Data-driven prediction of EVAR with confidence in time-varying datasets

Citation: Axelrod, Allan, et al. "Data-Driven Prediction of EVAR with Confidence in Time-Varying Datasets." 2016 IEEE 55th Conference on on Decision and Control (CDC), 12-14 December, 2016, Las Vegas, Nevada, IEEE, 2016, pp. 5833–38.

As Published: http://dx.doi.org/10.1109/CDC.2016.7799166

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/114854

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Data-driven Prediction of EVAR with Confidence in Time-varying Datasets

Allan Axelrod[1], Luca Carlone[2], Girish Chowdhary[1], Sertac Karaman[2]

***Abstract*— The key challenge for learning-based autonomous systems operating in time-varying environments is to predict when the learned model may lose relevance. If the learned model loses relevance, then the autonomous system is at risk of making wrong decisions. The entropic value at risk (EVAR) is a computationally efficient and coherent risk measure that can be utilized to quantify this risk. In this paper, we present a Bayesian model and learning algorithms to predict the state-dependent EVAR of time-varying datasets. We discuss applications of EVAR to an exploration problem in which an autonomous agent has to choose a set of sensing locations in order to maximize the informativeness of the acquired data and learn a model of an underlying phenomenon of interest. We empirically demonstrate the efficacy of the presented model and learning algorithms on four real-world datasets.**

I. Introduction

Autonomous agents must select actions that are appropriate to the environment in which they operate. Yet, without precise knowledge of the environment, autonomous agents are at risk of selecting suboptimal actions. Many autonomy and adaptive control architectures utilize online learning to improve the agent's environmental model. However, in time-varying environments, learned models may lose relevance over time. The key challenge facing autonomy in time-varying environments is therefore to predict and quantify how quickly environmental models lose relevance, and to take appropriate data-gathering actions that minimizes the risk of making wrong decisions due to inaccurate environmental models. In order to address these challenges, we must first answer what uncertainty measure best captures changes in the environment and design data-driven stochastic models that define how this uncertainty measure evolves over time.

In this paper, our goal is to lay the modeling and algorithmic foundations for autonomous agents to predict the risk of making the wrong decision in the face of time variations. We use an exploration problem as a motivating example: our agent is operating in a field with N sensing locations which are measuring some underlying time-varying phenomenon of interest (e.g., temperature, rainfall). Due to sensing constraints, the agent can only acquire information from $\kappa < N$ of these sensing locations at each time step. Therefore, the goal of the agent is to learn a predictive model that can help identify the subset of sensing locations which are most likely to help the agent maintain an up-to-date model of the environment. If the agent could achieve this, then the risk of the agent overestimating in a time-varying environment can be *proactively* mitigated.

In quantifying predictive uncertainty, it is natural to use variance (VAR) or tail probabilities. While such measures describe an expected distance of a sample from the mean of the current stochastic model, neither address how the stochastic model changes as a result of new observations. By contrast, uncertainty measures such as the information gain [1], the entropic value at risk (EVAR) [2], and the conditional value at risk (CVAR) [3] all describe uncertainty in how the model will change due to new observations.

While both CVAR and EVAR provide an intuitive bound on how the expectation of our model changes due to new observations, EVAR is more computationally efficient than CVAR and is the tightest upper bound on CVAR [2]. As we will show, EVAR can also incorporate the information gain, and so we examine EVAR as a comprehensive measure of the uncertainty in our model.

Our main contribution is a Bayesian model and learning algorithms for predicting the spatio-temporal evolution of the EVAR (in our running example, this is our risk of over-estimation at each sensing location). We first present a generalization of the Poisson Exposure Distribution (Ped) [4], which we call the *Poisson exposure process* (Pep). Then, we use Pep to model the evolution of the information gain term in the EVAR.

The key advantage of the Pep is that it allows us to overcome the assumption of identically distributed samples utilized in algorithms such as Predicted Information Gain (PIG) [5]. We develop an EVAR-variant of our Real-time Adaptive Prediction of Time-varying and Obscure Rewards (RAPTOR) algorithm [6] called (EVAR-RAPTOR). We show that the EVAR-RAPTOR algorithm outperforms the EVAR exploration variant of PIG, as well as sequential and random search, in simulations over four real-world datasets.

II. Preliminaries

This section reviews the definition of information gain (Section II-A), which is used in the definition of the entropic value at risk (Section II-B).

A. Information Gain and Exploration

Exploration can be viewed as an information-collecting task, hence a means of quantifying the information collected is essential to guide decision-making for exploration. Exploration strategies conventionally use some form of the posterior variance [7] or the information-entropy [8] to quantify the information gained by a given action. While both strategies make intuitive sense, they are mathematically imprecise descriptors of the information gain which, due to

[1]Allan Axelrod and Girish Chowdhary are with the school of Mechanical and Aerospace Engineering at Oklahoma State University, Stillwater OK, {allanma,girish.chowdhary}@okstate.edu.

[2]Luca Carlone and Sertac Karaman are with the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology, Cambridge MA, {lcarlone,sertac}@mit.edu.

a theorem of uniqueness in information theory [1], is defined as

$$I(P(Y)||P_0(Y)) = \int\limits_{P(Y)\ll P_0(Y)} P(y)\ln\frac{P(y)}{P_0(y)}dy. \quad (1)$$

Here, Y is a random variable, $P_0(Y)$ is a given prior distribution, $P(Y)$ is the posterior distribution (after the exploration action is undertaken), and $P(Y) \ll P_0(Y)$ denotes that $P_0(Y=c)=0$ implies that $P(Y=c)=0$.

B. Entropic Value at Risk (EVAR) Risk Measure

A *risk measure* assigns a real value to a random variable Y that quantifies the uncertainty associated with a variable. Examples of risk measures include value-at-risk (VaR), variance (VAR), and Conditional Value at Risk (CVAR). Ahmadi-Javid [2] introduced the *Entropic Value at Risk* (EVAR) measure that addresses computational and coherency shortcomings of VAR and CVAR.

The EVAR of a real-valued random variable Y, with *confidence level* $1-\gamma$ (with $\gamma \in (0,1]$), is defined as

$$\mathrm{EVAR}_{1-\gamma}(Y) \doteq \inf_{\theta>0}\{\theta^{-1}\ln\left(\mathbb{E}_{P_0}\left[e^{\theta Y}\right]/\gamma\right)\}. \quad (2)$$

The importance of the EVAR lies in the fact that, with confidence level $1-\gamma$, it upper bounds the value of the posterior expectation [2]:

$$\mathrm{EVAR}_{1-\gamma}(Y) = \sup_{\substack{P(Y)\ll P_0(Y)\\ I(P(Y)||P_0(Y)\leq -\ln(\gamma)}} \{\mathbb{E}_P(Y)\}. \quad (3)$$

Therefore, predicting the EVAR would help to quantify the uncertainty in a time-varying environment.

We use the Donsker-Varadhan Variational formula in Lemma 3.1 of [2] to yield the equivalent, yet simpler, form of the EVAR:

$$\mathrm{EVAR}_{1-\gamma}(Y) = \sup_{\substack{P(Y)\ll P_0(Y)\\ I(P(Y)||P_0(Y)\leq -\ln(\gamma)}} \ln\left(\mathbb{E}_{P_0}\left[e^{Y}\right]\right) - \ln(\gamma). \quad (4)$$

III. Formulation: Exploration as Multi-Play N-Armed Restless Bandits

In this paper we consider the following exploration problem: there exists a set of N sensing locations in a spatial domain from which we may observe a time-varying phenomenon. We denote the set of all sensing locations as $\Omega \doteq \{1,\ldots,N\}$. In particular, each sensing location $i \in \Omega$ is associated to a stochastic process Y_i that we want to estimate. At each time t, sensor i observes a *sample* $y_i(t)$ of the process Y_i. However, due to resource constraints (e.g., cost), we cannot acquire samples from all the N sensors but only from a subset of them, i.e., we can only observe the outcome of $\kappa \leq N$ sensors, where κ is a given upper bound. Therefore, the problem we address in this paper is learning how to select the subset of Ω, having cardinality κ, where the phenomena Y_i exhibit the greatest change, as quantified by EVAR. Note that while for clarity of presentation we tailor our problem formulation to a sensing task, the formulation can be extended to model risk-aware applications in mobile robotics, or asset selection in portfolio optimization.

A. Multi-Play N-Armed Restless Bandit Formulation

In our formulation, we focus on the exploration task, where only κ of $N \geq \kappa$ sensing locations may be sampled at any given time instant t (episode) within the sensor set Ω. Feedback from the selected subset of sensors $\mathcal{S}^\star(t) \subset \Omega$ is used to update the models of each time-varying process Y_i. We also associate a second stochastic process to each sensing location, which models the evolution of the information gain at each location. We denote this information gain process with the symbol Z_i, for all $i \in \Omega$. For simplicity, each sensor location $i \in \Omega$ is assumed to be statistically independent.

Similar to previous work [6], [9]–[12] we model the exploration problem as a multi-play N-armed restless bandit problem. In this formulation, each *arm* of the bandit corresponds to a sensor $i \in \Omega$. Therefore, the overall goal is to select the best set of κ bandit arms such that some measure on exploration reward is optimized. We call $\mathcal{S}^\star(t)$ the set of sensing locations (arms) selected at time t.

The key difference with respect to our previous work is that we learn to select the subset of arms with the most entropic value at risk (EVAR), rather than selecting the subset with the largest information gain [6], [9]–[11]. Intuitively, since the sensing locations observe time-varying phenomena, the EVAR at a given location changes over time. Therefore, our goal is to use a data-driven approach that can dynamically learn which sensing location is best to sample.

More formally, the objective of our exploration policy is to maximize the total EVAR obtained in each episode of the exploration task. This can be achieved by visiting the sensing locations which are expected to have the most EVAR:

$$\begin{aligned}\mathcal{S}^\star(t) \doteq \quad & \underset{\mathcal{S}\subset\Omega}{\operatorname{argmax}} \quad \textstyle\sum_{i\in\mathcal{S}} \mathbb{E}\left[\mathrm{EVAR}_{1-\gamma_i}(Y_i)\right] \\ & \text{subject to} \quad \mathrm{Card}(\mathcal{S}) = \kappa,\end{aligned} \quad (5)$$

where $\mathrm{Card}(\cdot)$ denotes the cardinality of a set. In Section IV we go one step further and we provide probabilistic bounds that assess the quality of our EVAR estimate at each location. This allows discussing when the "informed policy" of eq. (5) is expected to perform better than a naive policy (e.g., random or sequential sensor selection).

B. Data-Driven EVAR

In this section we provide a computational model for the EVAR and write the informed policy (5) in a more explicit form. Conventionally, the EVAR is calculated using an assumed confidence level γ. However, for time-varying environments, it is critical that a data-driven confidence level be implemented. Intuitively, the more data we collect, the more accurately we can predict the posterior. Therefore, we set the confidence level to be

$$\gamma^\star \doteq e^{-I(P(Y)||P_0(Y))} \leq \gamma, \quad (6)$$

based on the lower-bound in (3). Inserting $\gamma^\star$ (3) yields:

$$\mathrm{EVAR}_{1-\gamma^\star}(Y) = \ln\left(\mathbb{E}_{P_0}\left[e^{Y}\right]\right) - \ln(\gamma^\star) \leq \mathrm{EVAR}_{1-\gamma}(Y). \quad (7)$$

Substituting the expression of $\gamma^\star$, $\mathrm{EVAR}_{1-\gamma^\star}$ simplifies to:

$$\mathrm{EVAR}_{1-\gamma^\star}(Y) = \ln\left(\mathbb{E}_{P_0}\left[e^{Y}\right]\right) + I(P(Y)||P_0(Y)). \quad (8)$$

It is now clear that, in order to estimate the $\mathrm{EVAR}_{1-\gamma^\star}$, we need to learn a model of the information gain $I(P(Y)||P_0(Y))$. In the following section we show how to model the information gain as a Lévy process $Z = I(P(Y)||P_0(Y))$; since in our case we have N processes (one for each sensing location), we define $Z_i = I(P(Y_i)||P_0(Y_i))$, and, we rewrite the policy (5) as:

$$\begin{aligned} \mathcal{S}^\star(t) \doteq \quad & \underset{\mathcal{S}\subset\Omega}{\operatorname{argmax}} \quad \mathbb{E}\Big[\textstyle\sum_{i\in\mathcal{S}}\left(Z_i + \ln\left(\mathbb{E}_{P_0}(e^{Y_i})\right)\right)\Big] \\ & \text{subject to} \quad \mathrm{Card}(\mathcal{S}) = \kappa, \end{aligned} \tag{9}$$

C. Modeling the Information Gain

The data received as feedback $y_i \sim Y_i$ from the selected subset of sensors, $\mathcal{S}^\star(t)$, informs our model on the data process Y_i. Similarly, the information gain z_i of the feedback y_i informs our model on the information process Z_i. Formally, we use the following hierarchical model

$$\begin{aligned} y_i &\sim Y_i \\ z_i &\sim Z_i|y_i\,. \end{aligned} \tag{10}$$

We model Z_i as a stochastic random variable to form a predictive model on what the information gain of our next sample will be at location i. In Section III-C.2 we describe the stochastic model that we use to describe the evolution of Z_i. Our model is based on the Poisson Exposure Distribution (Ped), which we recall in Section III-C.1.

1) Poisson Exposure Distribution (Ped) Model: The Ped was introduced by [4] to model a continuous and monotonically increasing output resulting from a Poisson-distributed input. Since the Ped has a similar domain to the information gain and its expectation equivalently represents the average information gain used in [5], the Ped will be integral in forming a comparable baseline algorithm.

Definition 1. *The probability density function of the Ped is defined as*

$$f(z) = C_\Lambda \frac{\Lambda^z e^{-\Lambda}}{\Gamma(z+1)}, \tag{11}$$

where $\Lambda > 0$ is the distribution mean, C_Λ is the normalizing constant, and $z \in \mathbb{R}$ is the random variable.

Moreover, the Ped has an analytical Bayesian update.

Fact 1. *The Gamma distribution is a conjugate prior of the Ped such that*

$$G\left(\Lambda^*|\alpha+z,\beta+1\right) \propto Ped\left(z|\Lambda\right) G\left(\Lambda|\alpha,\beta\right). \tag{12}$$

where $G(\Lambda|\alpha,\beta)$ is the Gamma distribution with shape parameter α and rate parameter β.

Proof. See Proposition 1 in [6] for details. □

2) Poisson Exposure Process (Pep) Model: Since we are interested in modeling a time-varying information gain at each sensing location in our environment, we must generalize the Ped to incorporate time as a dependent argument. The hidden Markov model for the Ped as proposed in [4] parallels our problem where we cannot observe so-called informative events arriving at each location. However, we depart from [4] to extend the Ped to model phenomena that are dependent upon non-constant time intervals.

Definition 2. *The probability density function of the Pep is defined as*

$$f(z|\Lambda(t)) = C_{\Lambda(t)} \frac{(\Lambda(t))^z e^{-\Lambda(t)}}{\Gamma(z+1)}, \tag{13}$$

where $\Lambda(t)$ is the time-varying mean, $C_\Lambda(t)$ is the normalizing constant, and $z \in \mathbb{R}$ is the random variable. A Pep is called homogeneous if $\frac{d}{dt}\Lambda(t) = \lambda\ \forall\ t$, otherwise the Pep is called inhomogeneous.

In [4], the maximum likelihood estimate of the Poisson exposure distribution, and therefore the Pep by extension, was found to be highly nonlinear. However, as Definition 2 mathematically is similar in form to the Poisson process in (13), we show that the conjugate prior of the Pep is identical to that of the Poisson process as shown in Fact 2.

Fact 2. *The gamma distribution is a conjugate prior of the homogeneous Poisson exposure process (Pep) such that*

$$G\left(\lambda^* t|\alpha+z,\beta+t\right) \propto Pep\left(z|\lambda t\right) G\left(\lambda t|\alpha,\beta\right). \tag{14}$$

Proof. See Corollary 5.1 in [6] for details. □

IV. Algorithms and Probabilistic Guarantees

While Facts 1 and 2 provide a simple analytical update for the homogeneous Pep, we still need to demonstrate conditions when the homogeneous Pep model will be provably accurate. Until we have a provably accurate regression, we cannot rely on information-driven (informed) exploration using the Pep. Hence, we adopt the strategy of uninformed-to-informed exploration, where we start performing the exploration task with a sequential search strategy (an uninformed policy) and transition to informed policy once provable accuracy guarantees are available.

To prove probabilistic accuracy guarantees, we leverage the Chebyshev and Bienaymé-Chebyshev inequalities. The guarantees offered by the Bienaymé-Chebyshev inequality are desirable, namely that we now have error bounds on our next sample. Moreover, the independent and identically distributed condition is satisfactory for time-invariant distributions such as the Ped, which we show in Section IV-A. However, we are also interested in error bounds for our next sample in a time-varying environment. Therefore, we theoretically derive incremental variable changes (which are ideally suited for the homogeneous Pep) for the Chebyshev inequality, hence relaxing the assumption of identically distributed samples.

A. The Time-invariant Case: Ped-based Exploration

The regression in the Predicted Information Gain (PIG) algorithm [5] calculates the information gain for a sensing location i as the expectation of the random variable Z_i which models our belief over the information gain at each location:

$$\mathbb{E}[Z_i] = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_i(t_j) \tag{15}$$

where (15) is a time-invariant expectation. As only the expectation in (15) is used to drive the PIG algorithm to explore, we may equivalently use the Ped regression from Fact 1 in place of (15). To provide a probabilistic bound for the Ped analytical update shown in Fact 1, we leverage the Bienaymé-Chebyshev inequality.

Theorem 3 (Chebyshev Inequality for PIG, $k = \Lambda$)**.** *Let $Z(t_1), ..., Z(t_n)$ be independent and identically distributed Poisson exposure distribution trials. Let $\bar{Z}(t_n) = \frac{1}{n}\sum_{j=1}^{n} Z(t_j)$ and $\Lambda = \mathbb{E}[\bar{Z}(t_n)]$. Then,*

$$Pr\left(|\bar{Z} - \Lambda| \geq \Lambda\right) \leq \frac{1}{n\Lambda}. \tag{16}$$

Proof. This proof proceeds along the line of the proof of Theorem 6 in Appendix A, but where $k = \lambda$ and noting that $\beta = n$. □

Using the results from Theorem 3, we develop an uninformed-to-informed exploration bound for PIG in Corollary 4.

Corollary 4 (Accuracy Bound for PIG)**.** *If the environment statistics are time-invariant, then, given an accuracy threshold $0 < c \leq 1$, it holds $Pr\left(|\bar{Z} - \Lambda| \geq \Lambda\right) < c$ if the following condition is satisfied:*

$$n_b > \frac{1}{\Lambda_b c}, \qquad \textit{with} \qquad b = \underset{i}{\operatorname{argmax}} \frac{1}{n_i \Lambda_i}, \tag{17}$$

where n_i is the number of samples at location i and Λ_i is the Ped mean.

Proof. This proof proceeds along the line of the proof of Corollary 7 in Appendix B, but where the left hand side is $Pr(|\bar{Z} - \Lambda| \geq \Lambda)$ and $\beta = n$. □

Algorithm 1 EVAR-PIG*

Input: sensor set $\Omega = \{1, \ldots, N\}$, subset size κ
Initialize $(\alpha_i, \beta_i, \Lambda_i, n_i) \leftarrow 0 \ \forall \ i$
for each time t and for each sensor i **do**
  $\mathbb{E}[\text{EVAR}_{1-\gamma_i^*}] \leftarrow \Lambda_i + \ln\left(\mathbb{E}_{n_i}\left[e^{cY}\right]\right)$
  if Corollary 4 **then**
    $\mathcal{S}(t) \leftarrow$ (9)
  else
    $\mathcal{S}(t) \leftarrow$ Sequential sampling
  end if
  if $i \in \mathcal{S}(t)$ and n_i=1 **then**
    Initialize belief on Y_i
  else
    Update belief on Y_i (22) $\ \forall \ i \in \mathcal{S}(t)$
    if $\min_i \ n_i \geq 2$ **then**
      Update belief Λ_i (12) $\ \forall \ i \in \mathcal{S}(t)$
    end if
  end if
  $n_i \leftarrow n_i + 1 \ \forall \ i \in \mathcal{S}(t)$
end for

Corollary 4 is the accuracy condition that we implement for the uninformed-to-informed variant of the PIG algorithm, shown in Algorithm 1. Note that the informed exploration variant of PIG is simply Algorithm 1 for this class of bandit problems when the accuracy threshold is assumed to be satisfied upon the initialization of the PIG algorithm [5].

B. The Time-Varying Case: Pep-based Exploration

Since Corollary 4 is only valid in an environment with time-invariant statistics, we are compelled to develop an analogous result which is valid in time-varying environments. We now consider a so-called bi-homogeneous Pep with $\Lambda(\Delta t, \Delta n) = \lambda t n$, where λ is the *information gain per unit-time per sample* (i.e., $\frac{d}{dt}\frac{d}{dn}\Lambda(t, n) = \lambda \ \forall \ t, n > 0$). We further propose a gamma posterior distribution model for the bi-homogeneous Pep under consideration as

$$G\left(\lambda \mid \alpha_i = \frac{\alpha_{(i,1)} + \alpha_{(i,2)}}{2}, \beta = \beta_{(i,1)}\beta_{(i,2)}\right), \tag{18}$$

where an analytical update for the parameters is given in Fact 5.

Fact 5. *The parameters for the proposed gamma distribution in (18) are analytically updated such that*

$$\begin{aligned} &G\left(\lambda\Delta t|\alpha_{(i,1)} + z, \beta_{(i,1)} + \Delta t\right) \\ &\quad \propto Pep\left(z|\lambda\Delta t\right) G\left(\lambda\Delta t|\alpha_{(i,1)}, \beta_{(i,1)}\right) \\ &G\left(\lambda\Delta n|\alpha_{(i,2)} + z, \beta_{(i,2)} + \Delta n\right) \\ &\quad \propto Pep\left(z|\lambda\Delta n\right) G\left(\lambda\Delta n|\alpha_{(i,2)}, \beta_{(i,2)}\right) \end{aligned} \tag{19}$$

Proof. The proof follows directly from Fact 2. □

The key benefit of the bi-homogeneous Pep here is that we can derive a tighter inequality which does not require identically distributed samples, as stated in Theorem 6.

Theorem 6 (Chebyshev Inequality for RAPTOR)**.** *Let $\Delta Z(t_1), ..., \Delta Z(t_n) > 0$ be independent Poisson exposure process increments. Let $\bar{Z}(t_n) = \frac{1}{nt_n}\sum_{j=1}^{n} \Delta Z(t_j)$ and $\frac{d}{dt}\frac{d}{dn}\Lambda(\Delta t, \Delta n) = \lambda = \mathbb{E}[\bar{Z}(t_n)]$. Then,*

$$Pr\left(\left|\bar{Z} - \lambda\right| \geq \lambda^{\frac{3}{4}}\right) \leq \frac{1}{\sqrt{\alpha\beta}} = \frac{1}{\sqrt{\alpha n(t_n - t_1)}}, \tag{20}$$

where α and β are defined in (18) and (5).

Proof. See Appendix A for details. □

While our sample and time inequality in Theorem 6 provides an error bound for learning a single bi-homogeneous Pep, it does not yet provide a condition for the transition between uninformed and informed exploration across all bandit arms. This is given in the following corollary.

Corollary 7 (Accuracy Bound for RAPTOR)**.** *If the environment statistics are time-varying, then, given an accuracy threshold $0 < c \leq 1$, it holds $Pr\left(|\bar{Z} - \lambda| \geq \lambda^{\frac{3}{4}}\right) < c$ if the following conditions are satisfied:*

$$t_{n_b} > t_{1_b} + \frac{1}{n_b \alpha_b c^2}, \quad \text{and} \quad \frac{1}{\sqrt{\alpha_b \beta_b}} < c \tag{21}$$

where $b = \underset{i}{\operatorname{argmax}} \frac{1}{\sqrt{\alpha_i \beta_i}}$, n_i is the number of samples at location i, α_i is the total increase in the information gain

at location i, and β_i is the total duration that location i has been observed.

Proof. See Appendix B for details. □

The bound in Theorem 6 extends the Chebyshev inequality such that a result analogous to the Bienaymé-Chebyshev inequality is achieved, but where samples need not be identically distributed. Moreover, guarantees on the bi-homogeneous Pep regression by Theorem 6 provide a principled condition for transitioning from uninformed exploration to informed exploration in Corollary 7.

Algorithm 2 EVAR-RAPTOR

Input: sensor set $\Omega = \{1, \ldots, N\}$, subset size κ
Initialize $(\alpha_i, \beta_i, \lambda_i, n_i, \Delta t_i) \leftarrow 0 \ \forall \ i$
for each epoch at time t and for each sensor i **do**
  $\mathbb{E}[Z_i|\Delta t_i, \Delta n_i] \leftarrow \lambda_i \Delta t_i \Delta n_i + Z_i(t_{n_i}) + \ln\left(\mathbb{E}_{n_i}\left[e^{cY}\right]\right)$
  if Corollary 7 **then**
    $\mathcal{S}(t) \leftarrow$ (9)
  else
    $\mathcal{S}(t) \leftarrow$ Sequential sampling
  end if
  if $i \in \mathcal{S}(t)$ and n_i=1 **then**
    Initialize belief on Y_i
  else
    Update belief on Y_i (22) $\forall \ i \in \eta_t$
    if $\min_i \ n_i \geq 2$ **then**
      Update belief λ_i (19) $\forall \ i \in \eta_t$
    end if
  end if
  $n_i \leftarrow n_i + 1 \ \forall \ i \in \mathcal{S}(t)$
  $\Delta t_i \leftarrow \Delta t_i + 1 \ \forall \ i \in \mathcal{S}(t)$
end for

V. Experimental Evaluation

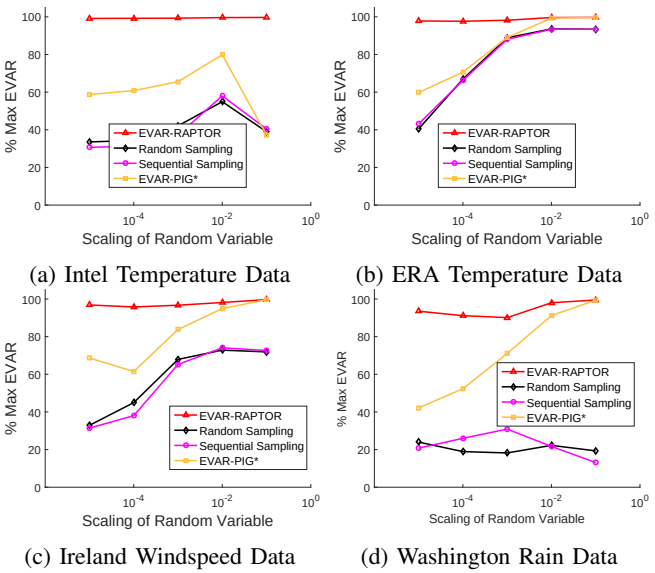
The empirical results in this section quantify the advantage of our proposed risk-aware strategy. The risk-aware performance of each algorithm is assessed in terms of the maximum available EVAR at each time step of the simulation; this maximum value is assessed in an omniscient post-processing of the data.

A. Evaluation Setup

For our experiments, we use the Intel Berkeley temperature dataset [13], the European Research Area (ERA) temperature dataset [14], the Ireland windspeed dataset [15], and the Washington rainfall dataset [16]. Each dataset reports the evolution of an environmental variable (e.g., temperature) at a N geographic locations. Details are given in the following. We create additional simulated datasets, by applying a scaling factor d to each random variable in the four real-world datasets mentioned above.

B. Assumptions on Prior and Posterior Distributions

The environmental variables Y_i are modeled using a normal distribution at each sensing location i. For each sample



(a) Intel Temperature Data (b) ERA Temperature Data

(c) Ireland Windspeed Data (d) Washington Rain Data

Fig. 1: The EVAR-seeking performance of each algorithm is shown across four real world datasets, with the environmental variable scaled from 10^{-5} to 10^{-1}. For Figures 1a, 1b, 1c, and 1d we used $\kappa = 6$, $\kappa = 6$, $\kappa = 2$, and $\kappa = 2$ as our testing conditions, respectively.

received, we calculate the frequentist variance σ_i^2 of the normal distribution. Using the calculated variance, we apply the following Bayesian update for the normal distribution

$$\mu_{(p,i)} = \frac{\frac{n_i}{\sigma_i^2}\mu_{(P_{0_0},i)} + \frac{y_i}{\sigma^2_{(P_{0_0},i)}}}{\frac{n_i}{\sigma_i^2} + \sigma^{-2}_{(P_{0_0},i)}}, \ \sigma^2_{P_0} = \left(\frac{\sigma^{-2}}{n_i} + \sigma_{P_0}^{-2}\right)^{-1}, \tag{22}$$

where $\mu_{(p,i)}$ is the posterior mean, $\mu_{(p_0,i)}$ is the prior mean, $\sigma^2_{(p,i)}$ is the posterior variance, $\sigma^2_{(p_0,i)}$ is the prior variance, and y_i is the sample mean [17]. The information gain of the posterior normal distribution is then computed as

$$I(P(Y)||P_0(Y)) = 0.5\Bigg[\ln\left(\frac{\sigma^2_{P_0}}{\sigma^2_P}\right) + tr\left[\left(\sigma^2{}_{P_0}\right)^{-1}\sigma^2{}_P\right] - 1 + \left(\mu_P - \mu_{P_0}\right)^T\left(\sigma^2{}_{P_0}\right)^{-1}\left(\mu_P - \mu_{P_0}\right)\Bigg], \tag{23}$$

where 1 is the dimensionality of the data which we use for experiments [18]. We then calculated the EVAR as

$$\text{EVAR}_{1-\gamma^\star} = \ln\left(\mathbb{E}_{P_0}(e^{dY})\right) + I(P(dY)||P_0(dY)), \tag{24}$$

where $\gamma^\star = e^{-I(P(dY)||P_0(dY))}$ is the lower bound of γ and d is the scaling parameter.

C. EVAR-Seeking Algorithms and Results

This section discusses the performance of the proposed learning algorithms: EVAR-PIG* (Algorithm 1) and EVAR-RAPTOR (Algorithm 2). We compare our algorithms against two baseline approaches: sequential and random sampling.

The average performance of the algorithms in each dataset for different scaling values is shown in Figure 1. Here, the regressions and derived bounds used for (23) distinguish the

empirical performance of EVAR-PIG* and EVAR-RAPTOR. While in (24) a change in the scaling parameter d affects the value of $\ln\left(\mathbb{E}\left[e^{dY}\right]\right)$, it does not affect the value of the information gain, $I(P(dY)||P_0(dY))$ [19]. As $\ln\left(\mathbb{E}\left[e^{dY}\right]\right)$ is an expectation based on the prior in (22), it is deterministic and given. Hence, as the scaling parameter d increases, the proportion of EVAR that is stochastic decreases. Resultantly, if the regression for the information gain is accurate, then we should see a consistent EVAR prediction performance across a range of values for the scaling parameter d in Figure 1.

In summary, EVAR-PIG* appears to be about as good, if not better, than random and sequential sampling. However, the performance of EVAR-PIG* is inconsistent across the scaling d. On the other hand, EVAR-RAPTOR outperforms all baselines and is consistent in performance, despite changes in the scaling parameter d.

VI. Conclusion

In the context of exploration, we mitigate the risk of data-driven models losing relevance in time-varying environments as quantified by the entropic value at risk (EVAR), which helps to quantify how a model changes as a result of new observations. While EVAR is conventionally calculated with a predetermined confidence value γ, here we use a data-driven approach to determine γ. Moreover, we propose the EVAR-RAPTOR algorithm which learns to predict the EVAR available at different sensing locations with probabilistic accuracy guarantees. These probabilistic guarantees allow us to accurately learn and predict (EVAR) values in real-world datasets, even when the random environmental variable is scaled by a constant. Empirical results on four real-world datasets demonstrate empirically that EVAR-RAPTOR has consistently superior performance in predicting EVAR values for data generated by time-varying distributions.

References

[1] Arthur Hobson. A new theorem of information theory. *Journal of Statistical Physics*, 1(3):383–391, 1969.

[2] Amir Ahmadi-Javid. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123, 2012.

[3] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[4] Taemin Kim, Ara V. Nefian, and Michael J. Broxton. Photometric recovery of ortho-images derived from apollo 15 metric camera imagery. In *Advances in Visual Computing*, pages 700–709. Springer, 2009.

[5] Daniel Y Little and Friedrich T Sommer. Learning and exploration in action-perception loops. *Frontiers in neural circuits*, 7, 2013.

[6] Allan Max Axelrod. Learning to exploit time-varying heterogeneity in distributed sensing using the information exposure rate. Master's thesis, Oklahoma State University, 2015.

[7] Michel Tokic. Adaptive ε-greedy exploration in reinforcement learning based on value differences. In *KI 2010: Advances in Artificial Intelligence*, pages 203–210. Springer, 2010.

[8] Huanyu Ding and David A Castañón. Optimal solutions for adaptive search problems with entropy objectives. *arXiv preprint arXiv:1508.04127*, 2015.

[9] Allan M Axelrod, Sertac A Karaman, and Girish V Chowdhary. Exploitation by informed exploration between isolated operatives for information-theoretic data harvesting. In *Conference on Decision and Control*, volume 54, Osaka, JP, 2015. CDC.

[10] Allan Axelrod and Girish Chowdhary. Uninformed-to-informed exploration in unstructured real-world environments. In *Self-Confidence In Autonomous Systems*, Washington D.C., USA, 2015. AAAI.

[11] Allan Axelrod and Girish Chowdhary. A hybridized bayesian parametric-nonparametric approach to the pure exploration problem. In *Bayesian Nonparametrics: The Next Generation*, Montreal, Canada, 2015. Neural Information Processing Systems.

[12] Allan Axelrod and Girish Chowdhary. *The Explore-Exploit Dilemma in Nonstationary Decision Making under Uncertainty*, chapter The Explore-Exploit Dilemma in Nonstationary Decision Making under Uncertainty. 2198-4182. Springer international publishing, 1 edition, 2015.

[13] Peter Bodik, Wei Hong, Carlos Guestrin, Sam Madden, Mark Paskin, and Romain Thibaux. Intel lab data. Technical report, Intel Berkely Research Lab, Feb 2004.

[14] Paul Berrisford, DPKF Dee, K Fielding, M Fuentes, P Kallberg, S Kobayashi, and S Uppala. The era-interim archive, 2009.

[15] John Haslett and Adrian E. Raftery. Ireland wind data set. Technical report, Trinity College and University of Washington, 1961-1978.

[16] Martin Widmann and Christopher S Bretherton. Validation of mesoscale precipitation in the ncep reanalysis using a new gridcell dataset for the northwestern united states. *Journal of Climate*, 13(11):1936–1950, 2000.

[17] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis*. CRC press, 3rd edition, 2013.

[18] John Duchi. Derivations for linear algebra and optimization. *Berkeley, California*, 2007.

[19] Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.

Appendix

A. Proof of Theorem 6

From Chebyshev's inequality, we know that

$$Pr(|\bar{Z} - \mu| \geq k) \leq \frac{\mathrm{VAR}(\bar{Z})}{k^2}. \tag{25}$$

Inserting the mean ($\lambda = \frac{\alpha}{\beta}$) and variance ($\frac{\alpha}{\beta^2}$) of our gamma distribution model on homogeneous Pep into (25) yields

$$Pr(|\bar{Z} - \lambda| \geq k) \leq \frac{\lambda}{\beta k^2}. \tag{26}$$

In assigning, $\lambda^{\frac{3}{4}} = k$ and noting that $\beta = n(t_n - t_1)$ due to (18), we resolve the proof:

$$Pr(|\bar{Z} - \lambda\beta| \geq \lambda^{\frac{3}{4}}) \leq \frac{1}{\sqrt{\alpha\beta}} = \frac{1}{\sqrt{\alpha n(t_n - t_1)}}. \tag{27}$$

B. Proof of Corollary 7

When the right-hand side of (27) is between 1 and 0, we have meaningful guarantees on the error of our regression; i.e., when

$$\frac{1}{\sqrt{\alpha_b\beta_b}} < c, \tag{28}$$

where $0 < c \leq 1$ and $\beta = n(t_n - t_1)$. Then meaningful guarantees are available at a sensing location i when

$$t_{n_i} > t_{1_i} + \frac{1}{n_i\alpha_i c^2}. \tag{29}$$

Guarantees are available across the entire sensing domain once $t_{n_i} > t_{1_i} + \frac{1}{n_b\alpha_b c^2}$, where $b = \underset{i}{\mathrm{argmax}}\ \frac{1}{n_i\alpha_i c^2}$.