Bayesian Estimation With Imprecise Likelihoods: Random Set Approach

Branko Ristic

***Abstract*—In many practical applications of statistical signal processing, the likelihood functions are only partially known. The measurement model in this case is affected by two sources of uncertainty: stochastic uncertainty and imprecision. Following the framework of random set theory [1], the paper presents the optimal Bayesian estimator for this problem. The resulting Bayes estimator in general has no analytic closed form solution, but can be approximated, for example, using the Monte Carlo method. A numerical example is included to illustrate the theory.**

***Index Terms*—Bayesian estimation, Bayesian robustness, random set theory.**

I. Introduction

BAYESIAN estimation heavily depends on models: the model of the prior density and the model of the likelihood function [2]. In many practical applications, however, the specification of a precise mathematical model is difficult or impossible. For example, in cellular networks, due to the multipath and shadowing, the path-loss exponent is known only to be an interval value [3]. The likelihood function of the received signal strength depends on the path-loss exponent and as a consequence, its model is imprecise.

Bayesian estimation using imprecise likelihoods have attracted some interest recently. Following Walley [4], some authors represent imprecise likelihoods, and probability densities in general, by convex sets of probability densities [5], [6]. Bayesian inference using convex sets of probability densities, however, is quite involved and still under development. An alternative approach to Bayesian estimation using imprecise likelihoods has been recently proposed by Mahler [1] in the framework of random set theory. This is a very simple and elegant approach which we further develop and demonstrate in this paper.

II. Problem Description

Let $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ be the state space and $\mathbf{x} \in \mathcal{X}$ an unknown random parameter vector which we want to estimate. Prior knowledge of $\mathbf{x}$, before any measurement is processed, is represented by prior probability density function (PDF) π_0, that is, initially $\mathbf{x} \sim \pi_0(\mathbf{x})$.

Manuscript received December 22, 2010; revised May 01, 2011; accepted May 02, 2011. Date of publication May 10, 2011; date of current version May 19, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thia Kirubarajan.

The author is with ISR Division, Defence Science and Technology Organisation, Melbourne, VIC 3207, Australia (e-mail: branko.ristic@dsto.defence.gov.au).

Digital Object Identifier 10.1109/LSP.2011.2152392

Suppose N independent measurements of $\mathbf{x}$ are available for estimation. The measurement space is $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$, while the measurements are denoted $\mathbf{z}_i \in \mathcal{Z}$, for $i = 1, \ldots, N$. The assumption is that each measurement z_i is related to the (hidden) parameter vector as follows:

$$\mathbf{z}_i = \mathbf{h}_i(\mathbf{x}; \boldsymbol{\theta}_i) + \mathbf{v}_i \tag{1}$$

where

- $\mathbf{h}_i$ is a known nonlinear mapping $\mathbf{h}_i : \mathcal{X} \to \mathcal{Z}$;
- $\boldsymbol{\theta}_i \in \mathbb{R}^{n_\theta}$ is a known parameter vector of the transformation $\mathbf{h}_i$;
- $\mathbf{v}_i$ is additive measurement noise of known PDF $p_{\mathbf{v}}$.

This is a standard problem, for which we can specify immediately the *precise* likelihood function as follows $\ell_{\boldsymbol{\theta}_i}(\mathbf{z}_i|\mathbf{x}) = p_{\mathbf{v}}(\mathbf{z}_i - \mathbf{h}_i(\mathbf{x}; \boldsymbol{\theta}_i))$.

The problem we are interested in the paper is when the parameter of transformation $\mathbf{h}_i$, that is $\boldsymbol{\theta}_i$, is known only partially. Let us assume that all we know is that $\boldsymbol{\theta} \in [\boldsymbol{\theta}]_i$, where $[\boldsymbol{\theta}] = [\underline{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}]$ denotes an interval (connected subset) of $\mathbb{R}^{n_\theta}$. Following (1) we can attempt to write in this case

$$\mathbf{z}_i \stackrel{?}{=} \mathbf{h}_i(\mathbf{x}; [\boldsymbol{\theta}]_i) + \mathbf{v}_i \tag{2}$$

where $[\theta]_i$, $i = 1, \ldots, N$ are known intervals. The problem is, however, that $\mathbf{h}_i(\mathbf{x}; [\boldsymbol{\theta}]_i)$ is *not* a function. This is because $\mathbf{h}_i(\mathbf{x}; [\boldsymbol{\theta}]_i)$ defines one-to-many mapping, that is a single value of the argument, $\mathbf{x}$, maps into a subset of the measurement space $\mathcal{Z}$. Moreover, this subset is *infinite*. For this reason (2) is meaningless.

In a similar manner, one could attempt to express the likelihood function in this case as: $\ell_{[\boldsymbol{\theta}]_i}(\mathbf{z}_i|\mathbf{x}) = p_{\mathbf{v}}(\mathbf{z}_i - \mathbf{h}_i(\mathbf{x}; [\boldsymbol{\theta}]_i))$, bearing in mind that this construct does not have a precise mathematical interpretation. One can at best think of it as a representation of an infinite number of precise likelihoods $\ell_{\boldsymbol{\theta}_i}(\mathbf{z}_i|\mathbf{x})$. We refer to $\ell_{[\boldsymbol{\theta}]_i}(\mathbf{z}_i|\mathbf{x})$ as to the *imprecise* likelihood function[1].

The imprecise likelihood function expresses the uncertainty due to two sources [8]: *randomness* (or stochastic uncertainty) is the consequence of the measurement error $\mathbf{v}_i$; *imprecision* is the consequence of the partial knowledge of the transformation parameter, i.e., $\boldsymbol{\theta} \in [\underline{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}]$. There are many practical situations where model (2) is appropriate; an illustrative example will be studied in Section IV.

Let the stacked measurement vector be denoted $\mathbf{z}_{1:N} = [\mathbf{z}_1^\intercal, \ldots, \mathbf{z}_N^\intercal]^\intercal$. Formally, in the Bayesian framework the problem is to compute the posterior PDF $\pi(\mathbf{x}|\mathbf{z}_{1:N})$, given

[1]The term *imprecision* here is borrowed from the artificial intelligence jargon [7].

the prior $\pi_0(\mathbf{x})$ and the measurement vector $\mathbf{z}_{1:N}$ modeled according to the imprecise model (2).

Two remarks are in order here. First, the described problem is related, but different from the one considered in [9], where the measurements were imprecise (intervals on $\mathcal{Z}$), but the likelihood was precise. Second, vector $\boldsymbol{\theta}$ is considered as a nuisance parameter and is not estimated.

III. The Solution in the Framework of Random Set Theory

In order to simplify notation, let us temporarily drop the subscript i. The key idea of the proposed approach ([1, Chs. 4–7]) is to realize that the measurement set resulting from the one-to-many mapping

$$\mathbf{h}(\mathbf{x};[\boldsymbol{\theta}]) + \mathbf{v} \tag{3}$$

can be modeled by a random closed set $\Sigma_{\mathbf{x}}$ which takes values as closed intervals of $\mathcal{Z}$, i.e., $\Sigma_{\mathbf{x}} \subseteq \mathcal{Z}$.

A random set is a random element whose value is a set. As such it is a generalization of the familiar concept of a random variable. The probability laws of random sets on $\mathcal{Z}$ can be specified via distribution functions of random sets. This has been done rigorously, see for example [10], [11]. Loosely speaking, a random closed set Σ is a mapping from the sample space Ω to some collection $\mathcal{I}(\mathcal{Z})$ of closed subsets of $\mathcal{Z}$, which is $\mathcal{A}$-$\mathcal{B}$ measurable, where $\mathcal{A}$ and $\mathcal{B}$ are σ fields on Ω and $\mathcal{I}(\mathcal{Z})$, respectively. If $\mathbb{P}$ is the probability measure on $\mathcal{A}$, then the probability law of Σ is the probability measure:

$$P(B) = \mathbb{P}(\{\omega : \Sigma(\omega) \in B\}) = \mathbb{P}(\Sigma^{-1}(B)) \tag{4}$$

for $B \in \mathcal{B}$.

Going back to the random closed set $\Sigma_{\mathbf{x}}$ which represents the model $\mathbf{h}(\mathbf{x};[\boldsymbol{\theta}]) + \mathbf{v}$, the first question is how to state that a particular (point) measurement $\mathbf{z} \in \mathcal{Z}$ matches the model. Mahler ([1, Sec. 6.3]) takes the approach that a measurement matches the model if it does not contradict it, i.e., if $\mathbf{z} \cap \Sigma_{\mathbf{x}} \neq \emptyset$. This translates to the following definition of the measurement likelihood function (referred to as the *generalized* likelihood):

$$g(\mathbf{z}|\mathbf{x}) = \Pr\{\mathbf{z} \cap \Sigma_{\mathbf{x}} \neq \emptyset\} \tag{5}$$

$$= \Pr\{\mathbf{z} \in \Sigma_{\mathbf{x}}\}. \tag{6}$$

Note that the generalized likelihood does not integrate to 1. For the case of additive noise as in (3), the generalized likelihood (6) can be expressed as follows:

$$g(\mathbf{z}|\mathbf{x}) = \int_{\underline{\mathbf{h}}_{\mathbf{x}}}^{\bar{\mathbf{h}}_{\mathbf{x}}} p_{\mathbf{v}}(\mathbf{z}-\mathbf{h})\, d\mathbf{h}. \tag{7}$$

where

$$\underline{\mathbf{h}}_{\mathbf{x}} = \min\{\mathbf{h}(\mathbf{x};\underline{\boldsymbol{\theta}}), \mathbf{h}(\mathbf{x};\bar{\boldsymbol{\theta}})\} \tag{8}$$

$$\bar{\mathbf{h}}_{\mathbf{x}} = \max\{\mathbf{h}(\mathbf{x};\underline{\boldsymbol{\theta}}), \mathbf{h}(\mathbf{x};\bar{\boldsymbol{\theta}})\}. \tag{9}$$

define the limits of an n_z dimensional interval on $\mathcal{Z}$ which contains all points $\mathbf{h}(\mathbf{x};\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in [\boldsymbol{\theta}]$. The integral in (7) is n_z dimensional, which is in contrast to the standard Bayesian analysis, where marginalization over nuisance parameters involves n_θ dimensional integration.

Next we work out analytically the generalized likelihood for the important case of a Gaussian additive noise $\mathbf{v}$. Let $\mathcal{N}(\mathbf{y};\boldsymbol{\mu},\mathbf{P})$ denote a Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance $\mathbf{P}$. Its cumulative distribution function is $\phi(\mathbf{u};\boldsymbol{\mu},\mathbf{P}) = \int_{-\infty}^{\mathbf{u}} \mathcal{N}(\mathbf{y};\boldsymbol{\mu},\mathbf{P})\, d\mathbf{y}$. Assuming that measurement noise $\mathbf{v}$ in (6) is zero-mean white Gaussian with covariance $\mathbf{R}$, i.e., $p_{\mathbf{v}}(\mathbf{v}) = \mathcal{N}(\mathbf{v};0,\mathbf{R})$, expression (7) simplifies to

$$g(\mathbf{z}|\mathbf{x}) = \int_{\underline{\mathbf{h}}_{\mathbf{x}}}^{\bar{\mathbf{h}}_{\mathbf{x}}} \mathcal{N}(\mathbf{h};\mathbf{z},\mathbf{R})\, d\mathbf{h} \tag{10}$$

$$= \phi(\bar{\mathbf{h}}_{\mathbf{x}};\mathbf{z},\mathbf{R}) - \phi(\underline{\mathbf{h}}_{\mathbf{x}};\mathbf{z},\mathbf{R}) \tag{11}$$

$$= 1 - \phi(\mathbf{z};\bar{\mathbf{h}}_{\mathbf{x}},\mathbf{R}) - (1 - \phi(\mathbf{z};\underline{\mathbf{h}}_{\mathbf{x}},\mathbf{R})) \tag{12}$$

$$= \phi(\mathbf{z};\underline{\mathbf{h}}_{\mathbf{x}},\mathbf{R}) - \phi(\mathbf{z};\bar{\mathbf{h}}_{\mathbf{x}},\mathbf{R}). \tag{13}$$

Note that the generalized likelihood $g(\mathbf{z}|\mathbf{x})$ in (13) effectively defines a fuzzy membership function on $\mathcal{Z}$. As the measurement noise $\mathbf{v}$ is reduced (i.e., when measurement uncertainty is dominated by imprecision rather than randomness), the fuzzy membership function (13) approaches the indicator function. This can be seen by considering the case with small covariance ($\mathbf{R} \to 0$) in (13): the cumulative distribution can then be approximated as $\phi(\mathbf{z};\underline{\mathbf{h}}_{\mathbf{x}},\mathbf{R}) \approx H(\mathbf{z}-\underline{\mathbf{h}}_{\mathbf{x}})$, where H is the Heaviside function, and from (13) we have

$$g(\mathbf{z}|\mathbf{x}) \approx H(\mathbf{z}-\underline{\mathbf{h}}_{\mathbf{x}}) - H(\mathbf{z}-\underline{\mathbf{h}}_{\mathbf{x}}) = \mathbb{I}_{[\mathbf{h}_{\mathbf{x}}]}(\mathbf{z}) \tag{14}$$

where $\mathbb{I}_{[\mathbf{h}]}(\mathbf{z}) = 1$ if $\mathbf{z} \in [\mathbf{h}]$ and zero otherwise, is the indicator function.

The generalized likelihood function provides a useful relationship with the Dempster–Shafer theory [12]: the generalized likelihood function (6) represents the *plausibility function* on singletons.

Fig. 1 illustrates the generalized likelihood $g(\mathbf{z}|\mathbf{x})$ for one-dimensional measurement ($n_z = 1$). The limits of integration are $\underline{\mathbf{h}}_{\mathbf{x}} = 45$ and $\bar{\mathbf{h}}_{\mathbf{x}} = 60$. Three values of variance $\mathbf{R}$ are considered: 4, 1, and 0.0001. For a very small value of $\mathbf{R}$, the generalized likelihood in accordance with (14) approaches the indicator function $\mathbb{I}_{[\mathbf{h}_{\mathbf{x}}]}(\mathbf{z})$.

The optimal Bayes estimator for imprecise likelihoods in now defined as ([1], (5.6))

$$\pi(\mathbf{x}|\mathbf{z}_{1:N}) = \frac{g(\mathbf{z}_{1:N}|\mathbf{x}) \cdot \pi_0(\mathbf{x})}{\int_{\mathcal{X}} g(\mathbf{z}_{1:N}|\mathbf{x}) \cdot \pi_0(\mathbf{x})\, d\mathbf{x}} \tag{15}$$

where $g(\mathbf{z}_{1:N}|\mathbf{x})$ is the joint generalized likelihood function. Assuming that measurements are conditionally independent, $g(\mathbf{z}_{1:N}|\mathbf{x}) = \prod_{i=1}^{N} g(\mathbf{z}_i|\mathbf{x})$. All quantities of interest related to $\mathbf{x}$ can be computed from the posterior PDF $\pi(\mathbf{x}|\mathbf{z}_{1:N})$.

The posterior PDF $\pi(\mathbf{x}|\mathbf{z}_{1:N})$ in general cannot be found in the closed-form. However, Monte Carlo methods can be applied easily and with a reasonable computational expense [13]. This will be illustrated next.

IV. Example: Localization Using RSS

Received signal strength (RSS) is often used for localization of energy emitting sources [14]. Since the emitting power of the source is in general unknown, the parameter vector for localization in the two-dimensional Cartesian coordinate system can be

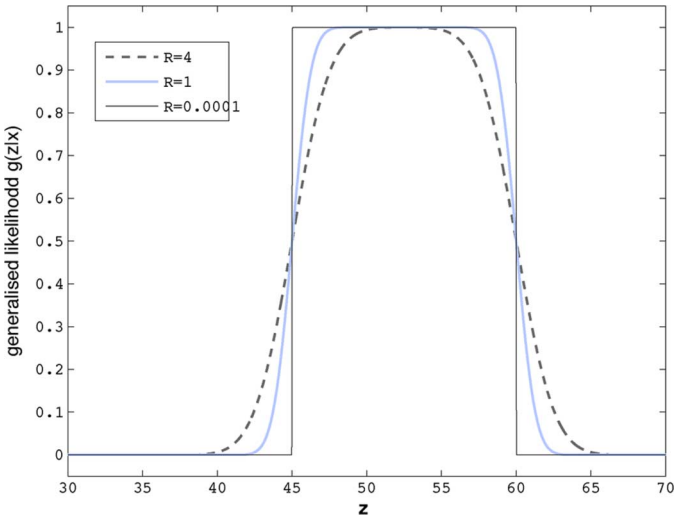

Fig. 1. Illustration of the generalized likelihood function (13) for $n_z = 1$ and additive zero-mean Gaussian measurement noise with variance $\mathbf{R}$.

specified as $\mathbf{x} = [x_0 \quad y_0 \quad A_0]^\intercal$, where $(x_0, y_0) \in \mathbb{R}^2$ is the source position and $A_0 \in \mathbb{R}^+$ is the received power (in dB) at a short reference distance d_0 from the source. A measurement z_i of RSS (in dB) at known location (x_i, y_i) can be modeled as [3]:

$$z_i = A_0 - 10 \cdot \theta_i \cdot \log \frac{d_i}{d_0} + v_i \tag{16}$$

for $i = 1, \ldots, N$, where

$$d_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2},$$

is the distance between the source and sensor i; θ_i is the propagation loss from the source to sensor i, and v_i is zero-mean white Gaussian noise with a known standard deviation $\sigma = 2$ dB. The propagation loss θ_i in free space is 2. In real world, however, due to multipath and shadowing [3], [14], θ_i is only partially known: experimentally it has been found that it can be anywhere in the interval [2, 4].

The scenario we consider is shown in Fig. 2(a): there are $N = 12$ sensors placed on a circle centred at (50, 50), with a radius 50 (sensors 1, 2, and 3 are labeled; anticlockwise labeling continued for the remaining sensors). The (unknown) true source parameters are $x_0 = 30$, $y_0 = 60$ and $A_0 = 35$, with a (known) reference distance $d_0 = 10$. The (unknown) propagation loss factors are set to $\theta = 2.3$ for sensors $1, 2, \ldots, 6$ and $\theta = 3.5$ for sensors $7, 8, \ldots, 12$. The prior PDF used in estimation is $\pi_0(\mathbf{x}) = \mathcal{U}_{[5,95]}(x_0) \cdot \mathcal{U}_{[5,95]}(y_0) \cdot \mathcal{U}_{[25,65]}(A_0)$, where $\mathcal{U}_{[a,b]}(c)$ is the uniform density with support $[a, b]$.

The Bayes density estimator (15) was implemented using a rather simplistic variant of the importance sampling method [15], [16]. The first step is to draw a random sample from the prior PDF: $\mathbf{x}_j \sim \pi_0(\mathbf{x})$ for $j = 1 \ldots, S_0$. An importance weight $w_j \propto \prod_{i=1}^{N} g(z_i|\mathbf{x}_j)$ is then associated with each sample $\mathbf{x}_j$. This is carried out by first computing the unnormalized weights using (13): $\tilde{w}_j = \prod_{i=1}^{N} g(z_i|\mathbf{x}_j)$. Then the weights are normalized to obtain: $w_j = \tilde{w}_j / \sum_{\ell=1}^{S_0} \tilde{w}_\ell$. The final step consists of resampling $S = S_0/10$ times from the random measure $\{\mathbf{x}_j, w_j; j = 1, \ldots, S_0\}$ to obtain S equally weighted samples $\{\mathbf{x}_j^*, w_j = (1)/(S); j = 1, \ldots, S\}$ which approximate the posterior PDF $\pi(\mathbf{x}|\mathbf{z}_{1:N})$. The result is shown in Fig. 2(a) in (x_0, y_0) space and Fig. 2(b) in A_0 space, using $S = 5000$ samples.

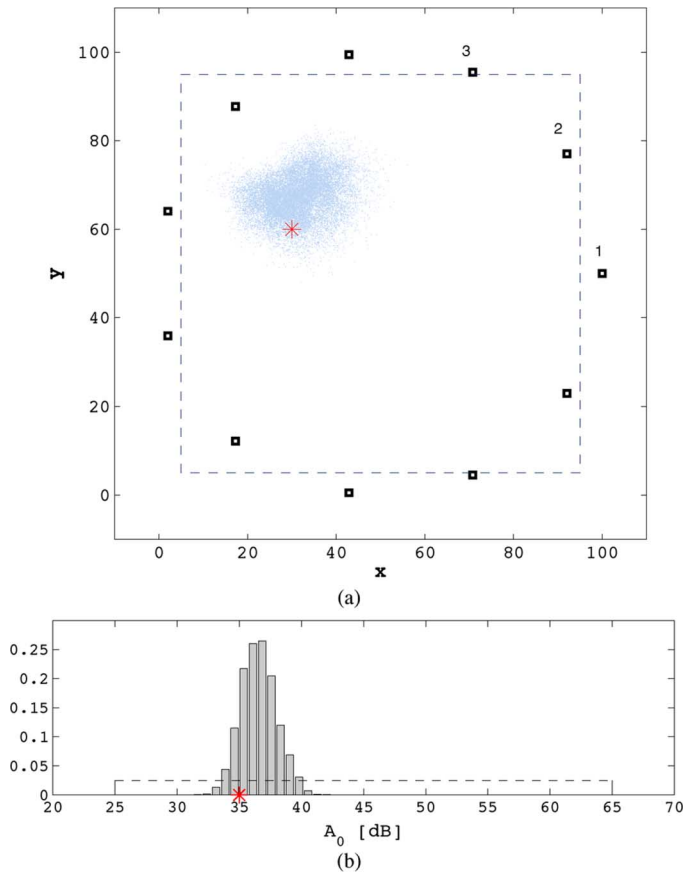

Fig. 2. (a) Estimation setup: 12 sensors, marked by □, placed on a circle; the source location (marked by asterisk) is (30, 60); dashed lines mark the limits of the prior in the (x_0, y_0) space; the cloud of dots represents the random sample approximation of the posterior PDF in the (x_0, y_0) space, $\pi(x_0, y_0|\mathbf{z}_{1:N})$; (b) The histogram of random samples approximating the posterior PDF $\pi(A_0|\mathbf{z}_{1:N})$; the prior PDF $\pi_0(A_0)$ is shown by a dashed line; the true value is $A_0 = 35$.

Few observations can be made from this simple example. First, the Bayes estimator has reduced the uncertainty (the support of the posterior PDF is much smaller than the support of the prior). Second, the true parameter $\mathbf{x}$ is included in the support of the posterior PDF. Third, the point Bayes estimates, such as the mean or the maximum of the posterior PDF, in general would be biased.

The Bayes optimality of the density estimator (15) guarantees that the resulting posterior is the most concentrated PDF whose support contains the true $\mathbf{x}$. Next we verify numerically that the support of the density estimator (15) indeed contains the true $\mathbf{x}$. For this we need the concept of a credible set ([2], p. 140).

V. Credible Set and Inclusion

A credible set $C(\alpha) \subseteq \mathcal{X}$ associated with the posterior PDF is defined via $\int_{C(\alpha)} \pi(\mathbf{x}|\mathbf{z}_{1:N})\, d\mathbf{x} = \alpha$, where $\alpha \in [0, 1]$ is the probability. A credible set at $\alpha \to 1$ represents the support of $p(\mathbf{x}|\mathbf{z}_{1:N})$. Inclusion criterion $\rho(\alpha)$ for the posterior PDF is defined as: $\rho(\alpha) = 1$ if true parameter vector $\mathbf{x} \in C(\alpha)$, and zero otherwise. Next we want to establish via Monte Carlo simulations the percentage of time the inclusion criterion is satisfied

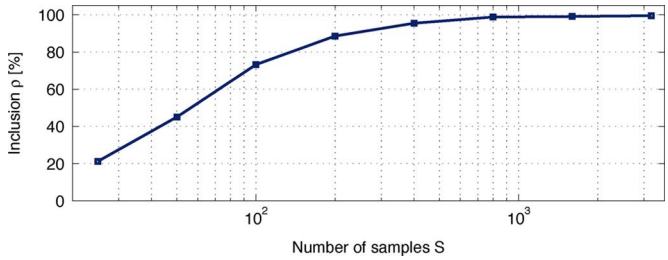

Fig. 3. Percentage of 1000 Monte Carlo runs in which the inclusion was satisfied, shown as a function of the number of samples S.

for the proposed Bayes estimator. For this we use the simulation setup described above, with one difference: the propagation loss factors are now random in each run and for each sensor: $\theta_i \sim \mathcal{U}_{[2,4]}(\theta)$, $i = 1, \ldots, N$. The kernel density estimation (KDE) method [17] is used to establish if the inclusion is satisfied in each run. Assuming the posterior PDF is represented by equally weighted samples $\{\mathbf{x}_j^*, w_j = (1)/(S); j = 1, \ldots, S\}$, the KDE approximation of the posterior PDF is [17]:

$$\pi(\mathbf{x}|\mathbf{z}_{1:N}) \approx \tilde{\pi}(\mathbf{x}) = \frac{1}{SW^{n_x}} \sum_{j=1}^{S} \phi\left(\frac{\mathbf{x} - \mathbf{x}_j^*}{W}\right) \tag{17}$$

where $\phi(\mathbf{x})$ is the kernel and W is the kernel width parameter. For convenience $\phi(\mathbf{x})$ is adopted to be a Gaussian PDF with zero-mean and covariance matrix $\mathbf{P}$. The optimal fixed bandwidth (under the assumption that the underlying PDF is Gaussian) for the Gaussian kernel $\phi(\mathbf{x})$ is [17]: $W^* = A \cdot S^{(1)/(n_x+4)}$, where $A = [4/(n_x + 2)]^{(1)/(n_x+4)}$. The covariance $\mathbf{P}$ is estimated as a sample covariance. Using the KDE approximation (17) it would be possible to estimate the boundary of the credible set $C(\alpha \to 1)$. The computation involved, however, would be prohibitively expensive, and we adopt a simpler approximation of the inclusion criterion as follows:

$$\widehat{\rho(1)} = \begin{cases} 1, & \text{if } \tilde{\pi}(\mathbf{x}) \geq \min\limits_{j=1,\ldots,S} \tilde{\pi}(\mathbf{x}_j^*) \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

where $\mathbf{x}$ is the true value of the parameter vector and $\tilde{\pi}(\mathbf{x})$ was defined in (17). The value of $\min_{j=1,\ldots,S} \tilde{\pi}(\mathbf{x}_j^*)$ in (18) effectively approximates the boundary of the credible set at $\alpha \to 1$.

Fig. 3 shows the percentage of 1000 Monte Carlo runs in which the inclusion was satisfied $(\widehat{\rho(1)} = 1)$, as a function of the number of samples S. Observe that as S is increased, and therefore the Monte Carlo approximation of the posterior PDF is more accurate, the inclusion percentage grows. For a sufficiently good approximation, i.e., for S above a certain threshold, the inclusion is guaranteed (i.e., satisfied 100%).

VI. Summary

In many practical applications of statistical signal processing, the likelihood functions are only partially known. Adopting the random set theoretical approach developed by Mahler [1], the paper presented the optimal Bayesian estimator for this case. The key role played the generalized likelihood function, which captured two sources of uncertainty, the stochastic uncertainty and imprecision. The optimal Bayesian estimator for partially known likelihoods in general cannot be found in the closed-form, however Monte Carlo methods can be applied easily and with a reasonable computational expense. The optimal Bayesian estimator for imprecise likelihoods was illustrated by a numerical example where the problem was to localize an emitting source in the presence of partially known propagation losses. It has been verified by Monte Carlo simulations that the support of an accurate approximation of the posterior PDF is guaranteed to contain the true value of the parameter vector.

References

[1] R. Mahler, *Statistical Multisource Multitarget Information Fusion*. Norwell, MA: Artech House, 2007.
[2] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer, 1985, Series in Statistics.
[3] N. Patwari, J. N. Ash, S. Kyperountas, R. L. M. A. O. Hero III, and N. S. Correal, "Locating the nodes," *IEEE Signal Process. Mag.*, pp. 54–68, July 2005.
[4] P. Walley, *Statistical Reasoning With Imprecise Probabilities*. London, U.K.: Chapman & Hall, 1991.
[5] B. Noack, V. Klumpp, D. Brunn, and U. D. Hanebeck, "Nonlinear Bayesian estimation with convex sets of probability densities," in *Proc. 11th Int. Conf. Information Fusion (Fusion 2008)*, Cologne, Germany, July 2008, pp. 1–8.
[6] A. Benavoli, M. Zaffalon, and E. Miranda, "Robust filtering through coherent lower previsions," *IEEE Trans. Automat. Contr.*, 2011, to be published.
[7] P. Smets, "Imperfect information: Imprecision and uncertainty," *Uncertainty Manag. Inf. Syst.*, pp. 225–254, 1996.
[8] G. J. Klir and M. J. Wierman, *Uncertainty-Based Information: Elements of Generalized Information Theory*, 2nd ed. New York: Physica-Verlag, 1999.
[9] L. Jaulin, "Computing minimal-volume credible sets using interval analysis: Application to Bayesian estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3632–3636, 2006.
[10] I. R. Goodman, R. P. S. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*. Norwell, MA: Kluwer, 1997.
[11] I. Molchanov, *Theory of Random Sets*. New York: Springer, 2005.
[12] , R. R. Yager and L. Liu, Eds., *Classic Works of the Dempster Shafer Theory of Belief Functions*. New York: Springer, 2008.
[13] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer, 2004.
[14] J. Figueiras and S. Frattasi, *Mobile Positioning and Tracking*. Hoboken, NJ: Wiley, 2010.
[15] , A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001.
[16] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwell, MA: Artech House, 2004.
[17] B. W. Silverman, *Density Estimation for Statistical and Data Analysis*. London, U.K.: Chapman & Hall, 1986.