple ways, but here we take it to mean that the particle filter converges in the L^p -sense. That is, for any bounded Borel test function ϕ

$$\mathbb{E} \left[\left| \int \phi(x_k) p(x_k | y_{1:k}) dx_k - \sum_{i=1}^N w_k^{(i)} \phi(x_k^{(i)}) \right|^p \right] \rightarrow 0, \quad (7)$$

when $N \rightarrow \infty$. Provided that this holds for $p = 4$, we also typically get a convergence of the empirical measure (see, e.g., Crisan and Doucet, 2000; Doucet et al., 2001; Bain and Crisan, 2009; Mbalawata and Särkkä, 2015, and references therein) via the Borel–Cantelli argument. The required conditions for the convergence are summarized in the following theorem.

Theorem 1. (L^p -convergence of particle filters I). Let us assume that (a) the dynamic and measurement model densities are bounded and (b) the unnormalized importance weights

$$\tilde{w}_k^{(i)} = \frac{p(y_k | x_k^{(i)}) p(x_k^{(i)} | x_{k-1}^{(i)})}{q(x_k^{(i)} | x_{k-1}^{(i)}, y_{1:k})} \quad (8)$$

are bounded from above over $(x_k^{(i)}, x_{k-1}^{(i)})$, then

$$\mathbb{E} \left[\left| \int \phi(x_k) p(x_k | y_{1:k}) dx_k - \sum_{i=1}^N w_k^{(i)} \phi(x_k^{(i)}) \right|^p \right] \leq c_k \frac{\|\phi\|^p}{N^{\frac{p}{2}}} \quad (9)$$

for some constant c_k and hence the particle filter converges in L^p -sense for all $p \geq 2$.

In fact, the above is for a particle filter version where we resample at every step, and the resampling method should satisfy certain conditions, but in any case the above assumptions are required to guarantee convergence.

The assumptions above are sufficient conditions for convergence, but they are not necessary. In fact, finding necessary conditions for convergence of particle filters is an open problem in general. However, different kind of sufficient conditions can also be used. For example, the result of Mbalawata and Särkkä (2015) (Mbalawata, 2015, see also) states the following.

Theorem 2. (L^p -convergence of particle filters II). Let us assume that (a) the dynamic and measurement model densities are bounded and (b) the unnormalized importance weights satisfy

$$\int \left(\frac{p(y_k | x_k) p(x_k | x_{k-1}^{(i)})}{q(x_k | x_{k-1}^{(i)}, y_{1:k})} \right)^p q(x_k | x_{k-1}^{(i)}, y_{1:k}) dx_k < \infty, \quad (10)$$

then

$$\mathbb{E} \left[\left| \int \phi(x_k) p(x_k | y_{1:k}) dx_k - \sum_{i=1}^N w_k^{(i)} \phi(x_k^{(i)}) \right|^p \right] \leq c'_k \frac{\|\phi\|^p}{N^{\frac{p}{2}}} \quad (11)$$

for some constant c'_k and hence the particle filter converges in L^p -sense for the given $p \geq 2$.

Although the above theorems (Theorems 1,2) might not provide necessary conditions for convergence, we may interpret the conditions of the theorems as practical requirements for convergence, since convergence is not guaranteed if they do not hold. Also, intuitively, if the unnormalized weights are not bounded or at least if their moments are not bounded either, then few particles with high weights may dominate the estimates.

3. IMPORTANCE DISTRIBUTIONS FOR THE POISSON REGRESSION MODEL

In this section, we first show that Gaussian importance distributions, including the Laplace approximation, for the Poisson regression lead to importance weights that are not bounded and hence the particle filter is not guaranteed to converge. We suggest a split-Gaussian modification to the Laplace-approximation based importance distribution and show that this leads to bounded weights and hence to a convergent particle filter.

3.1 Gaussian approximation based importance distribution

Let us now assume that we are using the following Gaussian importance distribution (at step k) for our particle filter:

$$q(x_k | x_{k-1}, y_{1:k}) = N(x_k | m_x, P_x), \tag{12}$$

where m_x and P_x are some functions of x_{k-1} and $y_{1:k}$. We could now write down the unnormalized weight expression and analyze its boundedness as such, but the task turns out to be much easier when we notice that it only depends on the scalar quantity $s_k = \beta_k^T x_k$.

Lemma 3. (Expression for Gaussian importance weights). The expression for the unnormalized importance weights for the case of Gaussian approximation based importance distribution has the form

$$\begin{aligned} \log \tilde{w}_k &= -e^{s_k} + \left[\frac{1}{2P_s} - \frac{1}{2\tilde{Q}} \right] s_k^2 + \left[y_k + \frac{s_k^-}{\tilde{Q}} - \frac{\hat{s}_k}{P_s} \right] s_k \\ &+ C(x_{k-1}) + D, \end{aligned} \tag{13}$$

where $\hat{s}_k = \beta_k^T m_x$, $P_s = \beta_k^T P_x \beta_k$, $s_k^- = \tilde{A}_k x_{k-1}$, $\tilde{A}_k = \beta_k^T A_k$, $\tilde{Q}_k = \beta_k^T Q_k \beta_k$, $C(x_{k-1}) = -\frac{(s_k^-)^2}{2\tilde{Q}} + \frac{1}{2} \log P_s + \frac{\hat{s}_k^2}{2P_s}$ does not depend on x_k and D is a constant with respect to both x_{k-1} and x_k .

Proof. Result follows by elementary manipulation and by collecting all s_k -independent terms to the constants.

Theorem 4. (Unboundedness of Gaussian weights). The Gaussian importance weights are unbounded if $P_s < \tilde{Q}$ or $P_s = \tilde{Q} \wedge \hat{s}_k > s_k^- + \tilde{Q} y_k$.

Proof. First notice that the boundedness of the weights is determined by the behaviour at $s_k \rightarrow \pm\infty$. We get the following:

- When $s_k \rightarrow \infty$ the term e^{s_k} always dominates and the logarithm goes to $-\infty$. This implies that the weight goes to zero and thus it remains bounded in this limit.

- When $s_k \rightarrow -\infty$ the term with s_k^2 dominates if $P_s \neq \tilde{Q}$ and we get:
 - If $P_s < \tilde{Q}$, then the logarithm goes to ∞ and thus the weight also grows without a bound.
 - If $P_s > \tilde{Q}$, then the logarithm goes to $-\infty$ and thus the weight remains bounded.
 - If $P_s = \tilde{Q}$, then the convergence at $s_k \rightarrow -\infty$ depends on whether the coefficient that multiplies s_k is nonnegative. By solving the inequality, we get convergence if $\hat{s}_k \leq s_k^- + \tilde{Q} y_k$.

The property $P_s \geq \tilde{Q}$ that is required for bounded weights is problematic in that intuitively using the information in the measurement y_k should decrease the variance compared to the dynamic model distribution. Indeed, we show that the variance always decreases for importance distributions based on non-linear Kalman filters (Theorem 5) and Laplace approximation to the optimal importance distribution (Theorem 9).

Theorem 5. (Unbounded Kalman weights). If the importance distribution is formed by using a one-step iteration of a non-linear Kalman filter, the unnormalized importance weights are not bounded.

Proof. With non-linear Kalman filters (Särkkä, 2013), the posterior covariance has the form

$$P_x = Q_k - Q_k H_k^T \Sigma_k^{-1} H_k Q_k, \tag{14}$$

where H_k and $\Sigma_k > 0$ depend on the method. The variance of the scalar quantity s_k is then

$$P_s = \tilde{Q} - \beta_k^T Q_k H_k^T \Sigma_k^{-1} H_k Q_k \beta_k < \tilde{Q}, \tag{15}$$

and thus the weights are unbounded by Theorem 4.

3.2 Laplace approximation

One way to form the Gaussian approximation is the Laplace approximation to the optimal importance distribution. Here we show that in our case, the Laplace approximation may also be formed using a marginal approach where the Laplace approximation is formed to the unidimensional $p(s_k | x_{k-1}, y_k)$, where $s_k = \beta^T x_k$. Then, we show that the Laplace approximation leads to unbounded weights.

Definition 6. (Laplace approximation). The Laplace approximation (see, e.g., Cappé et al., 2005) to the optimal importance distribution is a Gaussian distribution with the mean set to the mode of $\log p(x_k | y_k, x_{k-1}^{(i)})$ and the variance set to the inverse negative Hessian of $\log p(x_k | y_k, x_{k-1}^{(i)})$ evaluated at the mode.

Theorem 7. (Marginal Laplace approximation). Assume that the importance distribution is constructed as follows:

- $s_k \sim \hat{p}(s_k | x_{k-1}, y_k)$, the marginal Laplace approximation.
- $x_k \sim x_k | x_{k-1}, s_k$, the dynamic model distribution conditional on s_k .

Then, it is equal to the joint Laplace approximation to the optimal importance distribution.

Proof. If $\beta_k = 0$, both approximations are equal to the prior $p(x_k | x_{k-1})$. We may without loss of generality

assume $\beta_{1,k} \neq 0$. Then, x_k and $\tilde{x}_k := (s_k, x_{k,2:n})$ have a linear bijective correspondence, i.e., $\tilde{x}_k = Ax_k$ where A is of full rank. Thus, we may equivalently show that both constructions lead to the same distribution for \tilde{x}_k . In the following, we first argue that joint Laplace approximation of $p(x_k | x_{k-1}, y_k)$ transformed to a distribution over \tilde{x}_k is equivalent to forming the Laplace approximation directly to $p(\tilde{x}_k | x_{k-1}, y_k)$. Then, we show that the Laplace approximation to $p(\tilde{x}_k | x_{k-1}, y_k)$ equals the distribution of \tilde{x}_k under the marginal Laplace construction.

Applying the Jacobian transformation, we have $p(\tilde{x}_k | x_{k-1}, y_k) = |A|^{-1} p(x_k | x_{k-1}, y_k)$. Then, the logarithms of the densities are equal up to an additive constant. Thus, both densities have the same mode. Since $H_{Ax} \log p_{Ax}(Ax) = H_{Ax} p_x(A^{-1}Ax) = A^{-1}(H \log p(x))(A^{-1})^T$, both log-densities have the same Hessian at the mode and thus the same variance. Since by definition both are Gaussian, this concludes the first part.

Now, we need to show that the joint Laplace approximation to $p(\tilde{x}_k | x_{k-1}, y_k) = |A|^{-1} p(x_k | x_{k-1}, y_k)$ corresponds to the marginal Laplace approximation. Since both distributions are Gaussian, showing equality of modes and variances suffices. The measurement and state are conditionally independent given s_k , and thus we have

$$\begin{aligned} & \log p(\tilde{x}_k | x_{k-1}, y_k) \\ &= \log p(s_k | x_{k-1}, y_k) + \log p(x_{2:n,k} | s_k, x_{k-1}). \end{aligned} \quad (16)$$

Since the prior $p(\beta_k^T x_k, x_{2:n} | x_{k-1})$ is Gaussian, $p(x_{2:n,k} | s_k, x_{k-1})$ is a multivariate Gaussian with constant variance (independent of s_k) and therefore the second term attains its maximum whenever $x_{2:n,k} = E(x_{2:n,k} | s_k, x_{k-1})$. Thus, the expression is maximized whenever the first term is maximized with respect to s_k and $x_{2:n,k}$ is set to its conditional mean. This implies that both approximations have the same mode. To show that both approximations have the same variance, we show that their log-densities have the same Hessian evaluated at the mode. Recall that by definition the Hessian of the joint Laplace approximation equals the Hessian of the true log-density at the mode,

$$\begin{aligned} & H(\log p(s_k | x_{k-1}, y_k) + \log p(x_{2:n,k} | s_k, x_{k-1})) \\ &= H(\log p(s_k | x_{k-1}, y_k)) + H(\log p(x_{2:n,k} | s_k, x_{k-1})). \end{aligned} \quad (17)$$

By construction, the first term equals the Hessian of the log-density of the marginal Laplace approximation,

$$\begin{aligned} &= H(\log \hat{p}(s_k | x_{k-1}, y_k)) + H(\log p(x_{2:n,k} | s_k, x_{k-1})) \\ &= H(\log \hat{p}(s_k | x_{k-1}, y_k) + \log p(x_{2:n,k} | s_k, x_{k-1})), \end{aligned} \quad (18)$$

where the last expression is the Hessian of the log-density over \tilde{x}_k under the marginal Laplace construction. This concludes the proof.

Theorem 8. (Properties of the Laplace approximation).

The (marginal) Laplace approximation for s_k is $\hat{p}(s_k | y_k, x_{k-1}) = \mathcal{N}(s_k | \hat{s}_k, P_s)$, where

$$\tilde{Q} y_k + \tilde{s}_k^- = \hat{s}_k + \tilde{Q} \exp(\hat{s}_k) \quad (19)$$

$$P_s = \left(\exp(\hat{s}_k) + \tilde{Q}^{-1} \right)^{-1} \quad (20)$$

Proof. Follows from differentiating $\log p(s_k | y_k, x_{k-1})$ once and twice and manipulating the results.

Theorem 9. (Unbounded Laplace weights). The Laplace approximation to the optimal importance distribution leads to unbounded weights.

Proof. From Theorem 8, with the Laplace approximation, $P_s < Q$. Together with Theorem 4 this implies unbounded weights.

Theorem 10. (Unbounded Laplace moments). For any measurement y_k , covariates β_k and any index $p \geq 2$, there exists a previous state $x_{k-1}^{(i)}$ such that the moment condition of Theorem 2 is not satisfied.

Proof. Note that the moment integral may also be expressed as

$$\begin{aligned} & \int \left(\frac{p(y_k | s_k) p(s_k | x_{k-1})}{q(s_k | x_{k-1}, y_k)} \right)^p q(s_k | x_{k-1}) ds_k \\ &= \int \frac{p(y_k | s_k)^p p(s_k | x_{k-1})^p}{q(s_k | x_{k-1})^{p-1}} ds_k \end{aligned} \quad (21)$$

By substituting all distributions and ignoring constant factors independent of s_k , we obtain

$$\begin{aligned} & \propto \int \frac{e^{s_k y_k p} e^{-p e^{s_k}} e^{-\frac{p}{2Q} (s_k - \tilde{s}_k^-)^2}}{e^{-\frac{p-1}{2P_s} (s_k - \hat{s}_k)^2}} ds_k \\ &= \int e^{-p e^{s_k} + \left(\frac{p-1}{2P_s} - \frac{p}{2Q}\right) s_k^2 + \left(p y_k + \frac{p}{Q} \tilde{s}_k^- - \frac{p-1}{2P_s} \hat{s}_k\right) s_k + C} ds_k, \end{aligned} \quad (22)$$

where C is a constant independent of s_k . If $\frac{p-1}{2P_s} > \frac{p}{2Q}$, the limit of the exponent when $s_k \rightarrow -\infty$ is ∞ since the s_k^2 term dominates the limit. When the exponent approaches ∞ , the integrand approaches ∞ and thus the integral is ∞ , as well. From Theorem 8, we see that with small enough values of \hat{s}_k , P_s becomes arbitrarily small, and any value of \hat{s}_k is obtained with a suitable s_k^- . Thus, there always exists a x_{k-1} such that $\frac{p-1}{2P_s} > \frac{p}{2Q}$ and the moment is then infinite.

3.3 Split-Gaussian Importance Distribution

The split-Gaussian distribution was proposed by Geweke (1989) as an importance distribution in importance sampling. Guo and Wang (2006) used it within sequential quasi-Monte Carlo sampling. For properties of the distribution, see the article by Villani and Larsson (2006). In this subsection, we introduce the split-Gaussian distribution and define a suitable split-Gaussian importance distribution for the Poisson regression model such that the unnormalized importance weights become bounded.

Definition 11. (Split-Gaussian distribution). A scalar random variable s follows the one-dimensional split-Gaussian distribution $SN(\mu, P, q, r)$ if it is constructed as follows:

$$\eta \sim \mathcal{N}(0, P) \quad (23)$$

$$s := \begin{cases} \mu + r \eta, & \eta < 0 \\ \mu + q \eta, & \eta \geq 0. \end{cases} \quad (24)$$

Theorem 12. (Split-Gaussian density). The density of the split-Gaussian random variable $s \sim SN(\mu, P, q, r)$ is

$$p(s) = \begin{cases} \frac{1}{r\sqrt{2\pi P}} \exp\left(-\frac{(s-\mu)^2}{2r^2 P}\right), & s < \mu \\ \frac{1}{q\sqrt{2\pi P}} \exp\left(-\frac{(s-\mu)^2}{2q^2 P}\right), & s \geq \mu. \end{cases} \quad (25)$$

Definition 13. (Split-Gaussian importance distribution.) First, the marginal split-Gaussian importance distribution $q(s_k | x_{k-1}, y_k)$ is constructed as follows:

- (1) Form the marginal Laplace approximation $\hat{p}(s_k | x_{k-1}, y_k) = \mathcal{N}(s_k | \hat{s}_k, P_s)$.
- (2) Let the marginal split-Gaussian importance distribution be $q(s_k | x_{k-1}, y_k) = \text{SN}(\hat{s}_k, P_s, 1, \sqrt{\frac{\tilde{Q}}{P_s}})$.

Then, the split-Gaussian importance distribution $q(s_k | x_{k-1}, y_k)$ results from drawing $s_k \sim q(s_k | x_{k-1}, y_k)$ and x_k from $x_k \sim x_k | x_{k-1}, s_k$.

Theorem 14. (Boundedness of split-Gaussian weights). The particle filter importance weights are bounded over (s_k, x_{k-1}) when the split-Gaussian importance distribution is used.

Proof. We show boundedness for the cases $s_k < \hat{s}_k$ and $s_k \geq \hat{s}_k$ separately. First, assume $s_k < \hat{s}_k$. In this interval, the marginal split-Gaussian importance distribution equals $\mathcal{N}(s_k | \hat{s}_k, \tilde{Q})$, and thus we may apply Lemma 3 to obtain

$$\begin{aligned} \log \tilde{w}_k &= D \\ &- e^{s_k} + \left[y_k + \frac{s_k^- - \hat{s}_k}{\tilde{Q}} \right] s_k - \frac{(s_k^-)^2}{2\tilde{Q}} + \frac{1}{2} \log \tilde{Q} + \frac{(\hat{s}_k)^2}{2\tilde{Q}}. \end{aligned} \quad (26)$$

$\frac{1}{2} \log \tilde{Q}$ is constant and $-e^{s_k} < 0$, thus we obtain

$$\leq E + \left[y_k + \frac{s_k^- - \hat{s}_k}{\tilde{Q}} \right] s_k - \frac{(s_k^-)^2}{2\tilde{Q}} + \frac{(\hat{s}_k)^2}{2\tilde{Q}}. \quad (27)$$

The first property of Theorem 8 implies that the factor of s_k is positive, which with the assumption $s_k < \hat{s}_k$ gives the inequality

$$\leq E + \left[y_k + \frac{s_k^- - \hat{s}_k}{\tilde{Q}} \right] \hat{s}_k - \frac{(s_k^-)^2}{2\tilde{Q}} + \frac{(\hat{s}_k)^2}{2\tilde{Q}} \quad (28)$$

$$= E + y_k \hat{s}_k - \frac{(s_k^- - \hat{s}_k)^2}{2\tilde{Q}}. \quad (29)$$

Since $y_k \geq 0$, the last expression is bounded from above if \hat{s}_k is bounded from above, and thus the boundedness depends on the limit $\hat{s}_k \rightarrow \infty$. Substitute $s_k^- = \hat{s}_k + \tilde{Q} e^{\hat{s}_k} - \tilde{Q} y$ (Theorem 8):

$$\begin{aligned} \lim_{\hat{s}_k \rightarrow \infty} E + y_k \hat{s}_k - \frac{(s_k^- - \hat{s}_k)^2}{2\tilde{Q}} \\ = \lim_{\hat{s}_k \rightarrow \infty} E + y_k \hat{s}_k - \frac{1}{2} \tilde{Q} (e^{\hat{s}_k} - y_k)^2 = -\infty, \end{aligned} \quad (30)$$

which concludes the proof of the $s_k < \hat{s}_k$ case.

Now, assume $s_k \geq \hat{s}_k$. In this half-line, the marginal importance distribution equals $\mathcal{N}(s_k | \hat{s}_k, P_s)$. By Lemma 3, the unbounded weights are

$$\begin{aligned} \log w_k &= D - e^{s_k} + \left(\frac{1}{2P_s} - \frac{1}{2\tilde{Q}} \right) s_k^2 \\ &+ \left(y_k + \frac{s_k^- - \hat{s}_k}{\tilde{Q}} - \frac{\hat{s}_k}{P_s} \right) s_k - \frac{(s_k^-)^2}{2\tilde{Q}} + \frac{1}{2} \log P_s + \frac{\hat{s}_k^2}{2P_s}. \end{aligned} \quad (31)$$

Next, note that the $\frac{1}{2} \log P_s$ is bounded above by the constant $\frac{1}{2} \log \tilde{Q}$ and for the remaining P_s -terms substitute $P_s := \left(\exp(\hat{s}_k) + \tilde{Q}^{-1} \right)^{-1}$ (Theorem 8):

$$\begin{aligned} \leq E - e^{s_k} + \frac{1}{2} e^{\hat{s}_k} \hat{s}_k^2 + y_k s_k + \frac{s_k^- s_k}{\tilde{Q}} - \frac{\hat{s}_k s_k}{\tilde{Q}} \\ - \hat{s}_k s_k e^{\hat{s}_k} - \frac{(s_k^-)^2}{2\tilde{Q}} + \frac{1}{2} e^{\hat{s}_k} \hat{s}_k^2 + \frac{\hat{s}_k^2}{2\tilde{Q}}. \end{aligned} \quad (32)$$

Then substitute $s_k^- = \hat{s}_k + \tilde{Q} e^{\hat{s}_k} - \tilde{Q} y_k$ (Theorem 8), remove terms negating each other and perform some manipulations to obtain

$$\begin{aligned} = E + e^{\hat{s}_k} \left(-e^{s_k - \hat{s}_k} + \frac{1}{2} (s_k - \hat{s}_k)^2 + s_k - \hat{s}_k \right) \\ - \frac{1}{2} \tilde{Q} (e^{\hat{s}_k} - y)^2. \end{aligned} \quad (33)$$

The last term is obviously non-positive. We are assuming $s_k \geq \hat{s}_k$ and $e^x > x + \frac{1}{2}x^2$ when $x > 0$, wherefore the second term is also non-positive. Thus we have obtained an upper bound for the logarithm of the importance weight in the $s_k \geq \hat{s}_k$ case.

Corollary 15. (Moment condition for split-Gaussian importance weights.) The moment condition of Theorem 2 holds for the split-Gaussian importance weights with any $p \geq 2$.

Proof. Due to Theorem 14, the weights are bounded from above. Since the weights are by definition positive, finiteness of all moments directly follows.

4. NUMERIC EXPERIMENT

Consider the Poisson regression model (Eq.1) with two-dimensional state, initial state distribution $m_0 = (1 \ 1)$, $P_0 = I_2$ $Q = 0.1I_2$, observations $y_{0:2} = (0 \ 0 \ 14)$ and regressors $\beta = ((1 \ 1)^T \ (1 \ 1)^T \ (1 \ 4)^T)$. We run both filters with varying number of particle and investigated the fourth central moment of $\exp(-\beta_2^T x_2) | y_{0:2}$. For comparison, we also performed batch MCMC estimation using a Metropolis-Hastings sampler drawing independent proposals from the prior and computed the estimate based on last half of a single chain of 2×10^7 steps. The estimated test quantities as a function of number of particles as well as the corresponding MCMC estimate are shown in Figure 1. There is visibly more variation in the Laplace filter estimates, and compared to the MCMC estimate, the Laplace filter seems to typically underestimate the test quantity.

5. CONCLUSION AND DISCUSSION

In this paper, we have considered particle filter algorithms for a Poisson regression model with time-varying coefficients. The main contribution was to show that with a Gaussian approximation based importance distribution, the importance weights are unbounded, which is a problem for the convergence of the particle filter. To solve this problem, we proposed a split-Gaussian distribution where the tail of the importance distribution is fattened in the direction where the original Gaussian importance distribution would have unbounded weights. In the numerical example, we demonstrated that the split-Gaussian importance distribution may be perform better than the Gaussian importance distribution.

Our findings suggest that the weight conditions of the convergence results may have practical relevance when design-

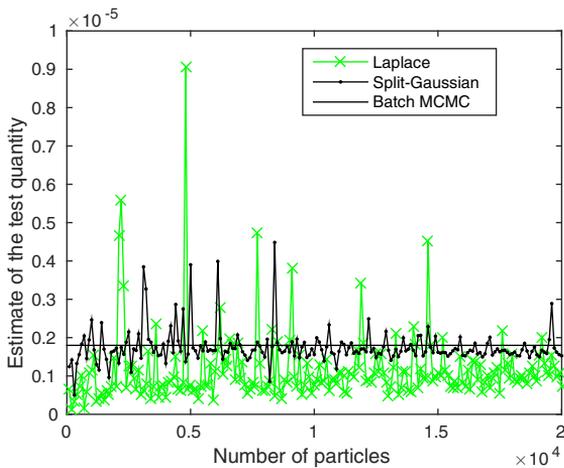


Fig. 1. Fourth central moment of $\exp(-\beta_2^T x_2)$ estimated with Laplace and split-Gaussian particle filters, varying the number of particles from 1000 to 20,000. The black horizontal line is the MCMC estimate of the same quantity. Note that the MCMC estimate is shown only for comparison and it is obtained from a single run, that is, it is not a function of the number of particles.

ing importance distributions for particle filters. Considering the behavior of the weights of importance distributions based on various approximations and their impact to the convergence of particle filters in general non-linear state-space models is potentially an interesting future research topic. We believe that the Poisson regression model and the corresponding split-Gaussian distribution studied in the present paper serve as a useful first example of this concept.

ACKNOWLEDGEMENTS

We thank Aki Vehtari and Enrique Lelo de Larrea Andrade for introducing the split-Gaussian distribution to us.

REFERENCES

- Aktekin, T., Soyer, R., and Xu, F. (2013). Assessment of mortgage default risk via Bayesian state space models. *The Annals of Applied Statistics*, 7(3), 1450–1473. doi: 10.1214/13-AOAS632.
- Bain, A. and Crisan, D. (2009). *Fundamentals of stochastic filtering*, volume 60. Springer.
- Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer.
- Crisan, D. and Doucet, A. (2000). Convergence of sequential Monte Carlo methods. Technical Report CUEDIF-INFENGrrR38, Signal Processing Group, Department of Engineering, University of Cambridge.
- Crisan, D. and Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions Signal Processing*, 50(3), 736–746.
- Del Moral, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer.

- Del Moral, P. (2013). *Mean field simulation for Monte Carlo integration*. Chapman & Hall/CRC.
- Del Moral, P. and Guionnet, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *The Annals of Applied Probability*, 9(2), 275–297.
- Doucet, A., Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer.
- Doucet, A., Godsill, S.J., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3), 197–208.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society*, 1317–1339.
- Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F (Radar and Signal Processing)*, 140(2), 107–113.
- Guo, D. and Wang, X. (2006). Quasi-Monte Carlo filtering in nonlinear dynamic systems. *IEEE Transactions on Signal Processing*, 54(6), 2087–2098.
- Hu, X., Schön, T.B., and Ljung, L. (2008). A basic convergence result for particle filtering. *IEEE Transactions on Signal Processing*, 56(4), 1337–1348.
- Hu, X., Schön, T.B., and Ljung, L. (2011). A general convergence result for particle filtering. *IEEE Transactions on Signal Processing*, 59(7), 3424–3429.
- Mbalawata, I.S. and Särkkä, S. (2014a). On the L^4 convergence of particle filters with general importance distributions. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 8048–8052.
- Mbalawata, I.S. and Särkkä, S. (2014b). Weight moment conditions for L^4 convergence of particle filters for unbounded test functions. In *Proceedings of European Signal Processing Conference (EUSIPCO)*, 1207–1211.
- Mbalawata, I.S. and Särkkä, S. (2015). Moment conditions for convergence of particle filters with unbounded importance weights. *Signal Processing*, to appear.
- Mbalawata, I.S. (2015). *Adaptive Markov chain Monte Carlo and Bayesian filtering for state space models*. Doctoral dissertation, Lappeenranta University of Technology.
- Särkkä, S. (2013). *Bayesian filtering and smoothing*. Cambridge University Press.
- van der Merwe, R., Doucet, A., de Freitas, N., and Wan, E. (2000). The unscented particle filter. In *Proceedings of NIPS*, 584–590.
- Villani, M. and Larsson, R. (2006). The multivariate split normal distribution and asymmetric principal components analysis. *Communications in Statistics: Theory & Methods*, 35(6), 1123 – 1140.
- Šimandl, M. and Straka, O. (2007). Sampling densities of particle filter: A survey and comparison. In *American Control Conference, 2007. ACC '07*, 4437–4442.