DISTRIBUTED BAYESIAN ESTIMATION OF ARRIVAL RATES IN ASYNCHRONOUS MONITORING NETWORKS

Angelo Coluccia and Giuseppe Notarstefano

Department of Engineering, Università del Salento
{angelo.coluccia, giuseppe.notarstefano}@unisalento.it

ABSTRACT

In this paper we consider a network of agents monitoring a spatially distributed traffic process. Each node measures the number of arrivals seen at its monitoring point in a given time-interval. We propose an asynchronous distributed approach based on a hierarchical Bayes model with unknown hyperparameter, which allows each node to compute the minimum mean square error (MMSE) estimator of the local arrival rate by suitably fusing the information from the whole network. Simulation results show that the distributed scheme improves the estimation accuracy compared to a purely decentralized setup and is reliable even in presence of limited local data. An ad-hoc algorithm with reduced complexity is also proposed, which performs very closely to the optimal MMSE estimator.

Index Terms — distributed estimation, Empirical Bayes, consensus, traffic network, monitoring.

1. INTRODUCTION

Rate estimation in arrival processes, like those arising in traffic networks, is an important problem with several practical applications. Typical scenarios are the estimation of traffic flows in packet-switched networks [1], both wired and wireless, and more in general queue analysis [2]. Recently, increasing attention has been devoted to Intelligent Transportation Systems (ITS) for smart cities [3]. In several application scenarios, like e.g. traffic jam avoidance, safety at road intersections, etc., networks of sensor devices are used to collect information concerning the operating environment [4, 5].

Distributed estimation has received a widespread attention in the distributed computation literature, especially as a natural application of linear (average) consensus algorithms, see, e.g. [6] for a recent survey. Nodes typically perform local estimation based on local data, and then interact iteratively with their neighbors to carry out the estimation task [7–9]. In alternative (dynamic) methods, the nodes keep collecting new measurements while interacting with each other [10, 11].

Maximum-Likelihood (ML) approaches for distributed estimation have been considered for parameter estimation [12]. In [13, 14] consensus-based algorithms have been developed to compute the solution of a ML optimization problem. In [15] consensus-based algorithms for parameter estimation have been developed in a linear Bayesian framework. In [7] a distributed Alternating Direction Method of Multipliers (ADMM) has been developed for distributed ML estimation of vector parameters in a wireless sensor network. In [16] we have introduced a distributed estimator for binary event probabilities, based on a hierarchical Bayesian approach and a dual decomposition distributed optimization algorithm.

Following the idea in [16], we propose a distributed estimation scheme based on the Empirical Bayes approach, in which the arrival rates are treated as random variables whose prior distribution is parametrized by a suitable unknown hyperparameter. We show that the ML estimator of the hyperparameter is obtained by a separable optimization problem that can be solved in a distributed way over the network. Moreover, we propose an alternative ad-hoc distributed estimator that, although suboptimal, performs comparably to the optimal one. The resulting algorithm has a simple updated rule based on linear consensus protocols and exhibits the same appealing convergence properties as consensus protocols.

2. MONITORING NETWORK AND ARRIVAL-RATE ESTIMATION PROBLEM

We consider a network of monitors having sensing, communication and computation capabilities. Each monitor can measure the number of arrivals in a given fixed timescale interval (e.g., 1 second or 1 minute), share local data with neighboring agents, and perform local computations on its own and its neighbors’ data. The objective is to fuse the data in order to improve the estimation of the arrival rates. Each node collects measurements asynchronously in a certain period of time, i.e. in an observation window, over which the underlying process can be assumed to be stationary. Accordingly, for each monitor \( i \in \{1, \ldots, N\} \) we introduce the following variables:

- \( \lambda_i \) unknown arrival rate
- \( y_{i,\ell} \) the \( \ell \)-th measurement of the number of arrivals in the observation window, i.e. \( y_{i,\ell} | \lambda_i \sim \text{Poisson}(\lambda_i) \) i.i.d.
- \( n_i \geq 1 \) number of measurements \( y_{i,\ell} \) collected in the observation window.
We assume that the network evolution is triggered by a universal slotted time, \( t \in \mathbb{Z}_{\geq 0} \), not necessarily known by the monitors. The monitors communicate according to a time-dependent directed communication graph \( t \mapsto G(t) = (\{1, \ldots, N\}, E(t)) \), where \( \{1, \ldots, N\} \) are the monitor identifiers and the edge set \( E(t) \) describes the communication among monitors; \((i,j) \in E(t)\) if monitor \( i \) communicates to \( j \) at time \( t \in \mathbb{Z}_{\geq 0} \). For each node \( i \), the nodes sending information to \( i \) at time \( t \), i.e., the set of \( j \in N \) such that \((i,j) \in E(t)\) are called in-neighbors of \( i \) at time \( t \). The set of in-neighbors of \( i \) at time \( t \) is denoted by \( N_i(t) \). We make the following minimal assumption on the graph connectivity:

**Assumption 2.1 (Uniform joint strong connectivity).** There exists an integer \( Q \geq 1 \) such that the graph \((\{1, \ldots, N\}, u_{\tau=t}^{(t+1)Q-1} E(\tau)) \) is strongly connected \( \forall t \geq 0 \).

### 3. DISTRIBUTED ARRIVAL-RATE ESTIMATION VIA EMPIRICAL BAYES

In a decentralized set-up, in which nodes do not communicate, each node could estimate \( \lambda_i \) based on the sample \( \{y_{i,\ell}, \ell = 1, \ldots, n_i\} \) simply computing the empirical mean of the available measures. That is, the decentralized estimator is \( \hat{\lambda}_{r,l} = \frac{1}{n_i} \sum_{\ell=1}^{n_i} y_{i,\ell} = \frac{\sigma_{i,r}}{n_i} \), where \( \sigma_{i,r} = \sum_{\ell=1}^{n_i} y_{i,\ell} \) is the cumulative sum of the measurements at node \( i \).

Notice that, the decentralized estimator turns out to be the Maximum Likelihood estimator of \( \lambda_i \) when node \( i \) can use only its own data. However, decentralized estimation yields reliable estimates only when the number of samples \( n_i \) is large enough. In our heterogeneous set-up, it may happen that some nodes satisfy such a condition, while other ones do not have enough data, thus resulting in a poor estimation. In this paper we propose a distributed estimation scheme in which every node, especially the ones with fewer measurements, can take advantage from communicating with neighboring nodes.

#### 3.1. Empirical Bayes approach in monitoring networks

In applying a Bayesian estimation approach to a network context, the assumption that the prior is known to all monitors may be too strong. To avoid this drawback we adopt the Empirical Bayes approach in which only the class of the prior is known, while the parameters need to be estimated.

Formally, in our network set-up, we assume that each \( \lambda_i \) is also a random variable. In particular, the \( \lambda_i \)s are independent identically distributed (i.i.d) Gamma variables, i.e., \( \lambda_i \sim \text{Gamma}(a, b) \), where the shape parameter \( a \) is known, while the scale parameter \( b \) is unknown.

The hyperparameter \( b \) can be estimated via a Maximum Likelihood (ML) procedure. To this aim, it is necessary to derive the joint distribution of all the measures \( \{y_{i,\ell}\}_{\ell=1}^{n_i} \) for each agent \( i \). Since these are independent, the likelihood function for the estimate of the hyperparameter \( b \) is the product of the marginal distributions of all agents:

\[
L(y_1, \ldots, y_N | b) = \prod_{i=1}^{N} p(y_i | b) \tag{1}
\]

where \( y_i = [y_{i,1} \cdots y_{i,n_i}]^T \). The marginal distribution of agent \( i \) is derived from the joint distribution of \( y_i \) and \( \lambda_i \).

\[
p(y_i | b) = \int_{0}^{\infty} \left( \prod_{\ell=1}^{n_i} f(y_{i,\ell} | \lambda_i, b) \right) p(\lambda_i | b) d\lambda_i
\]

By using eq. (2) into eq. (1) the likelihood is rewritten as:

\[
L(y_1, \ldots, y_N | b) \propto 1^{b \cdot N} \prod_{i=1}^{N} \left( \frac{b}{n_i b + 1} \right)^{\sigma_{i,a}} \tag{3}
\]

from which the ML estimator \( \hat{b} \) of \( b \) can be found by solving the following optimization problem:

\[
\hat{b} = \arg \min_{b \in \mathbb{R}_+} \sum_{i=1}^{N} \left( a N \log b - (\sum_{i=1}^{N} \sigma_i + a) \log \left( \frac{b}{n_i b + 1} \right) \right) \tag{4}
\]

which can be shown to be strictly convex.

The problem can be solved in closed-form only for the homogeneous case where all \( n_i \)'s are equal. This means that \( n_i = n/N \), being \( n \) the total number of measurements. In this case the ML estimator of \( b \) based on the entire set of measurements is given by

\[
\hat{b}_\text{hom} = \frac{1}{an} \sum_{i=1}^{N} \sigma_i = \frac{\sigma}{an} \tag{5}
\]

where \( n \overset{\text{def}}{=} \sum_{i=1}^{N} n_i \), and \( \sigma \overset{\text{def}}{=} \sum_{i=1}^{N} \sigma_i \).

After obtaining an estimate for \( b \), the Empirical Bayes estimator of the arrival rate \( \lambda_i \) that minimizes the Mean Square Error (MMSE) can be obtained by computing the conditional mean of the posterior distribution \( p(\lambda_i | y_i, b) \). The latter is given by the ratio between the joint pdf \( p(y_i, \lambda_i | b) \) and the marginal pdf \( p(y_i | b) \) as from eq. (2), i.e.:

\[
p(\lambda_i | y_i, b) = \frac{\lambda_i^{\sigma_i+a-1} e^{-\lambda_i (n_i b + 1)} \Gamma(\sigma_i + a)}{\Gamma(\sigma_i + a)} \left( \frac{b}{n_i b + 1} \right)^{-\sigma_i - a} \tag{6}
\]

that our model is more general than the classical Gamma-Poisson hierarchy, since the \( n_i \)'s are different (non-homogeneous sample).

---

\(^1\)The assumption that \( a \) is known, while only \( b \) is unknown, says that only the shape of the Gamma distribution (determined by the parameter \( a \)) is known, while the scaling is not. This assumption is reasonable in many applications, since it is a way to embed a rough information on the phenomenon, and is customary for the sake of mathematical tractability [17]. Notice though...
The distributed estimation algorithm is as follows. At each iteration, the chosen distributed optimization algorithm guarantees that all nodes reach consensus for the parameters of the chosen distributed optimization algorithm, and Eq. (6) is a Gamma pdf with parameters \((\lambda_i + a, b/(\eta_i + 1))\), hence the Empirical Bayes MMSE estimator of each \(\lambda_i\) is

\[
\hat{\lambda}_i = E[\lambda_i | y_i, \hat{b}] = \frac{\hat{b}}{\eta_i + 1 + (a + \lambda_i)}. \tag{7}
\]

In the following we will also consider an ad-hoc estimator obtained by using \(\hat{b}\) instead of the optimal \(\hat{b}\), i.e.:

\[
\hat{\lambda}_i^{\text{ad-hoc}} = \frac{\hat{b} \cdot \hat{b}_i}{n_\text{hom}(\lambda_i + 1)} (a + \lambda_i) = \frac{\sigma}{an + \sigma n_i} (a + \lambda_i).
\]

### 3.2. Distributed Empirical Bayes estimator

From eq. (7) it is clear that each agent can compute the Empirical Bayes MMSE estimator provided it knows \(\hat{b}\). The optimization problem (4) giving the ML estimator of \(b\) has a separable cost function (i.e., the total cost is the sum of \(N\) local costs), hence it can be solved by using available distributed optimization algorithms for unconstrained optimization. An example of algorithm working on general asynchronous networks (under Assumption 2.1 of uniform joint strong connectivity) is [18]. We assume that each node implements the local update rule of the chosen distributed optimization algorithm (e.g., equation (1) in [18]). We let each node \(i\) have a local state \(x_i\) and an estimate \(\hat{b}_i\) of \(\hat{b}\). At each \(t \in \mathbb{N}\) the node runs the local update rule

\[
(\hat{b}_i(t+1), x_i(t+1)) = \text{update}_\text{local} \left(\hat{b}_i(t), x_i(t), \{x_j(t)\}_{j \in N_i^a(t)}; \gamma(t)\right),
\]

where \(\{x_j(t)\}_{j \in N_i^a(t)}\) is the collection of states of the in-neighbors of node \(i\), update\_local is the local update of the chosen distributed optimization algorithm, and \(\gamma\) is an algorithm parameter as, e.g., a step-size. The distributed optimization algorithm guarantees that all node reach consensus on the minimizer of (4). That is,

\[
\lim_{t \to \infty} \hat{b}_i(t) = \hat{b}, \quad \text{for all } i \in \{1, \ldots, N\}.
\]

The distributed estimation algorithm is as follows. At each \(t \in \mathbb{N}\), each agent \(i\) stores a local state \(x_i(t)\), an estimate \(\hat{b}_i(t)\) of \(\hat{b}\) and an estimate \(\hat{\lambda}_i(t)\) of \(\hat{\lambda}_i\). The node initializes its local state \(x_1\) according to the chosen distributed optimization algorithm, sets \(\hat{b}_i(0) = \sigma_i/(an_i)\) (which would be the solution of (4) if \(i\) were the only agent) and \(\hat{\lambda}_i(0) = \hat{b}_i(0)/n_\text{hom}(\lambda_i + 1)\). Then it updates its estimate of \(\hat{b}\) by using (8) and updates the current estimate \(\hat{\lambda}_i(t)\) by using (7).

From the convergence properties of the chosen distributed optimization algorithm it follows immediately that the proposed distributed estimator asymptotically computes at each node \(i\) the Empirical Bayes MMSE estimator of \(\lambda_i\). However, most of the available distributed optimization algorithms, as the ones in [18, 19], need the tuning of a global parameter (we denoted it \(\gamma\)). Also, typically they exhibit a sub-exponential convergence even in static graphs. To overcome these drawbacks we propose an alternative ad-hoc distributed estimator that, although suboptimal, performs comparably to the optimal one. The algorithm is defined as follows.

For each \(t \in \mathbb{N}\), each node \(i \in \{1, \ldots, N\}\) stores in memory two local states \(s_i(t)\) and \(\eta_i(t)\), an estimate \(\hat{b}_i^\text{hom}(t)\) of \(\hat{b}\), and an estimate \(\hat{\lambda}_i^\text{ad-hoc}(t)\) of \(\hat{\lambda}_i\). For \((i, j) \in E(t)\) let \(w_{ij}(t) \in \mathbb{R}_+\) be a set of weights, satisfying \(\sum_{i=1}^n w_{ij}(t) = 1\).

**Initialization:** \(s_i(0) = \sigma_i, \eta_i(0) = n_i, \hat{b}_i^\text{hom}(0) = \sigma_i/(an_i), \hat{\lambda}_i^\text{ad-hoc}(0) = \hat{b}_i^\text{hom}(0)/(an_i + 1)\). 

**Iterate:**

\[
\begin{align*}
\hat{s}_i(t+1) &= \sum_{j \in N_i^a(t) \cup \{i\}} w_{ij}(t) s_j(t) \\
\hat{n}_i(t+1) &= \sum_{j \in N_i^a(t) \cup \{i\}} w_{ij}(t) n_j(t) \\
\hat{b}_i^\text{hom}(t+1) &= \frac{s_i(t+1)}{\eta_i(t+1)} \\
\hat{\lambda}_i^\text{ad-hoc}(t+1) &= \frac{\hat{b}_i^\text{hom}(t+1)}{\hat{b}_i^\text{hom}(t+1)/(an_i + 1) + (a + \sigma_i)}.
\end{align*}
\]

**Proposition 3.1.** Assume that Assumption 2.1 holds. Then the distributed algorithm (9) computes the ad-hoc estimator, i.e.,

\[
\lim_{t \to \infty} \hat{b}_i^\text{hom}(t) = \hat{b}_i^\text{hom}, \quad \lim_{t \to \infty} \hat{\lambda}_i^\text{ad-hoc}(t) = \hat{\lambda}_i^\text{ad-hoc}.
\]

Following [20], the proof relies on: (i) bounding \(\max_{i \in \{1, \ldots, N\}} |\hat{b}_i^\text{hom}(t) - \hat{b}| \leq \max_{i \in \{1, \ldots, N\}} |\hat{b}_i^\text{hom}(0)| f(t)\), where \(f(t)\) is a decreasing function of time, and (ii) showing the weak ergodicity of the weight matrices, by using Assumption 2.1. The proof is omitted due to lack of space.

**Remark 3.2.** The distributed algorithm (9) exhibits the exponential convergence of linear consensus protocols, as opposed to, for example, the \(O((\log t)/\sqrt{t})\) rate of [18].

**Remark 3.3.** We want to stress that \(t\) is a universal time that the monitors do not need to know to implement the estimation algorithm above. Thus, the algorithm is well-suited for a completely asynchronous implementation. Simply, it is enough to assume that when node \(i\) is not active, then there are no edges going out nor coming in agent \(i\) (i.e., \(N_i^a(t) = \emptyset\)).

### 4. PERFORMANCE ANALYSIS

The Cramer-Rao (lower) Bound (CRB) for the estimation of \(b\) turns out to be \(\text{CRB} = \frac{b^2}{\sum_{i=1}^n n_i b_i^2}\), which for \(n_i = n/N\) becomes \(\text{CRB}^\text{hom} = \frac{n b}{n b + 1}\). The variance of the homo-
A formal theoretical analysis of the estimator performance is much simpler to implement and exhibits faster convergence. A more insightful performance assessment is reported in Fig. 2. The root mean square error (RMSE) of the proposed ad-hoc estimator stays always very close to the optimal MMSE estimator. This is a merit of the particular, we have developed an ad-hoc distributed algorithm especially when the node local information is insufficient. In this work and generally outperforms the decentralized estimator, which performs closely to the optimal MMSE estimator, but is much simpler to implement and exhibits faster convergence.

Conversely, the decentralized estimator provides reliable estimates only for sufficiently large $n_i$ — several values are reported in the figure, with dotted horizontal lines highlighting the inability of $\hat{\lambda}_i^{\text{dec}}$ to gain information from the network. The figure shows that the estimation accuracy improves as more agents are available, since even a single measurement per-node will increase the aggregate $n$ hence, in turn, the performance of the distributed estimators.

5. CONCLUSION

In this paper we have proposed a novel distributed scheme, based on a hierarchical Bayes approach, for the estimation of arrival rates in asynchronous monitoring networks. The proposed distributed approach gains information from the network and generally outperforms the decentralized estimator, especially when the node local information is insufficient. In particular, we have developed an ad-hoc distributed algorithm that performs closely to the optimal MMSE estimator, but is much simpler to implement and exhibits faster convergence.

A formal theoretical analysis of the estimator performance is part of our ongoing work.

---

2 Notice that values (slightly) below the CRB can occur due to the possible (small) bias for finite sample size which may trade-off some variance, while the homogeneous estimator is always unbiased.
6. REFERENCES


