

REVIEW ARTICLE

A Nonmathematical Review of Optimal Operator and Experimental Design for Uncertain Scientific Models with Application to Genomics

Edward R. Dougherty*

Department of Electrical and Computer Engineering, College Station, Texas A&M University - TX, USA

Abstract: Introduction: The most basic aspect of modern engineering is the design of operators to act on physical systems in an optimal manner relative to a desired objective – for instance, designing a control policy to autonomously direct a system or designing a classifier to make decisions regarding the system. These kinds of problems appear in biomedical science, where physical models are created with the intention of using them to design tools for diagnosis, prognosis, and therapy.

Methods: In the classical paradigm, our knowledge regarding the model is certain; however, in practice, especially with complex systems, our knowledge is uncertain and operators must be designed while taking this uncertainty into account. The related concepts of intrinsically Bayesian robust operators and optimal Bayesian operators treat operator design under uncertainty. An objective-based experimental design procedure is naturally related to operator design: We would like to perform an experiment that maximally reduces our uncertainty as it pertains to our objective.

Results & Discussion: This paper provides a nonmathematical review of optimal Bayesian operators directed at biomedical scientists. It considers two applications important to genomics, structural intervention in gene regulatory networks and classification.

Conclusion: The salient point regarding intrinsically Bayesian operators is that uncertainty is quantified relative to the scientific model, and the prior distribution is on the parameters of this model. Optimization has direct physical (biological) meaning. This is opposed to the common method of placing prior distributions on the parameters of the operator, in which case there is a *scientific gap* between operator design and the phenomena.

ARTICLE HISTORY

Received: October 26, 2018
Revised: December 05, 2018
Accepted: December 10, 2018

DOI:
10.2174/1389202919666181213095743

Keywords: Nonmathematical review, Uncertain scientific models, Applications, Genomics, Optimization, Regulatory regimes.

1. INTRODUCTION

Scientific knowledge is translated into practical knowledge by applying some operation to the scientific model in an effort to achieve the desired objective. The basic paradigm consists of four parts: (1) a scientific (mathematical) model describing the physical system, (2) a family of operators from which to choose, (3) a cost function measuring how well the objective is achieved, and (4) optimization to find an operator possessing minimum cost. In biomedical science, models are created with the intention of using them for diagnosis, prognosis, and therapy. This review considers two genomic applications. In the first, the model is a gene regulatory network, the operator family consists of certain types of network interventions that in principle could be accomplished by drugs, the cost function is the probability of evolving into a pathological state, and the intervention that minimizes the cost is selected. In the second application, the model is a feature-label distribution, the operator family consists of classifiers to discriminate between two classes of objects, the cost function is the classifier error, and the classifier with minimum error is selected.

As stated, the classical paradigm assumes that the model of the physical system is assumed to be known with certainty;

however, often this assumption is not warranted, especially with complex systems, where there are many parameters to determine. If the model is a regulatory network, it may be that some of the regulations are unknown. This changes the optimization paradigm to one in which intervention optimization must consider both the objective and the state of knowledge regarding the network. This lack of certainty will be reflected in the loss of performance because we will not be able to optimize for the actual system, but instead will optimize so that the intervention works best on average relative to all possible systems that fit our knowledge. One might think of designing a drug that is to be applied to a population possessing different regulatory regimes rather than applying it to a smaller group consisting of a single regulatory regime. While the latter approach would be desirable, rarely is such an assumption practical.

In this work, we assume that the scientific model is uncertain and belongs to an uncertainty class of possible models. The aim is to design an operator that is optimal relative to both the objective and the uncertainty class. For instance, in classification, the scientific model is given by a feature-label distribution, which herein is considered to be uncertain. We desire a classifier with a minimum average error over all possible feature-label distributions. It is critical to recognize that, whereas Bayesian methods are often applied with the uncertainty placed on the possible operators (say, classifiers), we place the uncertainty on the scientific knowledge, which is where it actually arises.

*Address correspondence to this author at the Department of Electrical and Computer Engineering, College Station, Texas A&M University, TX, USA; Tel/Fax: 979-845-6259; E-mail: edward@ece.tamu.edu

It is useful to quantify our uncertainty, not in general, but relative to our intervention objective; that is, how much worse do we expect to do because we are applying a procedure that is optimal on average rather than optimal to the specific case at hand. If it turns out that there is too much uncertainty for satisfactory performance, then we can lessen the uncertainty by designing and performing appropriate experiments. The issue then is to determine which experiment (or experiments) is optimal: which experiment maximally reduces our uncertainty relative to the achievement of our objective? In the case of regulatory networks, we wish to choose an experiment that provides the regulatory knowledge that maximally reduces the pertinent uncertainty in our model.

As the title of the article states, this is a nonmathematical review. Hence, the actual mathematical form of many operations will not be specified explicitly and there might be some loss of rigor. In all cases, the relevant references are cited for the interested reader. Moreover, the methodologies discussed in this paper are part of an overall structure for designing optimal operators under model uncertainty, and one can also refer to [1] for mathematical details.

2. OPTIMAL STRUCTURAL INTERVENTION IN GENE REGULATORY NETWORKS

We first treat optimal structural intervention in gene regulatory networks with the aim being to alter the regulatory structure in order to beneficially change the long-run behavior of the system away from disease states.

2.1. Gene Regulatory Network Model

A Gene Regulatory Network (GRN) is a mathematical model comprised of a set of entities called “genes” and a regulatory structure that governs their behavior over time. GRNs can be finely detailed, as with differential-equation models, or coarse-grained, with discrete expression levels transitioning over discrete time. Coarse models need not closely represent actual molecular structure; rather, their purpose is to model interaction at a high level in order to serve as a framework for studying the regulation and to provide rough models that can be used to develop strategies for controlling aberrant cell behavior, such as finding optimal drug treatments.

We consider Boolean networks, in which each gene can have logical values 1 or 0, corresponding to expressing or not expressing, respectively, and regulation is specified by logical operations among genes [2]. Thus, the functional relationships between genes can be specified by a truth table. While the Boolean model is very coarse, it does model the thinking of cancer biologists when they speak of a gene being on or off under different conditions. Moreover, although the original formulation is two-valued, 0 or 1, the concept applies to any number of discrete gene values.

A Boolean network is defined by k binary variables, x_1, x_2, \dots, x_k , where the value x_i of gene g_i at time $t + 1$ is determined by the values of some regulator genes at time t via a Boolean function f_i operating on the regulator genes. There are k such Boolean functions, one for each gene, and together they determine the dynamic evolution of the system over

time. The state of the network is defined by the vector $\mathbf{x} = (x_1, x_2, \dots, x_k)$ of binary expression values. Given an initial state, a Boolean network will eventually reach a set of states, called an *attractor cycle*, through which it will cycle endlessly. Each initial state corresponds to a unique attractor cycle and the set of initial states leading to a specific attractor cycle is known as the *basin of attraction* of the attractor cycle.

Randomness is introduced into a Boolean network *via* perturbations. For each gene, there is some small perturbation probability that it will randomly switch values. This is practical because there is random variation in the amount of mRNA and protein produced. Perturbations allow a network to jump out of an attractor cycle and eventually transition to a new attractor. We utilize a Boolean network with perturbation (BNp) for regulatory modeling.

2.2. Structural Intervention in Gene Regulatory Networks

If every gene has a positive perturbation probability, then for any state \mathbf{x} , the probability that the network is in state \mathbf{x} , in the long run, is independent of the initial state. The collection of all such probabilities is called the *steady-state distribution*. Assuming the existence of a steady-state distribution, structural intervention in a gene regulatory network involves a one-time change of the regulatory structure, the aim being to minimize the sum of the steady-state probabilities corresponding to the undesirable states, which means minimizing the probability of being in an undesirable state [3].

To illustrate structural intervention, we consider a mammalian cell cycle Boolean network with perturbation ($p = 0.01$) based on a regulatory model proposed in [4]. We employ a structural intervention that models small interfering RNA (siRNA) interference in regulatory relationships: an intervention blocks the regulation between two genes in the network.

The cell cycle involves a sequence of events resulting in the duplication and division of the cell. It occurs in response to growth factors and under normal conditions, it is a tightly controlled process. The model in [4] contains 10 genes: CycD, Rb, p27, E2F, CycE, CycA, Cdc20, Cdh1, UbcH10, and CycB, with genes numbered in this order. The cell cycle in mammals is controlled *via* extra-cellular stimuli. Positive stimuli activate Cyclin D (CycD) in the cell, thereby leading to cell division. CycD inactivates the Rb protein, which is a tumor suppressor. When gene p27 and either CycE or CycA are active, the cell cycle stops, because Rb can be expressed even in the presence of cyclins. States in which the cell cycle continues even in the absence of stimuli are associated with cancerous phenotypes. For this reason, states with down-regulated CycD, Rb, and p27 ($x_1 = x_2 = x_3 = 0$) are undesirable.

The regulatory model in Fig. (1) has blunt and normal arrows representing suppressive and activating regulations, respectively. Genes are assumed to be regulated according to the majority vote rule. At each time point, a gene takes the value 1 if the majority of its regulator genes are activating and the value 0 if the majority of the regulator genes are suppressive; otherwise, it remains unchanged.

A structural intervention removes an arrow from the regulatory graph because it blocks a regulation between two

genes. Using the optimization methods of Qian and Dougherty [3], it is determined that the structural intervention that maximally lowers undesirable steady-state probability blocks the regulatory action from gene CycE to p27 and reduces total undesirable steady-state probability from 0.3401 to 0.2639 [5].

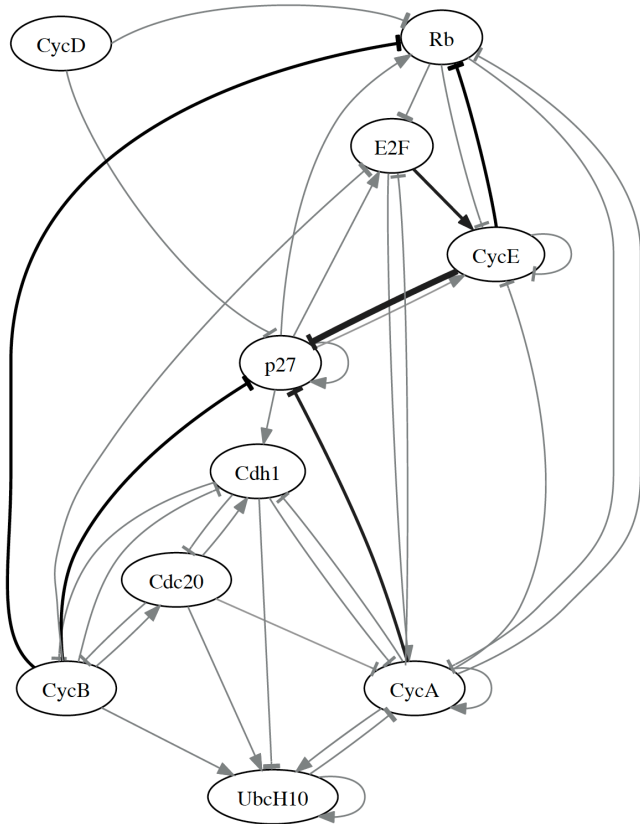


Fig. (1). Mammalian cell cycle network: normal arrows represent activations and blunt arrows represent suppressing effects. Given full knowledge of the network, the optimal intervention strategy is to block the regulation from CycE to p27 (shown in bold). Other regulations to be blocked according to the intrinsically Bayesian robust intervention strategies in the absence of full regulatory information are shown in black.

Note the four steps in the intervention paradigm: (1) model the cell cycle by a BNp; (2) an intervention operator blocks a single regulation between two genes; (3) the cost is the total steady-state probability of the undesirable states; and (4) an optimal intervention is found *via* the method of Qian and Dougherty [3]. Although we have not done so, one may wish to constrain the optimization procedure by avoiding states known to be associated with carcinogenesis or states that do not typically occur in healthy cells [6].

3. OPERATOR DESIGN IN THE PRESENCE OF MODEL UNCERTAINTY

Whereas the preceding analysis assumes that the regulatory model is fully known, we now suppose that there is model uncertainty. In this case, the true model belongs to an uncertainty class of possible models. Each model in the uncertainty class is associated with a parameter vector θ and the uncertainty class is denoted by Θ , which consists of all

the parameter vectors. For a gene regulatory model, some regulations are unknown, so that Θ consists of all possible parameter vectors corresponding to the unknown regulations. Let C be a cost function and Ψ be a family of operators on the model whose performances are measured by the cost function. This means that for each operator $\psi \in \Psi$ there is a cost $C_\theta(\psi)$ of applying ψ on the model θ .

For example, suppose Ψ consists of some number of drugs, meaning that each operator acts by applying a drug. Suppose the goal of the drug treatment is to reduce the expression of a particular gene g associated with metastasis in breast cancer and that the gene regulatory network being used is uncertain, so that there is an uncertainty class Θ of models. The cost function might be the steady-state probability of g , that being the sum of all steady-state probabilities for which g is on. Then $C_\theta(\psi)$ is the steady-state probability of g when the drug ψ is applied to model θ . Since the full network model is unknown, one would like to choose a drug whose performance works well over the uncertainty class.

To be precise, an Intrinsically Bayesian Robust (IBR) operator is an operator ψ_{IBR} such that the expected value (average) over Θ of the cost $C_\theta(\psi)$ is minimized by ψ_{IBR} , the expected value being with respect to a prior probability distribution $\pi(\theta)$ over Θ [5, 7]. An IBR operator is robust in the sense that on average it performs well over the whole uncertainty class. A prior probability distribution on the space of possible models quantifies our prior knowledge regarding the likelihood of the possible models being the true model. If there is no prior knowledge beyond the uncertainty class itself, then the prior distribution is taken to be uniform.

3.1. IBR Structural Intervention in Gene Regulatory Networks

We now consider intrinsically Bayesian robust structural intervention for the mammalian cell cycle network. Assume that there are D pairs of genes for which the existence of a regulatory relationship is known but the type of relationship, activating or suppressing, is unknown. The uncertainty class consists of 2^D possible networks, where each vector $\theta \in \Theta$ corresponds to a specific assignment of regulation types to the D uncertain edges. We assume a uniform prior distribution, meaning that we have no knowledge concerning model likelihood and all uncertain parameter vectors have prior probability $1/2^D$. As before, a structural intervention blocks the regulatory action between a pair of genes in the network, and the cost function is the total undesirable steady-state probability; however, now the optimization involves averaging the cost of each possible structural intervention (using the analytic methods provided in [3]) over the 2^D possible networks, and selecting the intervention with the minimal average.

Simulations in [5] incrementally increase the number of edges with unknown regulation from $D = 1$ to $D = 10$. In each case, 50 uncertain networks are created by randomly selecting uncertain edges while keeping the regulatory information for the remaining edges. Grouping the models with 1 to 5 uncertain edges, 54.0% of the time the IBR structural intervention is the actual optimal intervention, which blocks the regulation from CycE to p27. This reduces the total undesirable steady-state probability to 0.2639. The se-

cond most selected IBR intervention (41.6% of the time) blocks the regulation from CycE to Rb and reduces the total undesirable steady-state probability to 0.2643.

One must keep in mind that the IBR intervention works best on average over the uncertainty class and may perform poorly on the actual network. In this simulation, blocking regulation between CycB and p27 is selected 2.0% of the time and only reduces the undesirable steady-state probability to 0.3244.

With 6 to 10 uncertain edges, blocking CycE to p27 or blocking CycE to Rb accounts for 88.8% of the IBR interventions, as opposed to 95.6% of the IBR interventions for 1 to 5 uncertain edges. This change reflects greater uncertainty.

4. OPTIMAL EXPERIMENTAL DESIGN

Based on a cost function, an IBR an operator is optimal over an uncertainty class relative to a prior distribution reflecting our scientific knowledge; however, it will not likely be optimal relative to the true model. This loss of performance is the cost of uncertainty. To quantify this cost, for each model θ in the uncertainty class, let ψ_θ be an optimal operator for the model θ . Owing to the optimality of the IBR operator over the uncertainty class, on average it performs better than ψ_θ , meaning that the expected cost of ψ_{IBR} over the uncertainty class is less than or equal to the expected cost of ψ_θ . However, there is a loss of performance relative to applying ψ_{IBR} on the model θ because, since ψ_θ is optimal for model θ , $C_\theta(\psi_\theta) \leq C_\theta(\psi_{IBR})$.

For any θ , the Objective Cost of Uncertainty (OCU) is the cost differential between an IBR operator and an optimal operator for θ applied on θ : $OCU(\theta) = C_\theta(\psi_{IBR}) - C_\theta(\psi_\theta)$. We would like to compute the objective cost of uncertainty for the true model, but we do not know the true model. Our knowledge only concerns the uncertainty class. Hence, we compute the Mean Objective Cost of Uncertainty, $MOCU(\Theta)$, which is the expected value of $OCU(\theta)$, that is, the average cost of applying an IBR operator. $MOCU$ provides an objective-based quantification of uncertainty [5].

The more knowledge we have regarding the scientific model, the tighter will be the prior distribution around the true value of θ . If there is a set of experiments that can supply information relating to the unknown parameters, which should be performed first? A classical approach is to choose an experiment that minimizes the entropy of the prior distribution [8]. As opposed to entropy, $MOCU$ quantifies the uncertainty in our knowledge with respect to our objective, which is what matters, rather than a general reduction in uncertainty. For instance, it may be that determining a certain unknown gene regulation provides the greatest reduction in entropy (overall uncertainty), but the regulation has little or no effect on the disease.

Hence, we choose an experiment that yields the minimum expected (remaining) $MOCU$ given the experiment [9]. For each possible experiment, compute $MOCU$ for every possible outcome of the experiment, average these $MOCU$ values, and then take the minimum of these averages over all possible experiments. This can be done sequentially, either greedily by repeating the procedure after the preceding experiment has determined a regulation, and continuing until

some stopping criterion has been reached [9], or *via* dynamic programming [10]. In either case, the result is objective-based optimal experimental design.

To illustrate the greedy sequential experimental design, we randomly generate BNps, each containing six genes with each gene having three regulators, and perturbation probability $p = 0.001$ [9]. Simulations are performed with 50 different sets of k regulations, $k = 5$, being randomly selected from the network and their regulatory information assumed to be unknown. An uncertain parameter equals 1 for an activating relation and 0 for a suppressive relation. With k uncertain relations, the uncertainty class contains 2^k networks. Experimental design selects the parameter to determine. States with gene 1 activated are assumed to be undesirable. 1000 synthetic BNps are generated.

A practical issue in evaluating experimental design using synthetic networks is controllability. Unlike real networks, which are controllable to a certain extent, many randomly generated networks may not be controllable. Hence, regardless of the intervention, the shift in the steady-state distribution may be negligible. For such networks, the difference between optimal and suboptimal experiments may be insignificant. Thus, to examine the practical impact of experimental design, we must take controllability into account. We require that the percentage decrease of total steady-state mass in undesirable states after optimal structural intervention exceeds 40%.

Fig. (2) provides a performance comparison based on a sequence of experiments. It shows the average cost of robust intervention after performing the sequence of experiments determined by the design strategy and the average cost after performing randomly selected experiments. This kind of Sequential-design curve is typical. One gets large gain with the early experiments, which is precisely the goal of sequential experimental design. If experiments are costly, one need only perform a small number of experiments. Note that the curves meet when all unknown parameters have been determined.

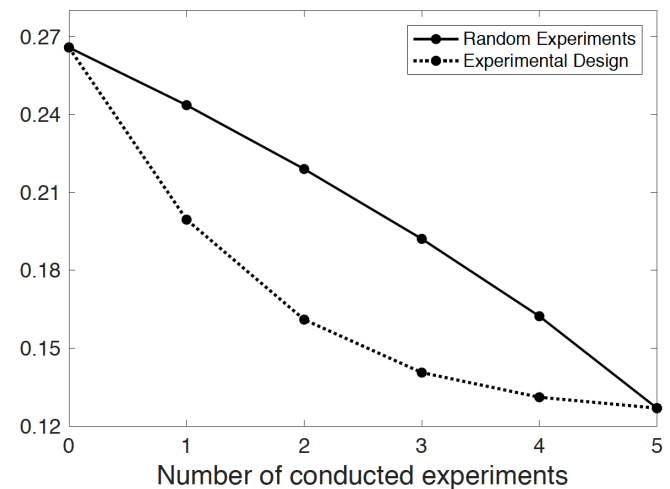


Fig. (2). Performance comparison based on a sequence of experiments. The average cost of robust intervention after performing the sequence of experiments predicted by the experimental design strategy and the average cost after performing randomly selected experiments.

5. CLASSIFICATION

In pattern classification, features are calculated on objects from two different populations and, based on a feature vector, a classifier predicts which population an object belongs to. For cancer medicine, classification can be between different kinds of cancer, stages of tumor development, or prognoses. For instance, gene expressions are measured for k genes and based on the measurements it is decided which drug should be administered. Classification has been a staple of bioinformatics since the inception of high-throughput expression measurements [11]. A feature vector belongs to one of two classes, and the model consists of feature-label pairs (\mathbf{X}, Y) , where $\mathbf{X} = (X_1, X_2, \dots, X_k)$ and $Y = 0$ or $Y = 1$. A binary classifier ψ is a function on the set of feature vectors: $\psi(\mathbf{X}) = 0$ or $\psi(\mathbf{X}) = 1$.

For classification, the model consists of two class-conditional distributions $f(\mathbf{x}|0)$ and $f(\mathbf{x}|1)$, which are the probability distributions governing the behavior of feature vectors in class 0 and class 1, respectively. The model also requires the probability c_0 that a randomly selected object comes from class 0, which automatically gives the probability c_1 that it comes from class 1, since $c_1 + c_0 = 1$. Taken together, $f(\mathbf{x}|0)$, $f(\mathbf{x}|1)$, and c_0 provide the *feature-label distribution* $f(\mathbf{x}, y)$ governing the feature-label vectors.

The error of classifier ψ is the probability of erroneous classification, $\epsilon[\psi] = P(\psi(\mathbf{X}) \neq Y)$, which can be found from the feature-label distribution. An optimal classifier from the collection of all classifiers is one having a minimal error. It is called a Bayes classifier and we denote it by ψ_{Bay} . Assuming $c_0 = c_1 = 1/2$ for simplicity, a Bayes classifier is defined by $\psi_{\text{Bay}}(\mathbf{x}) = 1$ if $f(\mathbf{x}|0) \leq f(\mathbf{x}|1)$, and $\psi_{\text{Bay}}(\mathbf{x}) = 0$ if $f(\mathbf{x}|0) > f(\mathbf{x}|1)$. The error of a Bayes classifier is known as the *Bayes error* and is denoted by ϵ_{Bay} . While there may be many Bayes classifiers for a feature-label distribution, the Bayes error is unique.

Considering features and labels as physical measurements, the feature-label distribution represents knowledge of the variables X_1, X_2, \dots, X_k, Y . Given a feature-label distribution, one can in principle find a Bayes classifier and the Bayes error; however, for important models, only in rare cases have these been analytically derived from the feature-label distribution, but they can be approximated by numerical methods.

5.1. Optimal Bayesian Classification

Model uncertainty arises when full knowledge of the feature-label distribution is lacking. Knowledge must come from existing scientific knowledge regarding the features and labels or be estimated from data. Since accurate estimation of distributions requires a huge amount of data, the amount increasing rapidly with dimension and distributional complexity, full knowledge of the feature-label distribution is rare. With model uncertainty, there is an uncertainty class Θ of parameter vectors corresponding to feature-label distributions. In this setting, an Intrinsically Bayesian Robust (IBR) classifier minimizes the expected error across the uncertainty class.

This minimization is analogous to the minimization for determining an IBR structural intervention in a gene regula-

tory network except that, whereas for structural intervention in the mammalian cell cycle network one can compute a finite number of operator costs (undesirable steady-state probabilities) and take the least, for IBR classification there may be an infinite number of classifiers to consider.

This problem is solved by using effective class-conditional distributions for the uncertainty class [12]. These are the expected values, relative to the prior distribution, of the individual class-conditional distributions over the uncertainty class. The class-0 and class-1 effective class-conditional densities $f_{\theta}(\mathbf{x}|0)$ and $f_{\theta}(\mathbf{x}|1)$ are the expected values at point \mathbf{x} of the class-conditional density values for classes 0 and 1, respectively. If we do not constrain the classifiers from which we are allowed to choose, then an IBR classifier is found in exactly the same manner as a Bayes classifier, except that the effective class-conditional densities are used: again assuming $c_0 = c_1 = 1/2$, $\psi_{\text{IBR}}(\mathbf{x}) = 1$ if $f_{\theta}(\mathbf{x}|0) \leq f_{\theta}(\mathbf{x}|1)$, and $\psi_{\text{IBR}}(\mathbf{x}) = 0$ if $f_{\theta}(\mathbf{x}|0) > f_{\theta}(\mathbf{x}|1)$. An IBR classifier is a Bayes classifier for the feature-label distribution determined by the effective class-conditional densities.

In addition to a prior distribution coming from existing knowledge, if one has a sample data set S , then a posterior distribution $\pi^*(S) = \pi(\theta|S)$, the prior distribution conditioned on the sample, can be constructed. All preceding definitions and propositions go through with the posterior in place of the prior, and with the optimal IBR classifier being known as the Optimal Bayesian Classifier (OBC). An OBC is an IBR classifier relative to the posterior distribution, and an IBR classifier is an OBC relative to a null sample. Thus, in some sense, they represent equivalent formulations.

The posterior distribution incorporates all of our knowledge, prior knowledge plus data. Under rather general conditions, an OBC is a consistent classification rule, meaning that the OBC converges to a Bayes classifier for the true feature-label distribution [13]. This, however, is not the main advantage of the OBC; rather, owing to the use of prior knowledge, it can provide good classification with small samples. The small-sample problem has bedeviled genomic classification since the early days of expression-based classification [14].

Digressing for a moment, we note that classification is typically studied under the assumption of random sampling, meaning that random sample points are collected independently and each is identically distributed with the true feature-label distribution. This is not always true in practice. Moreover, nonrandom sampling can be beneficial for classifier design [15]. In the case of the OBC, optimal sampling has been considered under different scenarios [1, 16].

An OBC (IBR classifier) provides the best performance on average over the uncertainty class but is usually not optimal for any specific feature-label distribution, for which a Bayes classifier is optimal. We define the objective cost of uncertainty in the same manner as before, but with the cost function being the classification error: For any θ , the OCU is the difference between the classification error of the IBR classifier and the Bayes classifier for θ relative to θ . We use MOCU as a measure of uncertainty and consider optimal MOCU-based experimental design.

In a similar vein, while the OBC is optimal on average across the posterior distribution, it need not outperform some other classifier for any particular feature-label distribution in the uncertainty class. If the prior distribution is concentrated in the vicinity of the true feature-label distribution, then OBC performance tends to be close to the performance of the Bayes classifier, and therefore, it will rarely be outperformed by some other classifier. In fact, under very general conditions (satisfied by both discrete and Gaussian models), as the sample size increases to infinity, the OBC will converge to the Bayes classifier for the true feature-label distribution [13]. But one must be prudent when selecting a prior distribution. If it is tight and concentrated away from the true feature-label distribution, then the results can be poor for small samples. Performance comparison relative to various kinds of prior assumptions is considered in [13]. Correct knowledge helps; incorrect knowledge hurts. Prior construction is important, and we consider it in the next section.

We illustrate IBR classification using a two-dimensional Gaussian model, meaning that the class-conditional distributions are two-dimensional Gaussian distributions, and, on account of uncertainty, the unknown parameters of the models are governed by prior distributions. We refer the interested reader to [13] for details. For any particular state, the Bayes classifier is quadratic. We also consider a plug-in classifier, which is the Bayes classifier assuming the expected value of each parameter. This classifier is linear. The average true errors are 0.2078 for the plug-in and 0.2007 for the IBR. The classifiers are depicted in Fig. (3), in which the level curves for the class-conditional distributions corresponding to the expected parameters are shown in gray dashed lines. Note that, whereas for any particular feature-label distribution in the uncertainty class, the Bayes classifier is quadratic, the IBR classifier is not quadratic. Its shape depends not only on the individual Gaussian models in the uncertainty class but also on the prior distribution.

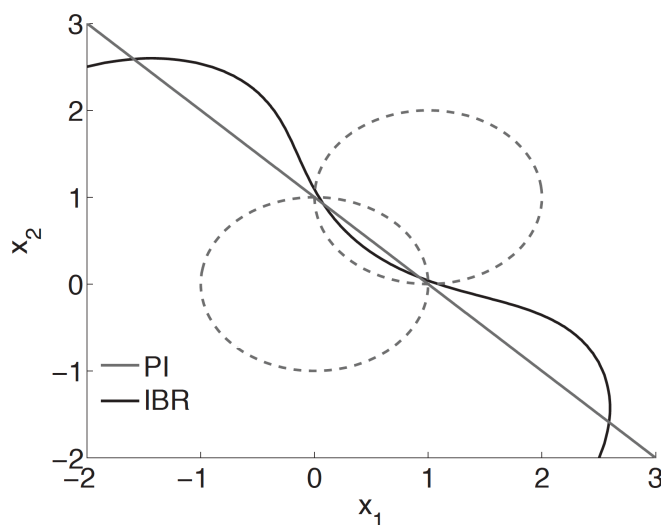


Fig. (3). IBR and Plug-in Classifiers for a Gaussian model with two features.

In [12, 13], Gaussian and discrete models are considered for which the OBC can be solved analytically. This is not generally the case. Markov-chain-Monte-Carlo (MCMC) OBC

methods were introduced in [17, 18] for RNA-Seq application, and are usually used in real-world settings where Gaussian models are not appropriate. Other applications include liquid chromatography-mass spectrometry data [19], selection reaction monitoring data [20], and classification based on dynamical measurements of single-gene expression measurements [21]. The OBC has been adapted to settings in which there are missing values [22]. This is important in real-world applications, in particular, in genomic classification, where the missing-value problem has been evident from the outset [23].

6. PRIOR CONSTRUCTION

In any Bayesian methodology, prior construction is a critical issue. In 1968, E.T. Jaynes remarked, “Bayesian methods, for all their advantages, will not be entirely satisfactory until we face the problem of finding the prior probability squarely.” [24]. Twelve years later, he added, “There must exist a general formal theory of determination of priors by logical analysis of prior information – and that to develop it is today the top priority research problem of Bayesian theory.” [25] For optimal operator design, this is an engineering problem, and it must be faced in the context of scientific knowledge and the transformation of that knowledge into prior distributions.

Historically, prior construction has tended to utilize general methodologies not targeting any specific type of prior information. Subsequent to the introduction of Jeffreys’ non-informative prior [26], objective-based methods were subsequently proposed, an early one being [27]. There appeared a series of information-theoretic approaches, including maximal data information priors [28].

The more a prior is constrained by scientific knowledge, the more confident one can be that the prior distribution is concentrated around the correct model; however, as noted previously, one must be prudent, since concentrating the prior away from the true model can result in very poor results. With optimal Bayesian classification in the context of phenotype classification, knowledge concerning genetic signaling pathways has been integrated into prior construction [29, 30]. In [17, 18], a hierarchical Poisson prior is employed that models cellular mRNA concentrations using a log-normal distribution and then models the sequencing as sampling the RNA concentrations through a Poisson process. A general procedure for prior construction in [31] uses a constrained information-based optimization, in which the constraints incorporate existing scientific knowledge augmented by slackness variables. The constraints tighten the prior distribution in accordance with prior knowledge, while at the same time avoiding inadvertent over restriction of the prior.

While we have concentrated our discussion on the classical situation in which sample data are collected from the true feature-label distribution, let us note that researchers are currently interested in augmenting limited data from the true system with more plentiful data from related systems. This is called transfer learning [32]. For instance, one may have limited data for human cancer but have a great deal of data for related animal cancer. A key issue with transfer learning is transferability, which refers to the relationship between the two data types. Transferability has been addressed by gener-

alizing the OBC framework to a joint prior distribution governing two feature-label distributions, with their relationship encoded into the joint statistics of the prior [33].

CONCLUDING REMARKS

Perhaps the most fundamental point regarding IBR operator design is that uncertainty is quantified relative to the scientific model, meaning that the prior distribution is on the physical parameters. This is opposed to the common method of placing prior distributions on the parameters of the operator. For instance, although we did not cover optimal Bayesian regression [34] in this paper, if we compare optimal Bayesian regression to existing Bayesian linear regression models [35-39], in the latter, the connection of the regression functions and prior assumptions with the underlying physical systems is vague. As noted in [34], there is a scientific gap in constructing operator models and making prior assumptions on the operator models. The actual uncertainty in the operator is derived from the uncertainty in the physical system *via* the optimization procedure that produces an optimal operator.

Let us close by noting that this kind of optimization has a long history in engineering. In control theory, where the problem is to apply inputs to a physical system over time in some optimal manner, it was recognized in the 1960s that knowledge might be limited so that the system is uncertain, and a Bayesian approach can be taken to design the optimal controller [40-42]. Full optimization was well beyond the computational capacity of the time, and computation remains problematic [43]. In signal processing, the issue arose in the 1970s and was treated *via* minimax optimization: find the filter that has the best worst-case performance over the uncertainty class [44-46]. Optimal signal processing given an uncertain signal model was subsequently treated suboptimally in a Bayesian framework [47] and later from an IBR perspective [7, 48]. Most recently, it has been applied to clustering [49].

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Dougherty, E.R. *Optimal Signal Processing Under Uncertainty*, SPIE Press: Bellingham, **2018**.
- [2] Kauffman, S. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, **1969**, 22(3), 437-467.
- [3] Qian, X.; Dougherty, E.R. Effect of function perturbation on the steady-state distribution of genetic regulatory networks: optimal structural intervention. *IEEE Trans. Sig. Process.*, **2008**, 56(10), 4966-4975.
- [4] Faure, A.; Naldi, A.; Chaaouiya, A.; Thieffry, D. Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics.*, **2006**, 22(14), 124-131.
- [5] Yoon, B.J.; Qian, X.; Dougherty, E.R. Quantifying the objective cost of uncertainty in complex dynamical systems. *IEEE Trans. Sig. Process.*, **2013**, 61(9), 2256-2266.
- [6] Qian, X.; Dougherty, E.R. Intervention in gene regulatory networks *via* phenotypically constrained control policies based on long-run behavior. *IEEE/ACM. Trans. Comput. Biol. Bioinform.*, **2012**, 9(1), 123-136.
- [7] Dalton, L.A.; Dougherty, E.R. Intrinsically optimal bayesian robust filtering. *IEEE Trans. Sig. Process.*, **2014**, 62(3), 657-670.
- [8] Yeang, C.H.; Mak, H. C.; McCuine, S.; Workman, C.; Jaakkola, T.; Ideker, T. Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome. Biol.*, **2005**, 6(7), 1974-1979.
- [9] Dehghannasiri, R.; Yoon, B.J.; Dougherty E.R. Optimal experimental design for gene regulatory networks in the presence of uncertainty. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **2015**, 14(4), 938-950.
- [10] Imani, M.; Dehghannasiri, R.; Braga-Neto, U.; Dougherty, E.R. Sequential experimental design for optimal structural intervention in gene regulatory networks based on the mean objective cost of uncertainty. *Cancer. Inform.*, **2018**, 17: (1176935118790247), 1-10.
- [11] Golub, T.R.; Slonim, D.K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J.P.; Coller, H.; Loh, M.L.; Downing, J.R.; Caligiuri, M.A.; Bloomfield, C.D.; Lander, E.S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science.*, **1999**, 286(5439), 531-537.
- [12] Dalton, L.A.; Dougherty, E.R. Optimal classifiers with minimum expected error within a bayesian framework – Part I: Discrete and gaussian models. *Pattern. Recognit.*, **2013**, 46(5), 1288-1300.
- [13] Dalton, L.A.; Dougherty, E.R. Optimal classifiers with minimum expected error within a bayesian framework – Part II: Properties and performance analysis. *Pattern. Recognit.*, **2013**, 46(5), 1301-1314.
- [14] Dougherty, E.R. Small sample issues for microarray-based classification. *Comp. Funct. Genom.*, **2001**, 2(1), 28-34.
- [15] Zollanvari, A.; Hua, J.; Dougherty, E.R. Analytic study of performance of linear discriminant analysis in stochastic settings. *Pattern. Recognit.*, **2013**, 46(11), 3017-3029.
- [16] Broumand, A.; Yoon, B.J.; Esfahani, M.S.; Dougherty, E.R. Discrete optimal bayesian classification with error-conditioned sequential sampling. *Pattern. Recognit.*, **2015**, 48(11), 3766-3782.
- [17] Knight, J.; Ivanov, I.; Dougherty, E.R. MCMC implementation of the optimal bayesian classifier for non-gaussian models: Model-based rna-seq classification. *BMC. Bioinform.*, **2014**, 15(1): 401.
- [18] Knight, J.; Ivanov, I.; Chapkin, R.; Dougherty E.R. Detecting multivariate gene interactions in rna-seq data using optimal bayesian classification. *IEEE/ACM. Trans. Comput. Biol. Bioinform.*, **2018**, 15(2), 484-493.
- [19] Nagaraja, K.; Braga-Neto, U. Bayesian classification of proteomics biomarkers from selected reaction monitoring data using an approximate bayesian computation-markov chain monte carlo approach. *Cancer. Inform.*, **2018**, 17(1):1-7.
- [20] Banerjee, U.; Braga-Neto, U. ABC-MCMC classification of liquid chromatography—mass spectrometry data. *Cancer. Inform.*, **2015**, 14(S5), 175-182.
- [21] Karbalayghareh, A.; Braga-Neto, U.M.; Dougherty, E.R. Intrinsically bayesian robust classifier for single-cell gene expression time series in gene regulatory networks. *BMC. Sys. Biol.*, **2018**, 12(Suppl 3); 23-37.
- [22] Dadaneh, S.Z.; Dougherty, E.R.; Qian, X. Optimal bayesian classification with missing values. *IEEE. Trans. Sig. Process.*, **2018**, 66(15), 4182-4192.
- [23] Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R.B. Missing value estimation methods for DNA microarrays. *Bioinformatics.*, **2001**, 17(6), 520-525.
- [24] Jaynes, E.T. Prior probabilities. *IEEE. Trans. Sys. Sci. Cybernet.*, **1968**, 4(3), 227-241.
- [25] Jaynes, E. What is the question? In: *Bayesian Statistics*; J.M. Bernardo, Ed.; Valencia University Press: Valencia, **1980**; Vol. 31, pp. 143-166.
- [26] Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. A, Math. Astro. Phy. Sci.*, **1946**, 186(1007), 453-461.
- [27] Kashyap, R. Prior probability and uncertainty. *IEEE Trans. Inf. Theory*, **1971**, 17(6), 641-650.

- [28] Zellner, A. *Economics and econometrics*, Department of Economics, University of Chicago Press: Chicago, **1995**.
- [29] Esfahani, M.S.; Knight, J.; Zollanvari, A.; Yoon, B.J.; Dougherty E. R. Classifier design given an uncertainty class of feature distributions *via* regularized maximum likelihood and the incorporation of biological pathway knowledge in steady-state phenotype classification. *Pattern. Recognit.*, **2013**, *46*(10), 2783-2797.
- [30] Esfahani, M.S.; Dougherty, E.R. An optimization-based framework for the transformation of incomplete biological knowledge into a probabilistic structure and its application to the utilization of gene/protein signaling pathways in discrete phenotype classification. *IEEE/ACM. Trans. Comput. Biol. Bioinform.*, **2015**, *12*(6), 1304-1321.
- [31] Boluki, S.; Esfahani, M.S.; Qian, X.; Dougherty, E.R. Incorporating biological prior knowledge for Bayesian learning *via* maximal knowledge-driven information priors. *BMC. Bioinform.*, **2017**, *18* (Suppl 14): 552.
- [32] Pan, S.J.; Yang, Q.A survey on transfer learning. *IEEE. Trans. Knowl. Data. Eng.*, **2010**, *22*(10), 1345-1359.
- [33] Karbalayghareh, A.; Qian, X.; Dougherty, E.R. Optimal Bayesian transfer learning. *IEEE Trans. Sig. Process.*, **2018**, *66*(14), 3724-3739.
- [34] Qian, X.; Dougherty, E.R. Bayesian regression with network prior: optimal bayesian filtering perspective. *IEEE. Trans. Sig. Process.*, **2016**, *64*(23), 6243-6253.
- [35] Gelman, A.; Carlin, J.; Stern, H.; Rubin, D. *Bayesian Data Analysis*, Chapman and Hall: Boca Raton, **1995**.
- [36] Bernardo, J.; Smith, A. *Bayesian Theory*, Wiley & Sons: Chichester, **2000**.
- [37] Bishop, C. *Pattern Recognition and Machine Learning*, Springer: New York, **2006**.
- [38] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer: New York, **2009**.
- [39] Murphy, K. *Machine Learning: A Probabilistic Perspective*, MIT Press: Cambridge, **2012**.
- [40] Silver, E.A. Markovian decision processes with uncertain transition probabilities or rewards. *Technical Report*, Defense Technical Information Center: Fort Belvoir, VA, **1963**.
- [41] Gozzolino, J.M.; Gonzalez-Zubieta, R.; Miller, R.L. Markovian decision processes with uncertain transition probabilities. *Technical Report*, Defense Technical Information Center: Fort Bevoir, VA, **1965**.
- [42] Martin, J.J. *Bayesian Decision Problems and Markov Chains*, John Wiley & Sons, Inc: New York, **1967**.
- [43] Yousefi, M.; Dougherty, E.R. A comparison study of optimal and suboptimal intervention policies for gene regulatory networks in the presence of uncertainty. *EURASIP J. Bioinfo. Sys. Biol.*, **2014**, *1*(6), 1-13.
- [44] Kuznetsov, V.P. Stable detection when the signal and spectrum of normal noise are inaccurately known. *Telecommun. Radio. Eng.*, **1976**, *30/31*(9), 58-64.
- [45] Kassam, S.A.; Lim, T.I. Robust wiener filters. *J. Franklin. Inst.*, **1977**, *304*(1), 171-185.
- [46] Poor, H. V. On robust wiener filtering. *IEEE. Trans. Auto. Control.*, **1980**, *25*(3), 531-536.
- [47] Grigoryan, A.M.; Dougherty, E.R. Bayesian robust optimal linear filters. *Sig. Process.*, **2001**, *81*(12), 2503-2521.
- [48] Dehghannasiri, R.; Esfahani, M.S.; Dougherty, E.R. Intrinsically Bayesian robust kalman filter: an innovation process approach. *IEEE. Trans. Sig. Process.*, **2017**, *65*(10), 2531-2546.
- [49] Dalton, L.A.; Benalcazar, M.E.; Dougherty, E.R. Optimal clustering under uncertainty. *Plos. One.*, **2018**, *13*(10): e0204627.