Estimation and Maintenance of Measurement Rates for Multiple Extended Target Tracking

Karl Granström
Division of Automatic Control
Department of Electrical Engineering
Linköping University, SE-581 83, Linköping, Sweden
Email: karl@isy.liu.se

Umut Orguner
Department of Electrical and Electronics Engineering
Middle East Technical University
06531, Ankara, Turkey
Email: umut@eee.metu.edu.tr

Abstract—In Gilholm *et al.*'s extended target model, the number of measurements generated by a target is Poisson distributed with measurement rate γ. Practical use of this extended target model in multiple extended target tracking algorithms requires a good estimate of γ. In this paper, we first give a Bayesian recursion for estimating γ using the well-known conjugate prior Gamma-distribution. In multiple extended target tracking, consideration of different measurement set associations to a single target makes Gamma-mixtures arise naturally. This causes a need for mixture reduction, and we consider the reduction of Gamma-mixtures by means of merging. Analytical minimization of the Kullback-Leibler divergence is used to compute the single Gamma distribution that best approximates a weighted sum of Gamma distributions. Results from simulations show the merits of the presented multiple target measurement-rate estimator. The Bayesian recursion and presented reduction algorithm have important implications for multiple extended target tracking, e.g. using the implementations of the extended target PHD filter.

Index Terms—Poisson distribution, Poisson rate, Gamma distribution, conjugate prior, Bayesian estimator, extended target, PHD filter, mixture reduction, Kullback-Leibler divergence.

I. INTRODUCTION

In target tracking, the assumption is often made that a target gives rise to at most one measurement per time step, see e.g. [1]. However, in extended target tracking this assumption is relaxed, and the extended targets are modeled as potentially giving rise to more than one measurement per time step. In an extended target tracking scenario it is therefore of interest to model the number of measurements that each target gives rise to.

One such model is given by Gilholm *et al.* [2], where the measurements are modeled as an inhomogeneous Poisson point process. At each time step, a Poisson distributed random number of measurements are generated, distributed around the target. Mahler has given an extended target Probability Hypothesis Density (PHD) filter under this model [3], and a Gaussian Mixture implementation of this filter, called the Extended Target Gaussian Mixture PHD (ET-GM-PHD) filter, has been presented [4], [5]. A Gaussian inverse Wishart implementation of [3], called the Gaussian inverse Wishart PHD (GIW-PHD) filter is presented in [6]. The measurement model [2] can be understood to imply that the extended target is sufficiently far away from the sensor for the measurements to resemble a cluster of points, rather than a geometrically

structured ensemble, see e.g. [4], [5] for simulation examples. However, the ET-GM-PHD filter and GIW-PHD filter have also been applied successfully to data from laser range sensors, which give (highly) structured measurements, see [5]–[7].

In the extended target PHD filter the Poisson rate γ is modeled as a function of the extended target state $\mathbf{x}$, i.e. $\gamma \triangleq \gamma(\mathbf{x})$. In the ET-GM-PHD filter γ is approximated as a function of the extended targets' state estimates [5]. It has been noted that having a good estimate $\hat{\gamma}$ of the true parameter is important when multiple targets are spatially close [5]. More specifically, under the assumption that the true rate is constant and equal for all targets, the true parameter must lie in the interval

$$\hat{\gamma} - \sqrt{\hat{\gamma}} \leq \gamma \leq \hat{\gamma} + \sqrt{\hat{\gamma}} \tag{1}$$

for the estimated cardinality to be correct [5]. However, in the most general case the rates are neither constant over time, nor equal for all extended targets. It might also be the case that the true function $\gamma(\cdot)$ is difficult to model, or even time-varying. All of these issues raise the need for a method to estimate individual Poisson rates for multiple extended targets.

In this paper we consider multiple extended targets under the measurement model [2]. The set of extended targets at time t_k is denoted

$$\mathbf{X}_k = \left\{ \mathbf{x}_k^{(i)} \right\}_{i=1}^{N_{x,k}}. \tag{2}$$

At each time step, the number of measurements generated by the i:th target is a Poisson distributed random variable with rate $\gamma_k^{(i)}$. The measurement set at time t_k, denoted

$$\mathbf{Z}_k = \left\{ \mathbf{z}_k^{(j)} \right\}_{j=1}^{N_{z,k}}, \tag{3}$$

is the union of all target measurements and the clutter measurements. The number of clutter measurements generated at each time step is assumed to be Poisson distributed with rate λ_k. Let $\mathbf{Z}^k$ denote all measurement sets up to, and including, time t_k. We assume the existence of an underlying multiple extended target tracker that estimates the target states $\mathbf{x}_k^{(i)}$, e.g. [4], [5]. The set of measurements that are used to update the state of the i:th target at time t_k is denoted $\mathbf{Z}_k^{(i)}$.

The first objective of this work is to estimate the set of measurement rates $\gamma_k^{(i)}$, given sequences of measurement sets

$$\mathbf{Z}^{k,(i)} \triangleq \left\{ \mathbf{Z}_0^{(i)}, \ldots, \mathbf{Z}_k^{(i)} \right\}, \ i = 1, \ldots, n_{x,k}. \tag{4}$$

To this end, in Section II we give a recursive Bayesian estimator for $\gamma_k^{(i)}$, with exponential forgetting for the prediction step. We also show how the predicted likelihood is affected when the measurement rates $\gamma_k^{(i)}$ are estimated in addition to estimating the target states $\mathbf{x}_k^{(i)}$.

In the multiple target case under clutter and missed detections, there might be multiple alternative measurement sets (corresponding to different association hypotheses)

$$\mathbf{Z}_k^{(i_1)}, \mathbf{Z}_k^{(i_2)}, \ldots, \mathbf{Z}_k^{(i_{N_i})} \tag{5}$$

that are used to update the i:th target state at time t_k. In this case, the state densities of the targets are represented by mixture densities. As time progresses, the number of mixture components grow. To obtain computationally tractable algorithms, hypothesis reduction must be performed, e.g. via pruning or merging.

The second objective of this work is to show how a mixture of γ estimates can be reduced. In Section III, we consider merging a weighted sum of measurement rate estimates by minimization of the Kullback-Leibler divergence, and we also give a criterion that is used to determine whether or not two components should be merged.

The proposed Bayesian estimator and merging method is evaluated in Simulations in Section IV, and the paper is finalized with concluding remarks in Section V.

II. BAYESIAN RECURSION FOR γ_k

In this section, we consider recursive estimation of the i:th target's measurement rate $\gamma_k^{(i)}$ from the sequence of measurement sets $\mathbf{Z}^{k,(i)}$. We also show how estimating the measurement rate affects the resulting extended target predicted likelihood. Since we consider only the i:th target, from this point on in this section, we suppress the superscript (i).

A. Measurement update and prediction

The conjugate prior to the Poisson distribution is well known to be the Gamma distribution, see e.g. [8]. Assume that at time t_k the prior distribution for the Poisson rate γ_k is a Gamma distribution,

$$p\left(\gamma_k|\mathbf{Z}^{k-1}\right) = \mathcal{GAM}\left(\gamma_k\,;\,\alpha_{k|k-1}, \beta_{k|k-1}\right) \tag{6}$$

$$= \frac{\beta_{k|k-1}^{\alpha_{k|k-1}}}{\Gamma\left(\alpha_{k|k-1}\right)} \gamma_k^{\alpha_{k|k-1}-1} e^{-\beta_{k|k-1}\gamma_k}. \tag{7}$$

Let the k:th measurement set $\mathbf{Z}_k$ contains $N_{z,k}$ elements, where $N_{z,k}$ is Poisson distributed with rate γ_k,

$$p\left(N_{z,k}|\gamma_k\right) = \mathcal{PS}\left(N_{z,k}\,;\,\gamma_k\right) \tag{8}$$

$$= \frac{\gamma_k^{N_{z,k}} e^{-\gamma_k}}{N_{z,k}!}. \tag{9}$$

The posterior distribution is

$$p\left(\gamma_k|\mathbf{Z}^k\right) = \mathcal{GAM}\left(\gamma_k\,;\,\alpha_{k|k-1}, \beta_{k|k-1}\right) \mathcal{PS}\left(N_{z,k}\,;\,\gamma_k\right) \tag{10a}$$

$$= \frac{\beta_{k|k-1}^{\alpha_{k|k-1}} \gamma_k^{\alpha_{k|k-1}+N_{z,k}-1} e^{-(\beta_{k|k-1}+1)\gamma_k}}{\Gamma\left(\alpha_{k|k-1}\right) N_{z,k}!} \tag{10b}$$

$$= \mathcal{GAM}\left(\gamma_k\,;\,\alpha_{k|k-1}+N_{z,k}, \beta_{k|k-1}+1\right) \tag{10c}$$

$$\times \frac{\Gamma\left(\alpha_{k|k-1}+N_{z,k}\right) \beta_{k|k-1}^{\alpha_{k|k-1}}}{\Gamma\left(\alpha_{k|k-1}\right) \left(\beta_{k|k-1}+1\right)^{\alpha_{k|k-1}+N_{z,k}} N_{z,k}!}$$

$$= \mathcal{GAM}\left(\gamma_k\,;\,\alpha_{k|k}, \beta_{k|k}\right)$$

$$\times \mathcal{L}_\gamma\left(\alpha_{k|k-1}, \beta_{k|k-1}, N_{z,k}\right), \tag{10d}$$

where the predicted likelihood $\mathcal{L}_\gamma\left(\cdot\right)$ is a negative binomial distribution, see e.g. [8].

In case the true parameter is known to be constant over time, the posterior distribution can be predicted as $p\left(\gamma_k|\mathbf{Z}^k\right) = p\left(\gamma_k|\mathbf{Z}^{k-1}\right)$. However, in the general case γ_k may change over time. We propose to use exponential forgetting with a forgetting factor $\frac{1}{\eta_k}$ for the prediction of γ_k,

$$\alpha_{k+1|k} = \frac{\alpha_{k|k}}{\eta_k}, \quad \beta_{k+1|k} = \frac{\beta_{k|k}}{\eta_k}, \tag{11}$$

where $\eta_k > 1$. This prediction has an effective window of length $w_e = \frac{1}{1-1/\eta_k} = \frac{\eta_k}{\eta_k-1}$. Using exponential forgetting prediction with window length w_e approximately means that we only "trust" the information that was contained in the measurements from the last w_e time steps.

The expected value and variance of γ_k are

$$\mathrm{E}\left[\gamma_k\right] = \frac{\alpha_{k|k}}{\beta_{k|k}}, \quad \mathrm{Var}\left(\gamma_k\right) = \frac{\alpha_{k|k}}{\beta_{k|k}^2}. \tag{12}$$

Note that the prediction (11) corresponds to keeping the expected value constant while increasing the variance with a factor $\eta_k > 1$.

B. Extended target predicted likelihood

The measurement update and corresponding predicted likelihood is an important part of any framework for multiple target tracking under uncertain association and clutter. Let ξ_k denote the augmented extended target state,

$$\xi_k = \left(\gamma_k, \mathbf{x}_k\right). \tag{13}$$

Given a set of measurements $\mathbf{Z}_k$ and a prior distribution $p\left(\xi_k|\mathbf{Z}^{k-1}\right)$, the posterior distribution is

$$p\left(\xi_k|\mathbf{Z}^k\right) = p\left(\mathbf{Z}_k|\xi_k\right) p\left(\xi_k|\mathbf{Z}^{k-1}\right) \tag{14a}$$

$$= p\left(\mathbf{Z}_k|\xi_k\right) p\left(\gamma_k|\mathbf{Z}^{k-1}\right) p\left(\mathbf{x}_k|\mathbf{Z}^{k-1}\right). \tag{14b}$$

Note that there is an implicit assumption here that the prior distribution $p\left(\xi_k|\mathbf{Z}^{k-1}\right)$ can be factorized as $p\left(\gamma_k|\mathbf{Z}^{k-1}\right) p\left(\mathbf{x}_k|\mathbf{Z}^{k-1}\right)$. This assumption neglects the dependence between the number of measurements and any extension parameters that are included in $\mathbf{x}_k$. However the probability density over the number of measurements, conditioned

on the target extension, is unknown in most applications, and we believe that this assumption is valid in most cases.

Assume also that the measurement likelihood can be decomposed as

$$p\left(\mathbf{Z}_k|\xi_k\right) = p\left(\mathbf{Z}_k, N_{z,k}|\xi_k\right) = p\left(N_{z,k}|\gamma_k\right) p\left(\mathbf{Z}_k|\mathbf{x}_k\right). \quad (15)$$

The validity of this assumption is also dependent on the considerations mentioned above. The posterior distribution and predicted likelihood is

$$\underbrace{p\left(\gamma_k|\mathbf{Z}^k\right) p\left(\mathbf{x}_k|\mathbf{Z}^k\right)}_{\text{posterior}}$$
$$\times \underbrace{\mathcal{L}_\gamma\left(\alpha_{k|k-1}, \beta_{k|k-1}, N_{z,k}\right) \mathcal{L}_\mathbf{x}\left(\check{\mathbf{x}}_{k|k-1}, \mathbf{Z}_k\right)}_{\text{predicted likelihood}}, \quad (16)$$

where $\check{\mathbf{x}}_{k|k-1}$ denotes the sufficient statistics of $\mathbf{x}_k$. Thus, any extended target tracking framework that estimates the states $\mathbf{x}_k$ can be augmented to also include estimates of the measurement rates γ_k.

In the results section below we give an example where we integrate γ_k estimation into the ET-GM-PHD filter [4], [5]. The posterior distribution for γ_k and the corresponding predicted likelihood $\mathcal{L}_\gamma(\,\cdot\,)$ are given in (10d). The details for the posterior distribution and predicted likelihood for $\mathbf{x}_k$, as well as the full filter recursion, can be found in [4], [5].

III. MULTI-TARGET MIXTURE REDUCTION

A straightforward way to model uncertainty in multiple target tracking is to use mixtures of distributions, see e.g. the Multi-hypothesis Tracking filter [9], or the Gaussian Mixture PHD-filters [4], [5], [10]. Let $p(\,\cdot\,)$ be a mixture of distributions,

$$p\left(\xi_k\right) = \sum_{j=1}^{J_{k|k}} w_j p_j\left(\xi_k\right) \quad (17)$$
$$= \sum_{j=1}^{J_{k|k}} w_j \mathcal{GAM}\left(\gamma_k\,;\,\alpha_{k|k}^{(j)}, \beta_{k|k}^{(j)}\right) p\left(\mathbf{x}_k\,;\,\check{\mathbf{x}}_{k|k}^{(j)}\right)$$

where each distribution p_j is called component. A common choice is to model the state $\mathbf{x}_k$ as Gaussian distributed, see e.g. [4], [5], [9], [10], which would give a Gamma Gaussian (GG) distributed extended target ξ_k. In Koch's random matrix framework [11], the extent is modeled as an inverse Wishart distributed random matrix X_k, and the kinematic parameters[1] are modeled as a random vector $\mathbf{x}_k$. In this case we have $\xi_k = (\gamma_k, \mathbf{x}_k, X_k)$, and (17) would be a mixture of Gamma Gaussian inverse Wishart distributions.

A natural consequence of the tracking frameworks [4], [5], [9], [10] is the increasing number of mixture components, or hypotheses. To keep the target tracking implementation at a tractable level, the mixture must be reduced regularly, which is typically performed via pruning or merging. The output of

---

[1]Position, velocity and acceleration.

mixture reduction is an approximate mixture,

$$\tilde{p}\left(\xi_k\right) = \sum_{j=1}^{\tilde{J}_{k|k}} w_j \tilde{p}_j\left(\xi_k\right) \quad (18)$$
$$= \sum_{j=1}^{\tilde{J}_{k|k}} w_j \mathcal{GAM}\left(\gamma_k\,;\,\tilde{\alpha}_{k|k}^{(j)}, \tilde{\beta}_{k|k}^{(j)}\right) p\left(\mathbf{x}_k\,;\,\tilde{\check{\mathbf{x}}}_{k|k}^{(j)}\right),$$

where $\tilde{J}_{k|k} < J_{k|k}$ and the difference between $p(\,\cdot\,)$ and $\tilde{p}(\,\cdot\,)$ is small by some measure. Here we address mixture reduction via component merging.

One approach to merging is to successively find component pairs that are close by some merging criterion, and merge them, see e.g. [4], [5], [10]. Different methods for merging of Gaussian mixtures are given in e.g. [12]–[16], a method for merging of Gaussian inverse Wishart mixtures is given in [17]. In Section III-A we give a theorem which is used to find the Gamma distribution $q(\,\cdot\,)$ that minimizes the Kullback-Leibler divergence between $\bar{w}q(\,\cdot\,)$ and the sum $p = \Sigma_{i\in L} w_i p_i$, where $\bar{w} = \Sigma_{i\in L} w_i$ and $L \subseteq \left\{1, \ldots, J_{k|k}\right\}$.

When the extended targets are modeled with a mixture (17), the merging criterion should consider both parts of the components, i.e. the distributions of both γ_k and $\mathbf{x}_k$. Different merging criteria for Gaussian distributions are given in e.g. [4], [5], [10], [12]–[16]. In Section III-B we give a merging criterion for mixtures of Gamma distributions.

A. Merging N Gamma components

The Kullback-Leibler divergence (KL-div),

$$\mathrm{KL}\left(p||q\right) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x, \quad (19)$$

is a measure of how similar two functions p and q are. The KL-div is well-known in the literature for its moment-matching characteristics, see e.g. [18], [19], and for probability distributions it is considered the optimal difference measure in a maximum likelihood sense [13]–[15]. Note that minimizing the KL-div between p and q w.r.t. q can be rewritten as a maximization problem,

$$\min_q \mathrm{KL}\left(p||q\right) = \max_q \int p(x) \log\left(q(x)\right) \mathrm{d}x. \quad (20)$$

Theorem 1: Let $p(\,\cdot\,)$ be a weighted sum of Gamma components,

$$p\left(\gamma\right) = \sum_{i=1}^{N} w_i \mathcal{GAM}\left(\gamma\,;\,\alpha_i, \beta_i\right) = \sum_{i=1}^{N} w_i p_i\left(\gamma\right), \quad (21)$$

where $\bar{w} = \sum_{i=1}^{N} w_i$. Let

$$q\left(\gamma\right) = \bar{w}\mathcal{GAM}\left(\gamma\,;\,\alpha, \beta\right) \quad (22)$$

be the minimizer of the KL-div between $p\left(\gamma\right)$ and $q\left(\gamma\right)$ among all Gamma distributions, i.e.

$$q\left(\gamma\right) \triangleq \underset{q(\gamma)\in\mathcal{GAM}}{\arg\min}\ \mathrm{KL}\left(p\left(\gamma\right)||q\left(\gamma\right)\right). \quad (23)$$

Then the parameter β is given by

$$\beta = \frac{\alpha}{\frac{1}{\bar{w}} \sum_{i=1}^{N} w_i \frac{\alpha_i}{\beta_i}}, \tag{24}$$

and the parameter α is the solution to

$$0 = \log \alpha - \psi_0(\alpha) + \frac{1}{\bar{w}} \sum_{i=1}^{N} w_i (\psi_0(\alpha_i) - \log \beta_i)$$
$$- \log \left(\frac{1}{\bar{w}} \sum_{i=1}^{N} w_i \frac{\alpha_i}{\beta_i} \right). \tag{25}$$

∎

Proof: Given in Appendix A.

Remarks: The expression for β (24) corresponds to matching the expected values under both distributions q and p,

$$\bar{w} \, \mathrm{E}_q[\gamma] = \sum_{i=1}^{N} w_i \, \mathrm{E}_{p_i}[\gamma]. \tag{26}$$

The expression for α (25) corresponds to matching the expected values of the logarithm under both distributions q and p,

$$\bar{w} \, \mathrm{E}_q[\log \gamma] = \sum_{i=1}^{N} w_i \, \mathrm{E}_{p_i}[\log \gamma]. \tag{27}$$

A value for the parameter α is easily obtained by applying a numerical root finding algorithm to (25), e.g. Newton's algorithm, see e.g. [20].

B. Merging criterion for Gamma components

In this section we derive a criterion that is used to determine whether or not two Gamma components should be merged. When reducing the number of components, it is preferred to keep the overall modality of the mixture. Thus, if the initial mixture $p(\cdot)$ has M modes, then the reduced mixture $\tilde{p}(\cdot)$ should have M modes.

The optimal solution to this problem is to consider every possible way to reduce $J_{k|k}$ components, compute the corresponding KL-div:s, and then find the best trade-off between low KL-div and reduction of $J_{k|k}$. For $J_{k|k}$ components, there are $B_{J_{k|k}}$ different ways to merge, where B_i is the i:th Bell number [21]. Because B_i increases rapidly with i, e.g. $B_5 = 52$ and $B_{10} = 115975$, the optimal solution can not be used in practice.

Instead a merging criterion must be used to determine whether or not a pair of Gamma components should be merged. As merging criterion the KL-div could be used, however because it is asymmetrical,

$$\mathrm{KL}(p||q) \neq \mathrm{KL}(q||p), \tag{28}$$

it should not be used directly. Instead we use the Kullback-Leibler difference (KL-diff), defined for two distributions $p(\gamma)$

and $q(\gamma)$ as

$$D_{\mathrm{KL}}(p(\gamma), q(\gamma))$$
$$= \mathrm{KL}(p(\gamma)||q(\gamma)) + \mathrm{KL}(q(\gamma)||p(\gamma)) \tag{29a}$$
$$= \int p(\gamma) \log \left(\frac{p(\gamma)}{q(\gamma)} \right) d\gamma + \int q(\gamma) \log \left(\frac{q(\gamma)}{p(\gamma)} \right) d\gamma. \tag{29b}$$

Let $p(\gamma)$ and $q(\gamma)$ be defined as

$$p(\gamma) = \mathcal{GAM}(\gamma; \alpha_1, \beta_1), \tag{30a}$$
$$q(\gamma) = \mathcal{GAM}(\gamma; \alpha_2, \beta_2). \tag{30b}$$

The KL-div between $p(\cdot)$ and $q(\cdot)$ is

$$\mathrm{KL}(p(\gamma)||q(\gamma))$$
$$= \alpha_1 \log \beta_1 - \log \Gamma(\alpha_1) + (\alpha_1 - 1)(\psi_0(\alpha_1) - \log \beta_1)$$
$$\quad - \alpha_1 - \alpha_2 \log \beta_2 + \log \Gamma(\alpha_2)$$
$$\quad - (\alpha_2 - 1)(\psi_0(\alpha_1) - \log \beta_1) + \beta_2 \frac{\alpha_1}{\beta_1} \tag{31a}$$
$$= \log \left(\frac{\beta_1^{\alpha_1}}{\beta_2^{\alpha_2}} \right) + \log \left(\frac{\Gamma(\alpha_2)}{\Gamma(\alpha_1)} \right)$$
$$\quad + (\alpha_1 - \alpha_2)(\psi_0(\alpha_1) - \log \beta_1) + \alpha_1 \left(\frac{\beta_2}{\beta_1} - 1 \right), \tag{31b}$$

and the KL-div between q and p is defined analogously. The KL-diff between p and q becomes

$$D_{\mathrm{KL}}(p(\gamma), q(\gamma))$$
$$= (\alpha_1 - \alpha_2) \left(\psi_0(\alpha_1) - \psi_0(\alpha_2) + \log \frac{\beta_2}{\beta_1} \right) \tag{32}$$
$$\quad + (\beta_2 - \beta_1) \left(\frac{\alpha_1}{\beta_1} - \frac{\alpha_2}{\beta_2} \right).$$

C. Merging of extended target components

When merging is used to reduce an extended target mixture (17), the merging criterion should be defined over both γ_k and $\mathbf{x}_k$. For example, the following merging criterion could be used

$$D_{\mathrm{KL}}(p_i(\xi_k), p_j(\xi_k)) < U, \tag{33}$$

where $D_{\mathrm{KL}}(\cdot)$ is the KL-diff between two extended target components. Owing to the assumed conditional independence of the distributions over γ_k and $\mathbf{x}_k$ in (14), the KL-diff can expressed as a sum

$$D_{\mathrm{KL}}(p_i(\xi_k), p_j(\xi_k)) = D_{\mathrm{KL}}^{\gamma}(i, j) + D_{\mathrm{KL}}^{\mathbf{x}}(i, j), \tag{34}$$

where $D_{\mathrm{KL}}^{\gamma}(i, j) = D_{\mathrm{KL}}(p_i(\gamma), p_j(\gamma))$ is given in (32) and $D_{\mathrm{KL}}^{\mathbf{x}}(i, j) = D_{\mathrm{KL}}(p_i(\mathbf{x}), p_j(\mathbf{x}))$. Thus, the following merging criterion could alternatively be used

$$\left(D_{\mathrm{KL}}^{\gamma}(i, j) < U_{\gamma} \right) \& \left(D_{\mathrm{KL}}^{\mathbf{x}}(i, j) < U_{\mathbf{x}} \right), \tag{35}$$

where $\&$ is the logical and operator. In case $\mathbf{x}_k$ is Gaussian distributed, possible merging criterions $D_{\mathrm{KL}}^{\mathbf{x}}(i, j)$ are given in e.g. [10], [12].
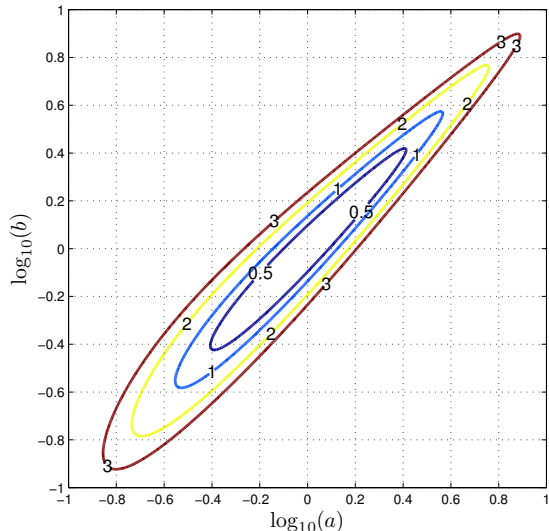
Fig. 1. KL-diff for two Gamma distributions, when $\alpha_2 = a\alpha_1 = a10$ and $\beta_2 = b\beta_1$. When $a \approx b$, the expected value is approximately the same for both distributions, and the the ratio of the variances is $1/b$. This explains the elongated shape of the KL-diff along $a \approx b$.
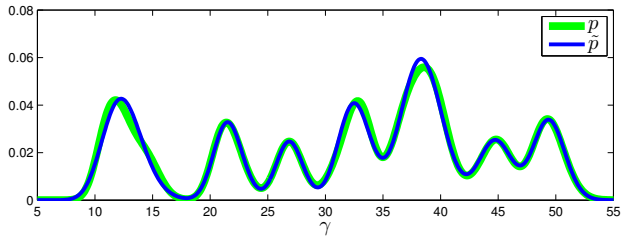

Fig. 2. Merging of 20 Gamma components, using the merging method and criterion of Section III. The reduced mixture has 7 components and has preserved the overall modality.

IV. RESULTS

A. Merging criterion

Letting $\alpha_2 = a\alpha_1$ and $\beta_2 = b\beta_1$, the KL-diff simplifies to

$$
\begin{aligned}
D_{\mathrm{KL}} &\left(p\left(\gamma\right), q\left(\gamma\right)\right) \\
=& \alpha_1 \left(1 - a\right) \left(\log b + \psi_0(\alpha_1) - \psi_0(a\alpha_1)\right) \\
&+ \alpha_1 \left(b - 1\right) \left(1 - \frac{a}{b}\right),
\end{aligned} \tag{36}
$$

i.e. it becomes independent of the specific value of β_1. It can be shown that, for given a and b, a larger α_1 means a larger KL-diff. For $\alpha_1 = 10$, the KL-diff is shown in Figure 1.

B. Comparison of merging algorithms

An intensity $p\left(\gamma\right)$ with 20 Gamma components was reduced using the merging method and criterion presented in Section III. The Gamma mixture parameters were sampled uniformly from the following intervals,

$$
w_i \in [0.05, 0.95], \; \alpha_i \in [50, 2500], \; \beta_i \in [5, 50], \tag{37}
$$

i.e. α_i and β_i were sampled such that the expected value and variance of γ belongs to $[10, 50]$ and $[1, 2]$, respectively.

(a) True γ_k

(b) $\eta_k = 1.10$, $w_e = 11$

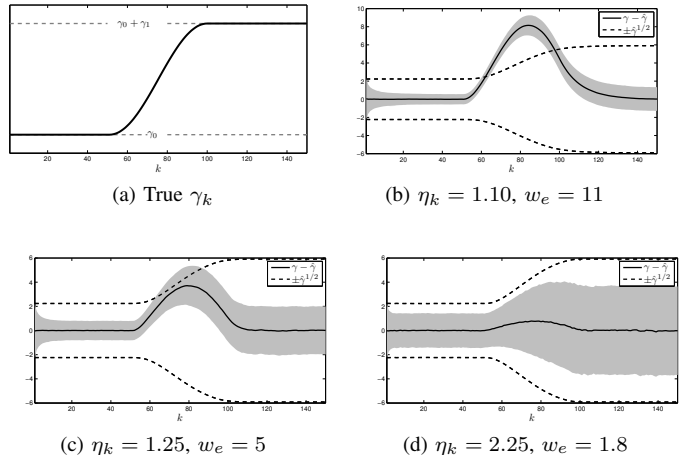(c) $\eta_k = 1.25$, $w_e = 5$

(d) $\eta_k = 2.25$, $w_e = 1.8$

Fig. 3. Single target results. (a): The true rate γ_k varied between $\gamma_0 = 5$ and $\gamma_0 + \gamma_1 = 35$. (b), (c) and (d): The solid line shows the average estimation error, the gray area is the average estimation error $\pm$ one standard deviation, and the dashed lines are the bounds $\pm\sqrt{\hat{\gamma}}$, c.f. (1). A higher η_k gives a lower average estimation error, however the estimation error also has much larger standard deviation.

The original mixture and the approximation are shown in Figure 2. The reduced mixture has 7 components, and manages to capture the overall modality of the original mixture.

C. Single target results

A single target with time varying γ was simulated for 150 time steps, the true measurement rate varied with time as shown in Figure 3a. The estimation error, averaged over 10^4 Monte Carlo runs, is shown in Figures 3b, 3c and 3d, for $\eta_k = 1.10$, $\eta_k = 1.25$ and $\eta_k = 2.25$, respectively. With a higher η_k, the estimate responds faster to changes in the true parameter, at the expense of being more sensitive to noise. As with any prediction and correction recursion, setting the parameter requires a trade off between noise cancellation and tracking capabilities.

D. Multiple target results

The Bayesian γ_k estimator was integrated into the Gaussian Mixture Probability Hypothesis Density (ET-GM-PHD) filter [4], [5]. A scenario with three targets was simulated for 100 time steps, the true Poisson rates were set to $\gamma_k^{(1)} = 5$, $\gamma_k^{(2)} = 15$ and $\gamma_k^{(3)} = 30$. Estimation results for $\eta_k = 1.25$ are shown in Figure 4a. The estimates

$$
\hat{\gamma}_{k|k}^{(i)} = \frac{\alpha_{k|k}^{(i)}}{\beta_{k|k}^{(i)}} \tag{38}
$$

are a bit noisy, however they remain within the bounds given by

$$
\gamma_k^{(i)} \pm \sqrt{\gamma_k^{(i)}}, \tag{39}
$$

i.e. the true mean $\pm$ one standard deviation. With $\eta_k = 1.01$ the estimation error is much smaller, see Figure 4b. However, as discussed previously, with a low η_k the response to changes in the true parameter would be slower.
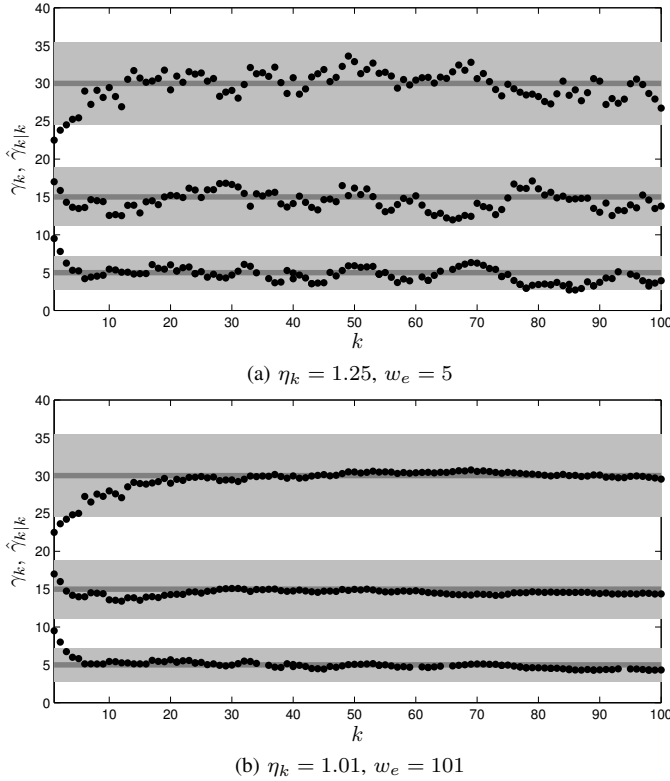
(a) $\eta_k = 1.25$, $w_e = 5$

(b) $\eta_k = 1.01$, $w_e = 101$

Fig. 4. Three targets with true rates $\gamma_k^{(1)} = 5$, $\gamma_k^{(2)} = 15$ and $\gamma_k^{(3)} = 30$, shown as dark gray lines. The estimates $\hat{\gamma}_{k|k}^{(i)}$, shown as black dots, remain within the bounds $\gamma_k^{(i)} \pm \sqrt{\gamma_k^{(i)}}$, i.e. the true mean $\pm$ one standard deviation, shown as light gray areas. The same sequence of measurement sets is used in (a) and (b), with different η_k.

V. CONCLUDING REMARKS

This paper presented a Bayesian estimator for the rate parameter γ of a Poisson distribution. The conjugate prior of γ is the Gamma distribution, and, using exponential forgetting prediction, it is possible to track a rate γ that changes over time. To manage multiple targets with different rates, a mixture of Gamma distributions is utilized. Mixture reduction is addressed, where components are merged via analytical minimization of the Kullback-Leibler divergence between a weighted sum of Gamma distributions and the single Gamma distribution that best approximates the sum. A simulation study was used to show the merits of the Poisson rate estimation framework.

In future work, we intend to integrate the rate estimation fully into the Gaussian mixture and Gaussian inverse Wishart extended target PHD filters. Having a good estimate of the measurement rate could have important implications for the performance, especially during the measurement partitioning step. Future work also includes improving upon the exponential forgetting prediction. The number of measurements generated can be affected by the extended target's position, as well as its shape and size. Including the estimated position, size and shape in the prediction step could possibly improve tracking of Poisson rates that change over time.

APPENDIX A
PROOF OF THEOREM 1

First we derive an expected value which is needed in the proof of Theorem 1.

A. Expected value of logarithm

Let y be a uni-variate random variable. The moment generating function for y is defined as

$$\mu_y(s) \triangleq \mathrm{E}_y\left[e^{sy}\right], \tag{40}$$

and the expected value of y is given in terms of $\mu_y(s)$ as

$$\mathrm{E}[y] = \left. \frac{\mathrm{d}\mu_y(s)}{\mathrm{d}s}\right|_{s=0}. \tag{41}$$

Let $y = \log\gamma$, where $\gamma \sim \mathcal{GAM}(\gamma; \alpha, \beta)$. The moment generating function of y is

$$\mu_y(s) = \mathrm{E}\left[\gamma^s\right] \tag{42a}$$

$$= \int \gamma^s \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma^{\alpha-1} e^{-\beta\gamma} \mathrm{d}\gamma \tag{42b}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(s+\alpha)}{\beta^{s+\alpha}} \int \mathcal{GAM}(\gamma; s+\alpha, \beta)\, \mathrm{d}\gamma \tag{42c}$$

$$= \frac{\Gamma(s+\alpha)}{\Gamma(\alpha)\beta^s}. \tag{42d}$$

The expected value of y is

$$\mathrm{E}[y] = \mathrm{E}[\log\gamma] \tag{43a}$$

$$= \left. \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\Gamma(s+\alpha)}{\Gamma(\alpha)\beta^s}\right)\right|_{s=0} \tag{43b}$$

$$= \left(\frac{\frac{\mathrm{d}}{\mathrm{d}s}\Gamma(s+\alpha)}{\Gamma(\alpha)\beta^s}\right)\Bigg|_{s=0} + \left(\frac{\Gamma(s+\alpha)}{\Gamma(\alpha)}\frac{\mathrm{d}}{\mathrm{d}s}\beta^{-s}\right)\Bigg|_{s=0} \tag{43c}$$

$$= \psi_0(\alpha) - \log\beta \tag{43d}$$

where $\psi_0(\cdot)$ is the digamma function (a.k.a. the polygamma function of order 0).

B. Proof

Proof: We have $q(\cdot)$ given as

$$q(\gamma) \triangleq \arg\min_q \mathrm{KL}(p||q) \tag{44a}$$

$$= \arg\max_q \int p(\gamma)\log(q(\gamma))\mathrm{d}\gamma \tag{44b}$$

$$= \arg\max_q \sum_{i=1}^N w_i \int p_i(\gamma)\log(q(\gamma))\,\mathrm{d}\gamma, \tag{44c}$$

where the i:th integral is

$$\int p_i(\gamma)\log(q(\gamma))\,\mathrm{d}\gamma$$
$$= \int p_i(\gamma)\left[\alpha\log\beta - \log\Gamma(\alpha) + (\alpha+1)\log\gamma - \beta\gamma\right]\mathrm{d}\gamma$$
$$= \alpha\log\beta - \log\Gamma(\alpha) + (\alpha-1)\mathrm{E}_i\left[\log\gamma\right] - \beta\,\mathrm{E}_i\left[\gamma\right]$$
$$= \alpha\log\beta - \log\Gamma(\alpha) + (\alpha-1)\left(\psi_0(\alpha_i) - \log\beta_i\right) - \beta\frac{\alpha_i}{\beta_i}. \tag{45}$$

Taking the derivative of the objective function with respect to β, equating the result to zero, and solving for β, we get

$$\beta = \frac{\alpha}{\frac{1}{\bar{w}} \sum_{i=1}^{N} w_i \frac{\alpha_i}{\beta_i}} . \tag{46}$$

Now, we take the derivative of the objective function with respect to α and equate the result to zero to obtain

$$0 = \sum_{i=1}^{N} w_i \left(\log \beta - \psi_0(\alpha) + \psi_0(\alpha_i) - \log \beta_i \right) \tag{47}$$

$$= \bar{w} \log \beta - \bar{w} \psi_0(\alpha) + \sum_{i=1}^{N} w_i \left(\psi_0(\alpha_i) - \log \beta_i \right) . \tag{48}$$

Inserting β and rearranging the terms we obtain

$$0 = \log \alpha - \psi_0(\alpha) + \frac{1}{\bar{w}} \sum_{i=1}^{N} w_i \left(\psi_0(\alpha_i) - \log \beta_i \right)$$
$$- \log \left(\frac{1}{\bar{w}} \sum_{i=1}^{N} w_i \frac{\alpha_i}{\beta_i} \right) . \tag{49}$$

∎

REFERENCES

[1] Y. Bar-Shalom and T. E. Fortmann, *Tracking and data association*, ser. Mathematics in Science and Engineering. San Diego, CA, USA: Academic Press Professional, Inc., 1987, vol. 179.

[2] K. Gilholm, S. Godsill, S. Maskell, and D. Salmond, "Poisson models for extended target and group tracking," in *Proceedings of Signal and Data Processing of Small Targets*, vol. 5913. San Diego, CA, USA: SPIE, Aug. 2005, pp. 230–241.

[3] R. Mahler, "PHD filters for nonstandard targets, I: Extended targets," in *Proceedings of the International Conference on Information Fusion*, Seattle, WA, USA, Jul. 2009, pp. 915–921.

[4] K. Granström, C. Lundquist, and U. Orguner, "A Gaussian mixture PHD filter for extended target tracking," in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.

[5] ——, "Extended Target Tracking using a Gaussian Mixture PHD filter," *IEEE Transactions on Aerospace and Electronic Systems*, 2012.

[6] K. Granström and U. Orguner, "A PHD filter for tracking multiple extended targets using random matrices," *IEEE Transactions on Signal Processing*.

[7] K. Granström, C. Lundquist, and U. Orguner, "Tracking Rectangular and Elliptical Extended Targets Using Laser Measurements," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 592–599.

[8] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, ser. Texts in Statistical Science. Chapman & Hall/CRC, 2004.

[9] Y. Bar-Shalom and X. Rong Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS, 1995.

[10] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[11] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, Jul. 2008.

[12] D. J. Salmond, "Mixture reduction algorithms for target tracking in clutter," in *Proceedings of SPIE Signal and Data Processing of Small Targets*, Orlando, FL, USA, Jan. 1990, pp. 434–445.

[13] J. L. Williams and P. S. Maybeck, "Cost-Function-Based Gaussian Mixture Reduction for Target Tracking," in *Proceedings of the International Conference on Information Fusion*, Cairns, Queensland, Australia, Jul. 2003.

[14] A. R. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, Jul. 2007.

[15] D. Schieferdecker and M. F. Huber, "Gaussian Mixture Reduction via Clustering," in *Proceedings of the International Conference on Information Fusion*, Seattle, WA, USA, Jul. 2009.

[16] D. F. Crouse, P. Willett, K. Pattipati, and L. Svensson, "A Look At Gaussian Mixture Reduction Algorithms," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011.

[17] K. Granström and U. Orguner, "On the Reduction of Gaussian inverse Wishart mixtures," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012.

[18] C. M. Bishop, *Pattern recognition and machine learning*. New York, USA: Springer, 2006.

[19] T. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, Jan. 2001.

[20] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed. New York: Springer-Verlag, 1993.

[21] G.-C. Rota, "The number of partitions of a set," *The American Mathematical Monthly*, vol. 71, no. 5, pp. 498–504, May 1964.