DISSERTATION

SLICED INVERSE APPROACH AND DOMAIN RECOVERY FOR STOCHASTIC INVERSE PROBLEMS

Submitted by

Jiarui Chi

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2021

Doctoral Committee:

Advisor: Haonan Wang
Co-Advisor: Don Estep

F. Jay Breidt
Simon Tavener
Wen Zhou

ABSTRACT

SLICED INVERSE APPROACH AND DOMAIN RECOVERY FOR STOCHASTIC INVERSE PROBLEMS

This dissertation tackles several critical challenges related to the Stochastic Inverse Problem (SIP) to perform scientific inference and prediction for complex physical systems which are characterized by mathematical models, e.g. differential equations. We treat both discrete and continuous cases. The SIP concerns inferring the values and quantifying the uncertainty of the inputs of a model, which are considered as random and unobservable quantities governing system behavior, by using observational data on the model outputs. Uncertainty of the inputs is quantified through probability distributions on the input domain which induce the probability distribution on the outputs realized by the observational data. The formulation of the SIP is based on rigorous measure-theoretic probability theory that uses all the information encapsulated in both the model and data. We introduce a problem in which a portion of the inputs can be observed and varied to study the hidden inputs, and we employ a formulation of the problem that uses all the knowledge in multiple experiments by varying the observable inputs.

Since the map that the model induces is typically not 1-1, an ansatz, i.e. an assumption of some prior information, is necessary to be imposed in order to determine a specific solution of the SIP. The resulting solution is heavily conditioned on the observable inputs and we seek to combine solutions from different values of the observable inputs in order to reduce that dependence. We propose an approach of combining the individual solutions based on the framework of the Dempster-Shafer theory, which removes the dependency on the experiments as well as the ansatz and provides useful distributional information about the unobservable inputs, more specifically, about the ansatz. We develop an iterative algorithm that updates the ansatz information in order to obtain a best form of the solution for all experiments. The philosophy of Bayesian approaches is

similar to that of the SIP in the sense that they both consider random variables as the model inputs and they seek to update the unobservable solution using information obtained from observations. We extend the classical Bayesian in the context of the SIP by incorporating the knowledge of the model.

The input domain is a pre-specified condition for the SIP given by the knowledge from scientists and is often assumed to be a compact metric space. The supports of the probability distributions computed in the SIP are restricted to the domain, and thus an inappropriate choice of domain might cause a massive loss of information in the solutions. Similarly, we combine the individual solutions from multiple experiments to recover a unique domain among many choices of domain induced by the distribution of the inputs in general cases. In particular, results on the convergence of the domain recovery in linear models are investigated.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Haonan Wang, and my co-advisor, Don Estep, for over six years of great support, guidance, and wisdom. I have been blessed with this wonderful team that has encouraged me and kept me motivated. I would also like to thank my committee - F. Jay Breidt, Simon Tavener, and Wen Zhou, for their belief in me, which inspired me to move forward and to continue pursuing my PhD. Finally, I would like to thank my fellow graduate students for their assistance, company, and advice during the life at Colorado State University.

DEDICATION

I would like to dedicate this thesis to my beloved parents.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

One of the important mathematical problems in science and engineering is the inverse problem (Tarantola, 2005; De Vito et al., 2005; Kaipio and Somersalo, 2006; Breidt et al., 2011; Butler et al., 2012, 2014, 2015). It is the process of studying the causal factors in a mathematical model, which is used to characterize a physical system of certain scientific and engineering discipline, from observed data on the model outputs. An inverse problem is the direct inverse of a forward problem that is the process of studying the behavior of the output from the causal factors in a physical system. For instance, we can observe data describing the flow of fluid through the ground and we seek to determine material properties from the data. Here, we commonly use a system of differential equations (porous media equations) to model the flow.

To be concrete, we consider a general model

$$y = Q(\boldsymbol{a}), \tag{1.1}$$

where Q is a map from the input domain Λ to the output range $\mathcal{D}$. In general, Q is a mathematical description of the behavior of a physical system, e.g. given by solving a system of differential equations that defines a dynamic system. Typically, Q is onto but not 1-1. Applications of (1.1) to practical problems involve the study of (1.1) as the input $\boldsymbol{a}$ varies over the domain Λ. Both the deterministic and stochastic cases are important. In this dissertation, we consider the stochastic case. Given (1.1), the deterministic forward problem refers to the process of computing the output y for a given $\boldsymbol{a}$, while the deterministic inverse problem is the process of determining $\boldsymbol{a}$ such that $Q(\boldsymbol{a}) = y$ for a given y. Stochastic versions of the problems are commonly considered for different purposes in practice, e.g. sensitivity analysis and uncertainty quantification. In the stochastic forward problem (SFP), $\boldsymbol{a}$ is considered as a random variable such that data of y can be observed through Q, which propagates stochastically. In other words, the uncertainty on the range

of the map is induced by the uncertainty on the input domain. Such computation is referred to as the "forward computation". The stochastic inverse problem (SIP) is the process of finding a probability distribution of $\boldsymbol{a}$ from the data given $\boldsymbol{x}$, i.e. the inverse process of the forward computation. The forward computation and its inverse can be explained by a simple example. The computation of probability of two sixes out of two rolls of a fair die is exactly the forward computation, and the inverse is computing the probability of a die being fair given that the two rolls are two sixes. They are also referred to as two types of probabilities: direct and inverse (Fisher, 1930; Aldrich et al., 1997; Senn et al., 2011). With a probability distribution on Λ, we can quantify the uncertainty of the input, determine regions of the input associated with probabilities that satisfy some threshold conditions, and make predictions about the output. To actually solve the SIP and find a distribution of the input, an approximate probability distribution of the output from the observational data is often used in practice. In the dissertation, we consider the SIP for a given probability distribution of the output.

As the inverse of the forward computation, solving the SIP involves a process of inverting the map Q. A significant complication is that Q is a "many-to-one" map. Rather than determining a single value for the input that reproduces the output, we consider all the values in the following inverse image of a value $y_0 \in \mathcal{D}$.

$$Q^{-1}(\{y_0\}) = \{\lambda \in \Lambda : Q(\lambda) = y_0\}.$$

Generally, $Q^{-1}(\{y_0\})$ is not a singleton. All points in $Q^{-1}(\{y_0\})$ are equivalent in the sense that they reproduce the same output value y_0. Thus, $Q^{-1}(\{y_0\})$ is an equivalence class, and the SIP has set-valued inverse solutions defined on the space of these equivalence classes. An analogous example to $Q^{-1}(\{y_0\})$ is the curve associated with a constant value in a contour map. Measure theory is often implemented to handle these set-valued inverse solutions, which also considers all the information in the model. However, the challenge is to extend the solutions to the physically meaningful domain Λ, since there is no information, additional to what given by the model, to distinguish points within an equivalence class. One approach, common in mathematics, is regu-

larization, which roughly speaking involves altering the map Q to get a 1-1 map $\hat{Q}$ and solving an altered inverse problem. That approach is inconsistent with probability theory. Alternatively, we can specify some "prior information" to obtain a unique solution. Such technique is referred to as the *disintegration*. A theoretical framework in recent years has been established through a series of papers (Butler et al., 2012, 2014, 2015) among which Butler et al. (2014) presents the most comprehensive work about extending the solutions to Λ by using the disintegration. They proposed an *ansatz* as an assumption of the distributions on the equivalence classes, which can be interpreted as an unbiased Bayesian prior. In particular, the uniform ansatz, under which the distribution on each equivalence class is uniform, is a "non-preferential" weighting determined by the underlying measure on the equivalence classes. In addition, it uses the least information and induces the largest support of the solution. Any other preferences could lead to inappropriate support of the solution. In the dissertation, we further show that the uniform ansatz induces the maximum entropy solution of the SIP.

The ansatz can be pre-specified according to one's belief. However, the resulting solution is heavily conditioned on the ansatz, opening the possibility of significant bias arising from the prior information. In the dissertation, we launch an observable input in the model (1.1) that can help mitigate the impact from the choice of ansatz. We consider a new model formulated as

$$y = Q(\boldsymbol{a}, \boldsymbol{x}), \tag{1.2}$$

where $\boldsymbol{a}$ is an unobservable input, $\boldsymbol{x}$ is an observable input, and Q is onto but not 1-1. This model in the context of an inverse problem describes the behavior of a physical system for different experiments. The role of $\boldsymbol{x}$ is analogous to that of a control variable by which the experiment can be governed to produce the corresponding output in the SFP. For multiple experiments, Butler et al. (2014) suggests an approach of solving the SIP measured on the product space of the collection of experiments. It is conducted by obtaining observational data from multiple experiments simultaneously, which means the data originate from the same realization of the unobservable input.

However, they are not feasible to obtain in many cases, such as time-to-event data. As a result of their approach, no additional ansatz information can be gained from adding up the experiments, and the solution is heavily conditioned on the experiments indexed by $\boldsymbol{x}$. Thus, we propose an approach that solves the SIP under the uniform ansatz for given observational data of the output in each individual experiment indexed by $\boldsymbol{x}$. The form of the SIP is now

$$Q^{-1}(\{y_1\}, \boldsymbol{x}) = \{\lambda \in \Lambda : Q(\lambda, \boldsymbol{x}) = y_1\},$$

where $y_1 \in Q(\Lambda, \boldsymbol{x})$ and $\boldsymbol{x}$ is varied in $\mathcal{X}$. Solutions from different experiments in which we constantly change the spaces of equivalence classes provide different "aspects" of the ansatz information. With respect to the geometry of the spaces of different equivalence classes, the solutions are considered to be "sliced" by $\boldsymbol{x}$.

In the dissertation, one of the two focuses is to reduce or remove the dependency of the SIP solution on the ansatz and the experiments indexed by $\boldsymbol{x}$. We provide a complete analysis in the discrete and continuous cases, i.e. regarding discrete and continuous domains of the inputs. We propose an approach that aggregates solutions from individual experiments using an "experimental average" to remove the dependency on $\boldsymbol{x}$. At the same time, the average reduces the effect of the ansatz in the solutions in some sense. We show that the average can be viewed as an extension of the Dempster-Shafer functions based on their framework (Yager and Liu, 2008; Zhang and Liu, 2011) for quantifying the uncertainty of the unobservable input. Most importantly, the average provides useful information about the generating distribution, essentially, about the ansatz on the equivalence classes. Thus, it can be used to update the ansatz iteratively in the solutions towards the "correct" one. We build a measure-theoretic algorithm and show convergence results under mild conditions.

In addition to the ansatz, the input domain Λ is also pre-specified by scientists given their related experience or knowledge and is commonly assumed to be contained in a compact metric space. In many cases, prior knowledge gives a domain that is much larger than the actual support of the distribution. All the computation are performed on Λ, and an inappropriate choice leads to a

high bias in the solution of which the support is restricted to Λ. Thus, determining a good approximation of the support is crucial to delivering an accurate result on the solution and interpreting it in the context of an application. For instance, by knowing the actual domain of the transmission rate of a disease studied in a region, we can directly identify the low-risk fractional parts where no transmission is ongoing. In the dissertation, we propose a general approach that defines a unique domain using the supports of the SIP solutions under the uniform ansatz from all possible experiments and we show the convergence results on the domain recovery in linear models.

Our approach used in solving the SIP can be viewed as an extension of the Bayesian paradigm to an inverse problem with random inputs. We also introduce several advantages by using a different methodology. Standard Bayesian approaches are based on Bayes' rule which is exploited to compute a posterior distribution that is given by a prior updated by the observational data. In contrast, we exploit disintegration of the observed distribution of the output to handle the inversion using all the information in Q. The ansatz is a non-biased prior that is used to construct the solution of the SIP, i.e. a *posteriori* conditioned on the observed distribution. Hence, the ansatz is a better form of prior, and in particular, the ansatz is objectively determined by the inverse images of Q and thus is "non-informative" about the inputs. Our approach to computing the *posteriori* by updating from the observed distribution is an extension of Bayesian method to the problem with random parameters. But using disintegration instead of Bayes' rule allows a truly nonparametric formulation and avoids non-computable scaling factors. Finally, our solution produces the observed distribution of the output, which is the metric we use to evaluate the solutions.

A related inverse problem in the field of statistics is Bayesian calibration (Sacks et al., 1989; Kennedy and O'Hagan, 2001) in which they treat the inputs of a deterministic model as random variables. Mathematical models are often used to describe physical processes and are typically implemented in computer models (or codes). To make predictions in a specific context by a computer model, it may be necessary to *calibrate* the model by adjusting the unknown inputs such that the outputs of the model *fit* the observed data of the physical process. Bayesian calibration considers the unknown inputs as parameters in calibration models which often adopt Gaussian processes to

represent the true process and the computer model, and uses the observed data to derive a posterior distribution of the parameters to quantify the uncertainty, as well as to find the best-fitting value of the parameters. However, this Bayesian approach completely treats the model as a "black box" (Kennedy and O'Hagan, 2001) and ignores the information about the mathematical model, i.e. Q in our context, for the benefit of a low computational complexity. Our approach exploits the structure of the model Q to its full potential such that we have better predictions for the future behavior of the process.

In a regression problem in which a model is adopted to study the relationship between the output responses and the input factors, implementing our approach needs additional treatment to the model. More specifically, Q in the SIP is a theoretically defined map that is often approximated by the established system of differential equations and solved by numerical methods in practice such that the approximation error from selecting a predictive model is minimized up to certain pre-specified threshold. A general regression problem often induces an unknown and abstract map that describes a relationship. It may be necessary to use a good approximation of Q such that the error of discovering the distribution of the parameters as well as the prediction error of the output can be controlled. For uncertainty in the parameters, Bayesian approaches to meta-analysis (Smith et al., 1995; Senn, 2000; Higgins et al., 2009) have provided a reasonable interpretation. In a random-effects meta-analysis, since the true effects in each clinical trial are not necessarily equal, they assume that the true effects are random observations from a common population distribution, and the random observations of the outcome deviate from the true effects due to the noise. Uncertainty of the observations of the outcome are truly propagated from uncertainty of the true effects, which coincides with the philosophy of the SIP. Thus, in the dissertation, we adapt this Bayesian methodology, in which they essentially use a family of distributions (e.g. Gaussian) to approximate the population distribution, to solve the SIP by incorporating the given map Q.

The remainder of the dissertation is organized as follows. In Chapter 2, we introduce the SIP in discrete domains and we propose a novel approach that uses all possible experiments in solving the SIP to find the best approximation of the distribution of unobservable inputs of a model.

In addition, we extend the a Bayesian approach and the DS theory in the context of the SIP. In Chapter 3, we switch the results in discrete domains to those in continuous domains for more practical methodologies. In Chapter 4, we devise an approach to recover the actual domain of the unobservable inputs by finding a unique domain that has some "mini-max" property, and we show results on rates of convergence of the domain recovery. Summary and future work are given in Chapter 6.

Chapter 2

Sliced Inverse Approach for Stochastic Inverse Problems: Discrete Distributions

2.1 Stochastic Inverse Problems in Discrete Domains

For illustration purpose, we consider two unobservable random variables a and b on the finite domain

$$\Lambda = \{(a,b) : a \in \{1,2,3\}, b \in \{1,2,3\}\} = \begin{bmatrix} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \\ (3,1) & (3,2) & (3,3) \end{bmatrix}, \tag{2.1}$$

and (a,b) is distributed according to

$$P_\Lambda = \{p_{i,j} : i \in \{1,2,3\}, j \in \{1,2,3\}\} = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} = \begin{bmatrix} 1/32 & 1/12 & 1/8 \\ 1/24 & 1/2 & 1/12 \\ 1/32 & 1/24 & 1/16 \end{bmatrix} \tag{2.2}$$

The distribution P_Λ is defined on $(\Lambda, \mathcal{P}_\Lambda)$ where $\mathcal{P}_\Lambda$ is the power set of Λ. Define

$$f(n) = \begin{cases} 1 & n \text{ is even,} \\ 0 & \text{otherwise,} \end{cases}$$

and further define

$$Q_x(a,b) = f(ax+b), \tag{2.3}$$

where x is an observable deterministic input taking values in $\mathcal{X} = \{1, 2\}$. Specifically,

$$Q_1(a, b) = f(a + b),$$
$$Q_2(a, b) = f(2a + b) = f(b).$$

Then, the output range of the map Q_x is $\mathcal{D} = \{0, 1\}$.

In the statistical literature, this model can be viewed as a generalized linear mixed model (GLMM), in which a is a random slope and b is the quantity representing the intercept as well as the confounder and measurement error. This can also be interpreted as a generalized factor model in which a and b are two unknown factors. In another viewpoint, the model (2.3) characterizes a data generating process in the sense that, given x and a probability distribution of (a, b), a probability distribution of the output $Q_{\boldsymbol{x}}$ can be uniquely determined. This is referred to as the *stochastic forward problem* (SFP; Butler et al. (2014), Butler et al. (2015)). Complementing the SFP, the *stochastic inverse problem* (SIP;Butler et al. (2014), Butler et al. (2015)) is the problem of recovering a probability distribution of (a, b) from the observed probability distribution of Q_x given x. In the following, we show theoretic results of solutions of the SIP and explicit examples according to the distribution (2.2) and the model (2.3).

2.1.1 Established Solutions of SIPs

In this section, we first describe the SIP in a general situation. Let $Q(\boldsymbol{x}, \boldsymbol{a})$ be a general system where Q is a measurable map, $\boldsymbol{a}$ is defined on a finite probability space $(\Lambda, \mathcal{P}_\Lambda, P_\Lambda)$, and $\boldsymbol{x}$ is defined on a finite space $\mathcal{X}$. In particular, Λ and $\mathcal{X}$ are finite metric spaces.

For a given $\boldsymbol{x} \in \mathcal{X}$, $Q_{\boldsymbol{x}}(\cdot) \equiv Q(\boldsymbol{x}, \cdot)$ denotes a measurable map, indexed by $\boldsymbol{x}$. For any $\boldsymbol{x}$, if we denote the range of $Q_{\boldsymbol{x}}$ by $\mathcal{D}_{\boldsymbol{x}} = Q_{\boldsymbol{x}}(\Lambda)$, then $Q_{\boldsymbol{x}}$ is a map from Λ to $\mathcal{D}_{\boldsymbol{x}}$. In this paper, we assume that $Q_{\boldsymbol{x}}$ is not one-to-one and the inverses of $Q_{\boldsymbol{x}}$ of two distinct points in $\mathcal{D}_{\boldsymbol{x}}$ are disjoint, i.e. $Q_{\boldsymbol{x}}^{-1}(y_1) \cap Q_{\boldsymbol{x}}^{-1}(y_2) = \emptyset$ where $y_1 \neq y_2$ and $y_1, y_2 \in \mathcal{D}_{\boldsymbol{x}}$. Here, the inverse map $Q_{\boldsymbol{x}}^{-1}$ can be defined as

$$Q_{\boldsymbol{x}}^{-1}(C) = \{\lambda \in \Lambda : Q_{\boldsymbol{x}}(\lambda) \in C\},$$

where $Q_{\boldsymbol{x}}^{-1}$ maps into $\mathcal{P}_\Lambda$ and C is any event in the power set $\mathcal{P}_{\mathcal{D}_{\boldsymbol{x}}}$ of $\mathcal{D}_{\boldsymbol{x}}$. This characterizes a geometrical property of Λ which will be explained later in this section.

Probabilistically speaking, a stochastic forward problem involves finding a probability distribution of the output $Q_{\boldsymbol{x}}$ for a given P_Λ through the following "forward computation". The probability distribution P_Λ on the domain Λ induces a probability distribution $P_{\mathcal{D}_{\boldsymbol{x}}}$ on the range $\mathcal{D}_{\boldsymbol{x}}$ through

$$P_{\mathcal{D}_{\boldsymbol{x}}}(C) = P_\Lambda(Q_{\boldsymbol{x}}^{-1}(C)). \tag{2.4}$$

In this paper, P_Λ is referred to as the *generating distribution* and $P_{\mathcal{D}_{\boldsymbol{x}}}$ is referred to as the *output distribution*.

Thus, given the distribution (2.2) and the model (2.3), we can obtain the probability distribution $P_{\mathcal{D}_x}$ of the output Q_x as follows. For $x = 1$,

$$\begin{bmatrix} P_{\mathcal{D}_1}(0) \\ P_{\mathcal{D}_1}(1) \end{bmatrix} = \begin{bmatrix} P_\Lambda(Q_1^{-1}(0)) \\ P_\Lambda(Q_1^{-1}(1)) \end{bmatrix} = \begin{bmatrix} p_{1,2} + p_{2,1} + p_{2,3} + p_{3,2} \\ p_{1,1} + p_{1,3} + p_{2,2} + p_{3,1} + p_{3,3} \end{bmatrix} = \begin{bmatrix} 1/4 \\ 3/4 \end{bmatrix},$$

and for $x = 2$,

$$\begin{bmatrix} P_{\mathcal{D}_2}(0) \\ P_{\mathcal{D}_2}(1) \end{bmatrix} = \begin{bmatrix} P_\Lambda(Q_2^{-1}(0)) \\ P_\Lambda(Q_2^{-1}(1)) \end{bmatrix} = \begin{bmatrix} p_{1,1} + p_{1,3} + p_{2,1} + p_{2,3} + p_{3,1} + p_{3,3} \\ p_{1,2} + p_{2,2} + p_{3,2} \end{bmatrix} = \begin{bmatrix} 3/8 \\ 5/8 \end{bmatrix},$$

where $Q_x^{-1}(y) = \{(a, b) : Q_x(a, b) = y\}$.

Then, a stochastic inverse problem is to find the probability distribution P_Λ from $P_{\mathcal{D}_{\boldsymbol{x}}}$, through inverting the measurable map $Q_{\boldsymbol{x}}$. Two distinct values of $\boldsymbol{a}$, say $\lambda_1, \lambda_2 \in \Lambda$, are *equivalently* relative to $\boldsymbol{x}$, denoted by $\lambda_1 \sim \lambda_2$, if there exists a $y \in \mathcal{D}_{\boldsymbol{x}}$ such that $\lambda_1, \lambda_2 \in Q_{\boldsymbol{x}}^{-1}(y)$. In addition, for $y_1, y_2 \in \mathcal{D}_{\boldsymbol{x}}$ and $y_1 \neq y_2$, we have $Q_{\boldsymbol{x}}^{-1}(y_1) \cap Q_{\boldsymbol{x}}^{-1}(y_2) = \emptyset$. Thus, $Q_{\boldsymbol{x}}^{-1}(y)$ can be viewed as an equivalence class indexed by $y \in \mathcal{D}_{\boldsymbol{x}}$, which contains all values of $\boldsymbol{a}$ corresponding to the same value of y. An equivalence class can also be depicted as a manifold in Λ, which is referred to as a *generalized contour* (Butler et al., 2014). Consequently, the domain Λ can be decomposed as the

union of equivalence classes

$$\Lambda = \bigcup_{y \in \mathcal{D}_{\boldsymbol{x}}} Q_{\boldsymbol{x}}^{-1}(y).$$

If $\mathcal{L}_{\boldsymbol{x}}$ denotes the quotient space $\Lambda/\sim$, then each point in $\mathcal{L}_{\boldsymbol{x}}$ represents a generalized contour in Λ. More specifically, for any $\lambda \in \Lambda$, $Q_{\boldsymbol{x}}^{-1}(Q_{\boldsymbol{x}}(\lambda)) \in \Lambda$ is a generalized contour in Λ, and corresponds to a point in $\mathcal{L}_{\boldsymbol{x}}$, denoted by $\mathcal{E}_\lambda$. Furthermore, for any $A \in \mathcal{P}_\Lambda$, let $\mathcal{E}_A$ be a collection of generalized contours in Λ, i.e.

$$\mathcal{E}_A = \{\mathcal{E}_\lambda : \lambda \in A\}.$$

Then the inverse map $Q_{\boldsymbol{x}}^{-1}$ is a one-to-one and onto map from $\mathcal{D}_{\boldsymbol{x}}$ to $\mathcal{L}_{\boldsymbol{x}}$. In the SIP, for a known probability space $(\mathcal{D}_{\boldsymbol{x}}, \mathcal{P}_{\mathcal{D}_{\boldsymbol{x}}}, P_{\mathcal{D}_{\boldsymbol{x}}})$, $Q_{\boldsymbol{x}}$ *uniquely* induces a probability distribution $P_{\mathcal{L}_{\boldsymbol{x}}}$ on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{P}_{\mathcal{L}_{\boldsymbol{x}}})$ through

$$P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = P_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad A \in \mathcal{P}_\Lambda.$$

Thus, we obtain a solution of the SIP specifically on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{P}_{\mathcal{L}_{\boldsymbol{x}}})$. This solution is uniquely determined by the map $Q_{\boldsymbol{x}}^{-1} : \mathcal{D}_{\boldsymbol{x}} \to \mathcal{L}_{\boldsymbol{x}}$ and induces $P_{\mathcal{D}_{\boldsymbol{x}}}$ in the SFP through the map.

In the previous example, the domain Λ in (2.1) can be decomposed according to the values in $\mathcal{D}_{\boldsymbol{x}}$ as

$$\Lambda = Q_x^{-1}(0) \cup Q_x^{-1}(1),$$

where the generalized contours $Q_x^{-1}(0)$ and $Q_x^{-1}(1)$ are two disjoint subsets of Λ. Specifically, we have

$$Q_1^{-1}(0) = \{(1,2),(2,1),(2,3),(3,2)\},$$
$$Q_1^{-1}(1) = \{(1,1),(1,3),(2,2),(3,1),(3,3)\},$$

and

$$Q_2^{-1}(0) = \{(1,1),(1,3),(2,1),(2,3),(3,1),(3,3)\},$$
$$Q_2^{-1}(1) = \{(1,2),(2,2),(3,2)\}.$$

In this case, each $Q_x^{-1}(\cdot)$ is an equivalence class in the sense that $Q_x(\lambda_1) = Q_x(\lambda_2)$ for any $\lambda_1, \lambda_2 \in \Lambda$. Let $\mathcal{L}_x$ denote the set of generalized contours, i.e. $\mathcal{L}_x = \{\ell_{\text{odd}}, \ell_{\text{even}}\}$ where ℓ_{odd} represents $Q_x^{-1}(0)$ and ℓ_{even} represents $Q_x^{-1}(1)$. Since $\mathcal{L}_x$ is isomorphic to $\mathcal{D}$, the probability distribution $P_{\mathcal{L}_x}$ on $(\mathcal{L}_x, \mathcal{P}_{\mathcal{L}_x})$ is the same as $P_{\mathcal{D}_x}$, i.e. $P_{\mathcal{L}_x}(\ell_{\text{odd}}) = P_{\mathcal{D}_x}(0)$ and $P_{\mathcal{L}_x}(\ell_{\text{even}}) = P_{\mathcal{D}_x}(1)$ where $x \in \mathcal{X}$.

The remaining problem, however, is that a probability distribution of $\boldsymbol{a}$ on Λ is of particular interest since each point in the domain Λ is physically and scientifically meaningful, not in the quotient space $\mathcal{L}_{\boldsymbol{x}}$. We seek solutions of the SIP on Λ.

Note that, for $A \in \mathcal{P}_\Lambda$, we have $\mathcal{E}_A \in \mathcal{P}_{\mathcal{L}_{\boldsymbol{x}}}$. The correspondence between λ and $\mathcal{E}_\lambda$ is defined as the *equivalence map* $\pi_{\mathcal{L}_{\boldsymbol{x}}} : \Lambda \to \mathcal{L}_{\boldsymbol{x}}$; i.e., for any $\lambda \in \Lambda$, $\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda) = \mathcal{E}_\lambda$. Consequently, we have $\pi_{\mathcal{L}_{\boldsymbol{x}}}(A) = \mathcal{E}_A$. In this case, we can embed $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{P}_{\mathcal{L}_{\boldsymbol{x}}})$ into $(\Lambda, \mathcal{P}_\Lambda)$. The issue is thus relating distributions on Λ to distributions on $\mathcal{L}_{\boldsymbol{x}}$. Since $Q_{\boldsymbol{x}}$ is not 1-1, we use the following decomposition of the generating distribution to extend the solution on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}})$ to $(\Lambda, \mathcal{B}_\Lambda)$.

Theorem 2.1.1.

$$P_\Lambda(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} P_\ell(\lambda), \quad \forall A \in \mathcal{P}_\Lambda, \tag{2.5}$$

where $\boldsymbol{x} \in \mathcal{X}$. The conditional probability $P_\ell(\lambda)$ is computed as

$$P_\ell(\lambda) = \frac{P_\Lambda(\lambda)}{P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)},$$

where $P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) > 0$ *and* $\ell = \pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)$*, and* $P_\ell(\lambda) \equiv 0$ *where* $P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = 0$ *and* $\ell = \pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)$*. In addition, the conditional probabilities satisfy*

$$\sum_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P_\ell(\lambda) = 1, \tag{2.6}$$

for $\ell \in \mathcal{L}_{\boldsymbol{x}}$ *and* $P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) > 0$*.*

Proof. Any event $A \in \mathcal{P}_\Lambda$ can be disjointly decomposed as

$$A = \bigcup_{\ell \in \mathcal{E}_A} (A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)).$$

Hence, we have

$$\begin{aligned}
\sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P_\ell(\lambda) &= \sum_{\ell \in \mathcal{E}_A - \mathcal{N}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P_\ell(\lambda) \\
&= \sum_{\ell \in \mathcal{E}_A - \mathcal{N}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} \frac{P_\Lambda(\lambda)}{P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} \\
&= \sum_{\ell \in \mathcal{E}_A - \mathcal{N}_A} \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P_\Lambda(\lambda) \\
&= \sum_{\ell \in \mathcal{E}_A - \mathcal{N}_A} \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P_\Lambda(\lambda) + \sum_{\ell \in \mathcal{N}_A} \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P_\Lambda(\lambda) \\
&= \sum_{\lambda \in A} P_\Lambda(\lambda),
\end{aligned}$$

where $\mathcal{N}_A = \{\ell \in \mathcal{L}_{\boldsymbol{x}} : P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = 0\}$ and $P_\Lambda(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)) = P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = 0$ for $\ell \in \mathcal{N}_A$. In addition, for $\ell \in \mathcal{L}_{\boldsymbol{x}}$ and $P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) > 0$, $P_\Lambda(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)) = P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) > 0$ and thus $\sum_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P_\ell(\lambda) = 1$. □

In the continuous probability case, such decomposition is known as the *disintegration*; see Butler et al. (2014) for more details. This result shows that the generating distribution P_Λ can be computed by iterated sums of a conditional probability P_ℓ on each generalized contour and a marginal probability $P_{\mathcal{L}_{\boldsymbol{x}}}$ on the set of generalized contours.

Returning to the example, we let $\pi_{\mathcal{L}_x}(i,j) = \ell_{\text{odd}}$ if $\{i,j\} \in Q_x^{-1}(0)$ and $\pi_{\mathcal{L}_x}(i,j) = \ell_{\text{even}}$ otherwise. Consequently, $\pi_{\mathcal{L}_x}^{-1}(\ell_{\text{odd}}) = Q_x^{-1}(0)$ and $\pi_{\mathcal{L}_x}^{-1}(\ell_{\text{even}}) = Q_x^{-1}(1)$. Then, any event A in $\mathcal{P}_\Lambda$ can be decomposed as

$$A = (A \cap \pi_{\mathcal{L}_x}^{-1}(\ell_{\text{odd}})) \cup (A \cap \pi_{\mathcal{L}_x}^{-1}(\ell_{\text{even}})) = (A \cap Q_x^{-1}(0)) \cup (A \cap Q_x^{-1}(1)),$$

where $x \in \mathcal{X}$. In this case, to compute the probability of any event A, we should first compute the probabilities of $\pi_{\mathcal{L}_x}^{-1}(\cdot)$ according to $P_{\mathcal{L}_{\boldsymbol{x}}}$, and then compute the conditional probabilities of $A \cap \pi_{\mathcal{L}_x}^{-1}(\cdot)$ in $\pi_{\mathcal{L}_x}^{-1}(\cdot)$, before summation of all probabilities. Thus, we have

$$P_\Lambda(A) = P_{\mathcal{L}_x}(\ell_{\text{odd}}) \sum_{(i,j)\in A\cap\pi_{\mathcal{L}_x}^{-1}(\ell_{\text{odd}})} P_{\ell_{\text{odd}}}(i,j) + P_{\mathcal{L}_x}(\ell_{\text{even}}) \sum_{(i,j)\in A\cap\pi_{\mathcal{L}_x}^{-1}(\ell_{\text{even}})} P_{\ell_{\text{even}}}(i,j), \quad \forall A \in \mathcal{P}_\Lambda.$$

where the conditional probabilities $\{P_\ell(i,j), i,j = 1,2,3\}$ along the generalized contours are computed as

$$x = 1\text{:} \begin{bmatrix} 1/24 & 1/3 & 1/6 \\ 1/6 & 2/3 & 1/3 \\ 1/24 & 1/6 & 1/12 \end{bmatrix}, \quad x = 2\text{:} \begin{bmatrix} 1/12 & 2/15 & 1/3 \\ 1/9 & 4/5 & 2/9 \\ 1/12 & 1/15 & 1/6 \end{bmatrix}. \tag{2.7}$$

In fact, (2.5) provides a way to obtain P_Λ given $\{P_\ell\}$ and $P_{\mathcal{L}_{\boldsymbol{x}}}$ in the SIP. We know the probability distribution for the set of generalized contours, $P_{\mathcal{L}_{\boldsymbol{x}}}$, but we have no information about the conditional probabilities $\{P_\ell\}$ since we have no information about P_Λ. In this case, we can specify any probability distribution for points in $\pi_{\mathcal{L}_x}^{-1}(\cdot)$. Such distributional assumption is referred to as the *ansatz*; see Butler et al. (2014). Then, any ansatz should satisfy (2.6), and different choices yield different solutions of P_Λ. This is a *fundamental characteristic* of the SIP and reflects the physical properties of the system being modeled by the map $Q_{\boldsymbol{x}}$.

Note that if we make an azsatz for the conditional probabilities $\{P_\ell\}$, we have a solution of the SIP of P_Λ, called an *(sliced) inverse distribution*, where the term "sliced" refers to the fact that the solution is conditioned on the specific $\boldsymbol{x}$ and thus on the generalized contours $\{Q_{\boldsymbol{x}}^{-1}(y)\}_{y\in\mathcal{D}_{\boldsymbol{x}}}$. Such approach from the SIP solutions is called the *sliced inverse* approach. There is no guarantee

that P_Λ and an inverse distribution are close for any $\boldsymbol{x}$. Butler et al. (2014) makes a probability argument to choose the uniform ansatz, that is,

$$\text{Uniform ansatz: } P_\ell^u(\lambda) = \frac{1}{\text{Card}(\pi_{\mathcal{L}_x}^{-1}(\ell))},$$

where $\lambda \in \pi_{\mathcal{L}_x}^{-1}(\ell)$ and $\text{Card}(\cdot)$ is the cardinality of a finite set. Essentially, in the absence of any information, each point in an equivalence class $\pi_{\mathcal{L}_x}^{-1}(\ell)$ is equally likely. Then, we have

Theorem 2.1.2. *For each $\boldsymbol{x} \in \mathcal{X}$,*

$$P_{\Lambda,\boldsymbol{x}}(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} P_\ell^u(\lambda), \quad \forall A \in \mathcal{P}_\Lambda, \tag{2.8}$$

is a probability distribution on $(\Lambda, \mathcal{P}_\Lambda)$ and induces $P_{\mathcal{D}_{\boldsymbol{x}}}$ through (2.4)*.*

Proof. We first have

$$P_{\Lambda,\boldsymbol{x}}(\Lambda) = \sum_{\ell \in \mathcal{E}_\Lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} P_\ell^u(\lambda) = P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_\Lambda) = P_\Lambda(\Lambda) = 1,$$

and we have $0 \leq P_{\Lambda,\boldsymbol{x}}(A) \leq 1$ for any $A \in \mathcal{P}_\Lambda$. Then we have, for any countable collection of pairwise disjoint subsets $\{A_i\}$ of Λ,

$$\begin{aligned}
P_{\Lambda,\boldsymbol{x}}\left(\bigcup_i A_i\right) &= \sum_{\ell \in \cup_i \mathcal{E}_{A_i}} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in \cup_i (A_i \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell))} P_\ell^u(\lambda) \\
&= \sum_i \left(\sum_{\ell \in \mathcal{E}_{A_i}} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in \cup_i (A_i \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell))} P_\ell^u(\lambda) \right) \\
&= \sum_i \left(\sum_{\ell \in \mathcal{E}_{A_i}} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A_i \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} P_\ell^u(\lambda) \right) \\
&= \bigcup_i P_{\Lambda,\boldsymbol{x}}(A_i),
\end{aligned}$$

since $A_j \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = \emptyset$ when $\ell \in \mathcal{E}_{A_i}$ and $i \neq j$. Finally, for any $C \in \mathcal{P}_{\mathcal{D}_{\boldsymbol{x}}}$, we have

$$\begin{aligned} P_{\Lambda,\boldsymbol{x}}(Q_{\boldsymbol{x}}^{-1}(C)) &= \sum_{\ell \in \mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)}} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in Q_{\boldsymbol{x}}^{-1}(C) \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^u_\ell(\lambda) = \sum_{\ell \in \mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)}} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^u_\ell(\lambda) \\ &= P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)}) = P_\Lambda(Q_{\boldsymbol{x}}^{-1}(C)) = P_{\mathcal{D}_{\boldsymbol{x}}}(C). \end{aligned}$$

This implies the inverse distribution induces the output distribution for the specific $\boldsymbol{x} \in \mathcal{X}$. □

This result shows that the inverse distribution for each $\boldsymbol{x} \in \mathcal{X}$ exactly reproduces the output distribution.

In the simple example, the uniform ansatz $\{P^u_\ell(i,j), i,j = 1,2,3\}$ is formulated as

$$x = 1\text{:} \begin{bmatrix} 1/5 & 1/4 & 1/5 \\ 1/4 & 1/5 & 1/4 \\ 1/5 & 1/4 & 1/5 \end{bmatrix}, \quad x = 2\text{:} \begin{bmatrix} 1/6 & 1/3 & 1/6 \\ 1/6 & 1/3 & 1/6 \\ 1/6 & 1/3 & 1/6 \end{bmatrix}.$$

In this case, we can obtain a solution to this SIP, that is a probability distribution of (a,b), as

$$P_{\Lambda,x}(A) = P_{\mathcal{L}_x}(\ell_{\text{odd}}) \sum_{(i,j) \in A \cap \pi^{-1}_{\mathcal{L}_x}(\ell_{\text{odd}})} P^u_{\ell_{\text{odd}}}(i,j) + P_{\mathcal{L}_x}(\ell_{\text{even}}) \sum_{(i,j) \in A \cap \pi^{-1}_{\mathcal{L}_x}(\ell_{\text{even}})} P^u_{\ell_{\text{even}}}(i,j), \quad \forall A \in \mathcal{P}_\Lambda \tag{2.9}$$

under $\{P^u_\ell(i,j), i,j = 1,2,3\}$. Specifically,

$$P_{\Lambda,1} = \begin{bmatrix} 3/20 & 1/16 & 3/20 \\ 1/16 & 3/20 & 1/16 \\ 3/20 & 1/16 & 3/20 \end{bmatrix}, \begin{bmatrix} \hat{P}_{\mathcal{D}_1}(0) \\ \hat{P}_{\mathcal{D}_1}(1) \end{bmatrix} = \begin{bmatrix} P_{\Lambda,1}(Q_1^{-1}(0)) \\ P_{\Lambda,1}(Q_1^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 1/4 \\ 3/4 \end{bmatrix} = \begin{bmatrix} P_{\mathcal{D}_1}(0) \\ P_{\mathcal{D}_1}(1) \end{bmatrix},$$

$$\begin{bmatrix} P'_{\mathcal{D}_2}(0) \\ P'_{\mathcal{D}_2}(1) \end{bmatrix} = \begin{bmatrix} P_{\Lambda,1}(Q_2^{-1}(0)) \\ P_{\Lambda,1}(Q_2^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 29/40 \\ 11/40 \end{bmatrix} \neq \begin{bmatrix} P_{\mathcal{D}_2}(0) \\ P_{\mathcal{D}_2}(1) \end{bmatrix},$$

$$P_{\Lambda,2} = \begin{bmatrix} 1/16 & 5/24 & 1/16 \\ 1/16 & 5/24 & 1/16 \\ 1/16 & 5/24 & 1/16 \end{bmatrix}, \begin{bmatrix} \hat{P}_{\mathcal{D}_2}(0) \\ \hat{P}_{\mathcal{D}_2}(1) \end{bmatrix} = \begin{bmatrix} P_{\Lambda,2}(Q_2^{-1}(0)) \\ P_{\Lambda,2}(Q_2^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 3/8 \\ 5/8 \end{bmatrix} = \begin{bmatrix} P_{\mathcal{D}_2}(0) \\ P_{\mathcal{D}_2}(1) \end{bmatrix},$$

$$\begin{bmatrix} P'_{\mathcal{D}_1}(0) \\ P'_{\mathcal{D}_1}(1) \end{bmatrix} = \begin{bmatrix} P_{\Lambda,2}(Q_1^{-1}(0)) \\ P_{\Lambda,2}(Q_1^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 13/24 \\ 11/24 \end{bmatrix} \neq \begin{bmatrix} P_{\mathcal{D}_1}(0) \\ P_{\mathcal{D}_1}(1) \end{bmatrix}.$$

In general, the inverse distribution $P_{\Lambda,\boldsymbol{x}}$ induces $P_{\mathcal{D}_{\boldsymbol{x}}}$ through the map $Q_{\boldsymbol{x}}$ for $\boldsymbol{x}$, but not for any other value in $\mathcal{X}$. For instance, $P_{\Lambda,1}$ induces $P_{\mathcal{D}_1}$ through the inverse of Q_1 but does not induces $P_{\mathcal{D}_2}$ through the inverse of Q_2. Thus, the inverse distribution is heavily conditioned on $\boldsymbol{x}$ in general. Figure 2.1 shows the plots of $P_\Lambda, P_{\Lambda,1}$, and $P_{\Lambda,2}$ in panels (a)-(c), respectively. In panels (b)-(c), the yellow lines indicate the points in generalized contours $Q_x^{-1}(1)$ where $x \in \mathcal{X}$, and the blue lines indicate the points in generalized contours $Q_x^{-1}(0)$ where $x \in \mathcal{X}$.

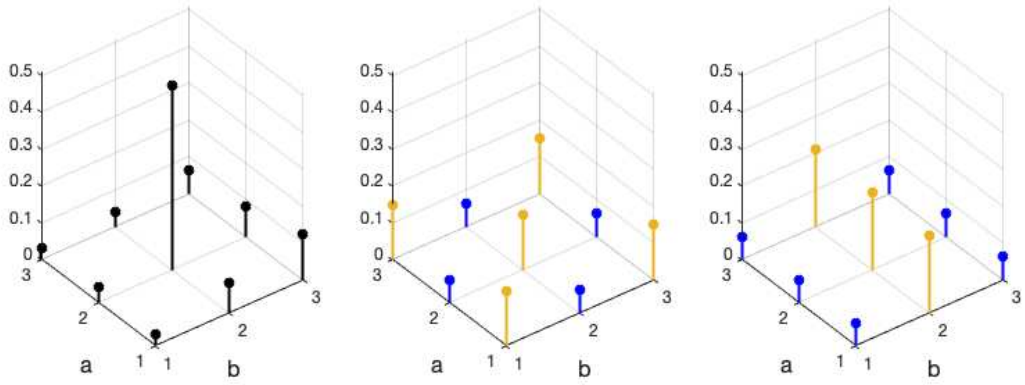


Figure 2.1: A graphical display of $P_\Lambda, P_{\Lambda,1}$, and $P_{\Lambda,2}$ in panels (a)-(c), respectively. In panels (b)-(c), the points in generalized contours $Q_x^{-1}(1)$ where $x \in \mathcal{X}$ are shown as yellow lines, while the points in generalized contours $Q_x^{-1}(0)$ where $x \in \mathcal{X}$ are shown as blue lines.

It is important to note that $P_{\Lambda,\boldsymbol{x}}$ in the sliced inverse approach is specifically constructed for a value of $\boldsymbol{x}$, which is not a desirable solution since the generating distribution of $\boldsymbol{a}$ does not depend on $\boldsymbol{x}$ in the SFP. In fact, this is due to the assumption of the uniform ansatz $\{P_\ell^u(i,j), i,j = 1,2,3\}$,

which is clearly different from the "true" conditional distributions $\{P_\ell(i,j), i,j = 1,2,3\}$. Bias from the choice of the ansatz leads to the dependency of inverse distributions on $\boldsymbol{x}$.

2.1.2 Uniform Ansatz and Maximum Entropy Solution

In this section, we explore the role of ansatz in the sliced inverse approach, namely, the distributions along the generalized contours defined by the map $Q_{\boldsymbol{x}}$. Recall that $Q_{\boldsymbol{x}}$ is a not 1-1, and thus, by Theorem 2.1.1, we have multiple solutions depending on the choice of the ansatz. Explicitly, given a value of $\boldsymbol{x}$ in the model and the domain Λ, an ansatz is a choice of

$$\mathcal{A}_{\mathcal{L}_{\boldsymbol{x}}} = \{\tilde{P}_\ell : \ell \in \mathcal{L}_{\boldsymbol{x}}\},$$

where a point ℓ in $\mathcal{L}_{\boldsymbol{x}}$ represents the generalized contour $\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)$ in Λ and $\tilde{P}_\ell$ can be any probability distribution along $\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)$.

Under the uniform ansatz, for each $\ell \in \mathcal{L}_{\boldsymbol{x}}$, $\tilde{P}_\ell$ is distributed according to the uniform distribution on the generalized contour $\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)$. Hence, the uniform ansatz can be written as

$$\mathcal{A}^u_{\mathcal{L}_{\boldsymbol{x}}} := \{\tilde{P}_\ell : \ell \in \mathcal{L}_{\boldsymbol{x}}, \tilde{P}_\ell \text{ is uniform on } \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\},$$

which is a set of conditional uniform distributions on generalized contours. Another choice of $\mathcal{A}_{\mathcal{L}_{\boldsymbol{x}}}$ is given by the binomial ansatz, which can be written as

$$\mathcal{A}^{bin}_{\mathcal{L}_{\boldsymbol{x}}} := \{\tilde{P}_\ell : \ell \in \mathcal{L}_{\boldsymbol{x}}, \tilde{P}_\ell \text{ is binomial on } \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \text{ with pre-specified success probability}\}.$$

Similarly, we can obtain an inverse distribution under the binomial ansatz,

$$P^{bin}_{\Lambda,\boldsymbol{x}}(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^{bin}_\ell(\lambda), \quad \forall A \in \mathcal{P}_\Lambda,$$

where $P_\ell^{bin} \in \mathcal{A}_{\mathcal{L}_{\boldsymbol{x}}}^{bin}$ for $\ell \in \mathcal{L}_{\boldsymbol{x}}$. Two inverse distributions, $P_{\Lambda,\boldsymbol{x}}$ and $P_{\Lambda,\boldsymbol{x}}^{bin}$, are both considered as solutions to the SIP in the sense that they induce the same output $P_{\mathcal{D}_{\boldsymbol{x}}}$ through (2.4) as P_Λ does in the SFP.

For instance, if we take the following binomial ansatz $\{P_\ell^{bin}(i,j), i,j = 1,2,3\}$ with success probability 0.5 for each generalized contour in the previous example as

$$x = 1\text{:} \begin{bmatrix} 3/8 & 3/8 & 1/4 \\ 3/8 & 1/16 & 1/8 \\ 1/16 & 1/8 & 1/4 \end{bmatrix}, \quad x = 2\text{:} \begin{bmatrix} 5/16 & 1/4 & 1/32 \\ 5/32 & 1/4 & 1/32 \\ 5/16 & 1/2 & 5/32 \end{bmatrix}.$$

Then, the corresponding inverse distributions $P_{\Lambda,1}^{bin}$ and $P_{\Lambda,2}^{bin}$ are

$$P_{\Lambda,1}^{bin} = \begin{bmatrix} 9/32 & 3/32 & 3/16 \\ 3/32 & 3/64 & 1/32 \\ 3/64 & 1/32 & 3/16 \end{bmatrix}, \begin{bmatrix} \hat{P}_{\mathcal{D}_1}^{bin}(0) \\ \hat{P}_{\mathcal{D}_1}^{bin}(1) \end{bmatrix} = \begin{bmatrix} P_{\Lambda|1}^{bin}(Q_1^{-1}(0)) \\ P_{\Lambda|1}^{bin}(Q_1^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 1/4 \\ 3/4 \end{bmatrix} = \begin{bmatrix} P_{\mathcal{D}_1}(0) \\ P_{\mathcal{D}_1}(1) \end{bmatrix}.$$

$$P_{\Lambda,2}^{bin} = \begin{bmatrix} 15/128 & 5/32 & 3/256 \\ 15/256 & 5/32 & 3/256 \\ 15/128 & 5/16 & 15/256 \end{bmatrix}, \begin{bmatrix} \hat{P}_{\mathcal{D}_2}^{bin}(0) \\ \hat{P}_{\mathcal{D}_2}^{bin}(1) \end{bmatrix} = \begin{bmatrix} P_{\Lambda|2}^{bin}(Q_2^{-1}(0)) \\ P_{\Lambda|2}^{bin}(Q_2^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 3/8 \\ 5/8 \end{bmatrix} = \begin{bmatrix} P_{\mathcal{D}_2}(0) \\ P_{\mathcal{D}_2}(1) \end{bmatrix}.$$

It can been seen that $P_{\Lambda,1}^{bin}$ and $P_{\Lambda,2}^{bin}$ are valid solutions to the SIP since they reproduce the output distributions $P_{\mathcal{D}_1}$ and $P_{\mathcal{D}_2}$, respectively. However, for example, $P_{\Lambda,1}$ and $P_{\Lambda,1}^{bin}$ provide distinct information about the location of the maximum probability, i.e. $P_{\Lambda,1}$ at $\{2,2\}$ and $P_{\Lambda,1}^{bin}$ at $\{1,1\}$, since the choices of the ansatz provide distinct distributional information along the generalized contours. Among many choices of the ansatz, the uniform ansatz is commonly used in practice when we have no distributional information for $\mathcal{A}_{\mathcal{L}_{\boldsymbol{x}}}$, and in fact, it is one of the most important and useful choices suggested by the principle of maximum entropy in the information theory. We show that the inverse distribution under the uniform ansatz is a solution to the SIP chosen by maximum entropy principle.

Theorem 2.1.3. *$P_{\Lambda,\boldsymbol{x}}$ under the uniform ansatz $\mathcal{A}^u_{\mathcal{L}_{\boldsymbol{x}}}$ has the maximum entropy compared to any other inverse distributions under different choices of ansatz in the finite domain Λ.*

Proof. We use the notion of the entropy as

$$h(P) = -\sum_{\lambda\in\Lambda} P(\lambda)\log P(\lambda),$$

where P is any distribution on $(\Lambda, \mathcal{P}_\Lambda)$. In addition, we use the relative entropy defined as

$$D(P||\tilde{P}) = \sum_{\lambda\in\Lambda} P(\lambda)\log\frac{P(\lambda)}{\tilde{P}(\lambda)},$$

where $P, \tilde{P}$ are distributions on $(\Lambda, \mathcal{P}_\Lambda)$ and $0\log\frac{0}{0} = 0\log\frac{0}{\tilde{P}} = 0$ and $P\log\frac{1}{0} = \infty$. In this case, the relative entropy, also known as the Kullback-Leibler divergence, is always non-negative. We denote the inverse distribution under any choice of the ansatz $\mathcal{A}^a_{\mathcal{L}_{\boldsymbol{x}}} = \{P^a_\ell : \ell \in \mathcal{L}_{\boldsymbol{x}}\}$ as $P^a_{\Lambda,\boldsymbol{x}}$.

Hence, we have

$$\begin{aligned}
D(P^a_{\Lambda,\boldsymbol{x}}||P_{\Lambda,\boldsymbol{x}}) &= \sum_{\lambda\in\Lambda}\left(\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^a_\ell(\tilde{\lambda})\log\frac{\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^a_\ell(\tilde{\lambda})}{\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^u_\ell(\tilde{\lambda})}\right)\\
&= -\sum_{\lambda\in\Lambda}\left(\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^a_\ell(\tilde{\lambda})\log\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^u_\ell(\tilde{\lambda})\right)\\
&\quad - h(P^a_{\Lambda,\boldsymbol{x}}) \geq 0.
\end{aligned}$$

Let $h_0 := -\sum_{\lambda\in\Lambda}\Big(\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^a_\ell(\tilde{\lambda})\log\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^u_\ell(\tilde{\lambda})\Big)$. Since $\Lambda = \cup_{\ell\in\mathcal{E}_\Lambda}\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)$, we have

$$
\begin{aligned}
h_0 &= -\sum_{\lambda\in\Lambda}\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^a_\ell(\tilde{\lambda})\log P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_\lambda)\\
&\quad -\sum_{\lambda\in\Lambda}\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^a_\ell(\tilde{\lambda})\log\frac{1}{\text{Card}(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell))}\\
&= -\sum_{\ell\in\mathcal{E}_\Lambda}\left(P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) + P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log\frac{1}{\text{Card}(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell))}\right)\\
&= -\sum_{\ell\in\mathcal{E}_\Lambda}\left(P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log\frac{P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)}{\text{Card}(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell))}\right).
\end{aligned}
$$

On the other hand, the entropy of the inverse distribution under the uniform ansatz is

$$
\begin{aligned}
h(P_{\Lambda,\boldsymbol{x}}) &= -\sum_{\lambda\in\Lambda}\left(\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^u_\ell(\tilde{\lambda})\log\sum_{\ell\in\mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\sum_{\tilde{\lambda}\in\lambda\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^u_\ell(\tilde{\lambda})\right)\\
&= -\sum_{\ell\in\mathcal{E}_\Lambda}\left(P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) + P_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log\frac{1}{\text{Card}(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell))}\right) = h_0.
\end{aligned}
$$

Hence, $h(P_{\Lambda,\boldsymbol{x}}) >= h(P^a_{\Lambda,\boldsymbol{x}})$ implies the inverse distribution under the uniform ansatz along generalized contours has the maximum entropy. □

While any inverse distribution under an appropriate chosen ansatz is a solution, we call the unique solution $P_{\Lambda,\boldsymbol{x}}$, the *maximum entropy inverse distribution* (MEID). The MEID is a specific representor we select based on the maximum entropy principle (Guiasu and Shenitzer, 1985). An advantage gained from using the uniform ansatz is that it uses the least information or assumptions about the model without any prior information about conditional distributions along the generalized contours. In the rest of the paper, the MEID is used for any implementation.

2.2 Feasible Generating Distributions

2.2.1 Equivalent Distributions

In Section 2.1, we introduce multiple solutions to the SIP for a single experiment indexed by a specific $\boldsymbol{x}$ under different choices of the ansatz, and the solutions are heavily depending on the specific experiment. Consequently, the solutions induce the output distribution $P_{\mathcal{D}_{\boldsymbol{x}}}$ through (2.4) for this specific experiment indexed by $\boldsymbol{x}$, but generally not for any other experiments. In this section, we consider making use of a collection of experiments, i.e. a collection of values in $\mathcal{X}$ to tackle P_Λ, to find a "global" solution to the SIP that is independent of $\boldsymbol{x}$ and induces distributions $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$ of all outputs $\{Q_{\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$ through (2.4).

One of the important difficulties in the SIP lies on the fact that various choices, alternative to the generating distribution P_Λ, may induce the same $P_{\mathcal{D}_{\boldsymbol{x}}}$, and in general, an SIP yields multiple solutions in the sense that they induce the same $P_{\mathcal{D}_{\boldsymbol{x}}}$, which in fact can be viewed as an equivalence class of distributions indexed by each given $\boldsymbol{x} \in \mathcal{X}$ and $P_{\mathcal{D}_{\boldsymbol{x}}}$. We consider probability distributions on Λ such that they can be decomposed into distributions along the contours, $\{\widetilde{P}_\ell\}$, and the distribution on the quotient space, $P_{\mathcal{L}_{\boldsymbol{x}}}$. Then we have the following equivalence class

$$\mathbb{P}_{\Lambda,\boldsymbol{x}} := \{\widetilde{P}_\Lambda : \widetilde{P}_\Lambda(A) = \sum_{\ell\in\mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda\in A\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} \widetilde{P}_\ell(\lambda), \quad \forall A \in \mathcal{P}_\Lambda\},$$

where each distribution in this class induces the same output distribution $P_{\mathcal{D}_{\boldsymbol{x}}}$ through the map $Q_{\boldsymbol{x}}$. Each element in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ is called a *locally feasible distribution* since distributions in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ are generally conditioned on the experiment $\boldsymbol{x}$. Note that the generating distribution P_Λ is also an element of $\mathbb{P}_{\Lambda,\boldsymbol{x}}$, and thus this equivalence class is non-empty. Clearly, the MEID is also contained in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$.

For each $\boldsymbol{x} \in \mathcal{X}$, there exists $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ indexed by $\boldsymbol{x}$, and the generating distribution P_Λ is contained in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ since P_Λ generates all the output distributions $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$. In the absence of assumption of dependencies of the collection of these equivalence classes $\{\mathbb{P}_{\Lambda,\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$, a natural approach to tackle P_Λ is to remove the effect of experiments by considering the intersection of $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ over all

possible $\boldsymbol{x}$. Let

$$\mathbb{P}_{\Lambda,\mathcal{X}} := \bigcap_{\boldsymbol{x}\in\mathcal{X}} \mathbb{P}_{\Lambda,\boldsymbol{x}},$$

which is nonempty since it contains the generating distribution P_Λ. All elements in $\mathbb{P}_{\Lambda,\mathcal{X}}$ are equivalent in the sense that they induce the same collection of output distributions $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$. In the rest of this paper, any element in $\mathbb{P}_{\Lambda,\mathcal{X}}$ is referred to as the *globally feasible generating distribution* (GFGD).

In general, $\mathbb{P}_{\Lambda,\mathcal{X}}$ has uncountably many elements. More specifically, $\mathbb{P}_{\Lambda,\mathcal{X}}$ is a convex set.

Theorem 2.2.1. *$\mathbb{P}_{\Lambda,\mathcal{X}}$ is convex: Any mixture defined as*

$$P_{Mix}(A) = wP_1(A) + (1-w)P_2(A), \quad \forall A \in \mathcal{P}_\Lambda,$$

where $P_1, P_2 \in \mathbb{P}_{\Lambda,\mathcal{X}}$ and $0 \le w \le 1$, is contained in $\mathbb{P}_{\Lambda,\mathcal{X}}$.

Proof. Since $P_1, P_2 \in \mathbb{P}_{\Lambda,\mathcal{X}}$, we have

$$P_1(A) = \sum_{\ell\in\mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda\in A\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^1_\ell(\lambda), P_2(A) = \sum_{\ell\in\mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda\in A\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^2_\ell(\lambda), \quad \forall A \in \mathcal{P}_\Lambda.$$

Hence, we have

$$P_{Mix}(A) = wP_1(A) + (1-w)P_2(A) = \sum_{\ell\in\mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda\in A\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} (wP^1_\ell(\lambda) + (1-w)P^2_\ell(\lambda)).$$

Similarly, we can show that P_{Mix} is a probability distribution on $(\Lambda, \mathcal{P}_\Lambda)$ following the proof of Theorem 2.1.2. Thus, any mixture is contained in $\mathbb{P}_{\Lambda,\mathcal{X}}$. In addition, for any $\boldsymbol{x} \in \mathcal{X}$ and $C \in \mathcal{P}_{\mathcal{D}_{\boldsymbol{x}}}$, we can verify that $P_{Mix}(Q^{-1}_{\boldsymbol{x}}(C)) = wP_1(Q^{-1}_{\boldsymbol{x}}(C)) + (1-w)P_2(Q^{-1}_{\boldsymbol{x}}(C)) = P_{\mathcal{D}_{\boldsymbol{x}}}(C)$. □

Alternative to the generating distribution P_Λ in (2.2), we show a GFGD as follows.

$$P_\Lambda^1 = \begin{bmatrix} 524/6077 & 591/5375 & 524/6077 \\ 165/10966 & 2243/5537 & 165/10966 \\ 524/6077 & 591/5375 & 524/6077 \end{bmatrix} \approx \begin{bmatrix} 0.0862 & 0.1100 & 0.0862 \\ 0.0150 & 0.4051 & 0.0150 \\ 0.0862 & 0.1100 & 0.0862 \end{bmatrix} \tag{2.10}$$

Here, P_Λ^1 is found by an iterative approach in Section 2.3.2. Any mixture of P_Λ and P_Λ^1 is also a viable GFGD that induces $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$. For instance, with weight $w = 0.5$, we have another GFGD as

$$P_\Lambda^2 = wP_\Lambda + (1-w)P_\Lambda^1 = \begin{bmatrix} 284/4835 & 596/6167 & 254/2405 \\ 132/4655 & 391/864 & 252/5123 \\ 284/4835 & 372/4907 & 219/2945 \end{bmatrix} \approx \begin{bmatrix} 0.0587 & 0.0966 & 0.1056 \\ 0.0284 & 0.4525 & 0.0492 \\ 0.0587 & 0.0758 & 0.0744 \end{bmatrix}.$$

In these three distinct distributions, the conditional distributions, $\{\widetilde{P}_\ell\}_{\ell\in\mathcal{L}_{\boldsymbol{x}}}$, along the contours are distinct and informative, which implies the choice of the ansatz plays an important role of finding a GFGD. In other words, we can find a GFGD by correcting the pre-specified ansatz. In subsequent discussions, we first propose an approach that removes the dependency of the $\boldsymbol{x}$ on the inverse distributions, and further develop an iterative approach that corrects the pre-specified ansatz to obtain a GFGD.

2.2.2 Degree of Beliefs and High-probability Regions in the SIP

The degree of beliefs is a concept widely used in Bayesian statistics to quantify the probabilities of events of $\boldsymbol{a}$. For instance, the degree of belief of event $A \in \mathcal{P}_\Lambda$ of $\boldsymbol{a}$ is defined as $P(A)$ where P is a probability distribution of $\boldsymbol{a}$. In the SIP, there are multiple solutions either for a specific experiment $\boldsymbol{x}$ or for a collection of experiments in $\mathcal{X}$. Thus, the degree of belief of event A that depends on the probability distribution of $\boldsymbol{a}$ varies. In Dempster (2008), the degree of belief can be defined on a convex set $\mathcal{C}$ of probability measures and the author suggests quantify the degree

of belief by considering the *lower* and *upper probabilities* over $\mathcal{C}$ as

$$P^*(A;\mathcal{C}) = \sup_{P\in\mathcal{C}} P(A),$$
$$P_*(A;\mathcal{C}) = \inf_{P\in\mathcal{C}} P(A).$$

In particular, we specify $\mathcal{C} = \mathbb{P}_{\Lambda,\mathcal{X}}$ to obtain the lower and upper probabilities for the degree of belief under the SIP framework as $P_*(A;\mathbb{P}_{\Lambda,\mathcal{X}})$ and $P^*(A;\mathbb{P}_{\Lambda,\mathcal{X}})$, respectively. Since the equivalence class is convex, we have the following corollary as a direct consequence of Theorem 2.2.1.

Corollary 2.2.1.1. *For any $A \in \mathcal{P}_\Lambda$, the set $\{P(A) : P \in \mathbb{P}_{\Lambda,\mathcal{X}}\}$ indexed by the elements in $\mathbb{P}_{\Lambda,\mathcal{X}}$ is convex.*

In this case, the lower and upper probabilities $P_*(A;\mathbb{P}_{\Lambda,\mathcal{X}})$ and $P^*(A;\mathbb{P}_{\Lambda,\mathcal{X}})$ can be naturally built into the following bound as

$$\left(P_*(A;\mathbb{P}_{\Lambda,\mathcal{X}}), P^*(A;\mathbb{P}_{\Lambda,\mathcal{X}})\right),$$

where each value can be found according to a distribution in $\mathbb{P}_{\Lambda,\mathcal{X}}$ due to the convexity. This bound considers all possible values among distributions in the class $\mathbb{P}_{\Lambda,\mathcal{X}}$ and characterizes the true range of the degree of belief of event A. Any value in this bound is of great importance to study the properties of $\boldsymbol{a}$.

To illustrate, we investigate three equivalent distributions P^1_Λ, P^2_Λ and P_Λ by considering an event $A = \{(1,2),(1,3),(2,2),(2,3)\}$ of (a,b) in the simple example in Section 2.1. In this case, these three equivalent distributions yield the following degree of beliefs of event A: $P^1_\Lambda(A) = 355/576 \approx 0.6163$, $P^2_\Lambda(A) = 811/1152 \approx 0.7040$ and $P_\Lambda(A) = 19/24 \approx 0.7917$. These numbers are clearly different, which implies that the established identifiability in traditional statistical approaches is not reliable when we acutally consider the data-generating process, i.e. the SFP. Another interesting fact can be observed that $wP_\Lambda(A) + (1-w)P^1_\Lambda(A) = P^2_\Lambda(A)$ with $w = 0.5$,

since P_Λ^2 is constructed by the weighted sum of P_Λ and P_Λ^1 with equal weights. This verifies the convexity of the set stated in Corollary 2.2.1.1.

The equivalent class $\mathbb{P}_{\Lambda,\mathcal{X}}$ is also crucial to finding a high-probability region in the SIP. The input $\boldsymbol{a}$ is generally a "physical" input in the SIP Butler et al. (2014) since it has physical meanings in the real-world problems, e.g. thermal conductivity of copper alloys. Thus, the region of $\boldsymbol{a}$ with high probability of occurrence is of particular interest in the SIP. The high-probability δ-region can be defined as

$$A_{\delta,P} := \{\lambda \in \Lambda : P(\lambda) \geq \delta > 0\},$$

where P is any discrete probability distribution on the finite domain Λ. Consequently, each distribution in $\mathbb{P}_{\Lambda,\mathcal{X}}$ yields a valid δ-region and $A_{\delta,P}$ is not unique. A better way of quantifying the high-probability δ-region in this case is through considering the minimum and maximum δ-regions as

$$A^* = \bigcap_{P\in\mathbb{P}_{\Lambda,\mathcal{X}}} A_{\delta,P},$$
$$A_* = \bigcup_{P\in\mathbb{P}_{\Lambda,\mathcal{X}}} A_{\delta,P}.$$

Note that these two optimal regions rely on the choice of δ and characterize the high-probability regions in the extreme cases, and each region $A_* \subset A_{\delta,P} \subset A^*$ for any $P \in \mathbb{P}_{\Lambda,\mathcal{X}}$.

Continuing to consider the GFGDs P_Λ^1, P_Λ^2 and P_Λ, we explore the high-probability δ-regions of the three distributions when $\delta = 0.1$. Then the 0.1-regions for these GFGDs are: $A_{0.1,P_\Lambda^1} = \{(2,2)\}$, $A_{0.1,P_\Lambda^2} = \{(1,3),(2,2)\}$ and $A_{0.1,P_\Lambda} = \{(1,3),(2,2)\}$. It can be seen that point $(2,2)$ is in each of the regions.

All these bounds defined by the equivalence class $\mathbb{P}_{\Lambda,\mathcal{X}}$ characterize the actual domains of the quantities of interest and any value in the bounds is informative. Thus, in the next section, we provide a way of finding a GFGD in the equivalence class.

2.3 Methodology and Example of Iterative Sliced Inverse Approach

2.3.1 Experimental Expectation of Inverse Distributions

In section 2.1, we find a unique representor in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$, i.e. the MEID. However, the MEID is a "local" solution for the SIP, that is, conditioned on specific experiment indexed by $\boldsymbol{x}$. In this section, we seek to find a solution that is not associated with any experiment $\boldsymbol{x}$. A probabilistic approach is to take the expectation of the MEID over $\boldsymbol{x}$ to remove the dependency. The *experimental expectation of inverses* (EEI) is defined as

$$\bar{P}(A) := \sum_{\boldsymbol{x}\in\mathcal{X}} P_{\Lambda,\boldsymbol{x}}(A)P_{\mathcal{X}}(\boldsymbol{x}), \quad \forall A \in \mathcal{P}_{\Lambda}, \tag{2.11}$$

where $P_{\mathcal{X}} > 0$ is the probability distribution on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$. In fact, the EEI is a probability distribution of $\boldsymbol{a}$.

Theorem 2.3.1. *$\bar{P}$ is a probability distribution on $(\Lambda, \mathcal{P}_{\Lambda})$.*

Proof. Since $P_{\Lambda,\boldsymbol{x}}$ is a probability distribution on $(\Lambda, \mathcal{P}_{\Lambda})$, we can simply have $0 \leq \bar{P} \leq 1$. On the other hand, we have

$$\bar{P}\left(\bigcup_i A_i\right) = \sum_{\boldsymbol{x}\in\mathcal{X}}\sum_i P_{\Lambda,\boldsymbol{x}}(A_i)P_{\mathcal{X}}(\boldsymbol{x}) = \sum_i\sum_{\boldsymbol{x}\in\mathcal{X}} P_{\Lambda,\boldsymbol{x}}(A_i)P_{\mathcal{X}}(\boldsymbol{x}) = \sum_i \bar{P}(A_i),$$

since all components in the summation are non-negative. □

By removing the dependency of $\boldsymbol{X}$ on the inverse distributions using the information from all possible experiments, the EEI generally provides a less biased result on $\boldsymbol{a}$ than an individual inverse distribution. More importantly, the EEI provides information about the generating distribution P_{Λ}, specifically, about the ansatz along generalized contours.

Returning to the simple example in Section 2.1, we obtain the EEI explicitly as

$$\begin{aligned}\bar{P}(A) &= \sum_{x\in\mathcal{X}} P_{\Lambda,x}(A)P_{\mathcal{X}}(x) \\ &= \sum_{x\in\mathcal{X}} P_{\mathcal{L}_x}(\ell_{\text{odd}}) \sum_{(i,j)\in A\cap\pi^{-1}_{\mathcal{L}_x}(\ell_{\text{odd}})} P^u_{\ell_{\text{odd}}}(i,j)P_{\mathcal{X}}(x) \\ &+ \sum_{x\in\mathcal{X}} P_{\mathcal{L}_x}(\ell_{\text{even}}) \sum_{(i,j)\in A\cap\pi^{-1}_{\mathcal{L}_x}(\ell_{\text{even}})} P^u_{\ell_{\text{even}}}(i,j)P_{\mathcal{X}}(x),\end{aligned} \tag{2.12}$$

for $P_{\mathcal{X}}(1) = 2/3, P_{\mathcal{X}}(2) = 1/3$ and $A \in \mathcal{P}_\Lambda$; see Figure 2.2. Specifically,

$$\bar{P} = \begin{bmatrix} 29/240 & 1/9 & 29/240 \\ 1/16 & 61/360 & 1/16 \\ 29/240 & 1/9 & 29/240 \end{bmatrix}, \begin{bmatrix} \bar{P}_{\mathcal{D},1}(0) \\ \bar{P}_{\mathcal{D},1}(1) \end{bmatrix} = \begin{bmatrix} \bar{P}(Q_1^{-1}(0)) \\ \bar{P}(Q_1^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 25/72 \\ 47/72 \end{bmatrix} \neq \begin{bmatrix} P_{\mathcal{D},1}(0) \\ P_{\mathcal{D},1}(1) \end{bmatrix},$$

$$\begin{bmatrix} \bar{P}_{\mathcal{D},2}(0) \\ \bar{P}_{\mathcal{D},2}(1) \end{bmatrix} = \begin{bmatrix} \bar{P}(Q_2^{-1}(0)) \\ \bar{P}(Q_2^{-1}(1)) \end{bmatrix} = \begin{bmatrix} 73/120 \\ 47/120 \end{bmatrix} \neq \begin{bmatrix} P_{\mathcal{D},2}(0) \\ P_{\mathcal{D},2}(1) \end{bmatrix}$$

Here, $\bar{P}$ provides more accurate distributional information about the generating distribution in

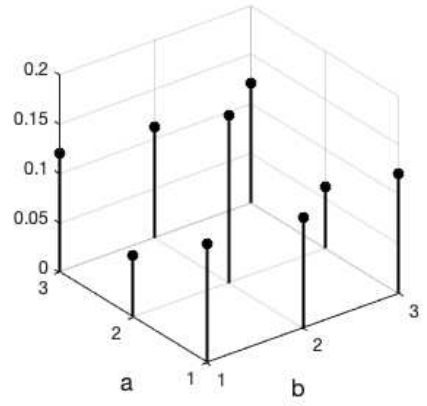


Figure 2.2: A graphical display of $\bar{P}$ in the simple example (2.3) according to the distribution (2.2).

panel (a) of Figure 2.1 than any other inverse distributions. Moreover, even though $\bar{P}$ is not equal

to the generating distribution P_Λ or a GFGD, $\bar{P}$ is closer to the generating distribution compared to any other individual inverse distributions, by showing the following distances of these probability distributions under L^1 metric,

$$|P_{\Lambda,1} - P_\Lambda|_1 = 47/60, |P_{\Lambda,2} - P_\Lambda|_1 = 3/4, |\bar{P} - P_\Lambda|_1 = 32/45. \tag{2.13}$$

Most importantly, the EEI provides more useful distributional information about the conditional probabilities $\{P_\ell(i,j), i,j = 1,2,3\}$ along the generalized contours in (2.7) than the uniform ansatz. This procedure essentially provides a way of removing the effect of $\boldsymbol{x}$ and simultaneously updates the ansatz information.

2.3.2 Iterative Approach to Finding GFGDs

In this section, we find a GFGD in $\mathbb{P}_{\Lambda,\mathcal{X}}$ through iteratively updating the ansatz information given by the ansatz information extracted from the EEI.

This iterative approach involves two main processes. Since the class $\mathbb{P}_{\Lambda,\mathcal{X}}$ is uniquely determined by the collection of probability distributions $\{P_{\mathcal{L}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$ on the quotient spaces $\{\mathcal{L}_{\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$, the collection $\{P_{\mathcal{L}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$ is crucial in this iterative approach. In each iteration, we first extract the new ansatz from the last EEI, and we use $\{P_{\mathcal{L}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$ and the new ansatz to compute the next batch of inverse distributions and the EEI. We describe the approach in detail in the following.

In the initial step, we choose the uniform ansatz $\{P^0_\ell\}_{\ell\in\mathcal{L}_{\boldsymbol{x}}}$ as an initial ansatz to compute inverse distributions for $\boldsymbol{x}\in\mathcal{X}$

$$\sum_{\ell\in\mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda\in A\cap\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^0_\ell(\lambda) = P^0_{\Lambda,\boldsymbol{x}}(A), \quad \forall A\in\mathcal{P}_\Lambda,$$

and the EEI

$$\bar{P}^0(A) = \sum_{\boldsymbol{x}\in\mathcal{X}} P^0_{\Lambda,\boldsymbol{x}}(A) P_{\mathcal{X}}(\boldsymbol{x}), \quad \forall A\in\mathcal{P}_\Lambda.$$

Starting the first iteration, we vary the choice of the ansatz based on the extracted distributions along the generalized contours from $\bar{P}^0$, which are obtained by the following decomposition of $\bar{P}^0$. For each $\boldsymbol{x} \in \mathcal{X}$,

$$\bar{P}^0(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^0}(\ell) \sum_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A} P_\ell^1(\lambda), \quad \forall A \in \mathcal{P}_\Lambda,$$

where $P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^0}$ is the unique probability distribution on the quotient space $\mathcal{L}_{\boldsymbol{x}}$ calculated through

$$P^0_{\mathcal{D}_{\boldsymbol{x}}}(C) = \bar{P}^0(Q_{\boldsymbol{x}}^{-1}(C)), \quad C \in \mathcal{P}_{\mathcal{D}_{\boldsymbol{x}}},$$
$$P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^0}(\mathcal{E}_A) = P^0_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad A \in \mathcal{P}_\Lambda,$$

and P_ℓ^1 is the (unique) conditional probability distribution along the generalized contour $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)$. Specifically, P_ℓ^1 is computed as

$$P_\ell^1(\lambda) = \frac{\bar{P}^0(\lambda)}{P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^0}(\ell)}, \quad \forall \lambda \in \Lambda,$$

where $P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^0}(\ell) > 0$ and $\ell = \pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)$, and $P_\ell^1(\lambda) \equiv 0$ where $P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^0}(\ell) = 0$ and $\ell = \pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)$.

Then we use $\{P_\ell^1\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ as the new ansatz to update inverse distributions for $\boldsymbol{x} \in \mathcal{X}$ as

$$\sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} P_\ell^1(\lambda) = P^1_{\Lambda, \boldsymbol{x}}(A), \quad \forall A \in \mathcal{P}_\Lambda,$$

and the updated EEI after the first iteration is

$$\bar{P}^1(A) = \sum_{\boldsymbol{x} \in \mathcal{X}} P^1_{\Lambda, \boldsymbol{x}}(A) P_{\mathcal{X}}(\boldsymbol{x}), \quad \forall A \in \mathcal{P}_\Lambda.$$

Iteratively, for any $A \in \mathcal{P}_\Lambda$ and $\boldsymbol{x} \in \mathcal{X}$, we have

$$\text{Decomposition of the EEI: } \bar{P}^i(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^i}(\ell) \sum_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} P^{i+1}_\ell(\lambda), \tag{2.14}$$

$$\text{Update inverse distributions: } \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^{i+1}_\ell(\lambda) = P^{i+1}_{\Lambda, \boldsymbol{x}}(A), \tag{2.15}$$

$$\text{Update the EEI: } \bar{P}^{i+1}(A) = \sum_{\boldsymbol{x} \in \mathcal{X}} P^{i+1}_{\Lambda, \boldsymbol{x}}(A) P_{\mathcal{X}}(\boldsymbol{x}), \tag{2.16}$$

where $i \geq 0$. The iteration proceeds under the hope $\{\bar{P}^i\}_{i \geq 0}$ converges. Note that in each iteration we update the ansatz $\{P^i_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}, i \geq 0$, and embed $\{P_{\mathcal{L}_{\boldsymbol{x}}}\}_{\boldsymbol{x} \in \mathcal{X}}$, more specifically, the data on the output, $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x} \in \mathcal{X}}$, to adjust the results.

In general, the sequence $\{\bar{P}^i\}_{i \geq 0}$ may not have a convergence. In the following, we explore one particular case in which the sequence has a convergence.

Theorem 2.3.2. *For each $i \geq 0$ and $\boldsymbol{x} \in \mathcal{X}$, the inverse distribution has a density function $P^i_{\Lambda, \boldsymbol{x}}(\lambda) = P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) P^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)$ where $\lambda \in \Lambda$.*

Proof. For any $\lambda \in \Lambda$, we simply have

$$\begin{aligned} P^i_{\Lambda, \boldsymbol{x}}(\lambda) &= \sum_{\ell \in \mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\tilde{\lambda} \in \lambda \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} P^i_\ell(\tilde{\lambda}) = \sum_{\ell \in \mathcal{E}_\lambda} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) P^i_\ell(\lambda) \\ &= P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_\lambda) P^i_{\mathcal{E}_\lambda}(\lambda) = P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) P^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda), \end{aligned}$$

where $i \geq 0$ and $\boldsymbol{x} \in \mathcal{X}$. □

Theorem 2.3.3. *If for every real number $\epsilon > 0$ there exists an positive integer $N \in \mathbb{N}$ such that for all positive integers $i, j > N$, we have $|P^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) - P^j_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)| < \epsilon P^0_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)$ where $\lambda \in \Lambda$, then $\bar{P}^i$ converges to a probability distribution $\bar{P}^\infty$ on Λ under the L^1 metric.*

Proof. We first obtain the density function of $\bar{P}^i$ as

$$\bar{P}^i(\lambda) = \sum_{\boldsymbol{x} \in \mathcal{X}} P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) P^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) P_{\mathcal{X}}(\boldsymbol{x}),$$

where $\lambda \in \Lambda$ and $i \geq 0$. Then we consider the complete metric space $(\mathcal{L}_D^1, d)$ where the space $\mathcal{L}_D^1$ contains probability distributions (i.e. density functions) on Λ and d is the L^1 metric defined as

$$d(P, \tilde{P}) = \sum_{\lambda \in \Lambda} |P(\lambda) - \tilde{P}(\lambda)|.$$

Then, for any real number $\epsilon > 0$ and $i, j > N$,

$$\begin{aligned} d(\bar{P}^i, \bar{P}^j) &= \sum_{\lambda \in \Lambda} \left| \sum_{\boldsymbol{x} \in \mathcal{X}} P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) P^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) P_{\mathcal{X}}(\boldsymbol{x}) - \sum_{\boldsymbol{x} \in \mathcal{X}} P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) P^j_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) P_{\mathcal{X}}(\boldsymbol{x}) \right| \\ &\leq \sum_{\lambda \in \Lambda} \sum_{\boldsymbol{x} \in \mathcal{X}} \left| P^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) - P^j_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \right| P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) P_{\mathcal{X}}(\boldsymbol{x}) \\ &< \epsilon \sum_{\lambda \in \Lambda} \sum_{\boldsymbol{x} \in \mathcal{X}} P^0_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) P_{\mathcal{X}}(\boldsymbol{x}) = \epsilon \sum_{\lambda \in \Lambda} \bar{P}^i(\lambda) = \epsilon. \end{aligned}$$

Hence, the sequence $\{\bar{P}^i\}_{i \geq 0}$ is Cauchy in $(\mathcal{L}_D^1, d)$, and $\bar{P}^i$ converges to a density function in $(\mathcal{L}_D^1, d)$ as $i \to \infty$. □

This result shows that $\{\bar{P}^i\}_{i \geq 0}$ has a convergence if the iterated ansatz can be controlled. We denote the limit of convergence of the EEIs by $\bar{P}^\infty$. In general, $\bar{P}^\infty$ is not contained in $\mathbb{P}_{\Lambda,\mathcal{X}}$. Note that each iteration in this approach can be characterized as a map $\mathcal{G}$ such that $\bar{P}^{i+1} = \mathcal{G}(\bar{P}^i)$. In the following, we show a sufficient and necessary condition for $\mathbb{P}_{\Lambda,\mathcal{X}}$ regarding the iteration map $\mathcal{G}$.

Theorem 2.3.4. *A sufficient and necessary condition for $\widetilde{P}$ in $\mathbb{P}_{\Lambda,\mathcal{X}}$ is state as follows:*

1. *$\widetilde{P}$ is a fixed point of map $\mathcal{G}$, i.e. $\widetilde{P} = \mathcal{G}(\widetilde{P})$;*
2. *the inverse distribution $\widetilde{P}_{\Lambda,\boldsymbol{x}}$ computed from $\widetilde{P}$ in the iterative approach in (2.14) and (2.15) by*

$$\textit{Decomposition of a distribution: } \widetilde{P}(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}, \widetilde{P}}(\ell) \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} \widetilde{P}_\ell(\lambda), \quad \forall A \in \mathcal{P}_\Lambda,$$

$$\textit{Compute inverse distributions: } \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} \widetilde{P}_\ell(\lambda) = \widetilde{P}_{\Lambda,\boldsymbol{x}}(A), \quad \forall A \in \mathcal{P}_\Lambda,$$

is not conditioned on $\boldsymbol{x}$, i.e. $\widetilde{P}_{\Lambda,\boldsymbol{x}} = \widetilde{P}_{\Lambda,\boldsymbol{x}'}$ if $\boldsymbol{x} \neq \boldsymbol{x}'$.

Proof. We start with the proof of the first direction. Since $\widetilde{P} \in \mathbb{P}_{\Lambda,\mathcal{X}}$, it can be decomposed as

$$\widetilde{P}(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} \widetilde{P}_{\ell}(\lambda) = \widetilde{P}_{\Lambda,\boldsymbol{x}}(A), \quad \forall A \in \mathcal{P}_{\Lambda},$$

for $\boldsymbol{x} \in \mathcal{X}$. It implies that the inverse distributions are the same as $\widetilde{P}$ and thus they are not conditioned on $\boldsymbol{x}$. Moreover,

$$\bar{P}(A) = \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_{\Lambda,\boldsymbol{x}}(A) P_{\mathcal{X}}(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}(A) P_{\mathcal{X}}(\boldsymbol{x}) = \widetilde{P}(A), \quad \forall A \in \mathcal{P}_{\Lambda},$$

which concludes $\widetilde{P} = \mathcal{G}(\widetilde{P})$.

Conversely, since $\widetilde{P}$ is a fixed point of $\mathcal{G}$, we have

$$\widetilde{P}(A) = \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_{\Lambda,\boldsymbol{x}}(A) P_{\mathcal{X}}(\boldsymbol{x}), \quad \forall A \in \mathcal{P}_{\Lambda}.$$

In addition, since the inverse distributions are not conditioned on $\boldsymbol{x}$, we have $\widetilde{P} = \widetilde{P}_{\Lambda,\boldsymbol{x}}$ for any $\boldsymbol{x} \in \mathcal{X}$. It implies

$$\widetilde{P}(A) = \sum_{\ell \in \mathcal{E}_A} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in A \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} \widetilde{P}_{\ell}(\lambda), \quad \forall A \in \mathcal{P}_{\Lambda}.$$

Hence, for any $\boldsymbol{x} \in \mathcal{X}$,

$$\begin{aligned}\widetilde{P}(Q^{-1}_{\boldsymbol{x}}(C)) &= \sum_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)}} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in Q^{-1}_{\boldsymbol{x}}(C) \cap \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} \widetilde{P}_{\ell}(\lambda) = \sum_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)}} P_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \sum_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} \widetilde{P}_{\ell}(\lambda) \\ &= P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)}) = P_{\Lambda}(Q^{-1}_{\boldsymbol{x}}(C)) = P_{\mathcal{D}_{\boldsymbol{x}}}(C), \quad \forall C \in \mathcal{P}_{\mathcal{D}_{\boldsymbol{x}}},\end{aligned}$$

which concludes $\widetilde{P} \in \mathbb{P}_{\Lambda,\mathcal{X}}$. □

The result shows that, if the limiting distribution $\bar{P}^{\infty}$ is a fixed point of $\mathcal{G}$ and inverse distributions computed from $\bar{P}^{\infty}$ are not conditioned on $\boldsymbol{x}$, then $\bar{P}^{\infty}$ is a GFGD in $\mathbb{P}_{\Lambda,\mathcal{X}}$.

2.3.3 Example of the Iterative Sliced Inverse Approach

In this section, we show the results of some steps in this iteration approach with the example in Section 2.1.

We first choose the EEI in (2.12) as the initial EEI $\bar{P}^0$. Then, we compute the output distributions that are induced by $\bar{P}^0$ as follows. For each $x \in \mathcal{X}$,

$$P_{\mathcal{L}_{\boldsymbol{x}},\bar{P}^0}(\ell_{\text{odd}}) = P^0_{\mathcal{D}_x}(0) = \bar{P}^0(Q_x^{-1}(0)),$$
$$P_{\mathcal{L}_{\boldsymbol{x}},\bar{P}^0}(\ell_{\text{even}}) = \bar{P}^0_{\mathcal{D}_x}(1) = \bar{P}^0(Q_x^{-1}(1)).$$

Then, by Theorem 2.1.1, we decompose the EEI as

$$\bar{P}^0(A) = P_{\mathcal{L}_{\boldsymbol{x}},\bar{P}^0}(\ell_{\text{odd}}) \sum_{\{i,j\}\in A\cap\pi^{-1}_{\mathcal{L}_x}(\ell_{\text{odd}})} P^1_{\ell_{\text{odd}}}(i,j) + P_{\mathcal{L}_{\boldsymbol{x}},\bar{P}^0}(\ell_{\text{even}}) \sum_{\{i,j\}\in A\cap\pi^{-1}_{\mathcal{L}_x}(\ell_{\text{even}})} P^1_{\ell_{\text{even}}}(i,j),$$

for any $A \in \mathcal{P}_\Lambda$, to obtain the new ansatz $\{P^1_\ell(i,j), i,j = 1,2,3\}$

$$x = 1\text{:} \begin{bmatrix} 87/470 & 8/25 & 87/470 \\ 9/50 & 61/235 & 9/50 \\ 87/470 & 8/25 & 87/470 \end{bmatrix}, \quad x = 2\text{:} \begin{bmatrix} 29/146 & 40/141 & 29/146 \\ 15/146 & 61/141 & 15/146 \\ 29/146 & 40/141 & 29/146 \end{bmatrix}.$$

Then, we replace the uniform ansatz with the new ansatz $\{P^1_\ell(i,j), i,j = 1,2,3\}$ to obtain the updated inverse distributions as

$$P_{\mathcal{L}_{\boldsymbol{x}}}(\ell_{\text{odd}}) \sum_{\{i,j\}\in A\cap\pi^{-1}_{\mathcal{L}_x}(\ell_{\text{odd}})} P^1_{\ell_{\text{odd}}}(i,j) + P_{\mathcal{L}_{\boldsymbol{x}}}(\ell_{\text{even}}) \sum_{\{i,j\}\in A\cap\pi^{-1}_{\mathcal{L}_x}(\ell_{\text{even}})} P^1_{\ell_{\text{even}}}(i,j) = P^1_{\Lambda,x}(A), \quad \forall A \in \mathcal{P}_\Lambda,$$

for each $x \in \mathcal{X}$, and we update the EEI as

$$\bar{P}^1(A) = \sum_{x\in\mathcal{X}} P^1_{\Lambda,x}(A) P_{\mathcal{X}}(x),$$

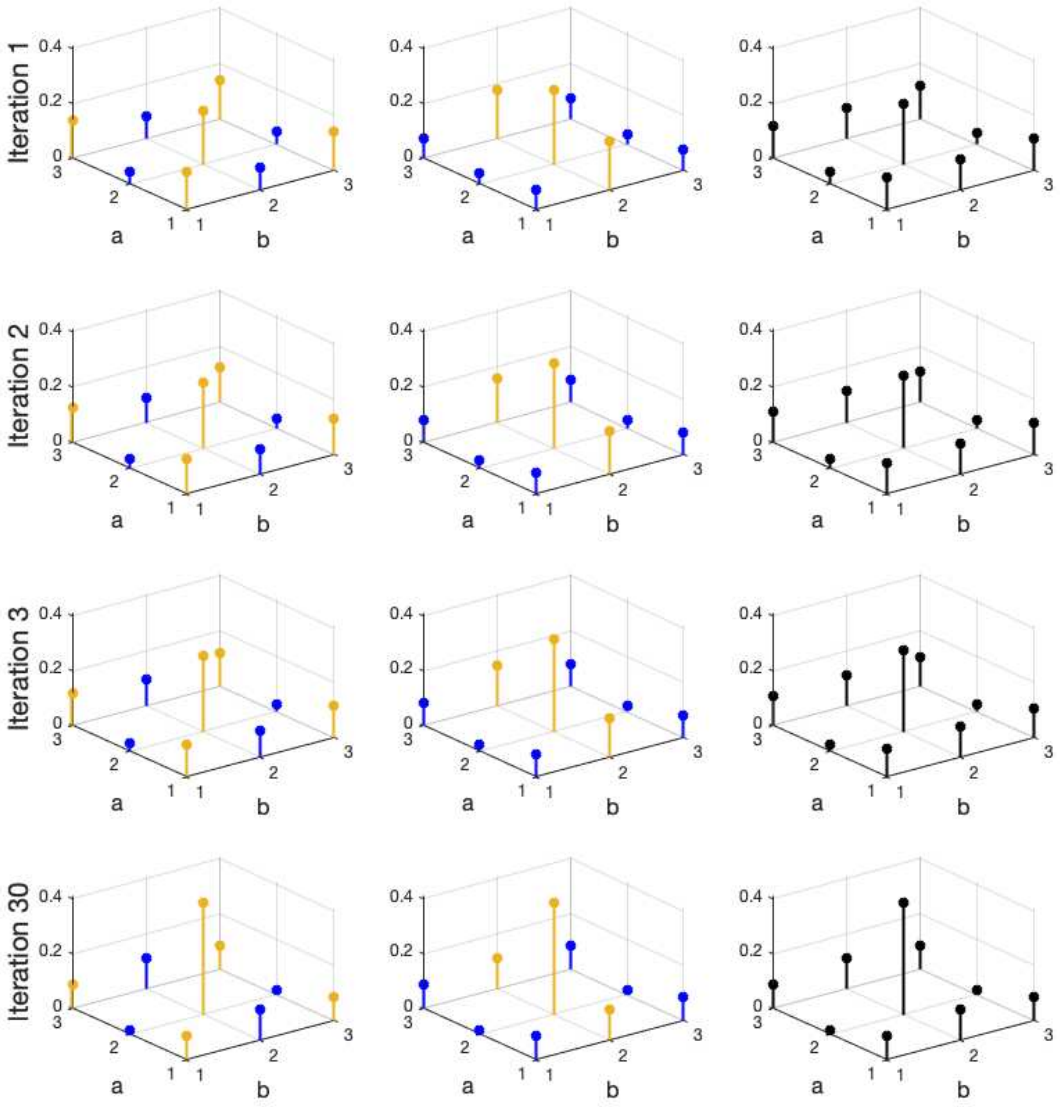


Figure 2.3: A graphical display of $P^i_{\Lambda,1}$, $P^i_{\Lambda,2}$ and $\bar{P}^i$ in columns 1-3, respectively, where each row indicates the number of iterations i. In the colored panels, the points in generalized contours $Q_x^{-1}(1)$ where $x \in \mathcal{X}$ are shown as yellow lines, while the points in generalized contours $Q_x^{-1}(0)$ where $x \in \mathcal{X}$ are shown as blue lines.

where $A \in \mathcal{P}_\Lambda$. Similarly, we decompose $\bar{P}^1$ to obtain the new ansatz $\{P^1_\ell(i,j), i,j = 1,2,3\}$ for the computation in the next iteration. In the following, we show the convergence results on the sequence $\{\bar{P}^i\}_{i\geq 0}$.

In the simulation runs, the convergence results on $P^i_{\Lambda,1}$, $P^i_{\Lambda,2}$ and $\bar{P}^i$ are shown in Figure 2.3 after $i = 1, 2, 3, 30$ iterations, in columns 1-3, respectively. Then, we have a convergence $\bar{P}^\infty$ after 30 iterations as

$$\bar{P}^\infty = \begin{bmatrix} 524/6077 & 591/5375 & 524/6077 \\ 165/10966 & 2243/5537 & 165/10966 \\ 524/6077 & 591/5375 & 524/6077 \end{bmatrix} \approx \begin{bmatrix} 0.0862 & 0.1100 & 0.0862 \\ 0.0150 & 0.4051 & 0.0150 \\ 0.0862 & 0.1100 & 0.0862 \end{bmatrix}.$$

One can verify that $\bar{P}^\infty$ induces the given output distribution $P_{\mathcal{D}_x}$ for each $x \in \mathcal{X}$, and thus $\bar{P}^\infty$ is a GFGD that is equivalent to the generating distribution. In addition, we observe that the L^1 distance of the limiting EEI and the generating distribution is

$$|\bar{P}^\infty - P_\Lambda|_1 = 411/899,$$

which is much smaller than that of the original EEI, $\bar{P}^0$, and the generating distribution in (2.13). This implies the iterative approach is essentially a process of reducing the distance of the EEI and the equivalence class of GFGDs.

2.4 Other Approaches under the SIP Setting

2.4.1 An Extension of the Classical Bayesian Approach

In the simple discrete example defined in Section 2.1, we focus on finding a probability distribution of (a, b) that can reproduce the output distributions, i.e. a GFGD. In this section, we introduce an alternative Bayesian approach specifically designed for random parameters in the domain. Suppose the generating distribution of (a, b) is a product of two independent distributions

as

$$P_\Lambda = \begin{bmatrix} 1/3 \\ 1/2 \\ 1/6 \end{bmatrix} \times \begin{bmatrix} 1/4 & 1/2 & 1/4 \end{bmatrix} = \begin{bmatrix} 1/12 & 1/6 & 1/12 \\ 1/8 & 1/4 & 1/8 \\ 1/24 & 1/12 & 1/24 \end{bmatrix} \approx \begin{bmatrix} 0.0833 & 0.1667 & 0.0833 \\ 0.1250 & 0.2500 & 0.1250 \\ 0.0417 & 0.0833 & 0.0417 \end{bmatrix},$$

and the sampling distribution on $(\mathcal{X}, \mathcal{P}_\mathcal{X})$ is $\{P_\mathcal{X}(1), P_\mathcal{X}(2)\} = \{1/2, 1/2\}$. The corresponding output distributions, that are induced by the map Q_x where $x \in \mathcal{X}$, are

$$\begin{bmatrix} P_{\mathcal{D}_1}(0) \\ P_{\mathcal{D}_1}(1) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} P_{\mathcal{D}_2}(0) \\ P_{\mathcal{D}_2}(1) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}. \tag{2.17}$$

We first assume that a and b are independently distributed according to two beta-binomial distributions with the following density functions,

$$d(a, 2; \alpha_1, \beta_1) = \binom{2}{a-1} \frac{B(2-\alpha_1, 2-(a-1)+\beta_1)}{B(\alpha_2, \beta_2)},$$
$$d(b, 2; \alpha_2, \beta_2) = \binom{2}{b-1} \frac{B(2-\alpha_2, 2-(b-1)+\beta_2)}{B(\alpha_2, \beta_2)},$$

respectively, where $B(\cdot, \cdot)$ is the density function of a beta distribution and $\boldsymbol{\omega} = (\alpha_1, \beta_1, \alpha_2, \beta_2)^\top$ is a vector of unknown hyperparameters. Then, the assumed joint distribution of (a, b) is

$$P^B_\Lambda(a, b|\boldsymbol{\omega}) = d(a, 2; \alpha_1, \beta_1) d(b, 2; \alpha_2, \beta_2).$$

The goal is to use $P^B_\Lambda(a, b|\boldsymbol{\omega})$ to approximate the generating distribution P_Λ or to find an equivalent GFGD by implementing the Bayesian technique to find an appropriate estimate of $\boldsymbol{\omega}$.

Conditional on $\boldsymbol{\omega}$, we obtain the likelihood functions of $Q_1|\boldsymbol{\omega}$ and $Q_2|\boldsymbol{\omega}$, denoted by $L(Q_1|\boldsymbol{\omega})$ and $L(Q_2|\boldsymbol{\omega})$, respectively. Note that $L(Q_x|\boldsymbol{\omega})$ is induced by the map Q_x and is computed by the forward computation using $P^B_\Lambda(a, b|\boldsymbol{\omega})$. Then, we obtain the likelihood function of the output of

Q_X as

$$L(Q_X|\boldsymbol{\omega}) = g_1 L(Q_1|\boldsymbol{\omega}) + g_2 L(Q_2|\boldsymbol{\omega}).$$

We further specify the following prior for $\boldsymbol{\omega}$ as

$$pr(\boldsymbol{\omega}) = \lambda_1\lambda_2\lambda_3\lambda_4 \exp(-\boldsymbol{\lambda}^\top \boldsymbol{\omega}),$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]^\top = [1/2, 1/2, 1/2, 1/2]^\top$ containing rates of exponential distributions. Consequently, the posterior of $\boldsymbol{\omega}$ can be computed as

$$P(\boldsymbol{\omega}|\{Q_X^i\}_{i=1}^n) \propto \Pi_{i=1}^n L(Q_X^i|\boldsymbol{\omega}) pr(\boldsymbol{\omega}),$$

where $\{Q_X^i\}_{i=1}^n$ are identically independent samples of Q_X with $n = 20$. Since $L(Q_X|\boldsymbol{\omega})$ has no closed form in general, we use its empirical estimates to implement the Metropolis-Hastings algorithm for approximating the mean of the posterior.

The resulting mean of $P(\boldsymbol{\omega}|\{Q_X^i\}_{i=1}^n)$ is $\hat{E}(\boldsymbol{\omega}) = [1.7532, 1.7803, 2.2680, 2.2793]^\top$, and the resulting estimated probability distribution of (a, b) is

$$\hat{P}_\Lambda^B = P_\Lambda^B(a, b|\hat{E}(\boldsymbol{\omega})) = \begin{bmatrix} 0.0916 & 0.1266 & 0.0908 \\ 0.1155 & 0.1597 & 0.1145 \\ 0.0893 & 0.1235 & 0.0885 \end{bmatrix}.$$

The corresponding output distributions induced by $P_\Lambda^B(a, b|\hat{E}(\boldsymbol{\omega}))$ are

$$\begin{bmatrix} P_{\mathcal{D}_1}^B(0) \\ P_{\mathcal{D}_1}^B(1) \end{bmatrix} = \begin{bmatrix} 0.4801 \\ 0.5199 \end{bmatrix}, \begin{bmatrix} P_{\mathcal{D}_2}^B(0) \\ P_{\mathcal{D}_2}^B(1) \end{bmatrix} = \begin{bmatrix} 0.5901 \\ 0.4099 \end{bmatrix}. \tag{2.18}$$

Even though $P_\Lambda^B(a, b|\hat{E}(\boldsymbol{\omega}))$ is not a GFGD, the results on the induced output distributions in (2.18) are close to the given output distributions in (2.17), and $P_\Lambda^B(a, b|\hat{E}(\boldsymbol{\omega}))$ provides distributional information about P_Λ, e.g. the mode at $(2, 2)$.

This particular Bayesian approach essentially uses a family of probability distributions to approximate the class of GFGDs. In general, there is no guarantee that this approach can have a good approximation to the GFGD. In this example, an assumption of independence of (a, b) is reasonable and feasible since the generating distribution is formed by two independent distributions. However, when a and b are dependent, the assumption of independence might be not feasible. In fact, no good Bayesian approximation has been found in this case, while the results in the iterative sliced inverse approach are still satisfying since the assumption of independence is not necessary. In addition, there is also an underlying ansatz assumption in this example which is induced by the assumption of beta-binomial distributions. The results from this Bayesian approach are highly dependent on the induced ansatz, while the iterative sliced inverse approach can reduce such bias introduced by the uniform ansatz.

2.4.2 Extension of the Dempster-Shafer Theory

The Dempster-Shafer (DS) theory introduced in Dempster (2008); Yager and Liu (2008); Zhang and Liu (2011) characterizes an upper bound and a lower bound for the probability of an event of random inputs. We compare the DS theory and the sliced inverse approach in the example in Section 2.1.

We first introduce the notion of the DS theory. Suppose there is a multivalued mapping Γ which assigns a subset $\Gamma_x \subset \Lambda$ to each $x \in \mathcal{X}$, and there is a probability distribution γ_{Γ_x} defined on $(\Gamma_x, \mathcal{P}_{\Gamma_x})$ such that

$$P_{DS}(A) = \int_{\mathcal{X}} \gamma_{\Gamma_x}(A \cap \Gamma_x) dP_{\mathcal{X}}(x) = \sum_{x=0}^{1} g_x \gamma_{\Gamma_x}(A \cap \Gamma_x),$$

where $A \in \mathcal{P}_\Lambda$, is a probability distribution defined on $(\Lambda, \mathcal{P}_\Lambda)$. Then, it has been shown that the following two functions,

$$P^*(A) = \frac{P_\mathcal{X}(\{x \in \mathcal{X} : \Gamma_x \cap A \neq \emptyset\})}{P_\mathcal{X}(\{x \in \mathcal{X} : \Gamma_x \neq \emptyset\})},$$
$$P_*(A) = \frac{P_\mathcal{X}(\{x \in \mathcal{X} : \emptyset \neq \Gamma_x \subset A\})}{P_\mathcal{X}(\{x \in \mathcal{X} : \Gamma_x \neq \emptyset\})},$$

are the upper and lower bounds for $P_{DS}(A)$, respectively. In the following, we show some explicit results to explain P_{DS}, P^*, and P_*, and illustrate the connection between the SIP solutions and the DS functions.

By first specifying $\Gamma_x = Q_x^{-1}(\mathcal{D}) = \Lambda$ and $\gamma_{\Gamma_x} = P_\Lambda$ in (2.2), we have $P_{DS} = P_\Lambda$, and specifically,

$$\text{True probability: } P_{DS}(\{(1,1),(2,2)\}) = P_\Lambda(\{(1,1),(2,2)\}) = 17/32,$$
$$\text{Upper bound: } P^*(\{(1,1),(2,2)\}) = 1,$$
$$\text{Lower bound: } P_*(\{(1,1),(2,2)\}) = 0.$$

The DS functions give a crude bound for the probability of the event $\{(1,1),(2,2)\}$ in Λ. Furthermore, since the generating distribution P_Λ is unknown in the SIP, we have the following result by specifying $\gamma_{\Gamma_x} = P_{\Lambda,x}$ in (2.9),

Approximated probability by inverse distributions:

$$\hat{P}_{DS}(\{(1,1),(2,2)\}) = \sum_{x=0}^{1} g_x P_{\Lambda,x}(\{(1,1),(2,2)\} \cap \Gamma_x) = 113/240. \quad (2.19)$$

Even though the probability of any event in Λ that is approximated by plugging in inverse distributions in (2.19) might not be any close to the true probability, it is still better than a crude bound $(0,1)$. Interestingly, the approximation $\hat{P}_{DS}$ is exactly the EEI $\bar{P}$ in (2.11), and thus (2.19) essentially uses the EEI to approximate P_Λ. In this case, the sliced inverse approach can be viewed as

an extension of the DS theory by decomposition of a probability distribution (i.e. by using inverse distributions).

Similarly, we investigate a more practical case in which the output data is observed. We specify $\Gamma_x = Q_x^{-1}(0)$ and

$$\gamma_{\Gamma_x}(A \cap \Gamma_x) = \frac{\sum_{x=0}^{1} P_\Lambda(A \cap Q_x^{-1}(0)) P_\mathcal{X}(x)}{\sum_{x=0}^{1} P_\Lambda(Q_x^{-1}(0)) P_\mathcal{X}(x)}, \tag{2.20}$$

where $A \in \mathcal{P}_\Lambda$. Note that $\gamma_{\Gamma_x}(A \cap \Gamma_x)$ in (2.20) is actually the conditional probability given the Borel sigma algebra of $(a, b)|Q_x = 0$. Then, we have

$$\text{True conditional probability: } P_{DS}(\{(1,1),(2,2)\}) = 1/28,$$
$$\text{Upper bound: } P^*(\{(1,1),(2,2)\}) = 1/3,$$
$$\text{Lower bound: } P_*(\{(1,1),(2,2)\}) = 0,$$

and by plugging in the SIP solutions $\{P_{\Lambda,x}\}_{x=0}^{1}$ in (2.20) as

$$\gamma_{\Gamma_x}(A \cap \Gamma_x) = \frac{\sum_{x=0}^{1} P_{\Lambda,x}(A \cap Q_x^{-1}(0)) P_\mathcal{X}(x)}{\sum_{x=0}^{1} P_{\Lambda,x}(Q_x^{-1}(0)) P_\mathcal{X}(x)}, \tag{2.21}$$

we have the following approximation of the true conditional probability as

Approximated conditional probability by inverse distributions:

$$\hat{P}_{DS}(\{(1,1),(2,2)\}) = \frac{\sum_{x=0}^{1} P_{\Lambda,x}(\{(1,1),(2,2)\} \cap Q_x^{-1}(0)) P_\mathcal{X}(x)}{\sum_{x=0}^{1} P_{\Lambda,x}(Q_x^{-1}(0)) P_\mathcal{X}(x)} = 1/14. \tag{2.22}$$

In this way, we extend the DS probabilities to the SIP solutions by decomposition of a probability distribution to have a concrete approximation of the true probability of any event (i.e. the probability computed from the generating distribution).

There is a case in which we do seek to find an upper and lower bound for the true probability of any event. The DS probabilities P_* and P^* might not be satisfying, as was shown in these examples,

is because they lose the information of the probability distribution of (a, b) and only focus on the information that comes from x. Thus, a better way of defining the upper and lower bounds for the true probability of an event $A \in \mathcal{P}_\Lambda$ is by $P_*(A; \mathbb{P}_{\Lambda,\mathcal{X}})$ and $P^*(A; \mathbb{P}_{\Lambda,\mathcal{X}})$ in Section 2.2.2. This bound gives accurate information about the true probability which is always contained in the bound.

Chapter 3

Sliced Inverse Approach for Stochastic Inverse Problems: Continuum Distributions

3.1 Stochastic Inverse Problems in Continuous Domains

3.1.1 A Simple Linear Example

To illustrate the ideas, we consider a simple linear model

$$y = a_1 x + a_2, \tag{3.1}$$

where $\boldsymbol{a} = (a_1, a_2)^\top$ is a vector of unobservable random inputs of the model on the continuous domain Λ, a compact set in $\mathbb{R}^2$, x is an observable deterministic input in $\mathbb{R}^1$, and y is the scalar output in $\mathbb{R}^1$ whose range depends on x. We assume that $\boldsymbol{a}$ has a probability distribution on $(\Lambda, \mathcal{B}_\Lambda)$ where $\mathcal{B}_\Lambda$ is the Borel sigma algebra of Λ. This distribution is of interest in this paper and is referred to as the *generating distribution* of $\boldsymbol{a}$. In the statistical literature, this model can be viewed as a random effect model, in which a_1 is a random slope and a_2 is the quantity representing the intercept as well as the confounder. This can also be interpreted as a factor model in which y is a linear combination of two unknown factors a_1 and a_2. The goal of the stochastic inverse problem is to recover the distribution of $\boldsymbol{a}$ from the observed distribution on y given x.

Before illustrating the concept of the stochastic inverse problem by concrete examples of $\boldsymbol{a}$, we introduce the concept of the stochastic forward problem which in turn defines the inverse problem. In (3.1), for a fixed x and a realization of $\boldsymbol{a}$, a value of y can be uniquely determined. The "forward computation" here is a simple example of *stochastic forward problem* (SFP; Butler et al. (2014), Butler et al. (2015)), which is exactly a data-generating process. For illustration purpose, we consider the following three examples:

(I) $\boldsymbol{a}$ has a uniform distribution on the unit square $[0,1] \times [0,1] = [0,1]^2$;

(II) $\boldsymbol{a}$ has a normal distribution truncated on the unit square $[0,1]^2$, with mean $(0.5, 0.5)^\top$ and covariance matrix $[0.1, 0.75\sqrt{0.005}; 0.75\sqrt{0.005}, 0.05]$.

(III) $\boldsymbol{a}$ has a uniform distribution on the marginals, a_1 or a_2, with a dependence structure characterized by an Archimedean copula,

$$C(u_1, u_2; \theta) = \psi^{[-1]}(\psi(u_1; \theta) + \psi(u_2; \theta); \theta),$$

where ψ is the Frank copula with $\theta = 1$, and $\psi^{[-1]}$ is the pseudo-inverse of ψ; see Nelsen (2007) for more details.

The density functions of $\boldsymbol{a}$ are depicted in panels (a)-(c) of Figure 3.1.

In the SFP, for a value of x, the distribution of y induced by the distribution of $\boldsymbol{a}$ is referred to as the *output (probability) distribution*. For instance, in panels (d)-(f) of Figure 3.1, the output density functions for $x = 1$ are depicted for each example. Complementing the SFP, the *stochastic inverse problem* (SIP; Butler et al. (2014), Butler et al. (2015)) is the problem of recovering a distribution of $\boldsymbol{a}$ given the distribution of y and a value of x. The recovered distribution and the generating distribution both induce the same distribution of the output y, for a specific x, through the forward computation in the SFP. Panels (g)-(i) of Figure 3.1 show the recovered distributions for Example I, II and III, respectively, given the output distributions of $y|x = 1$ in panels (d)-(f). The procedure of solving the SIP requires inverting (3.1). We emphasize that there are multiple solutions in general since the map from $\boldsymbol{a}$ to y is not one to one.

An important fact in the solutions is that, in each row of Figure 3.1, the recovered distribution and the generating distribution induce the same output distribution for $x = 1$, however, not for any other $x \in \mathbb{R}^1$ in general. This implies that the recovered distribution is a "local and conditioned" solution to the SIP with respect to the specific x. An interesting question arises, *besides the generating distribution, can we find a distribution that generates the same output distribution for all x?*

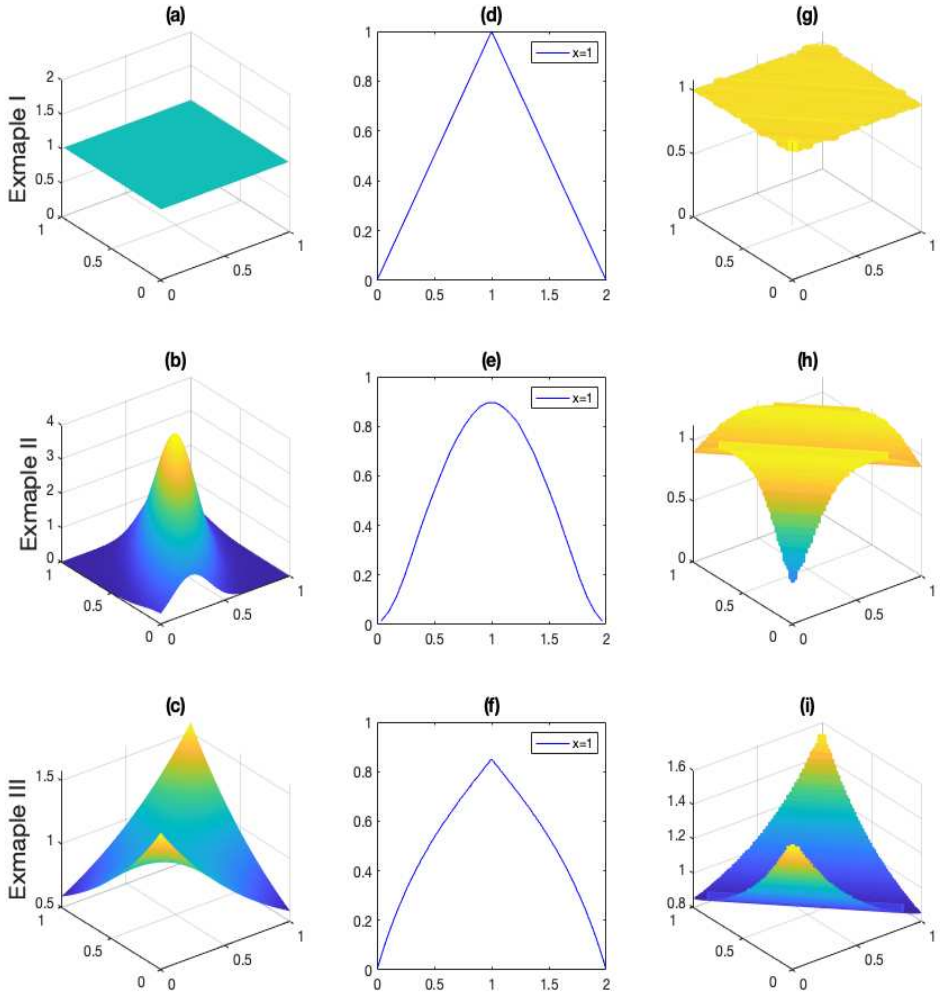


Figure 3.1: A graphical display of generating density functions of $\boldsymbol{a}$ (left), output probability density functions (middle) resulting from the SFP examples (I), (II) and (III), and inverse density functions solved for the SIP regarding the examples (I), (II) and (III). Panels (a)-(c) show generating distributions of $\boldsymbol{a}$ for examples (I), (II) and (III), respectively. The output probability distributions of y for $x = 1$ induced by the generating distributions in panels (a)-(c) are shown in panels (d)-(f), respectively. The results of solving the SIP given the output probability distributions of $y|x = 1$ in panels (d)-(f) are shown in panels (g)-(i), respectively. In particular, the generating distribution of $\boldsymbol{a}$ (left) and the inverse distribution (right) induce the same output probability distribution (middle) for $x = 1$ in each example.

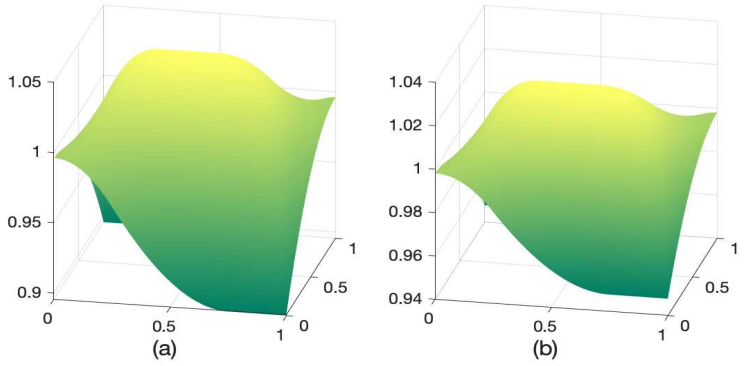


Figure 3.2: A graphical display of two distributions P_Λ^1 and P_Λ^2 in panels (a)-(b), respectively, alternative to the uniform generating distribution in Example I that induce the same output distribution for all $x \in (0, 1)$.

In this paper, the ultimate goal is to find a "global" solution, i.e., a distribution that induces the output distribution for all $x \in \mathbb{R}^1$ as the generating distribution does. In Figure 3.2, we show two alternative distributions P_Λ^1 and P_Λ^2 in panels (a)-(b), respectively, to the generating distribution in Example I such that all induce the output distribution for all $x \in (0, 1)$; see Section 3.2 for detailed formulation. Clearly, the distributions in Figure 3.2 are different from the uniform generating distribution. In fact, all three distributions are *equivalent* in the sense that they provide the exact same information of the output for all $x \in (0, 1)$. Any equivalent distributions of $\boldsymbol{a}$ are not identifiable in the SIP, and the distributional information they deliver are equally important.

In the next section, we provide a brief introduction to the stochastic inverse problem from a mathematical and probabilistic viewpoint to explain how we obtain the solutions in Figure 3.1 and the non-uniqueness issue in solving the SIP.

3.1.2 Background of the SIP

We describe the SIP in a model $y = Q(\boldsymbol{a}, \boldsymbol{x})$ characterized by a general measurable map Q. Here, $\boldsymbol{a}$ is the unobservable random input of interest in the domain Λ, and $\boldsymbol{x} \in \mathcal{X}$ is the observable deterministic input that governs the (scalar) output y of the model. Let Λ and $\mathcal{X}$ be metric spaces, and further let Λ be compact. For any given $\boldsymbol{x} \in \mathcal{X}$, $Q_{\boldsymbol{x}}(\cdot) \equiv Q(\boldsymbol{x}, \cdot)$ denotes a measurable map from Λ to $\mathcal{D}_{\boldsymbol{x}}$ which is the range of y given $\boldsymbol{x}$. In this paper, we assume that $Q_{\boldsymbol{x}}$ is not 1-1 and

continuously differentiable with respect to $\boldsymbol{a}$, and the Jacobian of $Q_{\boldsymbol{x}}$ has full rank except for a (Lebesgue) zero-measure set, see Section 3 of Butler et al. (2014) for details.

Then the SFP can be described by the following process: Any probability measure P_Λ on $(\Lambda, \mathcal{B}_\Lambda)$ induces a probability measure $P_{\mathcal{D}_{\boldsymbol{x}}}$ on $(\mathcal{D}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}})$ through

$$P_{\mathcal{D}_{\boldsymbol{x}}}(C) = P_\Lambda(Q_{\boldsymbol{x}}^{-1}(C)) \tag{3.2}$$

for any event C in the Borel sigma algebra $\mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}}$ generated by $\mathcal{D}_{\boldsymbol{x}}$. The inverse map $Q_{\boldsymbol{x}}^{-1}$ is defined as

$$Q_{\boldsymbol{x}}^{-1}(C) = \{\lambda \in \Lambda : Q_{\boldsymbol{x}}(\lambda) \in C\}.$$

We find $P_{\mathcal{D}_{\boldsymbol{x}}}$ for a given P_Λ through the "forward computation" (3.2) in the SFP, while we find a distribution of $\boldsymbol{a}$ that satisfies (3.2) given the observed $P_{\mathcal{D}_{\boldsymbol{x}}}$ in the SIP.

Because $Q_{\boldsymbol{x}}$, Λ, $\mathcal{D}_{\boldsymbol{x}}$ and $P_{\mathcal{D}_{\boldsymbol{x}}}$ are the only information we have in the SIP, we decompose Λ and P_Λ according to $\mathcal{D}_{\boldsymbol{x}}$ and $P_{\mathcal{D}_{\boldsymbol{x}}}$ through the inverse of $Q_{\boldsymbol{x}}$, respectively. Essentially, it is to decompose P_Λ based on the decomposition of Λ such that $P_{\mathcal{D}_{\boldsymbol{x}}}$ is embedded in the decomposition of P_Λ. Note that any points λ_1, λ_2 in $Q_{\boldsymbol{x}}^{-1}(y)$ are equivalent in the sense that $Q_{\boldsymbol{x}}(\lambda_1) = Q_{\boldsymbol{x}}(\lambda_2) = y$, which is exactly an equivalence relation. Hence, $Q_{\boldsymbol{x}}^{-1}(\cdot)$ is an equivalence class indexed by the point in the output. In Butler et al. (2014), $Q_{\boldsymbol{x}}^{-1}(\cdot)$ is considered as a manifold in Λ and thus is referred to as a *generalized contour*. Generalized contours indexed by distinct points in the output are disjoint due to the fact that the Jacobian of $Q_{\boldsymbol{x}}$ has full rank at every point in Λ. Then it is a straightforward result that Λ can be decomposed as a collection of generalized contours

$$\Lambda = \bigcup_{y \in \mathcal{D}_{\boldsymbol{x}}} Q_{\boldsymbol{x}}^{-1}(y).$$

Let $\mathcal{L}_{\boldsymbol{x}}$ be the quotient space of Λ under the equivalence relation, where each point in $\mathcal{L}_{\boldsymbol{x}}$ represents a generalized contour in Λ. Further let $\mathcal{E}_\lambda$ denote the point in $\mathcal{L}_{\boldsymbol{x}}$ that corresponds to the

generalized contour $Q_{\boldsymbol{x}}^{-1}(Q_{\boldsymbol{x}}(\lambda))$ for $\lambda \in \Lambda$. We have

$$\mathcal{E}_A = \{\mathcal{E}_\lambda : \lambda \in A\}$$

for any $A \in \mathcal{B}_\Lambda$, which contains a collection of points in $\mathcal{L}_{\boldsymbol{x}}$ corresponding to generalized contours that intersect A. Consequently, Λ is decomposed by $\mathcal{L}_{\boldsymbol{x}}$ as

$$\Lambda = \bigcup_{\lambda \in \Lambda} \mathcal{E}_\lambda.$$

In practice, $\mathcal{L}_{\boldsymbol{x}}$ can be any indexing manifold in Λ, e.g. a line in Λ that intersects each of the generalized contours once and only once. In this case, we switch the decomposition from $\mathcal{D}_{\boldsymbol{x}}$ to a manifold in Λ, where we can obtain a unique solution.

Since $Q_{\boldsymbol{x}}$ is a 1-1 and onto map from $\mathcal{L}_{\boldsymbol{x}}$ to $\mathcal{D}_{\boldsymbol{x}}$, $Q_{\boldsymbol{x}}$ *uniquely* induces a probability measure $P_{\mathcal{L}_{\boldsymbol{x}}}$ on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}})$ through

$$P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = P_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad A \in \mathcal{B}_\Lambda.$$

This is a solution to the SIP which is specifically defined on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}})$. Then we seek to extend $P_{\mathcal{L}_{\boldsymbol{x}}}$ to $(\Lambda, \mathcal{B}_\Lambda)$.

The *disintegration theorem* is one of the approaches suggested by Butler et al. (2014). Under general assumptions, for any $\boldsymbol{x}$, any distribution P_Λ can be disintegrated as

$$P_\Lambda(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda, \tag{3.3}$$

where $\mathcal{P}_{\boldsymbol{x}} = \{P_\ell : \ell \in \mathcal{L}_{\boldsymbol{x}}\}$ is a family of conditional probability measures on $(\Lambda, \mathcal{B}_\Lambda)$. In the disintegration of P_Λ, $\mathcal{P}_{\boldsymbol{x}}$ and $P_{\mathcal{L}_{\boldsymbol{x}}}$ are uniquely determined; see Chang and Pollard (1997); Butler et al. (2014) for more details. The theorem essentially describes a process that any d-dimensional probability measure P_Λ can be computed by iterated integrals of a $(d-1)$-dimensional *conditional measure* P_ℓ along each generalized contour and a one-dimensional *marginal measure* $P_{\mathcal{L}_{\boldsymbol{x}}}$ on the

set of generalized contours. It is important to note that the disintegration of P_Λ is performed by a choice of $\boldsymbol{x}$; In other words, P_Λ is "sliced" into sub-dimensions of Λ according to $\boldsymbol{x}$.

In turn, we can use the disintegration (3.3) to determine P_Λ given $P_{\mathcal{L}_{\boldsymbol{x}}}$. One may notice that there is no way to determine $\mathcal{P}_{\boldsymbol{x}}$ given the information in the SIP. One possible approach suggested by Butler et al. (2014) is to specify a probability measure on each of the generalized contours. Such distributional assumption is referred to as an *ansatz* by those authors. By specifying an azsatz of $\mathcal{P}_{\boldsymbol{x}}$ in (3.3), we have an approximation of P_Λ, denoted by $P_{\Lambda,\boldsymbol{x}}$. This approximation $P_{\Lambda,\boldsymbol{x}}$ is considered as a solution to the SIP for the given x, because $P_{\Lambda,\boldsymbol{x}}$ satisfies (3.2) and induces $P_{\mathcal{D}_{\boldsymbol{x}}}$ through the inverse map $Q_{\boldsymbol{x}}^{-1}$. In this paper, $P_{\Lambda,\boldsymbol{x}}$ is referred to as a *(sliced) inverse distribution* of the SIP. The notion of "sliced" means $P_{\Lambda,\boldsymbol{x}}$ is constructed by "slicing" the domain Λ according to the given x. Note that there is no guarantee that P_Λ and $P_{\Lambda,\boldsymbol{x}}$ are close for a given $\boldsymbol{x}$. Those authors also suggest the uniform ansatz under which each P_ℓ is a uniform distribution and points along a generalized contour are equally likely. In the next section, we provide information theoretic justification for the uniform ansatz.

We once again consider Examples I, II and III in (3.1) with $Q_{\boldsymbol{x}}(\boldsymbol{a}) = \boldsymbol{x}^\top \boldsymbol{a}$ for $\boldsymbol{a} = (a_1, a_2)^\top$ and a given $\boldsymbol{x} = (1/2, 1)^\top$. It can be seen that

$$Q_{\boldsymbol{x}}^{-1}(A) = \{\boldsymbol{a} \in \Lambda : \boldsymbol{x}^\top \boldsymbol{a} \in A\}, \quad A \in \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}},$$

which is a collection of parallel lines (or segments) with a common normal vector $\boldsymbol{x}$; see panel (a) of Figure 3.3. In panels (b), (c) and (d) of Figure 3.3, the quotient spaces $\mathcal{L}_{\boldsymbol{x}}$ are depicted in blue lines and the unique marginal measures $P_{\mathcal{L}_{\boldsymbol{x}}}$ are depicted in red lines for Examples I, II and III, respectively. Then the inverse distributions under the uniform ansatz are simply computed by "stretching out" the unique measures $P_{\mathcal{L}_{\boldsymbol{x}}}$ along the generalized contours (i.e. lines segments) by disintegration.

It is important to note that $P_{\Lambda,\boldsymbol{x}}$ is generally conditioned on $\boldsymbol{x}$ due to the pre-specified ansatz. For illustration purpose, we consider three values of $x = -1, 1, 2$ in Example II and show the inverse distributions under the uniform ansatz in panels (a)-(c) of Figure 3.4. All the inverse dis-

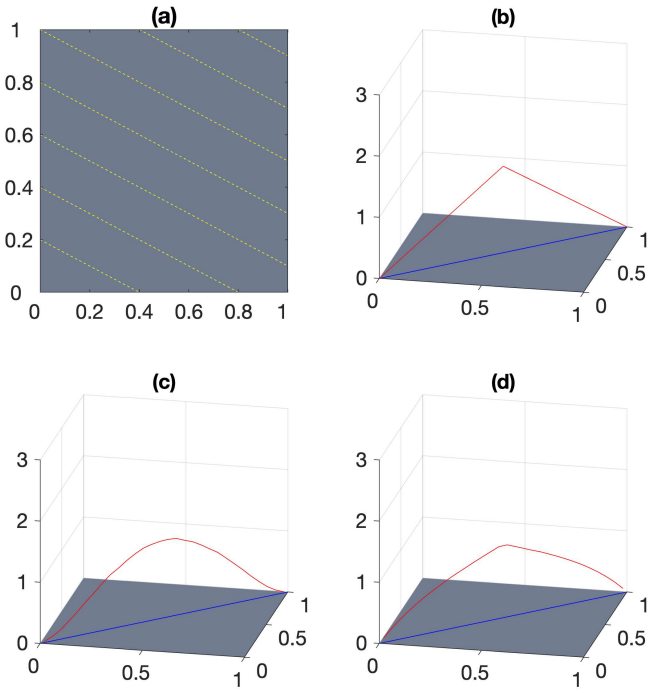


Figure 3.3: A graphical display of the generalized contours (dashed) on the domain of $\boldsymbol{a}$ in (3.1) for a specified $x = 1/2$ is shown in panel (a). Panels (b)-(d) show the unique probability measures $P_{\mathcal{L}_{\boldsymbol{x}}}$ on the quotient space $\mathcal{L}_{\boldsymbol{x}}$ for Example I, II and III, respectively.

tributions are highly dependent on $\boldsymbol{x}$, which is not intuitive given that the generating distribution does not depend on $\boldsymbol{x}$. The "true" conditional measure of $\boldsymbol{a}$ along each generalized contour disintegrated from the generating distribution is not equal to the pre-specified ansatz. Bias from the choice of the ansatz leads to the dependency of inverse distributions on $\boldsymbol{x}$.

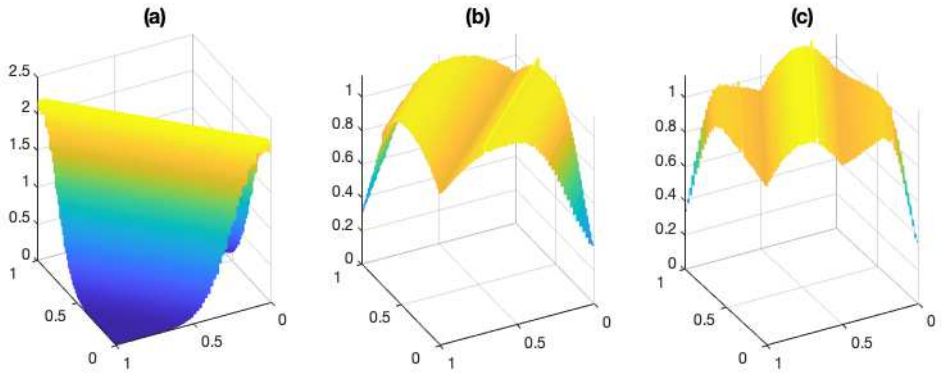


Figure 3.4: A graphical display of inverse distributions under the uniform ansatz for three different values of x, (a) $x = -1$, (b) $x = 1$ and (c) $x = 2$, in the simple linear model for Example II.

3.1.3 Uniform Ansatz and Maximum Entropy Solution

In this section, we explore the role of the ansatz. The ansatz refers to the assumption of a family of conditional probability measures assigned to $\boldsymbol{a}$ along the generalized contours, i.e. a choice of $\mathcal{P}_{\boldsymbol{x}}$. Hence, by the disintegration theorem, we have multiple solutions depending on the choice of the ansatz.

For instance, under the uniform ansatz, for each $\ell \in \mathcal{L}_{\boldsymbol{x}}$, P_ℓ is distributed according to the uniform distribution on the generalized contour $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)$. The uniform ansatz can be formally written as

$$\mathcal{P}_{\boldsymbol{x}}^{u} := \{P_\ell : \ell \in \mathcal{L}_{\boldsymbol{x}}, P_\ell \text{ is uniform on } \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\},$$

i.e. a set of conditional uniform distributions. Replacing $\mathcal{P}_{\boldsymbol{x}}$ in (3.3) with $\mathcal{P}^u_{\boldsymbol{x}}$, we obtain an inverse distribution $P_{\Lambda,\boldsymbol{x}}$ as

$$\int_{\ell\in\mathcal{E}_A}\int_{\lambda\in\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\cap A} dP^u_\ell(\lambda)dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = P_{\Lambda,\boldsymbol{x}}(A), \quad \forall A\in\mathcal{B}_\Lambda. \tag{3.4}$$

where $P^u_\ell \in \mathcal{P}^u_{\boldsymbol{x}}$ for $\ell\in\mathcal{L}_{\boldsymbol{x}}$. Another choice of $\mathcal{P}_{\boldsymbol{x}}$ is given by the truncated normal ansatz, which can be written as

$$\mathcal{P}^{tn}_{\boldsymbol{x}} := \{P_\ell : \ell\in\mathcal{L}_{\boldsymbol{x}}, P_\ell \text{ is truncated normal on } \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \text{ with pre-specified mean and variance}\}.$$

Similarly, we can obtain an inverse distribution

$$\int_{\ell\in\mathcal{E}_A}\int_{\lambda\in\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\cap A} dP^{tn}_\ell(\lambda)dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = P^{tn}_{\Lambda|\boldsymbol{x}}(A), \quad \forall A\in\mathcal{B}_\Lambda,$$

where $P^{tn}_\ell \in \mathcal{P}^{tn}_{\boldsymbol{x}}$ for $\ell\in\mathcal{L}_{\boldsymbol{x}}$. These two inverse distributions, $P_{\Lambda|\boldsymbol{x}}$ and $P^{tn}_{\Lambda|\boldsymbol{x}}$, are both considered as solutions to the SIP for the given $\boldsymbol{x}$ in the sense that they induce the same output $P_{\mathcal{D}_{\boldsymbol{x}}}$ through the forward computation (3.2) as P_Λ does.

We explore the above uniform and truncated normal ansatz in Example I in the simple linear example. The generalized contours corresponding to $x = 1/2$ are shown as dashed lines in panel (a) of Figure 3.5. Three viable choices of ansatz along the contours are considered, including the uniform ansatz and two versions of truncated normal ansatz, namely $\mathcal{P}^{tn1}_{\boldsymbol{x}}$ and $\mathcal{P}^{tn2}_{\boldsymbol{x}}$. Each truncated normal distribution in $\mathcal{P}^{tn1}_{\boldsymbol{x}}$ has a mean in Λ, and the means of all distributions in $\mathcal{P}^{tn1}_{\boldsymbol{x}}$ are located on an arc of circle, $\{\boldsymbol{a}\in\Lambda : (a_1-1)^2+a_2^2 = 1\}$, which is depicted as a dotted line in panel (c) of Figure 3.5. All distributions in $\mathcal{P}^{tn1}_{\boldsymbol{x}}$ have a common variance 0.01. Similarly, the dotted line $\{\boldsymbol{a}\in\Lambda : a_1^2+(a_2-1)^2 = 1\}$ in panel (d) of Figure 3.5 shows the location of means of all truncated normal distributions, with common variance 0.01, in $\mathcal{P}^{tn2}_{\boldsymbol{x}}$.

For illustration purpose, we show all the ansatz along a specific contour $\{\boldsymbol{a}\in\Lambda : y = a_1x+a_2\}$ with $y = 0.4$ in Figure 3.5. The contour is depicted as a thick dashed-line in panel (a). On the

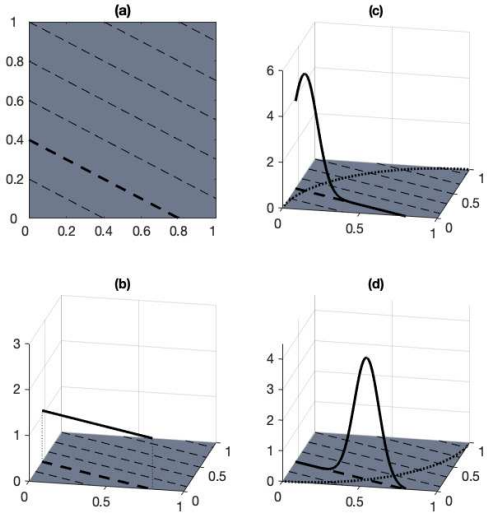


Figure 3.5: A graphical display of the generalized contours of Example I in panel (a) and three representative distributions in three choices of ansatz, including the uniform ansatz and two versions of truncated normal ansatz $\mathcal{P}_{\boldsymbol{x}}^{tn1}, \mathcal{P}_{\boldsymbol{x}}^{tn2}$, in panels (b)-(d), respectively. Distributions in $\mathcal{P}_{\boldsymbol{x}}^{tn1}$ and $\mathcal{P}_{\boldsymbol{x}}^{tn2}$ have means on the dotted lines in panels (c) and (d), respectively, with a common variance 0.01. The three representative distributions in panels (b)-(d) are investigated along a specific contour shown as a solid thick line in panel (a).

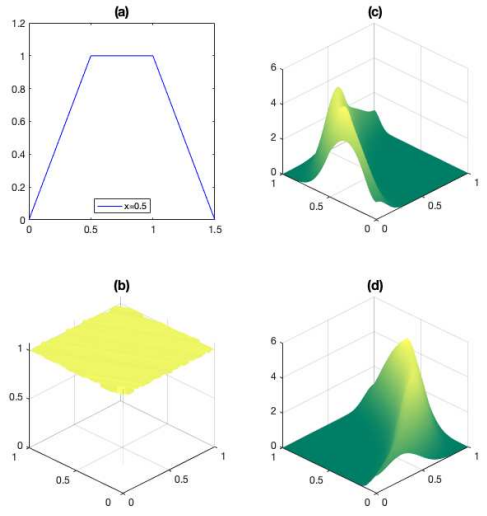


Figure 3.6: A graphical display of three inverse distributions in panels (b), (c) and (d) computed by the ansatz information provided in panels (b), (c) and (d) of Figure 3.5, respectively. Panel (a) shows the induced distribution $P_{\mathcal{D}_{\boldsymbol{x}}}$ of the output $y|x = 0.5$ from these three inverse distribution.

contour, the distribution in the uniform ansatz is shown in panel (b), the distribution in $\mathcal{P}_{\boldsymbol{x}}^{tn1}$ with a mean of $(0.069, 0.365)^\top \in \Lambda$ and a variance of 0.01 is depicted in panel (c), and the distribution in $\mathcal{P}_{\boldsymbol{x}}^{tn2}$ with a mean of $(0.515, 0.143)^\top$ and a variance of 0.01 is depicted in panel (d).

Figure 3.6 shows the given output distribution for $x = 1/2$ in panel (a), and the inverse distributions under certain ansatz in panels (c)-(d). Panel (b) depicts $P_{\Lambda,\boldsymbol{x}}$ under the uniform ansatz, while $P_{\Lambda,\boldsymbol{x}}^{tn1}$ and $P_{\Lambda,\boldsymbol{x}}^{tn2}$ under $\mathcal{P}_{\boldsymbol{x}}^{tn1}$ and $\mathcal{P}_{\boldsymbol{x}}^{tn2}$ are depicted in panels (c) and (d), respectively. The inverse distributions $P_{\Lambda|\boldsymbol{x}}$, $P_{\Lambda|\boldsymbol{x}}^{tn1}$ and $P_{\Lambda|\boldsymbol{x}}^{tn2}$ yield the same prediction for the output for the given $x = 1/2$; in other words, they induce the same output distribution through the forward computation (3.2). Hence, any distribution of $P_{\Lambda|\boldsymbol{x}}$, $P_{\Lambda|\boldsymbol{x}}^{tn1}$ and $P_{\Lambda|\boldsymbol{x}}^{tn2}$ can be chosen as a solution to this SIP. With different means in these solutions, the uniform ansatz is commonly used in practice and suggested by the maximum entropy principle when we have no preference over $\mathcal{P}_{\boldsymbol{x}}$.

We consider a general family of ansatzes, a family of absolutely continuous ansatzes, that contains both the uniform ansatz and the truncated normal ansatz. In probability theory, a natural choice of the dominating measure is the Lebesgue measure. In the SIP, the disintegrated measure μ_ℓ plays the role of a dominating measure along each generalized contour, which is defined in the following disintegration of the Lebesgue measure μ_Λ on Λ,

$$\mu_\Lambda(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A} d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda,$$

where $\mu_{\mathcal{L}_{\boldsymbol{x}}}$ is a measure on $\mathcal{L}_{\boldsymbol{x}}$ induced by μ_Λ through the following computation

$$\mu_{\mathcal{D}_{\boldsymbol{x}}}(C) = \mu_\Lambda(Q_{\boldsymbol{x}}^{-1}(C)), \quad C \in \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}},$$
$$\mu_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = \mu_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad A \in \mathcal{B}_\Lambda.$$

The Lebesgue measure μ_Λ induces $\mu_{\mathcal{D}_{\boldsymbol{x}}}$ on the output range $\mathcal{D}_{\boldsymbol{x}}$, and $\mu_{\mathcal{D}_{\boldsymbol{x}}}$ induces $\mu_{\mathcal{L}_{\boldsymbol{x}}}$ on $\mathcal{L}_{\boldsymbol{x}}$. Given μ_Λ and $\mu_{\mathcal{L}_{\boldsymbol{x}}}$, the disintegrated measures $\{\mu_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ are uniquely determined. Then the family

of absolutely continuous ansatzes can be defined as

$$\mathcal{P}_{\boldsymbol{x}}^{c} := \{\mathcal{P}_{\boldsymbol{x}} : P_\ell \in \mathcal{P}_{\boldsymbol{x}} \text{ is absolutely continuous with respect to } \mu_\ell \text{ on } \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \text{ for each } \ell \in \mathcal{L}_{\boldsymbol{x}}\},$$

where any ansatz in $\mathcal{P}_{\boldsymbol{x}}^{c}$ consists of absolutely continuous distributions along generalized contours.

Among choices in $\mathcal{P}_{\boldsymbol{x}}^{c}$, we choose the uniform ansatz because the maximum entropy principle implies it is least-informative among all choices when there is no information about $\mathcal{P}_{\boldsymbol{x}}$. This idea can be expressed and constructed in the following. Suppose the generating distribution P_Λ is absolutely continuous with respect to μ_Λ. Then we have

Theorem 3.1.1. *$P_{\Lambda,\boldsymbol{x}}$ under the uniform ansatz $\mathcal{P}_{\boldsymbol{x}}^{u}$ is the maximum entropy solution among all possible $P_{\Lambda|\boldsymbol{x}}^{c}$ under the ansatz in $\mathcal{P}_{\boldsymbol{x}}^{c}$.*

Proof. This result is proved in Chapter 5.1.1. □

An advantage gained from using the uniform ansatz is that it uses the least information or assumptions about the model, i.e. without any prior information about conditional distributions along generalized contours. While any inverse distribution under an appropriate chosen ansatz is a solution, we call the unique solution $P_{\Lambda,\boldsymbol{x}}$, the *maximum entropy inverse distribution* (MEID). The MEID is a specific representor we select by using the ansatz in $\mathcal{P}_{\boldsymbol{x}}^{c}$ based on the maximum entropy principle. A key difference of $P_{\Lambda,\boldsymbol{x}}$ and P_Λ is that the former is generally conditioned on the given $\boldsymbol{x}$, while the latter is not. In Section 3.2, we construct a family of solutions that are not conditioned on a specific $\boldsymbol{x}$, and in Section 3.4, we use the MEID to initiate an iterative procedure to determine a candidate of such solutions.

3.2 Feasible Generating Distributions

3.2.1 Equivalent Distributions

In Section 3.1, we introduce a solution, MEID, to the SIP for a single experiment indexed by $\boldsymbol{x}$. In this section, we employ a collection of experiments indexed by values in $\mathcal{X}$ where $\mathcal{X} \subset \mathbb{R}^d$

is the domain for $\boldsymbol{x}$ such that $y = Q_{\boldsymbol{x}}(\boldsymbol{a})$ is well-defined. Let the generating distribution P_Λ be absolutely continuous with respect to μ_Λ where $\Lambda \subset \mathbb{R}^d$ is compact and simply connected. In this case, P_Λ induces a probability distribution $P_{\mathcal{D}_{\boldsymbol{x}}}$ on the output y for any $\boldsymbol{x} \in \mathcal{X}$ through (3.2). Then we seek to tackle P_Λ by considering the given collections, $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$ and $\{Q_{\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$.

By varying the ansatz in the disintegration of P_Λ, we obtain multiple distributions on Λ that induce the same $P_{\mathcal{D}_{\boldsymbol{x}}}$ through the inverse map $Q_{\boldsymbol{x}}^{-1}$. We denote the collection of such distributions by

$$\mathbb{P}_{\Lambda,\boldsymbol{x}} := \{\widetilde{P}_\Lambda \ll \mu_\Lambda : \widetilde{P}_\Lambda(A) = \int_{\ell\in\pi_{\mathcal{L}_{\boldsymbol{x}}}(A)} \int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\cap A} d\widetilde{P}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = P_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \forall A \in \mathcal{B}_\Lambda, \widetilde{P}_\ell \ll \mu_\ell, \forall \ell \in \mathcal{L}_{\boldsymbol{x}}\},$$

where each distribution in this class is absolutely continuous with respect to μ_Λ and induces $P_{\mathcal{D}_{\boldsymbol{x}}}$ through $Q_{\boldsymbol{x}}^{-1}$. In this sense, $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ is an equivalence class indexed by $P_{\mathcal{D}_{\boldsymbol{x}}}$ or $\boldsymbol{x}$, and each distribution in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ is called a *locally feasible generating distribution*. In addition, $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ is not empty and contains $P_{\Lambda,\boldsymbol{x}}$ and P_Λ.

Then P_Λ is contained in any equivalence class in $\{\mathbb{P}_{\Lambda,\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$ since P_Λ induces all the output distributions $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$. By considering the intersection of $\{\mathbb{P}_{\Lambda,\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$ as

$$\mathbb{P}_{\Lambda,\mathcal{X}} := \bigcap_{\boldsymbol{x}\in\mathcal{X}} \mathbb{P}_{\Lambda,\boldsymbol{x}},$$

we obtain a non-empty equivalence class in which each distribution induces $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$. Any element in this family is referred to as a *globally feasible generating distribution* (GFGD).

Theorem 3.2.1. *$\mathbb{P}_{\Lambda,\mathcal{X}}$ is convex: Any mixture defined as*

$$P_{Mix}(A) = wP_1(A) + (1-w)P_2(A), \quad \forall A \in \mathcal{B}_\Lambda,$$

where $P_1, P_2 \in \mathbb{P}_{\Lambda,\mathcal{X}}$ and $1 \geq w \geq 0$, is contained in $\mathbb{P}_{\Lambda,\mathcal{X}}$.

Proof. See Chapter 5.1.2. □

The theorem shows that $\mathbb{P}_{\Lambda,\mathcal{X}}$ is convex and has uncountably many elements in general. In Figure 3.2 of Section 3.1.1, we show two equivalent distributions P^1_Λ and P^2_Λ belonging to $\mathbb{P}_{\Lambda,\mathcal{X}}$ in Example I of the simple linear model in which $\boldsymbol{a}$ have a uniform generating distribution P_Λ on the unit square. Specifically, P^1_Λ is computed by integrating out the inverse distributions $P_{\Lambda,\boldsymbol{x}}$ over $\boldsymbol{x} \in \mathcal{X}$ under a truncated normal ansatz in which the distributions have means located on the centers of the generalized contours and have standard deviation 1. Then P^2_Λ is a mixture of P_Λ and P^1_Λ with equal weights. Any mixture of P^1_Λ, P^2_Λ and P_Λ is also a viable GFGD that induces the distributions of the output, $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x}\in\mathcal{X}}$. In the distributions in $\mathbb{P}_{\Lambda,\mathcal{X}}$, the ansatzes are distinct and informative, which implies the choice of the ansatz plays an important role of finding a GFGD; in other words, we can find a GFGD by correcting the pre-specified ansatz.

3.2.2 Degree of Beliefs and High-probability Regions in the SIP

In this section, we propose an approach to quantify the degree of belief of any event of $\boldsymbol{a}$ given the non-identifiable distributions in $\mathbb{P}_{\Lambda,\mathcal{X}}$, which is also suggested by Dempster (2008). We quantify the degree of belief by considering the *lower* and *upper probabilities* over $\mathbb{P}_{\Lambda,\mathcal{X}}$ as

$$P^*(A;\mathbb{P}_{\Lambda,\mathcal{X}}) = \sup_{P\in\mathbb{P}_{\Lambda,\mathcal{X}}} P(A),$$
$$P_*(A;\mathbb{P}_{\Lambda,\mathcal{X}}) = \inf_{P\in\mathbb{P}_{\Lambda,\mathcal{X}}} P(A).$$

Since $\mathbb{P}_{\Lambda,\mathcal{X}}$ is convex, we have the following corollary as a direct consequence of Theorem 3.2.1.

Corollary 3.2.1.1. *For any $A \in \mathcal{B}_\Lambda$, the set $\{P(A) : P \in \mathbb{P}_{\Lambda,\mathcal{X}}\}$ is convex.*

The corollary shows that the degree of belief of A computed by any $P \in \mathbb{P}_{\Lambda,\mathcal{X}}$ is contained in the bound

$$\left(P_*(A;\mathbb{P}_{\Lambda,\mathcal{X}}), P^*(A;\mathbb{P}_{\Lambda,\mathcal{X}})\right).$$

This bound gives the actual range of the degree of belief in which any value is informative about $\boldsymbol{a}$.

Returning to P_Λ^1, P_Λ^2 and P_Λ in Example I in the simple linear model, we consider an event $A = \{\boldsymbol{a} \in \Lambda : (a_1 - 0.5)^2/0.2^2 + (a_2 - 0.5)^2/0.1^2 <= 1\}$, which has an elliptical shape. Then we have the following degrees of belief.

$$P_\Lambda(A) = 0.0629, P_\Lambda^1(A) = 0.0659, P_\Lambda^2(A) = 0.0644.$$

These numbers show that the probability distribution P_Λ^1 is more concentrated around the mode, i.e. the mean, than any other two.

Similarly, the equivalent class $\mathbb{P}_{\Lambda,\mathcal{X}}$ can be used to find a high-probability region in the SIP. The high-probability δ-region is defined as

$$A_{\delta,P} := \{\lambda \in \Lambda : P'(\lambda) \geq \delta > 0\},$$

where P is an absolutely continuous probability measure with respect to μ_Λ and $P' = dP/d\mu_\Lambda$ is the Radon-Nikodym derivative. Then we consider the minimum and maximum δ-regions over $\mathbb{P}_{\Lambda,\mathcal{X}}$ as

$$A^* = \bigcap_{P \in \mathbb{P}_{\Lambda,\mathcal{X}}} A_{\delta,P},$$
$$A_* = \bigcup_{P \in \mathbb{P}_{\Lambda,\mathcal{X}}} A_{\delta,P}.$$

Consequently, each high-probability δ-region contains A_* and is contained in A^*. In this case, A_* and A^* provide extremal information about the region of $\boldsymbol{a}$ with a high probability.

Continuing to consider $P_\Lambda, P_\Lambda^1, P_\Lambda^2$, we explore the high-probability δ-regions of the three distributions when $\delta = 1.02$. Note that P_Λ is the uniform distribution on a unit square, and thus $A_{\delta,P_\Lambda} = \emptyset$. Then the δ-regions for P_Λ^1 and P_Λ^2 are shown in panels (a) and (b) of Figure 3.2, respectively. The regions have a similar shape around the center (i.e. the mean).

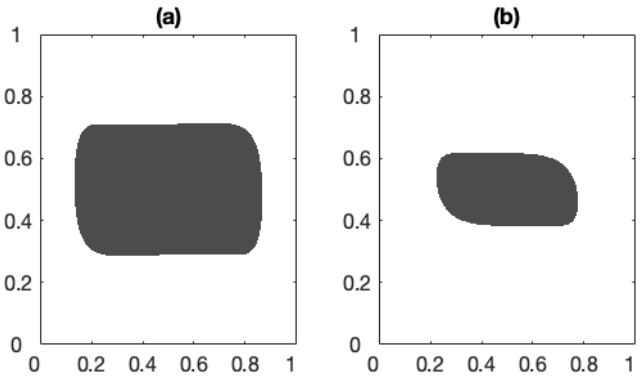


Figure 3.7: The display of two δ-regions when $\delta = 1.02$ for two equivalent distributions shown in panels (a)-(b) of Figure 3.2, respectively.

Due to the fact that the GFGDs in $\mathbb{P}_{\Lambda,\mathcal{X}}$ are equivalent and non-identifiable in the SIP, it is not sufficient to only consider one specific GFGD to make inference.

3.3 Dempster-Shafer Probabilities and Inverse Distributions

The DS theory, as a generalization of Bayesian inference, considers quantifying the degree of beliefs of $\boldsymbol{a}$ by varying $\boldsymbol{x}$, in a similar way to the SIP approach. In the theory, $\boldsymbol{x}$ is considered as an observable random input, and only its distributional information is considered to approximate the probability distribution of $\boldsymbol{a}$. This is a simplistic approach to quantifying the degree of beliefs when no information is given about $\boldsymbol{a}$. In this section, we find an approach that aggregates the inverse distributions from different experiments by varying $\boldsymbol{x}$ and we explore the connection between the DS theory and the SIP. In particular, we extend the simplistic approach in the DS theory by disintegration.

Consider a general model $y = Q(\boldsymbol{X}, \boldsymbol{a})$ where $\boldsymbol{X}$ is random on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_{\mathcal{X}})$ with a sampling distribution $P_{\mathcal{X}}$ that is absolutely continuous with respect to the Lebesgue measure $\mu_{\mathcal{X}}$ on $\mathcal{X}$ and has a positive density. For each $\boldsymbol{X} = \boldsymbol{x} \in \mathcal{X}$, the conditioned model $Q_{\boldsymbol{x}}(\boldsymbol{a})$ is defined in Section 3.2. In addition, the map $Q_{\boldsymbol{x}}(A)$ as a function of $\boldsymbol{x}$ is measurable for any $A \in \mathcal{B}_{\Lambda}$, and $Q_{\boldsymbol{x}}^{-1}(C)$ as a function of $\boldsymbol{x}$ is measurable for any set C in the Borel sigma algebra $\mathcal{B}_{\mathcal{D}}$ of $\mathcal{D} = \cup_{\boldsymbol{x}\in\mathcal{X}}\mathcal{D}_{\boldsymbol{x}}$. Furthermore, the inverse distribution $P_{\Lambda,\boldsymbol{x}}(A)$ is a measurable function of $\boldsymbol{x}$ for any $A \in \mathcal{B}_{\Lambda}$.

In the DS theory, the degree of beliefs are characterized by the belief and plausibility functions. To illustrate, we introduce an important type of functions in the DS theory: the *posterior belief function* (3.5) and the *posterior plausibility function* (3.6)

$$Bel_C(A) = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} I_{\{Q_{\boldsymbol{x}}^{-1}(C)\subset A\}}(\boldsymbol{x})dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} I_{\{Q_{\boldsymbol{x}}^{-1}(C)\neq\emptyset\}}(\boldsymbol{x})dP_{\mathcal{X}}}, \tag{3.5}$$

$$Pl_C(A) = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} I_{\{Q_{\boldsymbol{x}}^{-1}(C)\cap A\neq\emptyset\}}(\boldsymbol{x})dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} I_{\{Q_{\boldsymbol{x}}^{-1}(C)\neq\emptyset\}}(\boldsymbol{x})dP_{\mathcal{X}}}, \tag{3.6}$$

where $A \in \mathcal{B}_\Lambda$, $C \in \mathcal{B}_\mathcal{D}$ and $I_E(\boldsymbol{x}) := I(\boldsymbol{x} \in \mathcal{X} : \boldsymbol{x} \in E \subset \mathcal{X})$ is an indicator function. The DS functions Bel_C and Pl_C have often been called, respectively, the *lower* and *upper probabilities* of A given C, since the functions provides lower and upper bounds for the degree of belief of event A when the collection of generalized contours $Q_{\boldsymbol{x}}^{-1}(C)$ is contained in or intersects A. In particular, the functions are lower and upper bounds for the following quantity of interest introduced in Dempster (2008)

$$P_{DS}(A) = \int \gamma_{Q_{\boldsymbol{x}}^{-1}(C)}(A \cap Q_{\boldsymbol{x}}^{-1}(C))dP_{\mathcal{X}},$$

where $\gamma_{Q_{\boldsymbol{x}}^{-1}(C)}$ is a general member of certain class of probability measures of $\boldsymbol{a}$ given $\boldsymbol{X} = \boldsymbol{x}$ and $y \in C$.

In the following, we investigate the behavior of P_{DS} when we connect the SIP approach to the DS theory in two extreme cases of C. Consider an event $C = (y_0 - \epsilon, y_0 + \epsilon)$ with $y_0 \in \mathcal{D}$ and a sufficiently small $\epsilon > 0$ such that $P_{\Lambda,\boldsymbol{x}}(Q_{\boldsymbol{x}}^{-1}(C))$ is small and $P_{\Lambda,\boldsymbol{x}}(Q_{\boldsymbol{x}}^{-1}(C)) \neq 0$. With the specified probability measure $\gamma_{Q_{\boldsymbol{x}}^{-1}(C)}(A \cap Q_{\boldsymbol{x}}^{-1}(C)) = P_{\Lambda,\boldsymbol{x}}(A \cap Q_{\boldsymbol{x}}^{-1}(C))/P_{\Lambda,\boldsymbol{x}}(Q_{\boldsymbol{x}}^{-1}(C))$, we have

$$P_{DS}(A) = \int \gamma_{Q_{\boldsymbol{x}}^{-1}(C)}(A \cap Q_{\boldsymbol{x}}^{-1}(C))dP_{\mathcal{X}} \approx \int \frac{P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_{A\cap Q_{\boldsymbol{x}}^{-1}(C)})P_\ell^u(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\cap A)}{P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)})P_\ell^u(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell))}dP_{\mathcal{X}},$$

where $\ell \in \mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)}$. Note that the approximate equality here requires the continuity of P_ℓ^u in terms of ℓ for each $A \in \mathcal{B}_\Lambda$. This P_{DS} over a small C implies the following computation:

1. First compute the ratio of probabilities of event A mapped onto the quotient space $\mathcal{L}_{\boldsymbol{x}}$, i.e. $P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_{A\cap Q_{\boldsymbol{x}}^{-1}(C)})/P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)})$. In fact, this computes the weight of $\mathcal{E}_{A\cap Q_{\boldsymbol{x}}^{-1}(C)}$ over the entire $\mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)}$;

2. Then compute the ratio of probabilities of event A along the generalized contour, i.e. the weight of $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\cap A$ over the entire generalized contour $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)$, which is referred to as the ansatz;

3. Finally, integrate out the probabilities over all the experiments (indexed by $\boldsymbol{x}$).

Note that the first two steps are exactly the steps of numerically computing an inverse distribution, specifically by disintegration, described in Butler et al. (2014). The last step is of particular interest in this paper, since it provides an approach that aggregates the probabilities over a collection of experiments when multiple experiments are considered in the SIP. Thus, following the same procedure, we consider $C = \mathcal{D}$ and $\gamma_{Q_{\boldsymbol{x}}^{-1}(\mathcal{D})} = P_{\Lambda,\boldsymbol{x}}$ such that the DS quantity P_{DS} is conditioned on all the information of the output, and we have

$$P_{DS}(A) = \int_{\boldsymbol{x}\in\mathcal{X}}\int_{\ell\in\mathcal{E}_A}\int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\cap A} dP_\ell^u(\lambda)dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell)dP_{\mathcal{X}}, \quad \forall A\in\mathcal{B}_\Lambda. \tag{3.7}$$

Theorem 3.3.1. *P_{DS} is a probability measure on $(\Lambda, \mathcal{B}_\Lambda)$.*

Proof. See Chapter 5.1.3. □

In this case, we can use this DS quantity, which incorporates the local solutions in the SIP, to compute a distribution of $\boldsymbol{a}$ that is not conditioned on any experiment $\boldsymbol{x}$. This DS quantity gives an exact approximation to the degree of belief of an event $A\in\mathcal{B}_\Lambda$ by disintegration rather than a crude bound from the DS functions. We call this particular quantity P_{DS} in (3.7) the *experimental expectation of inverses* (EEI) and denote it by $\bar{P}$ in the rest of this paper.

Returning to Example II in (3.1), Figure 3.8 shows the density functions of the EEI, $\bar{P}$, over $\mathcal{X} = \mathbb{R}$ in panel (a) and the truncated normal generating distribution P_Λ in panel (b). In this example, the EEI and the generating distribution are both unimodal and have similar shape around

the mean. The EEI is very informative and useful to approximate a high-probability δ-region of $\boldsymbol{a}$. In addition, the ansatz along the contours in the EEI consists convex functions, similar to the shape of P_Λ, rather than constant functions in the uniform ansatz. Thus, the EEI in this example provides a better ansatz than the uniform ansatz. In panel (c), the given density function of the output $y|x = -1$ in the SIP is depicted in blue, and the predicted density function of the output $y|x = -1$ is depicted in red. The prediction is performed by the following forward computation

$$\hat{P}_{\mathcal{D}_{\boldsymbol{x}}}(A) = \bar{P}(Q_{\boldsymbol{x}}^{-1}(A)), \tag{3.8}$$

where $A \in \mathcal{B}_\Lambda$, $\boldsymbol{x} = (-1, 1)^\top$, and $\hat{P}_{\mathcal{D}_{\boldsymbol{x}}}$ is the prediction of the output by using $\bar{P}$. As depicted, these two distributions $P_{\mathcal{D}_{\boldsymbol{x}}}, \hat{P}_{\mathcal{D}_{\boldsymbol{x}}}$ are distinct, which means the EEI is not equivalent to the generating distribution in the sense that they induce different distributions of the output $y|x = -1$ through $Q_{\boldsymbol{x}}^{-1}$. In general, the EEI is not contained in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ or $\mathbb{P}_{\Lambda,\mathcal{X}}$.

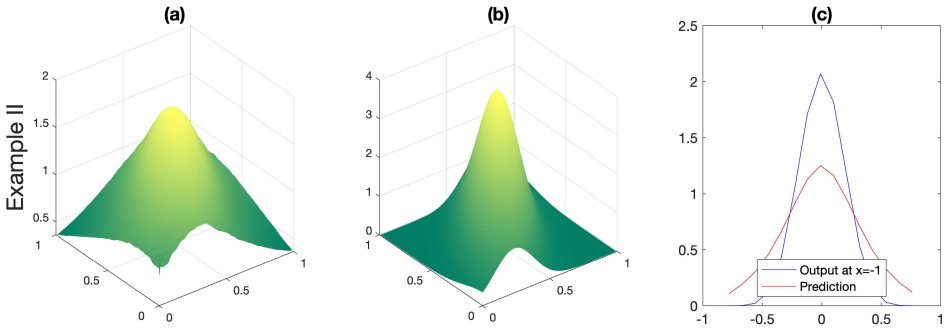


Figure 3.8: A graphical display of the EEI $\bar{P}$ in panel (a), the generating distribution in panel (b), and the output distributions at $x = -1$ of the observed and the predicted regarding Example II in the simple linear model, which implies the EEI provides information of the generating distribution but the EEI is not an element in $\mathbb{P}_{\Lambda,x=-1}$ or $\mathbb{P}_{\Lambda,\mathcal{X}}$.

The EEI actually provides a way of removing the effect of $\boldsymbol{x}$ and simultaneously updates the ansatz information. In particular, it reduces the bias introduced by the assumption of the uniform ansatz in each individual inverse distribution $P_{\Lambda,\boldsymbol{x}}$, and provides distributional information about $\boldsymbol{a}$, e.g. high, low and zero probability occurrence regions. We can also use the EEI to initiate an

approach to finding a GFGD in the equivalence class $P_{\Lambda,\mathcal{X}}$, since the EEI provides information about the generating distribution P_Λ, specifically, about the ansatz along generalized contours; see Section 3.4 for more details.

Similarly, if we extend the DS functions Bel_C and Pl_C by disintegration, i.e. inverse distributions, we have

Theorem 3.3.2. *The SIP posterior belief function is*

$$\widetilde{Bel}_C(A) = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell\in\mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)}} \int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} I_{\{Q_{\boldsymbol{x}}^{-1}(C)\subset A\}}(\boldsymbol{x}) dP_\ell^u(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{y\in C} dP_{\mathcal{D}_{\boldsymbol{x}}}(y) dP_{\mathcal{X}}},$$

and the SIP posterior plausibility function is

$$\widetilde{Pl}_C(A) = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell\in\mathcal{E}_{Q_{\boldsymbol{x}}^{-1}(C)\cap A}} \int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\cap A} I_{\{Q_{\boldsymbol{x}}^{-1}(C)\cap A\neq\emptyset\}}(\boldsymbol{x}) dP_\ell^u(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{y\in C} dP_{\mathcal{D}_{\boldsymbol{x}}}(y) dP_{\mathcal{X}}},$$

where $\widetilde{Pl}_C(A) \leq 1 - \widetilde{Bel}_C(A^c)$, $\widetilde{Pl}_C(A) \geq \widetilde{Bel}_C(A)$ *for any* $A \in \mathcal{B}_\Lambda$, $A^c = \Lambda - A$.

The formulation of the SIP functions is given in Chapter 5.1.4. In fact, the SIP plausibility function is a posterior distribution computed by Bayes' theorem under the SIP framework.

Theorem 3.3.3. *For* $A \in \mathcal{B}_\Lambda$ *and* $C \in \mathcal{B}_\mathcal{D}$, $\widetilde{Pl}_C(A) = P(\boldsymbol{a} \in A | y \in C)$.

The proof is given in Chapter 5.1.5. This result shows that Bel_C and Pl_C actually approximate the posterior of $\boldsymbol{a}$ given Css, and the generalized plausibility $\widetilde{Pl}_C$ is exactly the posterior of $\boldsymbol{a}$ given C. In this case, we only need the generalized plausibility, i.e. the SIP posterior, for inferences in practice. In conclusion, the SIP approach provides a concrete formula of computing the probability distribution of $\boldsymbol{a}$ by incorporating disintegration based on the framework of the DS theory, and thus the SIP approach can be viewed as an extension of the DS probabilities for approximating the probability distribution of unobservable inputs in a model.

3.4 Approximation of GFGDs

3.4.1 Iterative Approach to Updating Inverse Distributions

In section 3.3, the EEI gives information about the generating distribution of $\boldsymbol{a}$, more specifically, the ansatz information of $\boldsymbol{a}$ along generalized contours. However, the EEI is not an element in $\mathbb{P}_{\Lambda,\mathcal{X}}$ in general, and thus the EEI is not a solution to the SIP that induces the output. In this section, we propose an iterative approach to update the EEI through updating the ansatz information in inverse distributions to find a GFGD.

The iterative approach is described in the following. For any $A \in \mathcal{B}_\Lambda$ and $\boldsymbol{x} \in \mathcal{X}$,

$$\text{Disintegration of the EEI: } \bar{P}^i(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} dP^{i+1}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^i}(\ell), \tag{3.9}$$

$$\text{Update inverse distributions: } \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} dP^{i+1}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = P^{i+1}_{\Lambda,\boldsymbol{x}}(A), \tag{3.10}$$

$$\text{Compute the next EEI: } \bar{P}^{i+1}(A) = \int_{\boldsymbol{x} \in \mathcal{X}} P^{i+1}_{\Lambda,\boldsymbol{x}}(A) dP_{\mathcal{X}}, \tag{3.11}$$

where $i \geq 0$. In the initial step $i = 0$, we choose the uniform ansatz $\{P^0_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ as an initial ansatz to compute inverse distributions $\{P^0_{\Lambda,\boldsymbol{x}}\}$ and the EEI $\bar{P}^0$. Essentially, we first disintegrate $\bar{P}^i$ according to $P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^i}$ which is the unique probability measure on the quotient space $\mathcal{L}_{\boldsymbol{x}}$ calculated through

$$P^i_{\mathcal{D}_{\boldsymbol{x}}}(C) = \bar{P}^i(Q^{-1}_{\boldsymbol{x}}(C)), \quad C \in \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}},$$
$$P_{\mathcal{L}_{\boldsymbol{x}}, \bar{P}^i}(\mathcal{E}_A) = P^i_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad A \in \mathcal{B}_\Lambda,$$

to obtain the new ansatz $\{P^{i+1}_\ell\}$. Then we use $\{P^1_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ to update the old ansatz and compute the updated inverse distributions and the EEI.

In the following, we show a general results on the convergence of $\{\bar{P}^i\}_{i \geq 0}$.

Theorem 3.4.1. *The sequence $\{\bar{P}^i\}_{i \geq 0}$ has a subsequence converging weakly.*

Proof. This result is proved in Appendx 5.1.6. □

Furthermore, the sequence has a convergence under mild conditions on the iterated ansatzes. We show this result in terms of densities. Note that each GFGD in $P_{\Lambda,\mathcal{X}}$ is absolutely continuous with respect to μ_Λ and thus has a density, i.e. the Radon-Nikodym derivative. We assume the conditional probability measures $\{P_\ell^i\}_{i\geq 1}$ are absolutely continuous with respect to μ_ℓ for each $\ell \in \mathcal{L}_{\boldsymbol{x}}$, and each density function is denoted by $\rho_\ell^i = dP_\ell^i/\mu_\ell$. In addition, in an experiment indexed by $\boldsymbol{x}$, an element ℓ in $\mathcal{L}_{\boldsymbol{x}}$ can be expressed as $\ell = \pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)$ indexed by $\lambda \in \Lambda$. Then we have

Theorem 3.4.2. *For each $i \geq 0$ and $\boldsymbol{x} \in \mathcal{X}$, the inverse distribution has a density function $\rho^i_{\Lambda|\boldsymbol{x}} = dP^i_{\Lambda,\boldsymbol{x}}/d\mu_\Lambda = \rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda))$ almost everywhere in Λ, where $\rho_{\mathcal{L}_{\boldsymbol{x}}} = dP_{\mathcal{L}_{\boldsymbol{x}}}/d\mu_{\mathcal{L}_{\boldsymbol{x}}}$ is the Radon-Nikodym derivative.*

Proof. The result is proved in Chapter 5.1.7. □

We further denote the density function of $\boldsymbol{X}$ by $\rho_{\boldsymbol{X}} = dP_{\mathcal{X}}/\mu_{\boldsymbol{X}}$ and assume that $\rho^i_{\Lambda|\boldsymbol{x}}(\lambda)\rho_{\boldsymbol{X}}(\boldsymbol{x})$ is a measurable function on the product probability space $(\Lambda \times \mathcal{X}, \mathcal{B}_\Lambda \times \mathcal{B}_{\mathcal{X}})$ for $i \geq 0$.

Theorem 3.4.3. *The EEI has a density function with respect to μ_Λ for $i \geq 0$, denoted by $\bar{\rho}^i = d\bar{P}^i/d\mu_\Lambda$. If for every real number $\epsilon > 0$ there exists an positive integer $N \in \mathbb{N}$ such that for all positive integers $i, j > N$, we have $|\rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) - \rho^j_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)| < \epsilon\rho^0_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)$ almost everywhere in Λ and $\mathcal{X}$ where $\rho^0_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)$ is the density of the initial uniform ansatz with respect to μ_Λ, then $\bar{\rho}^i$ converges to a density function $\bar{\rho}^\infty$ on Λ under the L^1 metric.*

Proof. The result is proved in Chapter 5.1.8. □

This theorem shows that the sequence of probability measures has a convergence if the iterated ansatzes are Cauchy.

Let $\bar{P}^\infty(A) = \int \bar{\rho}^\infty d\mu_\Lambda$ for $A \in \mathcal{B}_\Lambda$. The limiting EEI $\bar{P}^\infty$ is generally not guaranteed to be a GFGD. We show a sufficient and necessary condition for $\bar{P}^\infty \in \mathbb{P}_{\Lambda,\mathcal{X}}$ in this iteration. Let $\mathcal{G}$ denote the map that characterizes the iteration $\bar{P}^{i+1} = \mathcal{G}(\bar{P}^i)$.

Theorem 3.4.4. *A sufficient and necessary condition for $\widetilde{P} \in \mathbb{P}_{\Lambda,\mathcal{X}}$ is stated as follows:*

1. *$\widetilde{P}$ is absolutely continuous with respect to μ_Λ on $(\Lambda, \mathcal{B}_\Lambda)$;*

2. *$\widetilde{P}$ is a fixed point of map $\mathcal{G}$, i.e. $\widetilde{P} = \mathcal{G}(\widetilde{P})$;*

3. *The inverse distribution $\widetilde{P}_{\Lambda,\boldsymbol{x}}$ computed in the iterative approach in (3.9) and (3.10) as*

$$\textit{Disintegration of a measure: } \widetilde{P}(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} d\widetilde{P}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}, \widetilde{P}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda,$$

$$\textit{Compute inverse distributions: } \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} d\widetilde{P}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = \widetilde{P}_{\Lambda,\boldsymbol{x}}(A), \quad \forall A \in \mathcal{B}_\Lambda,$$

   is not conditioned on $\boldsymbol{x}$, i.e. $\widetilde{P}_{\Lambda,\boldsymbol{x}} = \widetilde{P}_{\Lambda,\boldsymbol{x}'}$ if $\boldsymbol{x} \neq \boldsymbol{x}'$.

4. *$\widetilde{P}$ is absolutely continuous with respect to μ_ℓ for $\ell \in \mathcal{L}_{\boldsymbol{x}}$ and $\boldsymbol{x} \in \mathcal{X}$.*

The proof is given in Chapter 3.4.4. This result shows that, for $\bar{P}^\infty$ to be a GFGD in $\mathbb{P}_{\Lambda,\mathcal{X}}$, the limiting EEI $\bar{P}^\infty$ should be a fixed point of the iteration and the limiting inverse distributions for individual $\boldsymbol{x}$ should be the same.

In the next section, we show the intermediate steps of this iteration approach with a detailed example.

3.4.2 Intermediate Results by Monte Carlo Sampling

In the previous section, we propose an approach to approximate GFGDs via iteratively updating the inverse distributions and the EEIs. In this section, Monte Carlo is implemented to estimate the EEI in each iteration. We display some intermediate results in (3.1) with the generating distribution of $\boldsymbol{a}$ from Example II. All the results in this section are shown in Figure 3.9.

In (3.1), x is the slope of a generalized contour $\{\boldsymbol{a} \in \Lambda : a_2 = y - a_1 x\}$ given y. In this case, the slope can be written as $x = \tan(\theta)$ where θ is the angle between the contour and the horizontal axis. Suppose there is a collection of i.i.d. random samples $\{\Theta_j\}_{j=1}^n$ drawn from $((-\pi/2, \pi/2), \mathcal{B}_{(-\pi/2,\pi/2)}, P_\Theta)$ where $\mathcal{B}_{(-\pi/2,\pi/2)}$ is the Borel sigma algebra of $(-\pi/2, \pi/2)$ and P_Θ is the uniform distribution on $(-\pi/2, \pi/2)$. Clearly, the distribution P_Θ induces a probability

distribution of samples $\{X_j\}_{j=1}^n$ on $(\mathcal{X}, \mathcal{B}_\mathcal{X}, P_X)$. In the following, we use $\{X_j\}_{j=1}^n$ and P_X in the iterative approach for illustration.

Recall that in the initial step we compute a collection of inverse distributions $\{P^0_{\Lambda,X_j}\}_{j=1}^n$ using the initial uniform ansatz. Then, we compute the EEI, now using Monte Carlo,

$$\widehat{P}^0(A) = \frac{1}{n}\sum_{j=1}^{n} P^0_{\Lambda,X_j}(A).$$

Panels (a) and (b) are the results of two inverse distributions $P^0_{\Lambda,X_1=-1.66}$ and $P^0_{\Lambda,X_2=0.56}$ using the initial uniform ansatz, respectively, and panel (c) is the result of the computed EEI $\widehat{P}^0$. It can be seen that the distributions along the generalized contours are uniform in $P^0_{\Lambda,X_1=-1.66}$ and $P^0_{\Lambda,X_2=0.56}$, and the inverse distributions using the uniform ansatz strongly depend on different samples of X. The EEI $\widehat{P}^0$ is closer to the generating distribution P_Λ (i.e. a truncated normal distribution), compared to $P^0_{\Lambda,X_1=-1.66}$ and $P^0_{\Lambda,X_2=0.56}$. The step of computing the EEI removes the dependence from $\boldsymbol{X}$. In addition, the EEI computation reduces the effect of the uniform ansatz on individual inverse distributions in some sense, which is illustrated later in the two inverse distributions in panels (d) and (e).

Next, for each X_j ($j = 1, \ldots, n$), we compute the distributions $\{P^1_\ell\}$ along the generalized contours disintegrated from $\widehat{P}^0$ as

$$\widehat{P}^0(A) = \int_{\ell\in\pi_{\mathcal{L}_{X_j}}(A)} \int_{\lambda\in\pi^{-1}_{\mathcal{L}_{X_j}}(\ell)\cap A} dP^1_\ell(\lambda) dP_{\mathcal{L}_{X_j},\widehat{P}^0}(\ell), \quad \forall A \in \mathcal{B}_\Lambda,$$

where $\mathcal{L}_{X_j}$ is the quotient space indexed by X_j. The distributions $\{P^1_\ell\}$ are used to define the ansatz for the first iteration. The inverse distributions using the updated ansatz $\{P^1_\ell\}$ are in the form of

$$P^1_{\Lambda,X_j}(A) = \int_{\ell\in\pi_{\mathcal{L}_{X_j}}(A)} \int_{\lambda\in\pi^{-1}_{\mathcal{L}_{X_j}}(\ell)\cap A} dP^1_\ell(\lambda) dP_{\mathcal{L}_{X_j}}(\ell),$$

for $j = 1, \ldots, n$, and we have the first updated EEI as

$$\widehat{P}^1(A) = \frac{1}{n}\sum_{j=1}^{n} P^1_{\Lambda,X_j}(A), \quad \forall A \in \mathcal{B}_\Lambda.$$

Panels (d) and (e) show the updated inverse distributions $P^1_{\Lambda|X_1=-1.66}$ and $P^1_{\Lambda|X_2=0.56}$ using the updated ansatz from the last EEI $\widehat{P}^0$, respectively. The step of computing the EEI reduces the effect of the uniform ansatz on inverse distributions $P^0_{\Lambda|X_1=-1.66}$ and $P^0_{\Lambda|X_2=0.56}$ since the distributions along the generalized contours in panels (d) and (e) are more curved towards truncated normal distributions. Furthermore, the updated EEI $\widehat{P}^1$ in panel (f) is closer to the generating distribution P_Λ than $\widehat{P}^0$. Note that in general we do not expect the updated EEI to become closer to the generating distribution.

Likewise, for $i \geq 0, j = 1 \ldots, n$, and any $A \in \mathcal{B}_\Lambda$,

$$\text{Disintegration of the EEI: } \widehat{P}^i(A) = \int_{\ell \in \pi_{\mathcal{L}_{X_j}}(A)} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{X_j}}(\ell) \cap A} dP^{i+1}_\ell(\lambda) dP_{\mathcal{L}_{X_j},\widehat{P}^i}(\ell),$$

$$\text{Update inverse distributions: } \int_{\ell \in \pi_{\mathcal{L}_{X_j}}(A)} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{X_j}}(\ell) \cap A} dP^{i+1}_\ell(\lambda) dP_{\mathcal{L}_{X_j}}(\ell) = P^{i+1}_{\Lambda,X_j}(A),$$

$$\text{Compute the next EEI: } \widehat{P}^{i+1}(A) = \frac{1}{n}\sum_{j=1}^{n} P^{i+1}_{\Lambda,X_j}(A).$$

Panels (g), (h) and (i) show the inverse distributions $P^2_{\Lambda,X_1=-1.66}$, $P^2_{\Lambda,X_2=0.56}$ and the EEI $\widehat{P}^2$ after 2 iterations. Panels (j), (k) and (d) show the inverse distributions $P^{10}_{\Lambda,X_1=-1.66}$, $P^{10}_{\Lambda,X_2=0.56}$ and the EEI $\widehat{P}^{10}$ after 10 iterations. It can be seen that the ansatz is updated towards the distributions along the generalized contours disitegrated according to the generating distribution after more iterations. In Iteration 10, the fact that the two inverse distributions are almost the same as their EEI suggests the convergence of the ansatz and the EEI iteration.

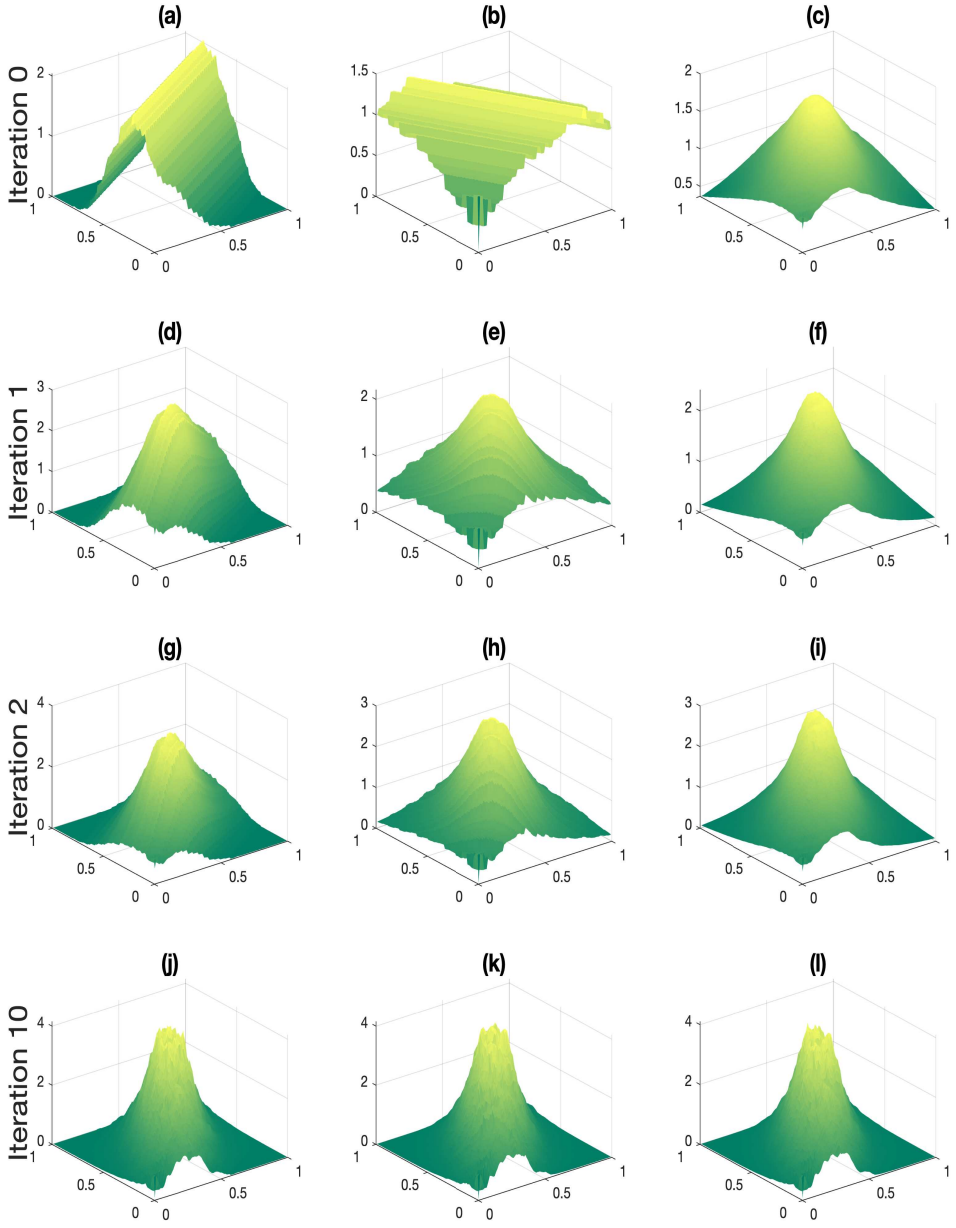


Figure 3.9: The intermediate results in the iterations of the EEI and individual inverse distributions in Example II of the simple linear model. Panels in each row display the results for two inverse distributions $P^i_{\Lambda,X_1=-1.66}$ and $P^i_{\Lambda,X_2=0.56}$, and the EEI $\bar{P}^i$ after $i = 0, 1, 2, 10$ iterations. In each row, $P^i_{\Lambda,X_1=-1.66}$ and $P^i_{\Lambda,X_2=0.56}$ are shown in the first and second columns, respectively, and $\bar{P}^i$ is shown in the last column.

3.5 Examples of Linear and Nonlinear Models

The following linear model is considered

$$\text{Model I}: \qquad y = Q_{\boldsymbol{x}}(\boldsymbol{a}) := a_1 x + a_2. \tag{3.12}$$

This simple linear model is the same as (3.1) stated in Section 3.1.1, and we continue to explore the converging results on Example I, II and III where $\boldsymbol{a} = [a_1, a_2]^\top \in \Lambda = [0,1]^2$ is distributed, in the data generating process, according to a uniform distribution, a truncated normal distribution, and a distribution characterized by an Archimedean copula, respectively. Let n be the number of generated samples of $\boldsymbol{a}$, K be the number of generated samples of x, N^2 be the number of grid cells that partition the two dimensional domain Λ, N_0 be the number of bins that partition the one dimensional range of y, and S be the steps of iteration.

The density functions of the resulting EEIs after S iterations for Example I, II and III are shown in Figure 3.10. The density functions of the predictions of the output distributions of y given two different values of x are also displayed. The predictions are computed through the forward computation (3.8) by using the EEIs after S iterations. The results on Example I are displayed in panels (a)-(c) where panel (a) shows the EEI computed under the setting $n = 10000, K = 1000, N^2 = 100^2, N_0 = 50, S = 0$, and panels (b)-(c) show the predictions of distributions of $y|x = \pm 10$. The predictions are marked in red, and the observed distributions of $y|x = \pm 10$, i.e. the distributions induced by the data generating distribution (uniform on Λ), are marked in blue. An interesting result is that the EEI without any iteration is exactly the data generating distribution since the uniform ansatz is the correct ansatz information. However, the EEI of Example II took $S = 40$ iterations to converge under the setting $n = 15000, K = 1000, N^2 = 100^2, N_0 = 50$, see panel (d). Panels (e) and (f) depict the predictions of $y|x = \pm 10$ and the observed distributions. The resulting EEI of Example III is shown in panel (g) computed under the setting $n = 10000, K = 1000, N^2 = 50^2, N_0 = 25, S = 30$. The predictions of $y|x = \pm 2$ are compared with the observed distributions in panels (h) and (i). It can be seen that all the resulting EEIs in Example I, II and III

are close to the corresponding generating distributions, and all predictions are almost the same to the observed distributions. This suggests the EEIs converge and their limits are contained in $\mathbb{P}_{\Lambda,\mathcal{X}}$. One of the sources of the approximation error in the predictions is the error of approximating each probability distribution in the algorithm by an empirical distribution of the data. Thus, better results on the predictions can be obtained by increasing n.

Then, we consider a nonlinear model that consists of the following ordinary differential equations

$$\text{Model II}: \qquad \begin{aligned} \frac{dS}{dt} &= -\beta S\frac{I}{N_e} + \gamma I, \\ \frac{dI}{dt} &= \beta S\frac{I}{N_e} - \gamma I. \end{aligned}$$

This model is referred to as the Susceptible-Infectious-Susceptible (SIS) model in the field of the epidemiology of infectious diseases, e.g. the common cold and influenza. In this model, $S(t) + I(t) = N_e$ is a constant at any time $t \geq 0$ where $S(t)$ is the number of the susceptible at time t and $I(t)$ is the number of the infectious at time t given the total size of a population N_e. The transmission variable is $\boldsymbol{a} = [\beta, \gamma]^\top$ where β is the average number of contacts that are sufficient for disease transmission of a person per unit time and γ is the fraction of infected individuals who recover and return to the susceptible class per unit time. Note that there is no entry or departure including deaths in this population.

This model is equivalent to the following equation

$$\frac{dI}{dt} = (\beta - \gamma)I - \beta I^2/N_e.$$

Given $I(0) = I_0 > 0$, the solution is a logistic function

$$I(t) = \frac{(1 - \gamma/\beta)N_e}{1 + ((1 - \gamma/\beta)N_e/I_0 - 1)e^{-(\beta-\gamma)t}}, \tag{3.13}$$

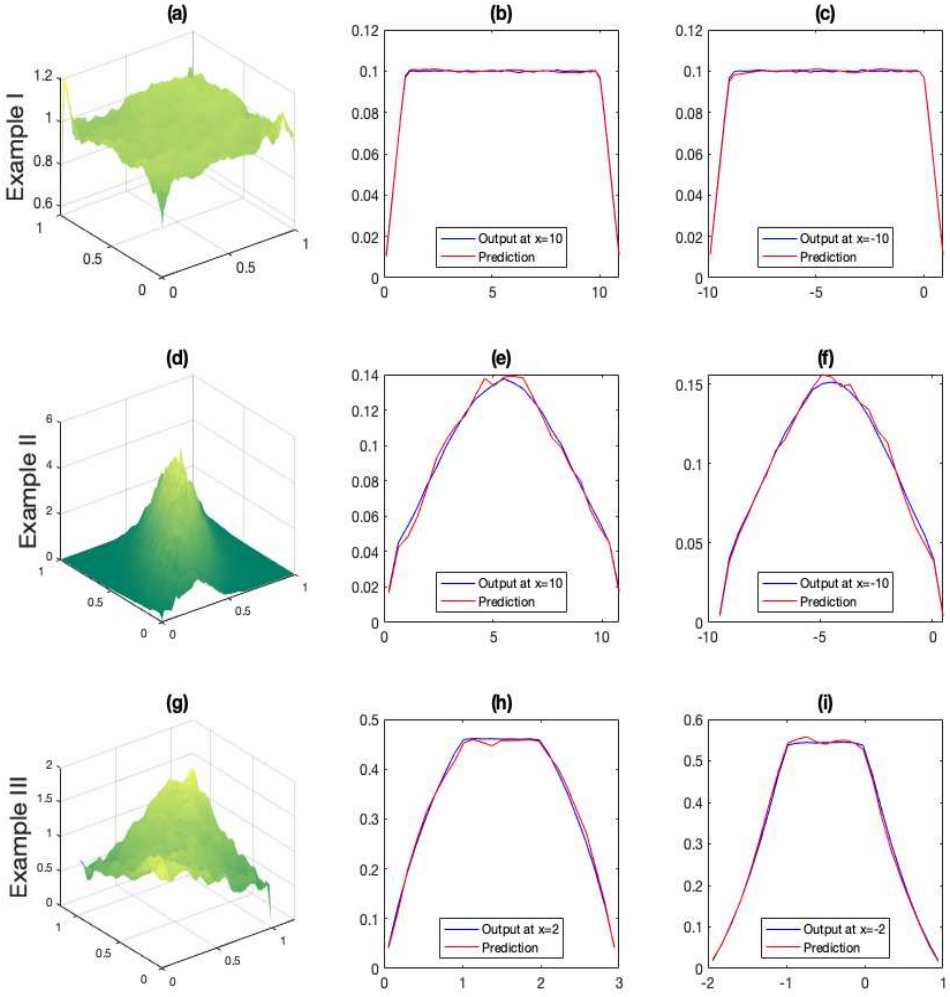


Figure 3.10: The resulting EEIs and predictions of distributions of $y|x$ in Example I, II and III in Model I. Panel (a) shows the resulting EEI of $\boldsymbol{a}$ in Example I. In panels (b) and (c), the predictions of distributions of $y|x = \pm 10$ computed by using the EEI in panel (a) are depicted in red color, and the observed distributions induced by the generating distribution are marked in blue. The same results on Example II are displayed in panels (d)-(f) when $x = \pm 10$, and the same results on Example III are shown in panels (g)-(i) when $x = \pm 2$.

when $\beta \neq \gamma$, and it is

$$I(t) = \frac{1}{1/I_0 + \beta t/N_e}, \tag{3.14}$$

when $\beta = \gamma$. Note that the function $I(t)$ is continuous with respect to β, γ since the limit of (3.13) as $\beta \to \gamma$ is the exact (3.14). Then, the solution of this model can be written as

$$I(\boldsymbol{a}; t) = \frac{(1 - \gamma/\beta)N_e}{1 + ((1 - \gamma/\beta)N_e/I_0 - 1)e^{-(\beta-\gamma)t}}, \tag{3.15}$$

for $\beta, \gamma > 0$. In this case, the SIP can be set up as $y = Q_t(\boldsymbol{a}) = I(\boldsymbol{a}; t)$ where $\boldsymbol{a}$ is random. In this model, we let $N_e = 1000, I_0 = 1$, and we investigate Example I, II and III in Section 3.1.1 with $\Lambda = [0.01, 1]^2$. Samples of t are randomly drawn according to the uniform distribution on $t \in [0, 1]$. It is important to note that the solution (3.15) is approximately a constant when $t > 1$, and thus it provides little information to the SIP. This is not true in the simple linear model in which each choice of x is equally likely important. In addition, due to the smoothness of the solution (3.15) with respect to t, we can predict distributions of the output y for $t > 1$ by using the resulting EEI computed under $t \in [0, 1]$.

Figure 3.11 shows the resulting EEIs after S steps of iteration and two predictions of the output distribution given two values $t = 0.5$ and $t = 2$. Panels (a)-(c) show the results on Example I under the setting $n = 10000, K = 1000, N^2 = 100^2, N_0 = 25, S = 1$. Similar to the results in the simple linear model, the result after one step of iteration and the result after $S = 20$ iterations have no noticeable difference, since the uniform ansatz is the true ansatz along the generalized contours. The results on Example II under the setting $n = 15000, K = 1000, N^2 = 100^2, N_0 = 50, S = 5$ are shown in panels (d)-(f). Panel (d) shows the resulting EEI after S=5 iterations, which is clearly different from a truncated normal distribution. The predictions of the output in panels (e)-(f) suggests that the resulting EEI is a GFGD. Similar results on Example III are shown on the last row of Figure 3.11 under the setting $n = 10000, K = 1000, N^2 = 50^2, N_0 = 25, S = 4$.

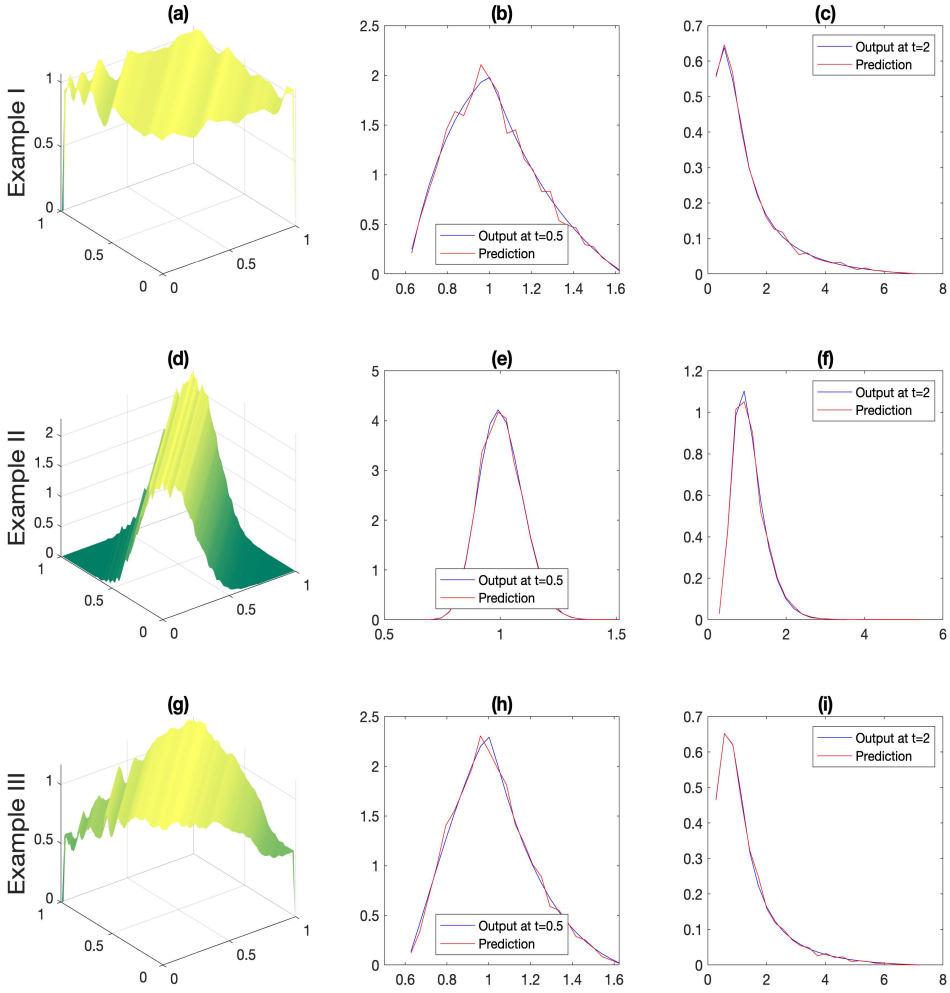


Figure 3.11: The resulting EEIs and predictions of the distributions of $y|t = 0.5$ and $y|t = 2$ in Example I, II and III in Model II on $\Lambda = [0.01, 1]^2$. Panel (a) shows the resulting EEI of $\boldsymbol{a}$ in Example I. In panels (b) and (c), the predictions of distributions of $y|t$ computed by using the EEI in panel (a) are depicted in red color, and the observed distributions induced by the generating distribution are marked in blue. The same results on Example II and Example III are displayed in panels (d)-(f) and panels (g)-(i), respectively.

Chapter 4

Domain Recovery for Stochastic Inverse Problems

4.1 Stochastic Inverse Problems

4.1.1 A Simple Linear Example in Different Domains

Consider a simple linear model

$$y = a_1 x + a_2, \tag{4.1}$$

where $\boldsymbol{a} = (a_1, a_2)^\top$ is a vector of unobservable random inputs on Λ, a compact set in $\mathbb{R}^2$, x is an observable deterministic input in $\mathbb{R}^1$, and y is the scalar output in the solution space $\mathcal{Y}$ in $\mathbb{R}^1$. We assume that $\boldsymbol{a}$ has a probability distribution P_Λ on $(\Lambda, \mathcal{B}_\Lambda)$ where $\mathcal{B}_\Lambda$ is the Borel sigma algebra and P_Λ is absolutely continuous with respect to the Lebesgue measure μ_Λ on $(\Lambda, \mathcal{B}_\Lambda)$. In this paper, P_Λ is referred to as the *generating (probability) distribution*. The set Λ is referred to as the *domain* of $\boldsymbol{a}$, which is generally pre-specified by scientists and contains the support of the generating distribution. The support of a probability distribution is defined topologically as the set of all points λ in Λ for which every open neighborhood N_λ of λ has positive probability measure:

$$supp(P_\Lambda) = \{\lambda \in \Lambda : P_\Lambda(N_\lambda) > 0 \text{ for every open neighborhood } N_\lambda \in \mathcal{B}_\Lambda \text{ of } \lambda\}.$$

Note that this support is compact in Λ and can be used to the define the support of the Radon-Nikodym density of P_Λ with respect to μ_Λ. Given the generating distribution of $\boldsymbol{a}$, we can obtain a probability distribution of the output y for a given $\boldsymbol{x}$ through (4.1). In this case, the model (4.1) characterizes a data-generating process, which is a simple example of *stochastic forward problem* for a random input (SFP; Butler et al. (2014, 2015)). For illustration purpose, we consider the following two examples:

(I) $\boldsymbol{a}$ has a uniform distribution supported on the unit square $\Lambda = [0,1] \times [0,1] = [0,1]^2$;

(II) $\boldsymbol{a}$ has a uniform distribution supported on a disk, with center at $(0.5, 0.5)$ and radius 0.2.

The density functions of $\boldsymbol{a}$ in Example I and II are depicted in panels (a)-(b) of Figure 4.1, respectively. For different values of x, the generating distribution of $\boldsymbol{a}$ induces different probability

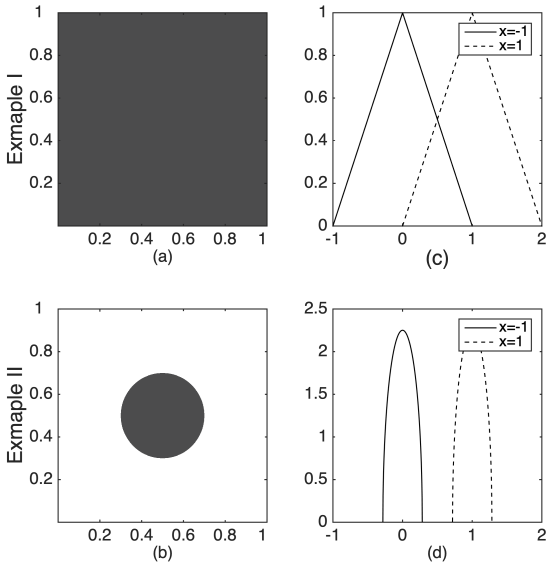


Figure 4.1: A graphical display of the density functions of $\boldsymbol{a}$ (left) and induced probability density functions (right) resulting from the stochastic forward problem examples (I) and (II). In panel (c)-(d), two induced probability density functions for $x = -1$ (solid) and $x = 1$ (dashed) are shown for each example. Panel (a) shows that $\boldsymbol{a}$ has a uniform distribution on the unit square and panel (c) shows two different induced density functions of y according to the uniform distribution of $\boldsymbol{a}$ for two values of x. Panel (b) shows $\boldsymbol{a}$ has a uniform distribution supported on a disk and its corresponding induced density functions of y are shown in panel (d).

distributions of y, which is referred to as the *output (probability) distribution*. For instance, in panels (c)-(d), the output probability density functions for $x = -1$ and $x = 1$ are depicted for each example. While clearly different, the two output distributions in each panel share similarities such as the shape-symmetry and the convexity of the support of the generating distribution of $\boldsymbol{a}$. Note that the output probability distributions are different between panels (c) and (d) because the support of $\boldsymbol{a}$ in its domain in Example II is different from the support $[0, 1]^2$ in Example I. In addition, the supports of the output distributions in each panel are uniquely determined by the support of $\boldsymbol{a}$. This characterizes an interesting case in practice. The set Λ represents the domain of all

physically meaningful input values. However, often the values in a given situation are restricted to a proper subset of Λ. For example, in an electron diffraction problem, Λ might consist of the set of all apertures with Lipschitz continuous boundaries, while an experiment might limit to the set of apertures that are ellipses. Thus, the observations of light are limited to the elliptical apertures.

The *stochastic inverse problem* (SIP; Butler et al. (2014, 2015)), as the inverse of the SFP, is the problem of recovering a distribution of $\boldsymbol{a}$ given a probability distribution of the output y and a value of x. A solution to the SIP is referred to as an *inverse distribution* of $\boldsymbol{a}$, and it should induce the output distribution of y for a given x in the SFP process. In general, there are multiple solutions to an SIP since the map from the model inputs to the model output is not 1-1. In the previous work of the SIP, their inverse distributions are computed on a pre-specified domain Λ and the support issue of inverse distributions is not addressed. However, figuring out a proper domain of $\boldsymbol{a}$ is crucial to characterizing the support of an inverse distribution, and more importantly, an improper domain might result in a wrong solution. To illustrate, we consider the two examples depicted in panels (a)-(b) of Figure 4.1. In Figure 4.2, panels (a) and (c) show two functions of $\boldsymbol{a}$ recovered from Example I on two different domains given the output distribution for $x = 1$. The function in panel (a) is recovered on the domain $[0, 1]^2$ of $\boldsymbol{a}$ that is exactly the support of $\boldsymbol{a}$, and thus it is a density function of a solution to the SIP. However, the function in panel (c) is recovered on an improper domain, i.e. a disk with center at $(0.5, 0.5)$ and radius 0.2. In this case, the recovered function is not a density of $\boldsymbol{a}$ or a solution to the SIP, since its support restricted to this domain is improper. In later sections, we show feasible supports for computing inverse distributions. Panels (b) and (d) show the density functions of inverse distribution recovered on $\Lambda = [0, 1]^2$ and a disk with center at $(0.5, 0.5)$ and radius 0.4, respectively, which contain the actual support of $\boldsymbol{a}$ in Example II. Both of the density functions are solutions to this SIP. Furthermore, as the domain grows, the distance (e.g. the L^1 distance) of the corresponding inverse distribution and the generating distribution of $\boldsymbol{a}$ increases. Such "bias" is caused by the choice of domain, and the goal in this chapter is to find a good domain of the unobservable inputs, by determining the support of the generating distribution,

so that the "bias" can be reduced. In addition, these examples also demonstrate the non-uniqueness of the SIP solution from the viewpoint of domains.

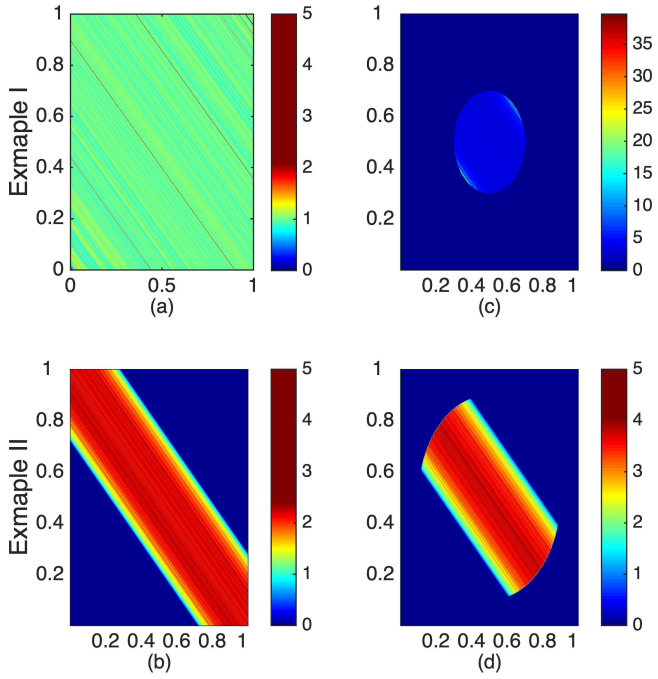


Figure 4.2: Plots of two inverse distributions computed on two different input domains from a distribution of $y|x = 1$ in Examples I and II. In Example I, panel (a) shows the first inverse distribution recovered on the exact domain of $\boldsymbol{a}$ which is equal to $\Lambda = [0, 1]^2$, and panel (c) shows a function recovered on a disk with center at $(0.5, 0.5)$ and radius 0.2. The function in panel (c) is not a valid density since its domain is improper. In Example II, panel (b) shows an inverse distribution recovered on $\Lambda = [0, 1]^2$ and panel (d) shows an inverse distribution recovered on a disk with center at $(0.5, 0.5)$ and radius 0.4.

In the next section, we introduce the mathematical background of the SIP and its theoretical solution with emphasis on the domain.

4.1.2 Solution of the SIP

Let Q be the measurable map induced by a mathematical model characterizing a physical system, $y = Q(\boldsymbol{x}, \boldsymbol{a})$. In the SFP process, we observe the behavior of the output y of the model for a given observable input $\boldsymbol{x}$ as unobservable input $\boldsymbol{a}$ is varied over its domain. In this case, the model for a given $\boldsymbol{x}$ is denoted by the measurable map $Q_{\boldsymbol{x}}(\cdot) \equiv Q(\boldsymbol{x}, \cdot)$ indexed by $\boldsymbol{x}$. We consider

a compact metric space Λ as the domain of $\boldsymbol{a}$, and an absolutely continuous generating distribution P_Λ of $\boldsymbol{a}$ with respect to the "volume" (Lebesgue) measure μ_Λ on $(\Lambda, \mathcal{B}_\Lambda)$. The support $\mathcal{K}_0$ of P_Λ is contained in Λ. The set of values of $\boldsymbol{x}$ is denoted by $\mathcal{X}$ in which each $Q_{\boldsymbol{x}}$ is well-defined. For a specific $\boldsymbol{x}$, if we denote the range of y by $\mathcal{D}_{\boldsymbol{x}} = Q_{\boldsymbol{x}}(\Lambda)$, then $Q_{\boldsymbol{x}}$ is a measurable map from Λ to $\mathcal{D}_{\boldsymbol{x}}$. In this paper, we assume that $Q_{\boldsymbol{x}}$ is differentiable with respect to $\boldsymbol{a}$, and the Jacobian of $Q_{\boldsymbol{x}}$ has full rank at every point in Λ. Then, the generating distribution P_Λ on the domain Λ induces a probability measure $P_{\mathcal{D}_{\boldsymbol{x}}}$ on the compact space $\mathcal{D}_{\boldsymbol{x}}$ through the following *forward computation*

$$P_{\mathcal{D}_{\boldsymbol{x}}}(C) = P_\Lambda(Q_{\boldsymbol{x}}^{-1}(C)), \tag{4.2}$$

for any event C in the Borel sigma algebra $\mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}}$ of $\mathcal{D}_{\boldsymbol{x}}$. Since the map $Q_{\boldsymbol{x}}$ is not 1-1 in general, the inverse map $Q_{\boldsymbol{x}}^{-1}$ is defined as

$$Q_{\boldsymbol{x}}^{-1}(C) = \{\boldsymbol{a} \in \Lambda : Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C\},$$

where $Q_{\boldsymbol{x}}^{-1}$ maps to a subset in $\mathcal{B}_\Lambda$. Note that $P_{\mathcal{D}_{\boldsymbol{x}}}$ is absolutely continuous with respect to the following dominating measure

$$\mu_{\mathcal{D}_{\boldsymbol{x}}}(C) = \mu_\Lambda(Q_{\boldsymbol{x}}^{-1}(C)), \quad \forall C \in \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}},$$

which is induced by μ_Λ through the same forward computation. Similarly, the support of $P_{\mathcal{D}_{\boldsymbol{x}}}$ is also propagated through (4.2). Suppose the map $Q_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \mathcal{X}$ is a closed map in the sense that it maps closed subsets of Λ to closed subsets of $\mathcal{D}_{\boldsymbol{x}}$, and $Q_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \mathcal{X}$ is an open map in the sense that it map open subsets of Λ to open subsets of $\mathcal{D}_{\boldsymbol{x}}$.

Theorem 4.1.1. *The support of $P_{\mathcal{D}_{\boldsymbol{x}}}$ is $Q_{\boldsymbol{x}}(\mathcal{K}_0)$.*

Proof. This result is proved in Chapter 5.2.1. □

The process of computing $P_{\mathcal{D}_{\boldsymbol{x}}}$ in (4.2) given P_Λ characterizes an SFP, and an SIP is to find the probability measure P_Λ from the given $P_{\mathcal{D}_{\boldsymbol{x}}}$. In solving an SIP, the decomposition of Λ and

the disintegration of P_Λ are involved. We first decompose the domain Λ based on the equivalence relation defined by $Q_{\boldsymbol{x}}^{-1}$. Let λ_1, λ_2 be two distinct points in Λ. For a specific $\boldsymbol{x}$, λ_1 and λ_2 are equivalent relative to $\boldsymbol{x}$, denoted by $\lambda_1 \sim \lambda_2$, in the sense that $Q_{\boldsymbol{x}}(\lambda_1) = Q_{\boldsymbol{x}}(\lambda_2)$. In other words, $Q_{\boldsymbol{x}}^{-1}(y)$ where $y \in \mathcal{D}_{\boldsymbol{x}}$ represents an equivalence class of points in Λ indexed by y, that is, $Q_{\boldsymbol{x}}(\lambda) = y$ for any $\lambda \in Q_{\boldsymbol{x}}^{-1}(y)$. The set $Q_{\boldsymbol{x}}^{-1}(y)$ is called a *generalized contour* (Butler et al., 2014, 2015), since it represents a locally smooth manifold in Λ. In addition, for $y_1, y_2 \in \mathcal{D}_{\boldsymbol{x}}$ and $y_1 \neq y_2$, we have $Q_{\boldsymbol{x}}^{-1}(y_1) \cap Q_{\boldsymbol{x}}^{-1}(y_2) = \emptyset$, by the assumption that $Q_{\boldsymbol{x}}$ has full rank at every point in Λ. Then, the domain Λ can be decomposed into a union of equivalence classes

$$\Lambda = \bigcup_{y \in \mathcal{D}_{\boldsymbol{x}}} Q_{\boldsymbol{x}}^{-1}(y).$$

We denote the space of generalized contours in Λ by $\mathcal{L}_{\boldsymbol{x}}$, and each point in $\mathcal{L}_{\boldsymbol{x}}$ corresponds to a generalized contour in Λ. Specifically, for a point $\lambda \in \Lambda$, the generalized contour $Q_{\boldsymbol{x}}^{-1}(Q_{\boldsymbol{x}}(\lambda))$ corresponds to a point in $\mathcal{L}_{\boldsymbol{x}}$, denoted by $\mathcal{E}_\lambda$. Then, we further define a measurable map from Λ to $\mathcal{L}_{\boldsymbol{x}}$ as $\pi_{\mathcal{L}_{\boldsymbol{x}}} : \Lambda \to \mathcal{L}_{\boldsymbol{x}}$ such that $\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda) = \mathcal{E}_\lambda$ for any $\lambda \in \Lambda$. In this case, the inverse of $Q_{\boldsymbol{x}}$ defines a one-to-one and onto map between $\mathcal{L}_{\boldsymbol{x}}$ and $\mathcal{D}_{\boldsymbol{x}}$, and we obtain the probability measure on $\mathcal{L}_{\boldsymbol{x}}$ induced by the $P_{\mathcal{D}_{\boldsymbol{x}}}$ through

$$P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = P_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad \forall A \in \mathcal{B}_\Lambda,$$

where $\mathcal{E}_A = \pi_{\mathcal{L}_{\boldsymbol{x}}}(A) = \{\mathcal{E}_\lambda : \lambda \in A\}$ denotes the event in the Borel sigma algebra $\mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}}$ of $\mathcal{L}_{\boldsymbol{x}}$. Similarly, we can obtain a dominating measure $\mu_{\mathcal{L}_{\boldsymbol{x}}}$ of $P_{\mathcal{L}_{\boldsymbol{x}}}$ by

$$\mu_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = \mu_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad \forall A \in \mathcal{B}_\Lambda.$$

In addition, the support of $P_{\mathcal{L}_{\boldsymbol{x}}}$ can also be uniquely determined by the support of P_Λ through the map $\pi_{\mathcal{L}_{\boldsymbol{x}}}$ under proper assumptions that are consistent with the map $Q_{\boldsymbol{x}}$. Supposing $\pi_{\mathcal{L}_{\boldsymbol{x}}}$ is a closed and open map from Λ to $\mathcal{L}_{\boldsymbol{x}}$ and $\pi_{\mathcal{L}_{\boldsymbol{x}}}$ is continuous on Λ, we have

Corollary 4.1.1.1. *The support of $P_{\mathcal{L}_{\boldsymbol{x}}}$ is $\pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0)$.*

This result is a direct consequence of Theorem 4.1.1. In Butler et al. (2014), a manifold in Λ indexing generalized contours is used to represent $\mathcal{L}_{\boldsymbol{x}}$, which is called a *transverse parameterization*. In the following, we use $\mathcal{L}_{\boldsymbol{x}}$ to denote any such transverse parameterization in Λ. With a proper transverse parameterization, the assumptions that $\pi_{\mathcal{L}_{\boldsymbol{x}}}$ is both closed and continuous can be satisfied. In this case, the probability measure $P_{\mathcal{L}_{\boldsymbol{x}}}$ on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}})$ is uniquely determined by $P_{\mathcal{D}_{\boldsymbol{x}}}$ and thus by P_Λ, and any Borel set of generalized contours is distributed according to $P_{\mathcal{L}_{\boldsymbol{x}}}$. Then, the remaining problem is to describe how $\boldsymbol{a}$ is distributed along each generalized contour. Since $Q_{\boldsymbol{x}}$ is not 1-1 and yields multiple inverse probability measures on $(\Lambda, \mathcal{B}_\Lambda)$, the disintegration theorem is used to extend the solution on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}})$ to $(\Lambda, \mathcal{B}_\Lambda)$; see Butler et al. (2014). For any $\boldsymbol{x}$, the generating distribution P_Λ on Λ can be disintegrated as

$$P_\Lambda(A) = \int_{\mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda, \tag{4.3}$$

where $P_{\mathcal{L}_{\boldsymbol{x}}}$ is the probability measure on $(\mathcal{L}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}})$ determined by $P_{\mathcal{D}_{\boldsymbol{x}}}$ and $\{P_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ is a family of conditional probability measures along generalized contours $\{\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ in Λ. Given P_Λ, $P_{\mathcal{L}_{\boldsymbol{x}}}$ can be determined by $P_{\mathcal{D}_{\boldsymbol{x}}}$ through (4.2), and then $\{P_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ can be (uniquely) determined through (4.3) $\mu_{\mathcal{L}_{\boldsymbol{x}}}$-almost everywhere; see Chang and Pollard (1997).

The Disintegration Theorem (4.3) provides a way to obtain P_Λ given $\{P_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ and $P_{\mathcal{L}_{\boldsymbol{x}}}$. In the SIP, the probability measure $P_{\mathcal{D}_{\boldsymbol{x}}}$ is given for a specific $\boldsymbol{x}$, and thus the probability measure $P_{\mathcal{L}_{\boldsymbol{x}}}$ can be obtained by $P_{\mathcal{D}_{\boldsymbol{x}}}$. However, we have no information about $\{P_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ in general. One possible approach, suggested by Butler et al. (2014), is to assume a specified measure along each generalized contour, which is referred to as an *ansatz*. Specifically, the uniform ansatz is suggested by those authors, which is defined as, for $\ell \in \mathcal{L}_{\boldsymbol{x}}$,

$$P^u_\ell(\cdot) = \frac{1}{\mu_\ell(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell))} \mu_\ell(\cdot),$$

where μ_ℓ is an underlying measure along the generalized contour and it is computed by the disintegration of μ_Λ as

$$\mu_\Lambda(A) = \int_{\mathcal{E}_A} \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A} d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda.$$

By plugging in the uniform ansatz, we obtain an inverse distribution of $\boldsymbol{a}$, denoted by $P_{\Lambda,\boldsymbol{x}}$. In fact, different choices of $\{P_\ell\}_{\ell \in \mathcal{L}_{\boldsymbol{x}}}$ yield different inverse distributions of P_Λ, and any inverse distribution is considered as a solution of the SIP in the sense that it induces the same $P_{\mathcal{D}_{\boldsymbol{x}}}$ through the forward computation (4.2) as P_Λ does. Among many choices of the ansatz, the uniform ansatz is used since it is considered as a "non-preferential" choice determined by μ_Λ and the resulting inverse distribution $P_{\Lambda,\boldsymbol{x}}$ has the maximum entropy.

Depending on different choices of the ansatz, inverse distributions might have different supports. We first show that the support of the inverse distribution under any ansatz is determined by $Q_{\boldsymbol{x}}(\mathcal{K}_0)$.

Corollary 4.1.1.2. *Let S denote the support of an inverse distribution under certain ansatz. Then, we have $Q_{\boldsymbol{x}}(C) = Q_{\boldsymbol{x}}(\mathcal{K}_0)$.*

This is a direct result of Theorem 4.1.1, since any inverse distribution induces $P_{\mathcal{D}_{\boldsymbol{x}}}$ through (4.2). This result also defines a *feasible support* for computing an inverse distribution. In particular, we show that the inverse distribution $P_{\Lambda,\boldsymbol{x}}$ using the uniform ansatz has a support in the following forms.

Theorem 4.1.2. *The support of $P_{\Lambda,\boldsymbol{x}}$ is $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0))$. Furthermore, since the map between $\mathcal{L}_{\boldsymbol{x}}$ and $\mathcal{D}_{\boldsymbol{x}}$ is one-to-one and onto, the support of $P_{\Lambda,\boldsymbol{x}}$ is $Q_{\boldsymbol{x}}^{-1}(Q_{\boldsymbol{x}}(\mathcal{K}_0))$.*

Proof. This result is proved in Chapter 5.2.2. □

Then, we have the following direct result on the support of any inverse distribution by combining Corollary 4.1.1.2 and Theorem 4.1.2.

Corollary 4.1.2.1. *The support of any inverse distribution is a subset of $Q_{\boldsymbol{x}}^{-1}(Q_{\boldsymbol{x}}(\mathcal{K}_0))$.*

This result shows that the support of $P_{\Lambda,\boldsymbol{x}}$ is the maximum among all possible supports of inverse distributions, including the generating distribution P_Λ. We call this particular support the *inverse support* of the SIP. In this paper, we use the inverse support to recover $\mathcal{K}_0$ in the absence of information about distributions along generalized contours.

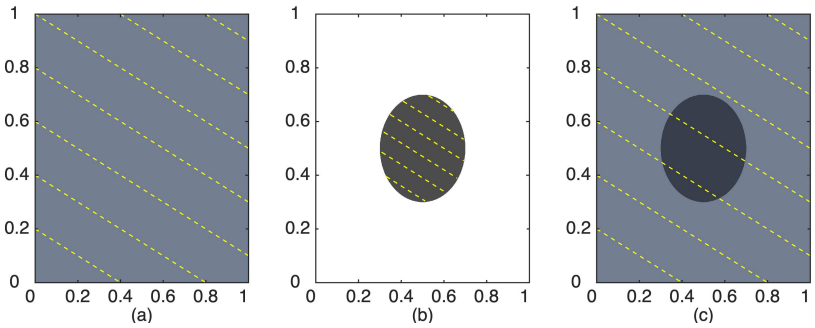


Figure 4.3: A graphical display of the equivalence classes, i.e., generalized contours (dashed), on the space of $\boldsymbol{a}$ with uniform distribution when $x = 1$. Panel (a) shows the generalized contours on $\Lambda_1 = [0,1]^2$ and panel (b) shows them on $\Lambda_2 = \{(a_1, a_2)^\top : (a_1 - 0.5)^2 + (a_2 - 0.5)^2 \leq 0.2^2\}$. Panel (c) shows the generalized contours on Λ_1 when the domain of $\boldsymbol{a}$ actually lies inside the disk.

We continue to consider Examples I and II in Section 4.1.1 for the simple linear model in which we make explicit argument about the domain and support of $\boldsymbol{a}$. The generalized contours in each example are parallel lines (or segments) with a common $x = 1$, which are shown as dashed lines in each panel of Figure 4.3. In Example I, $\boldsymbol{a}$ is uniformly distributed and supported on the domain $\Lambda = [0,1]^2$. For a fixed $x = 1$, the generalized contour is a line segment in $\mathbb{R}^2$ confined to the unit square; see the dashed lines in panel (a). In Example II, we consider the domain $\tilde{\Lambda} = \{(a_1, a_2)^\top : (a_1 - 0.5)^2 + (a_2 - 0.5)^2 \leq 0.2^2\}$, which is the exact support of $\boldsymbol{a}$. Then, the generalized contour is a line segment confined to the disk; see panel (b). Panels (a) and (b) of Figure 4.4 show the unique probability distributions $P_{\mathcal{L}_{\boldsymbol{x}}}$ for Examples I and II when $x = 1$, respectively. In addition, the corresponding inverse distributions under the uniform ansatz are shown in panels (a) and (b) of Figure 4.5. It can be seen that the support of $P_{\mathcal{L}_{\boldsymbol{x}}}$ is determined by the support of the generating distribution and the domain, and the inverse distributions under the

uniform ansatz exactly recovers the generating distributions when the domain is chosen according to the support of the generating distribution.

Furthermore, we consider an interesting scenario, namely Example IIa. Unlike Example II, even though $\boldsymbol{a}$ lies uniformly inside the disk, we choose a more *conservative* choice of domain, i.e. Λ. Here, Λ is considered as the space of scientifically admissible values of $\boldsymbol{a}$, and the support $\tilde{\Lambda}$ is an unknown subset of Λ in the SIP. As depicted in panel (c) of Figure 4.3, the generalized contours are similar to those in Example I. However, the probability distribution $P_{\mathcal{L}_x}$ shown in panel (c) of Figure 4.4 is identical to the one in Example II. This is due to the fact that the output distribution of $y|x = 1$ and its support remain the same in both Examples II and IIa. The corresponding inverse distribution under the uniform ansatz is shown in panel (c) of Figure 4.5, and it is quite different from the inverse distribution in Example II. The domain of the inverse distribution in Example IIa is much larger than that of the inverse distribution in Example II. This suggests that the choice of domain is crucial to solving the SIP, which leads to the discussion of domain recovery in the next section.

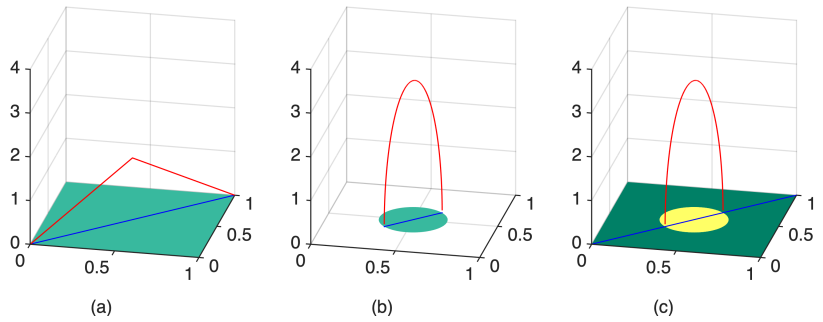


Figure 4.4: A graphical display of the density functions of $P_{\mathcal{L}_x}$ (red) induced by $P_{\mathcal{D}_x}$ on the set of the equivalence classes $\mathcal{L}_x$ (blue) when $x = 1$. Panels (a) and (b) show the unique $P_{\mathcal{L}_x}$ for Examples I and II, respectively. Panel (c) shows $P_{\mathcal{L}_x}$ in Example IIa on the unit square $[0, 1]^2$ with the actual domain of $\boldsymbol{a}$, a yellow colored area.

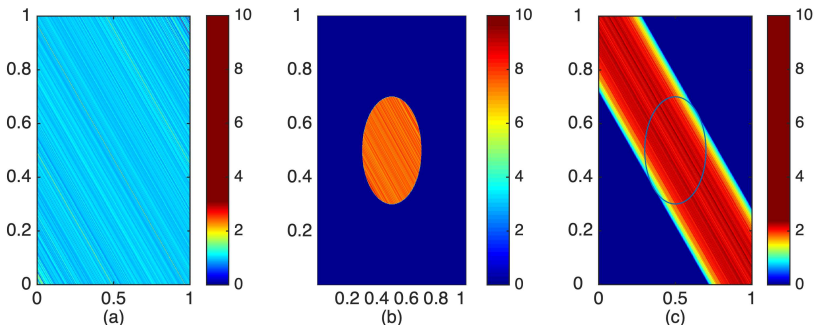


Figure 4.5: A graphical display of the inverse distributions $P_{\Lambda,\boldsymbol{x}}$ computed through (4.3) on the domain of $\boldsymbol{a}$ under the uniform ansatz for Examples I, II and IIa. Panels (a) and (b) show the inverse distributions for Examples I and II, respectively. Panel (c) shows the inverse distribution for Example IIa computed on $[0,1]^2$ whose domain contains the actual support of $\boldsymbol{a}$, i.e. the region inside the blue colored circle.

4.2 Feasible Supports in the SIP

In the previous work of the SIP (Butler et al., 2015), the domain Λ is given by the prior knowledge which is suggested by scientists in solving the inverse problems. However, this might cause bias and give misleading information because the domain might be much larger than the actual support of P_Λ.

In this section, we introduce some interesting aspects regarding the support in SIPs and propose an approach to recover the actual support. Consider a general physical system described by a map, $y = Q_{\boldsymbol{x}}(\boldsymbol{a})$, where $\boldsymbol{a} = (a_1, a_2, \ldots, a_d)^\top \in \Lambda \subset \mathbb{R}^d$ is an unobservable random input reflecting the physical state of the system and $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top \in \mathcal{X} \subset \mathbb{R}^d$ is an observable deterministic input. In addition, Λ is assumed to be compact and simply connected. In order to obtain the continuity of inverse distributions, we further assume that $Q_{\boldsymbol{x}}^{-1}$ is differentiable and that $Q_{\boldsymbol{x}}$ and the boundary of Λ, denoted by $\partial\Lambda$, satisfy certain smoothness assumptions; see Yang (2018) for more details. These assumptions are also satisfied in Section 4.3 for linear models. We provide more details when theoretical properties are established.

4.2.1 Feasible Generating Distributions

Recall that, a SFP describes a data-generating process; that is, given a probability distribution P_Λ, dominated by the Lebesgue measure μ_Λ in $\mathbb{R}^d$, of $\boldsymbol{a} \in \Lambda$ and a map $Q_{\boldsymbol{x}}$, a unique probability

distribution $P_{\mathcal{D}_{\boldsymbol{x}}}$ is induced. However, for any $\boldsymbol{x} \in \mathcal{X}$, various choices of P_Λ may induce the same $P_{\mathcal{D}_{\boldsymbol{x}}}$ which means a SIP may yield multiple solutions, e.g. of (4.2). In Chapter 3, an equivalence class of locally feasible solutions indexed by each given $\boldsymbol{x} \in \mathcal{X}$ and $P_{\mathcal{D}_{\boldsymbol{x}}}$ is defined as

$$\mathbb{P}_{\Lambda,\boldsymbol{x}} := \{\widetilde{P}_\Lambda \ll \mu_\Lambda : \widetilde{P}_\Lambda(A) = \int_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(A)} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\ P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = P_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \forall A \in \mathcal{B}_\Lambda\},$$

where μ_Λ is the Lebesgue measure. Clearly, each inverse distribution $P_{\Lambda,\boldsymbol{x}}$ is contained in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ and the generating distribution P_Λ is contained in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ for any $\boldsymbol{x} \in \mathcal{X}$. Then, a solution for all $\boldsymbol{x} \in \mathcal{X}$ simultaneously is to consider the intersection of $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ over all possible $\boldsymbol{x}$. Then we have

$$\mathbb{P}_{\Lambda,\mathcal{X}} = \bigcap_{\boldsymbol{x} \in \mathcal{X}} \mathbb{P}_{\Lambda,\boldsymbol{x}}, \tag{4.4}$$

which is nonempty since it contains the generating distribution P_Λ. All elements in $\mathbb{P}_{\Lambda,\mathcal{X}}$ are equivalent distributions of $\boldsymbol{a}$ in the sense that they induce the same family of output distributions $\{P_{\mathcal{D}_{\boldsymbol{x}}}\}_{\boldsymbol{x} \in \mathcal{X}}$ as P_Λ does. The family $\mathbb{P}_{\Lambda,\mathcal{X}}$ is a convex set.

The supports of distributions in this family are of concern in this paper. Note that distributions in $\mathbb{P}_{\Lambda,\mathcal{X}}$ may have different supports. We propose a maximal feasible support that contains all possible feasible supports in $\mathbb{P}_{\Lambda,\mathcal{X}}$.

4.2.2 Feasible Support and Inverse Support

In this section, we focus on the support of inverse distributions and address the recoverability issue of the supports in $\mathbb{P}_{\Lambda,\mathcal{X}}$. For any $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$, the support of $\widetilde{P}_\Lambda$ is denoted by

$$\mathcal{K}_0(\widetilde{P}_\Lambda) = \{\lambda \in \Lambda : \widetilde{P}_\Lambda(N_\lambda) > 0 \text{ for every open neighborhood } N_\lambda \in \mathcal{B}_\Lambda \text{ of } \lambda\}.$$

In the rest of this paper, $\mathcal{K}_0(\widetilde{P}_\Lambda)$ is referred to as the *feasible forward support* for $\widetilde{P}_\Lambda$. Note that, for a given $\boldsymbol{x}$, the image $Q_{\boldsymbol{x}}(\mathcal{K}_0(\widetilde{P}_\Lambda))$ of $\mathcal{K}_0(\widetilde{P}_\Lambda)$, is unique for any $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$ because $Q_{\boldsymbol{x}}(\mathcal{K}_0(\widetilde{P}_\Lambda))$

is the support of $P_{\mathcal{D}_{\boldsymbol{x}}}$ by (4.2) and Theorem 4.1.1. Unless it is important to note $\widetilde{P}_\Lambda$, we use $\mathcal{K}_0$ for simplicity. Back to $\mathbb{P}_{\Lambda,\mathcal{X}}$, we can discuss practically those $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$ in SFPs where the support boundary has measure zero $\mu_\Lambda(\partial\mathcal{K}_0(\widetilde{P}_\Lambda)) = 0$.

Let $B_{\boldsymbol{x}} = Q_{\boldsymbol{x}}(\mathcal{K}_0)$, which is an element in $\mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}}$. Then, for any $\boldsymbol{x} \in \mathcal{X}$, the collection of pre-images can be expressed as

$$\mathbb{B}_{\boldsymbol{x}}^{-1} := \{C \in \mathcal{B}_\Lambda : Q_{\boldsymbol{x}}(C) = B_{\boldsymbol{x}}\},$$

which is well defined for any given $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$ from the definition of $\mathcal{K}_0$ and $B_{\boldsymbol{x}}$. $\mathbb{B}_{\boldsymbol{x}}^{-1}$ is referred to as the *locally feasible support family* for a given $B_{\boldsymbol{x}}$ and it is also the set that contains all possible supports of locally feasible distributions in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$.

Next, we adopt a similar approach as (4.4) by considering the intersection of $\mathbb{B}_{\boldsymbol{x}}^{-1}$ over all $\boldsymbol{x} \in \mathcal{X}$. Define

$$\mathbb{B}_{\mathcal{X}}^{-1} := \bigcap_{\boldsymbol{x}\in\mathcal{X}} \mathbb{B}_{\boldsymbol{x}}^{-1}. \tag{4.5}$$

Any element (i.e., a set) in $\mathbb{B}_{\mathcal{X}}^{-1}$ supports all GFGDs in $\mathbb{P}_{\Lambda,\mathcal{X}}$, and hence is referred to as the *globally feasible support*. In addition, $\mathbb{B}_{\mathcal{X}}^{-1}$ is called the *globally feasible support family*. It can be seen that, for any $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$, $\mathcal{K}_0(\widetilde{P}_\Lambda)$ belongs to $\mathbb{B}_{\mathcal{X}}^{-1}$.

Let $\Lambda_{\boldsymbol{x}}$ be the support of $P_{\Lambda,\boldsymbol{x}}$ under the uniform ansatz, and specifically, $\Lambda_{\boldsymbol{x}} = Q_{\boldsymbol{x}}^{-1}(Q_{\boldsymbol{x}}(\mathcal{K}_0))$. It is a direct consequence from Corollary 4.1.2.1 that $\mathcal{K}_0 \subset \Lambda_{\boldsymbol{x}}$, and the inverse support is essentially the *maximal element* in $\mathbb{B}_{\boldsymbol{x}}^{-1}$.

Theorem 4.2.1. *For any $C \in \mathbb{B}_{\boldsymbol{x}}^{-1}$, we have $C \subset \Lambda_{\boldsymbol{x}} \in \mathbb{B}_{\boldsymbol{x}}^{-1}$.*

Since $\Lambda_{\boldsymbol{x}}$ is the maximal in the equivalence class, it is independent of the choice of $\widetilde{P}_\Lambda$ and unique. When we have no *a priori* knowledge regarding the inverse support, we choose the maximal element $\Lambda_{\boldsymbol{x}}$ in $\mathbb{B}_{\boldsymbol{x}}^{-1}$ to prevent potential information loss, i.e. to avoid assigning zero probability to events with nonzero probability. This is fairly important in some cases, e.g. when we try to locate certain region in which a hurricane occurs. Regions have the chance to encounter the hurricane should not be avoided and clear regions should not be counted in the forecast. Any other choices

in $\mathbb{B}_{\boldsymbol{x}}^{-1}$ require additional assumptions about the geometry that is not only determined by $Q_{\boldsymbol{x}}$, Λ, or $P_{\mathcal{D}_{\boldsymbol{x}}}$.

Returning to Example IIa as depicted in the panel (c) of Figure 4.5, a feasible forward support $\mathcal{K}_0$ in the SFP is shown as the disk, and the corresponding inverse support $\Lambda_{\boldsymbol{x}}$, i.e., the maximal element, is shown as a band or parallel slab. Figure 4.6 shows two elements in $\mathbb{B}_{\boldsymbol{x}}^{-1}$, different from $\mathcal{K}_0$ and $\Lambda_{\boldsymbol{x}}$, that yield the same range $Q_{\boldsymbol{x}}(\mathcal{K}_0)$ of y.

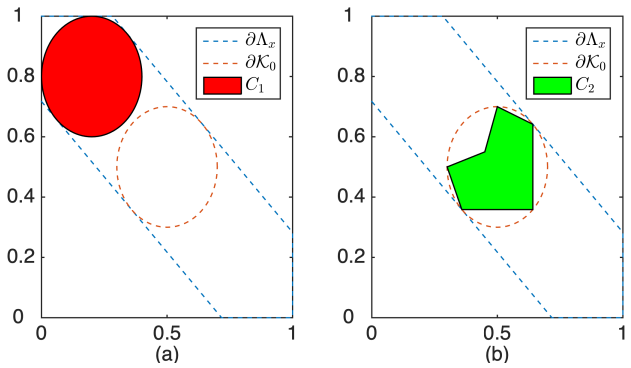


Figure 4.6: Two different elements C_1, C_2 in $\mathbb{B}_{\boldsymbol{x}}^{-1}$ are shown, where C_1 is a disk different from $\mathcal{K}_0$ and C_2 is a much smaller non-convex domain that does not preserve the shape of $\mathcal{K}_0$. These two elements are both contained in $\Lambda_{\boldsymbol{x}}$.

4.2.3 Recoverable Support

Since $\mathcal{K}_0 \subset \Lambda_{\boldsymbol{x}}$ for any $\boldsymbol{x} \in \mathcal{X}$, we consider the intersection of the collection of all maximal elements, i.e. $\{\Lambda_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$, denoted by

$$\mathcal{K}_{\mathcal{X}} := \bigcap_{\boldsymbol{x}\in\mathcal{X}} \Lambda_{\boldsymbol{x}},$$

when $\mathcal{X}$ is countable or $\mathcal{K}_{\mathcal{X}}$ is equal to the intersection under a countable subsequence. We further denote the collection of all intersections of the maximal supports $\{\Lambda_{\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$ by $\mathcal{R}_{\mathcal{X}}$ and it is referred to as the family of *recoverable supports* in the SIP algorithm. In the next theorem, we first show $\mathcal{K}_{\mathcal{X}}$ is the maximal element in $\mathbb{B}_{\mathcal{X}}^{-1}$.

Theorem 4.2.2. *For any $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$ and any $C \in \mathbb{B}_{\mathcal{X}}^{-1}$, we have $C \subset \mathcal{K}_{\mathcal{X}} \in \mathbb{B}_{\mathcal{X}}^{-1}$.*

Proof. See Chapter 5.2.3. □

It can be further seen that, for a given SFP or, more generally, for any GFGD, $\mathcal{K}_{\mathcal{X}}$ is unique and independent of $\widetilde{P}_{\Lambda}$, same to $\Lambda_{\boldsymbol{x}}$. In practice, it is also attainable by intersecting all $\Lambda_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathcal{X}$. In addition, $\mathcal{K}_{\mathcal{X}}$ is also a minimax in the sense that it is the smallest recoverable support among all recoverable supports in $\mathcal{R}_{\mathcal{X}}$ and thus $\mathcal{K}_{\mathcal{X}}$ is referred to as the *minimax support* in the SIP. Our research goal focuses on recovering $\mathcal{K}_{\mathcal{X}}$ instead of $\mathcal{K}_0$ which depends on the choice of the GFGD.

Continue to consider Example IIa in panel (c) of Figure 4.5. The feasible forward support is

$$\mathcal{K}_0 = \{(a_1, a_2)^\top \in \Lambda : (a_1 - 0.5)^2 + (a_2 - 0.5)^2 \leq 0.2^2\},$$

where $\boldsymbol{a}$ has a uniform distribution on the disk; see panel (a) of Figure 4.7. When $\boldsymbol{x} = (1, 1)^\top$, the corresponding inverse support $\Lambda_{\boldsymbol{x}}$ is a band as shown in panel (b) of Figure 4.7. In panel (c), we show the the intersection of inverse supports from three different values of $\boldsymbol{x}$. The shape of the intersection is a circumscribed polygon of $\mathcal{K}_0$. As the number of values of $\boldsymbol{x}$ increases and the collection of those values is dense in $\mathcal{X}$, the resulting intersection of inverse supports will converge to the exact $\mathcal{K}_0$. That is $\mathcal{K}_{\mathcal{X}}$ is the same as $\mathcal{K}_0$. Next, we consider a new example, namely Example

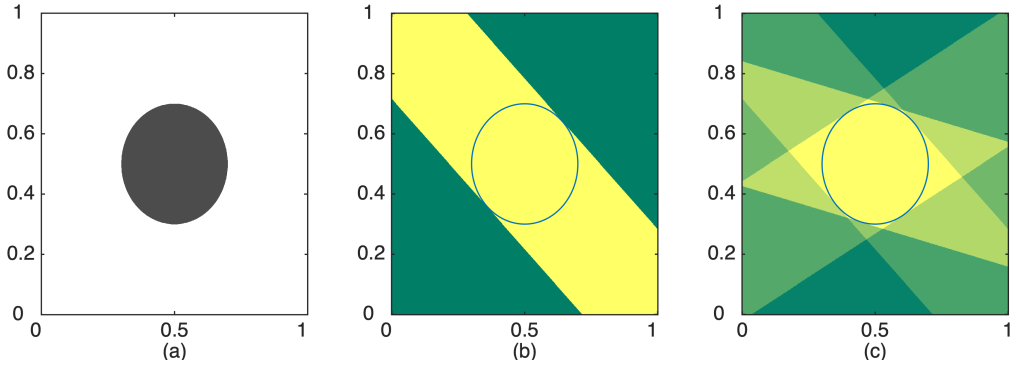


Figure 4.7: A graphical display of the feasible forward support and an inverse support in Example IIa. (a) The feasible forward support $\mathcal{K}_0$, (b) the inverse inverse support when $\boldsymbol{x} = (1, 1)^\top$, and (c) an intersection of three different inverse supports corresponding to three distinct values of $\boldsymbol{x}$. Note that, all inverse supports (the yellow bands) contain the feasible original support $\mathcal{K}_0$, and so is their intersection.

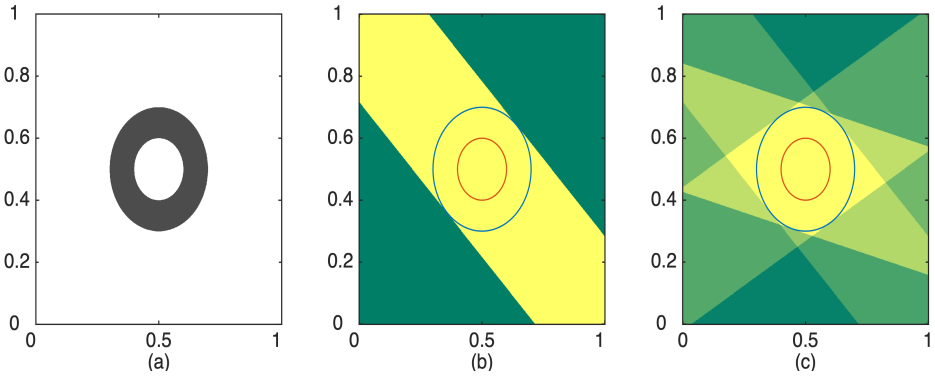


Figure 4.8: A graphical display of the feasible forward support and an inverse support in Example III. (a) The feasible forward support $\mathcal{K}_0$, (b) the inverse inverse support when $\boldsymbol{x} = (1, 1)^\top$, and (c) an intersection of three different inverse supports corresponding to three distinct values of $\boldsymbol{x}$.

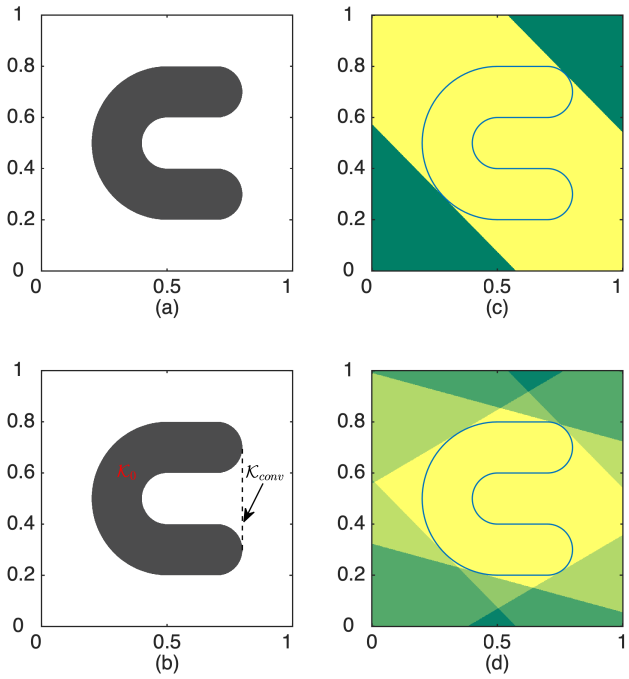


Figure 4.9: A graphical display of the feasible forward support and an inverse support in Example IV. (a) A U-shaped feasible forward support $\mathcal{K}_0$, (b) $\mathcal{K}_{conv}$, the convex hull of $\mathcal{K}_0$, (c) the inverse inverse support when $\boldsymbol{x} = (1, 1)^\top$, (d) an intersection of three different inverse supports corresponding to three distinct values of $\boldsymbol{x}$.

III, in which $\boldsymbol{a}$ has a uniform distribution on an annulus

$$\mathcal{K}_0 = \{(a_1, a_2)^\top \in \Lambda : 0.1^2 \leq (a_1 - 0.5)^2 + (a_2 - 0.5)^2 \leq 0.2^2\}$$

where $\Lambda = [0, 1]^2$. In addition, $Q_{\boldsymbol{x}}$ is the same simple linear model (4.1). In Figure 4.8, the feasible forward support is shown in panel (a). Interestingly, when $\boldsymbol{x} = (1, 1)^\top$, the inverse support $\Lambda_{\boldsymbol{x}}$ remains the same compared to that of Example IIa. In addition, as shown in panel (c), the intersection of three supports remains the shape of a polygon, which is circumscribed around the outer disk of $\mathcal{K}_0$. However, different lesson is learned here. As the number of $\boldsymbol{x}$-values increases, the intersection, always containing $\mathcal{K}_0$, will no longer converge to $\mathcal{K}_0$. Consequently, the smallest recoverable support $\mathcal{K}_{\mathcal{X}}$ is different from $\mathcal{K}_0$.

Finally, we consider a more challenging example, Example IV, in which $\boldsymbol{a}$ has a uniform distribution on a U-shaped $\mathcal{K}_0$ inside $\Lambda = [0, 1]^2$; see panel (a) of Figure 4.9. Again, we consider the simple linear model as described in (4.1). Similar figures, showing the inverse support $\Lambda_{\boldsymbol{x}}$ for a given $\boldsymbol{x} = (1, 1)^\top$ and the intersection of supports corresponding to three $\boldsymbol{x}$-values, are displayed in panels (c) and (d), respectively. The shape of the intersection is also a circumscribed polygon; in fact, as the number of $\boldsymbol{x}$-values increases, the intersection converges to the convex hull of $\mathcal{K}_0$, denoted by $\mathcal{K}_{conv}$; see panel (b) of Figure 4.9. In fact, for general linear models, the smallest recoverable support $\mathcal{K}_{\mathcal{X}}$ is exactly the convex hull of any feasible forward support of $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$. We prove this result in Section 4.3.

In summary, the importance of recovering $\mathcal{K}_{\mathcal{X}}$ is that $\Lambda_{\boldsymbol{x}}$ may be much larger than the minimax support $\mathcal{K}_{\mathcal{X}} \supset \mathcal{K}_0$ and thus having several negative consequences including assigning nonzero probability to events outside the feasible forward support which should have zero probability. A larger support might contain events with zero probability in the real problem which can be misleading and means we spend effort to discover useless information in the sense of probabilities, see Figure 4.5 panel (b) and (c). However, we observe that $\Lambda_{\boldsymbol{x}}$ always contains any $\mathcal{K}_0 \in \mathbb{B}_{\mathcal{X}}^{-1}$ for any value of $\boldsymbol{x}$ and describes some different geometry of the feasible forward support for different $\boldsymbol{x}$. Then it is natural to use the intersection of inverse supports to find the minimax support.

4.2.4 Recoverable Support in Nonlinear Examples

In the previous examples, we consider a simple linear model with various choices of feasible forward support $\mathcal{K}_0$. In this section, some examples are shown to describe $\mathcal{K}_\mathcal{X}$ in nonlinear models.

We consider two examples, namely, Example V and Example Va. In both examples, $\boldsymbol{a} = (a_1, a_2)^\top \in \Lambda = [0,1]^2$ is uniformly distributed on

$$\mathcal{K}_0 = \{(a_1, a_2)^\top : (a_1 - 0.5)^2 + (a_2 - 0.5)^2 \leq 0.2^2\}.$$

Two different nonlinear models are considered

$$\text{Example V}: \qquad y = Q_{\boldsymbol{x}}(\boldsymbol{a}) := (a_1 - 0.5)^2 x + (a_2 - 0.5)^2, \tag{4.6}$$

$$\text{Example Va}: \qquad y = Q_{\boldsymbol{x}}(\boldsymbol{a}) := (a_1 - 0.3)^2 x + (a_2 - 0.5)^2, \tag{4.7}$$

where $x \in \mathbb{R}$. A small change of the center is made in Example V to obtain the new Example Va. The nonlinear equations of them are similar but not the results of their inverse supports. In Figure 4.10, panel (a) shows the intersection of inverse supports $\Lambda_{\boldsymbol{x}}$ for three values of $\boldsymbol{x}$ in Example V and panel (b) shows the limiting intersection of inverse supports for many $\boldsymbol{x}$. Even though each inverse support might not be a convex support, $\mathcal{K}_\mathcal{X}$ is still a convex set. However, in Figure 4.11, some inverse supports $\Lambda_{\boldsymbol{x}}$ for different values of $\boldsymbol{x}$ in Example Va are shown in panel (a) and the limiting intersection of inverse supports is shown in panel (b), where $\mathcal{K}_\mathcal{X}$ is approximately two disks containing $\mathcal{K}_0$. This is an interesting scenario that shows a convex $\mathcal{K}_0$ might no longer be identifiable in nonlinear models.

In nonlinear models, the minimax support might not be a straightforward support that can identify the feasible forward support $\mathcal{K}_0$ in a region and can be rather complicated, e.g. Example Va, but it is still useful in the sense that it significantly reduces the size of the domain and thus has a better approximation of the inverse distributions as in linear models. In this paper, we focus on the linear models and show some properties of the inverse distributions computed on the recovered

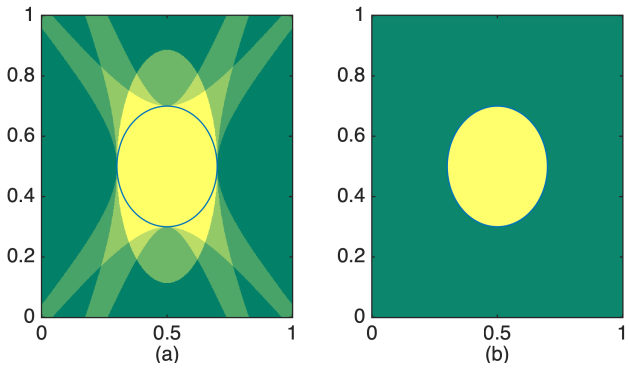


Figure 4.10: With regard to Example V, three different inverse supports indexed by three different values of $\boldsymbol{x}$ are displayed in panel (a) in different shading and panel (b) shows the limiting intersection, $\mathcal{K}_{\mathcal{X}}$, of inverse supports.

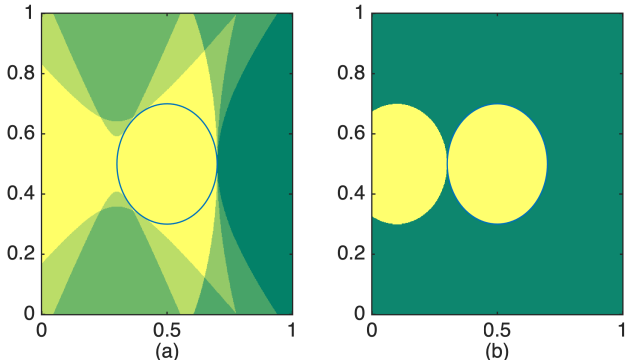


Figure 4.11: With regard to Example Va, three different inverse supports indexed by three different values of $\boldsymbol{x}$ are displayed in panel (a) in different shading and panel (b) shows the limiting intersection, $\mathcal{K}_{\mathcal{X}}$, of inverse supports.

support in the following section. Some of the properties can actually be generalized to nonlinear models.

4.3 Linear Models

4.3.1 Geometry of Inverse Supports in General Linear Models

In this section, we explore some properties of the inverse support and inverse distributions computed on the recoverable support and we show some results on the convergence of inverse supports in linear models. The main result is that the intersection of supports, $\Lambda_{\boldsymbol{x}}$, over $\boldsymbol{x} \in \mathcal{X}$, i.e.

$\mathcal{K}_{\mathcal{X}}$, is the convex hull of the support $\mathcal{K}_0(\widetilde{P}_\Lambda)$ for any $\widetilde{P}_\Lambda \in \mathbb{P}_{\Lambda,\mathcal{X}}$ when $\mathcal{X} = \mathbb{R}^d$, see Theorem 4.3.4. In particular, the intersection is denoted by $\mathcal{K}_{conv}$ in this case.

We first review some examples of the inverse support in the previous discussion. In Examples I and II, $\Lambda = \mathcal{K}_0$ for all $\boldsymbol{x}$ and the shape of the inverse support $\Lambda_{\boldsymbol{x}}$ is the same as $\mathcal{K}_0$; see panels (a)-(b) of Figure 4.5. However, in Example IIa, when $\Lambda \supset \mathcal{K}_0$, an interesting result is observed. Here, $\Lambda_{\boldsymbol{x}}$ is a parallel slab, restricted to $\Lambda = [0,1]^2$; see panel (c) of Figure 4.5. In the following theorem, we establish that $\Lambda_{\boldsymbol{x}}$ can be written as an intersection of a parallel slab and the domain Λ.

Consider a general linear model expressed as

$$y = Q_{\boldsymbol{x}}(\boldsymbol{a}) = \boldsymbol{a} \cdot \boldsymbol{x} = a_1 x_1 + a_2 x_2 + \cdots + a_d x_d, \tag{4.8}$$

where $\boldsymbol{a} = (a_1, a_2, \ldots, a_d)^\top \in \Lambda \subset \mathbb{R}^d$ is the unobservable input with a generating distribution in $\mathbb{P}_{\Lambda,\mathcal{X}}$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top \in \mathcal{X} \subset \mathbb{R}^d$ is the observable deterministic input. We use the inverse support $\Lambda_{\boldsymbol{x}}$ under the uniform ansatz for the rest of the paper.

Theorem 4.3.1. *Consider the general linear model in (4.8), $\Lambda_{\boldsymbol{x}}$ can be uniquely expressed as*

$$\Lambda_{\boldsymbol{x}} = \Lambda \cap S_{c_1,c_2},$$

where c_1, c_2 are some constants, and $S_{c_1,c_2} = \{\boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{a} \cdot \boldsymbol{x} \in [c_1, c_2]\}$ is a parallel slab in $\mathbb{R}^d$.

Proof. See Chapter 5.2.4. □

The shape of $\Lambda_{\boldsymbol{x}}$ is referred to as a generalized parallel slab. Here, $\boldsymbol{a} \cdot \boldsymbol{x} = c_1$ or $\boldsymbol{a} \cdot \boldsymbol{x} = c_2$ characterize the boundary of the support $\mathcal{K}_0$, and are referred to as the supporting hyperplanes. Note that each hyperplane yields two open half-spaces. Except for the boundary, the support $\mathcal{K}_0$ lies in exactly one half-space. In addition, $\mathcal{K}_0$ has some boundary points on the hyperplane.

It is shown in Section 4.2 that $\mathcal{K}_0$ is a subset of $\Lambda_{\boldsymbol{x}}$ and thus

$$\mathcal{K}_0 \subset \mathcal{K}_{\mathcal{X}} = \bigcap_{\boldsymbol{x} \in \mathcal{X}} \Lambda_{\boldsymbol{x}}.$$

Returning to Example IIa in Figure 4.12, three generalized parallel slabs, corresponding to $x = \tan(\frac{\pi}{4}), \tan(-\frac{\pi}{6}), \tan(\frac{5\pi}{12})$, are depicted. It can be seen that $\mathcal{K}_0$ lies inside the intersection, and is tangent to all supporting hyperplanes. It can be seen that the intersection provides a reasonable approximation to the support $\mathcal{K}_0$. As shown in Theorem 4.3.4, such intersection converges $\mathcal{K}_{conv}$ and in this particular case $\mathcal{K}_0 = \mathcal{K}_{conv}$. The main reason is that the underlying support $\mathcal{K}_0$ is convex and compact, and has smooth boundary.

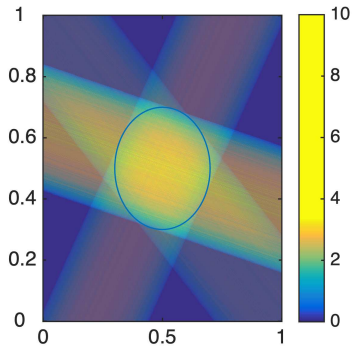


Figure 4.12: Three different density functions of $P_{\Lambda,\boldsymbol{x}}$ indexed by three different values of $\boldsymbol{x}$ displayed in different shading. The support $\Lambda_{\boldsymbol{x}}$ of all three inverses contain the support $\mathcal{K}_0$. The impact of the uniform ansatz on the inverse distribution is different in three different geometries.

In summary, when $\mathcal{K}_0$ is convex, it is recovered as the limit of the intersection of generalized parallel slabs corresponding to different $\boldsymbol{x} \in \mathbb{R}^d$ values because $\mathcal{K}_0 = \mathcal{K}_{conv}$ is recoverable. However, when $\mathcal{K}_0$ is non-convex, it is no longer recoverable; instead, the convex hull of $\mathcal{K}_0$, i.e. $\mathcal{K}_{conv}$, is recoverable that is of interest.

4.3.2 Inverse Distribution Computed on the Recoverable Support

We show that the inverse distribution computed on the convex hull $\mathcal{K}_{conv}$ in linear models is, in some sense, a "minimax" solution. To formulate this problem, let g_{conv} denote the density function of the inverse distribution computed on $\mathcal{K}_{conv}$ and g_c denote the density function of the inverse distribution on any convex support that contains $\mathcal{K}_0$ and $\mathcal{H}_c$ denote the class of such g_c. g_0 refers to the density function of the GFGDs in Section 4.2. Suppose we compute the inverse

distribution based on the observed value of $\boldsymbol{x}$ and the corresponding map $Q_{\boldsymbol{x}}$. If g_0 is unimodal, by the disintegration theorem on the domain $\Lambda \supset \mathcal{K}_{conv}$, we have

$$g_{conv} = \arg\inf_{g_c \in \mathcal{H}_c} \int_{\mathcal{L}_{\boldsymbol{x}}} \sup_{\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap \Lambda} (g_c - g_0) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell),$$

since g_{conv} on each equivalence class $\ell \in \mathcal{L}_{\boldsymbol{x}}$ is the minimum of

$$\sup_{\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap \Lambda} (g_c - g_0)$$

of all $g_c \in \mathcal{H}_c$ in the disintegration. This is because g_c is uniform on each $\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap \Lambda$ while g_0 is unimodal.

If g_0 is a general density function, then we have

$$g_{conv} = \arg\inf_{g_c \in \mathcal{H}_c} \int_{\mathcal{L}_{\boldsymbol{x}}} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap \mathcal{K}_{conv}} (g_c(\lambda) - g_0(\lambda)) d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell)$$

since

$$\int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap \Lambda} (g_c(\lambda) - g_0(\lambda)) d\mu_\ell(\lambda) = 0$$

for each ℓ and g_c by the disintegration.

This shows the inverse distribution on the convex hull is the most "concentrated" distribution in the class. In the next section, we show each individual inverse distribution on the convex hull along with the the experimental expectation of all inverse distributions can provide a fairly adequate approximation to the region of high probability of the SIP due to the "concentration" property here.

4.3.3 Experimental Expectation of Inverse Distributions on the Recoverable Support

The experimental expectation is defined as the expectation of inverse distributions over $\boldsymbol{x}$, see Section 3.3 of Chapter 3. In particular, when $\boldsymbol{X}$ is random on $(\mathcal{X}, \mathcal{B}_{\mathcal{D}_{\mathcal{X}}}, P_{\mathcal{X}})$, the experimental

expectation is defined as

$$\bar{P}(A) = \int_{\boldsymbol{x}\in\mathcal{X}} P_{\Lambda,\boldsymbol{x}}(A) f_{\mathcal{X}}(\boldsymbol{x}) d\mu_{\mathcal{X}}(\boldsymbol{x}), \quad \forall A \in \mathcal{B}_{\Lambda},$$

where $P_{\Lambda,\boldsymbol{x}}$ is an inverse distribution computed on Λ, $f_{\mathcal{X}}$ is the density function of $P_{\mathcal{X}}$ with respect to the Lebesgue measure $\mu_{\mathcal{X}}$ on $\mathcal{X}$. Supposing there is a collection of identical and independent samples of $\boldsymbol{X}$, denoted by $\{\boldsymbol{X}_i\}_{i=1}^{\infty}$, the intersection of inverse supports is defined as

$$\Lambda_n := \bigcap_{i=1}^{n} \Lambda_{\boldsymbol{X}_i},$$

where $\Lambda_{\boldsymbol{X}_i}$ is the inverse support of the inverse distribution $P_{\Lambda,\boldsymbol{X}_i}$, and the experimental expectation is defined as

$$\bar{P}(A;\Lambda_n) := \int_{\boldsymbol{x}\in\mathcal{X}} P_{\Lambda_n,\boldsymbol{x}}(A) f_{\mathcal{X}}(\boldsymbol{x}) d\mu_{\mathcal{X}}(\boldsymbol{x}),$$

where $P_{\Lambda_n,\boldsymbol{x}}$ is the inverse distribution computed on the domain Λ_n.

Now we investigate the convergence of inverse distributions and the experimental expectation assuming the convergence of inverse support (see Theorem 4.3.4), i.e. $\Lambda_n \to \mathcal{K}_{conv}$ pointwisely as $n \to \infty$ with probability 1. We show the individual inverse distributions converge almost surely, $P_{\Lambda_n,\boldsymbol{x}} \to P_{\mathcal{K}_{conv},\boldsymbol{x}}$ with probability 1, where $\boldsymbol{x} \in \mathcal{X} = \mathbb{R}^d$. In addition, the convergence of $\bar{P}(A;\Lambda_n)$ is implied by the convergence of inverse distributions.

Theorem 4.3.2. *Supposing $\Lambda_n \to \mathcal{K}_{conv}$ almost surely in the generalized linear model (4.8) with a generating distribution in $\mathbb{P}_{\Lambda,\mathcal{X}}$, we have $P_{\Lambda_n,\boldsymbol{x}} \to P_{\mathcal{K}_{conv},\boldsymbol{x}}$ with probability 1.*

Proof. See Chapter 5.2.5. □

Theorem 4.3.3. *Following Theorem 4.3.2, $\bar{P}(\cdot;\Lambda_n) \to \bar{P}(\cdot;\mathcal{K}_{conv})$ almost surely. Equivalently, with probability 1, $\bar{P}(A;\Lambda_n) \to \bar{P}(A;\mathcal{K}_{conv})$ for any $A \in \mathcal{B}_{\mathcal{K}_{conv}}$.*

Proof. See Chapter 5.2.6. □

We continue to explore Example IV about $\mathcal{K}_{conv}$ and the experimental expectation in this case, see Figure 4.13. Panel (a) shows the experimental expectation of inverses $\bar{P}(\cdot;\Lambda)$ without recovering the domain, that is, on the pre-specified domain Λ, and panel (b) shows the the experimental expectation of inverses $\bar{P}(\cdot;\mathcal{K}_{conv})$ after recovering the domain $\mathcal{K}_{conv}$.

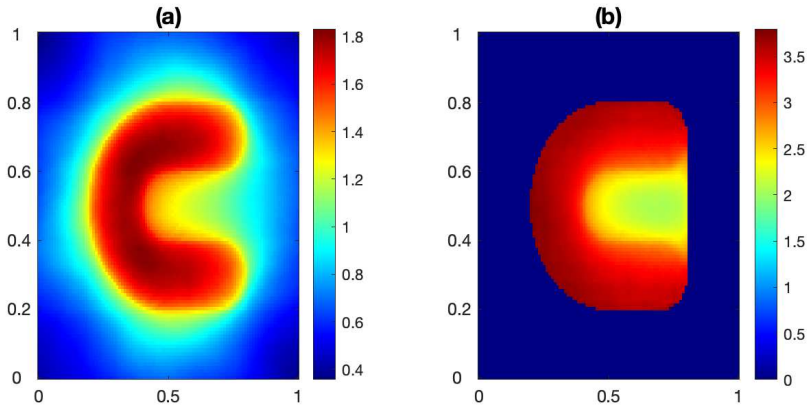


Figure 4.13: A graphical display of $\bar{P}(\cdot;\Lambda)$ and $\bar{P}(\cdot;\mathcal{K}_{conv})$ of Example IV in panels (a)-(b), respectively.

4.3.4 General Results of Convergence

In this section, we discuss the convergence of the intersection of the inverse support given a family of $\boldsymbol{x}$ values. We assume the domain Λ is a simply connected convex compact set with non-empty interior and $\mu_\Lambda(\partial\Lambda) = 0$. In the model (4.8), we assume $\boldsymbol{a}$ is on the probability space $(\Lambda, \mathcal{B}_\Lambda, P_\Lambda)$ with the support $\mathcal{K}_0$ of P_Λ and y in $(\mathcal{D}_{\boldsymbol{x}}, \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}}, P_{\mathcal{D}_{\boldsymbol{x}}})$. Let $\{\boldsymbol{X}_i\}_{i=1}^{\infty}$ be a collection of independent random points distributed according to $P_{\mathcal{X}}$ and let $\Lambda_{\boldsymbol{X}_i}$ denote the inverse support computed according to the point $\boldsymbol{X}_i$.

$\Lambda_{\boldsymbol{X}_i}$ describes the following procedure: once $\boldsymbol{X}_i = \boldsymbol{x}_i$ is sampled, we observe $Q_{\boldsymbol{x}_i}$ and compute the inverse support $\Lambda_{\boldsymbol{X}_i=\boldsymbol{x}_i}$. Therefore, $\Lambda_{\boldsymbol{X}_i}$ is also a random variable only due to the randomness when we sample $\boldsymbol{X}_i$. Since we are interested in a good approximation of $\mathcal{K}_0$ by the sequence $\{\bigcap_{i=1}^{n} \Lambda_{\boldsymbol{X}_i}\}_{n=1}^{\infty}$, we first show that the geometry of the limit is the convex hull of $\mathcal{K}_0$, namely $\mathcal{K}_{conv}$. Note that it is equivalent to compute $\bigcap_{\boldsymbol{x}\in\mathcal{X}} \Lambda_{\boldsymbol{x}}$ because the geometry only depends on the values as we discuss earlier.

Theorem 4.3.4. *Suppose there is a collection $\{\boldsymbol{x}_i\}_{i=1}^{\infty}$ which is a dense subset of $\mathcal{X} = \mathbb{R}^d$. Then, $\bigcap_{i=1}^{\infty} \Lambda_{\boldsymbol{x}_i}$ is the convex hull of $\mathcal{K}_0$.*

Proof. See Chapter 5.2.7. □

Since the support of a GFGD is unknown, we replace it by its convex hull in practice. Accordingly, we assume the support $\mathcal{K}_0$ is a convex compact set with non-empty interior for the rest of the paper unless otherwise specified. As for the case of generalized parallel slabs $\{\bigcap_{i=1}^{n} \Lambda_{\boldsymbol{X}_i}\}_{n=1}^{\infty}$, we use the Nikodym metric to measure the distance between two sets, defined as $d_N(U, V) = \mu_\Lambda(U \triangle V)$ where $U, V \in \mathcal{C}^d := \{$the set of convex bodies (i.e., of compact convex subsets with nonempty interior) in $\mathbb{R}^d\}$. In general measure spaces, d_N is a pseudometric and it becomes a metric in $\mathcal{C}^d$ under the Lebesgue measure.

Note that Theorem 4.3.4 yields the following consequence when a sequence of samples is drawn and $\partial\Lambda$ is of some general boundary class.

Corollary 4.3.4.1. *Suppose a collection $\{\boldsymbol{x}_i\}_{i=1}^{\infty}$ is drawn from a dense subset of $\mathcal{X} = \mathcal{R}^d$ such that $F_n = \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{x}_i \leq \boldsymbol{x}) \to F(\boldsymbol{x})$ where I is the indicator function and F has a positive continuous density w.r.t $\mu_{\mathcal{X}}$. Then $d_N\left(\bigcap_{i=1}^{n} \Lambda_{\boldsymbol{x}_i}, \mathcal{K}_0\right) \to 0$ as $n \to \infty$.*

4.3.5 Convergence Rate of Inverse Supports in the Plane

For future work, we rewrite the model in 4.1 as

$$y = Q_\theta(\boldsymbol{a}) = a_1 \tan\theta + a_2,$$

where Θ is a random variable on the probability space $(I_0 = (-\frac{\pi}{2}, \frac{\pi}{2}), \mathcal{B}_{I_0}, P_\Theta)$. Here, P_Θ has a positive continuous density f with respect to the Lebesgue measure on $(I_0, \mathcal{B}_{I_0})$ and θ is a realization of Θ. Let $\Theta_1, \Theta_2, \ldots,$ be a sequence of independent random points distributed according to P_Θ and let Λ_{Θ_i} denote the inverse support computed according to the point Θ_i. Then, $\Lambda_\theta := \{\lambda \in \mathcal{K}_0 : Q_\theta(\lambda) \in Q_\theta(\Lambda)\}$ where $\mathcal{K}_0$ is the support of $\boldsymbol{a}$ with non-empty interior which is

contained in Λ. Λ_θ is essentially a shape of generalized parallel slabs and θ is the angle of slope characterized by the horizontal axis.

Corollary 4.3.4.2. *Suppose there is a collection $\{\theta_i\}_{i=1}^{\infty}$ which is a dense subset of $I_0 = (-\frac{\pi}{2}, \frac{\pi}{2})$. Then, $\bigcap_{i=1}^{\infty} \Lambda_{\theta_i}$ is the convex hull of $\mathcal{K}_0$.*

Proof. This is equivalent to $\{\tan(\theta_i)\}_{i=1}^{\infty}$ being a dense subset in the image space since it is a continuous mapping. □

We measure the distance under some boundary conditions in this section and explore the convergence rate under specific conditions.

Condition (PL). $\partial(\mathcal{K}_0)$ of $\mathcal{K}_0$ is a (δ, m_0)-Lipschitz domain if for each $z \in \partial(\mathcal{K}_0)\backslash m_0$ there exist a neighborhood (open ball) $B(z, r)$ where $r > 0$ and a Lipschitz function $A_z : \mathbb{R} \to \mathbb{R}$ with slope $||A_z'||_\infty \leq \delta$ and A_z'' continuous and bounded from below by $\tau > 0$ such that, after a suitable rotation, $\mathcal{K}_0 \cap B(z, r) = \{(x, y) \in B(z, r) : y \geq A_z(x)\}$. $\delta > 0$ and m_0 is a countable measurable set w.r.t μ_Λ.

Condition $(S2)$. The boundary curve $\partial(\mathcal{K}_0)$ of $\mathcal{K}_0$ is of class C^2 and has positive curvature everywhere.

Lipschitz boundary basically means that it is locally some sub-graph of a Lipschitz function w.r.t some choice of orthogonal coordinates. Specifically, for each $z \in \partial(\mathcal{K}_0)$, there exists a $r > 0$ and certain rotation $\mathcal{O}$ such that $\mathcal{K}_0 \cap B(z, r)$ is the same as $z + \mathcal{O}(P)$ where $P = \{(x, y) \in B(z, r) : y \geq A_z(x)\}$ for some Lipschitz function A_z. Condition (PL) describes piecewise Lipschitz smooth curves with a countable set of discontinuities in the first and second derivatives (nowhere dense). In condition (S2), let $r_{\mathcal{K}_0}(\eta)$ denote the radius curvature (reciprocal curvature) of $\partial\mathcal{K}_0$ at point $x(\eta)$ with the angle $\eta \in [0, 2\pi)$ which is characterized by the outward unit normal vector from the positive direction of horizontal axis. For convenience, we extend Θ and the its density f periodically by $f(\theta + \pi) = f(\theta)$ and $\Lambda(\supset \mathcal{K}_0)$ is sufficiently large.

SCHNEIDER Schneider (1988) shows the convergence of random inscribed polygons to the compact convex set under (S2). Basically, random parallel slabs have the same behavior under this

condition. Thus, some of the theorems and proofs are useful to our application. Before the main theorem, we introduce a lemma which is presented in Schneider (1988) and Drobot et al. (1982).

Lemma 4.3.5. *Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables, uniformly distributed in $[0,1]$. For each n, let $Y_1(n), Y_2(n), \ldots, Y_n(n)$ be the non-decreasing order statistics of $Y_1, Y_2, \ldots, Y_n$. Define $U_j(n) = Y_{j+i}(n) - Y_j(n)$ $j = 1, 2, \ldots, n$, where $Y_{n+1}(n) = 1 + Y_1(n)$. Let g be a continuous real valued function on $[0,1]$ and let $p > 1$. Then with probability 1*

$$\lim_{n\to\infty} n^{p-1} \sum_{j=1}^{n} g(Y_j(n))(U_j(n))^p = \Gamma(p+1) \int_0^1 g(x)dx.$$

Note that $Y_{n+1}(n) = 1 + Y_1(n)$ is defined since we are dealing with closed curves.

Theorem 4.3.6. *Suppose the sequence of i.i.d. random points $\{\Theta_i\}_{i=1}^{\infty}$ in statements 1-3 is distributed according to uniform distribution and f in statement 4 is positive continuous on $[-\frac{\pi}{2}, \frac{\pi}{2}]$,*

1. $d_N\left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0\right) \to 0$ *a.s. in general;*
2. $d_N\left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0\right) = o(1/n^{\beta}), \beta < 2$ *a.s. under (PL);*
3. $d_N\left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0\right) = o_p((\log(n))^{\gamma}/n^2), \gamma \geq 3$ *and* $d_N(\bigcap_{i=1}^{n} \Lambda_{\Theta_i} = o((\log(n))^{\gamma}/n^2), \gamma > 3$ *a.s. under (PL);*
4. $\lim_{n\to\infty} n^2 d_N\left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0\right) = \frac{1}{4} \int_{(-\frac{\pi}{2}, \frac{\pi}{2})} \frac{r^2_{\mathcal{K}_0}(\theta + \frac{\pi}{2}) + r^2_{\mathcal{K}_0}(\theta + \frac{3\pi}{2})}{f^2(\theta)} d\theta$ *a.s. under (S2).*

Proof. See Chapter 5.2.8. □

This theorem shows that the intersection always converges to the convex support in any general boundary condition. Under condition (PL) the convergence rate is exactly slower than $o(1/n^2)$ and it can be rewritten as the form $o((\log(n))^3/n^2)$. The parameters β and γ depend on the geometry and the set of discontinuities. The consequence rate reaches $o(1/n^2)$ when the boundary is of class C^2 (S2).

Chapter 5

Technical Details

5.1 Proofs of Lemmas and Theorems in Chapter 3

5.1.1 Proof of Theorem 3.1.1

In this proof, we use the definition of differential entropy,

$$h = -\int_S f \log(f) d\mu,$$

for a density f and the corresponding support is $S = supp(f)$ as the closure of the set of points in the domain of f where f is positive. Note that the relative entropy defined as

$$D(f||g) := \int_S f \log\left(\frac{f}{g}\right) d\mu,$$

is always non-negative where f, g are densities of continuous distributions ($\ll \mu$) and $S = supp(f) \subset supp(g)$. Let u_ℓ denote the uniform ansatz on the generalized contour $\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)$, and according to the definition of the uniform ansatz,

$$u_\ell = \frac{1}{\mu_\ell(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell))},$$

for $\ell \in \mathcal{L}_{\boldsymbol{x}}$.

Lemma 5.1.1. *The absolute continuity of P_Λ with respect to μ_Λ induces the absolute continuity of $P_{\mathcal{D}_{\boldsymbol{x}}}$ with respect to $\mu_{\mathcal{D}_{\boldsymbol{x}}}$. In addition, the absolute continuity of $P_{\mathcal{D}_{\boldsymbol{x}}}$ with respect to $\mu_{\mathcal{D}_{\boldsymbol{x}}}$ induces the absolute continuity of $P_{\mathcal{L}_{\boldsymbol{x}}}$ with respect to $\mu_{\mathcal{L}_{\boldsymbol{x}}}$.*

Proof. This result has been proved by Butler et al. (2014). □

$P_{\mathcal{L}_{\boldsymbol{x}}}$ is absolutely continuous to with respect to $\mu_{\mathcal{L}_{\boldsymbol{x}}}$, and it has a Radon-Nikodym density $\rho_{\mathcal{L}_{\boldsymbol{x}}} = dP_{\mathcal{L}_{\boldsymbol{x}}}/d\mu_{\mathcal{L}_{\boldsymbol{x}}}$. Since each P_ℓ^c in $\mathcal{P}_{\boldsymbol{x}}^c$ is absolutely continuous with respect to μ_ℓ, it has a Radon-Nikodym density $\rho_\ell^c = dP_\ell^c/d\mu_\ell$. Then the inverse distribution can be calculated as

$$\int_{\ell\in\mathcal{E}_A}\int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\cap A}\rho_\ell^c(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)d\mu_\ell(\lambda)d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = P_{\Lambda|\boldsymbol{x}}^c(A),\quad A\in\mathcal{B}_\Lambda, \tag{5.1}$$

using the densities, and we only need to show the following

$$\mathcal{P}_{\boldsymbol{x}}^u = \underset{\mathcal{P}_{\boldsymbol{x}}\in\mathcal{P}_{\boldsymbol{x}}^c}{\arg\max} -\int_{\ell\in\mathcal{E}_\Lambda}\int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)}\frac{dP_\ell^c}{d\mu_\ell}(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log\left(\frac{dP_\ell^c}{d\mu_\ell}(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\right)d\mu_\ell(\lambda)d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell).$$

By incorporating the non-negative property of the relative entropy and the disintegration in (5.1), we have

$$\begin{aligned}
D(\rho_\ell\rho_{\mathcal{L}_{\boldsymbol{x}}}||u_\ell\rho_{\mathcal{L}_{\boldsymbol{x}}}) &= \int_{\mathcal{E}_\Lambda}\int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)}\rho_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(l)\log\left(\frac{\rho_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)}{u_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)}\right)d\mu_\ell(\lambda)d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell),\\
&= \int_{\mathcal{E}_\Lambda}\int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)}\rho_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log(\rho_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell))d\mu_\ell(\lambda)d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\\
&\qquad -\int_{\mathcal{E}_\Lambda}\int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)}\rho_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log(u_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell))d\mu_\ell(\lambda)d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell).
\end{aligned}$$

Let

$$h := -\int_{\mathcal{E}_\Lambda}\int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)}\rho_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\log(\rho_\ell(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell))d\mu_\ell(\lambda)d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell).$$

Then h is the entropy of the inverse distribution given any continuous distribution in the disintegration.

Let

$$
\begin{aligned}
h_0 := -\int_{\mathcal{E}_\Lambda} \left(\int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} \rho_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log(u_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell)) d\mu_\ell(\lambda) \right) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\
= -\int_{\mathcal{E}_\Lambda} \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} \rho_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log u_\ell(\lambda) d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \\
- \int_{\mathcal{E}_\Lambda} \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} \rho_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\
= -\int_{\mathcal{E}_\Lambda} \left(\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log \left(\frac{1}{\mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell))} \right) + \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \right) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell).
\end{aligned}
$$

The last equation is proved by the fact that $\mathcal{E}_\Lambda = \mathcal{L}_{\boldsymbol{x}}$.

We show that h_0 is the entropy of the MEID, i.e. the inverse distribution under the uniform ansatz, denoted by h_u in the following.

$$
\begin{aligned}
h_u = -\int_{\mathcal{E}_\Lambda} \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} u_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log \big(u_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \big) d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\
= -\int_{\mathcal{E}_\Lambda} \left(\int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)} u_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log u_\ell(\lambda) d\mu_\ell(\lambda) + u_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) d\mu_\ell(\lambda) \right) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\
= -\int_{\mathcal{E}_\Lambda} \left(\rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log \frac{1}{\mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell))} + \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \log \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \right) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = h_0.
\end{aligned}
$$

Thus, $D(\rho_\ell \rho_{\mathcal{L}_{\boldsymbol{x}}} || u_\ell \rho_{\mathcal{L}_{\boldsymbol{x}}}) = -h + h_u \geq 0$ implies $h_u \geq h$ for any given continuous distributions as the ansatz in the inverse distribution.

5.1.2 Proof of Theorem 3.2.1

We first investigate the absolute continuities among measures in the following equations of disintegration

$$
\begin{aligned}
P_\Lambda(A) &= \int_{\ell \in \mathcal{E}_A} P_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda, \\
\mu_\Lambda(A) &= \int_{\ell \in \mathcal{E}_A} \mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda.
\end{aligned}
$$

Lemma 5.1.2. *The absolute continuity of $P_{\mathcal{L}_{\boldsymbol{x}}}$ with respect to $\mu_{\mathcal{L}_{\boldsymbol{x}}}$ and the absolute continuity of P_ℓ with respect to μ_ℓ for $\mu_{\mathcal{L}_{\boldsymbol{x}}}$-almost everywhere imply the absolute continuity of P_Λ with respect to μ_Λ.*

Proof. If $\mu_\Lambda(A) = 0$ for $A \in \mathcal{B}_\Lambda$, then either $\mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A) = 0$ for $\mu_{\mathcal{L}_{\boldsymbol{x}}}$-almost everywhere or $\mu_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = 0$, which can be shown by contradiction. Since P_ℓ and $P_{\mathcal{L}_{\boldsymbol{x}}}$ are absolutely continuous, $P_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A) = 0$ for $\mu_{\mathcal{L}_{\boldsymbol{x}}}$-almost everywhere or $P_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{E}_A) = 0$. In either case, this implies $P_\Lambda(A) = 0$. □

Proof of Theorem 3.2.1. Since $P_1, P_2 \in \mathcal{P}_{\Lambda,\mathcal{X}}$, for any $\boldsymbol{x} \in \mathcal{X}$ and $A \in \mathcal{B}_\Lambda$ we have

$$P_i(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A} dP_\ell^i(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell),$$

where conditional probability measures $P_\ell^i, i = 1, 2$ exist $P_{\mathcal{L}_{\boldsymbol{x}}}$-almost everywhere. Any $\ell \to P_\ell^i$ is a measurable function in the sense that $\ell \to P_\ell^i(B)$ is a measurable function for each Borel-measurable set B in the Borel sigma algebra of $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)$, see Chang and Pollard (1997) for details. This implies $\ell \to (\omega P_\ell^1 + (1-\omega)P_\ell^2)$ is a measurable function for $\omega \geq 0$. Note that $P_{\mathcal{L}_{\boldsymbol{x}}}$ is absolutely continuous with respect to $\mu_{\mathcal{L}_{\boldsymbol{x}}}$ since $P_{\mathcal{D}_{\boldsymbol{x}}}$ is absolutely continuous with respect to $\mu_{\mathcal{D}_{\boldsymbol{x}}}$, see Butler et al. (2014) Theorem 4.2.

Thus, for any $A \in \mathcal{B}_\Lambda$ and $\omega \geq 0$,

$$\begin{aligned} P_{Mix}(A) &= \int_{\ell \in \mathcal{E}_A} (\omega \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A} dP_\ell^1(\lambda) + (1-\omega) \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A} dP_\ell^2(\lambda)) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\ &= \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap A} d(\omega P_\ell^1 + (1-\omega)P_\ell^2)(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell). \end{aligned}$$

Then by Lemma 5.1.2, P_{Mix} is absolutely continuous with respect to μ_Λ since $\omega P_\ell^1 + (1-\omega)P_\ell^2$ is absolutely continuous with respect to μ_ℓ for $\ell \in \mathcal{L}_{\boldsymbol{x}}$, and $P_{\mathcal{L}_{\boldsymbol{x}}}$ is absolutely continuous with respect to $\mu_{\mathcal{L}_{\boldsymbol{x}}}$ for any $\boldsymbol{x} \in \mathcal{X}$. Then, the result follows the fact that $\omega P_\ell^1 + (1-\omega)P_\ell^2$ is a probability measure when $P_\ell^i, i = 1, 2$ are probability measures defined on $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)$. □

5.1.3 Proof of Theorem 3.3.1

The first result is implied by the fact that $Q_{\boldsymbol{x}}^{-1}(D)$ contains the support of the probability measure $P_{\Lambda,\boldsymbol{x}}$ for each $\boldsymbol{x} \in \mathcal{X}$ for each $\boldsymbol{x} \in \mathcal{X}$, which is defined as $supp(P_{\Lambda,\boldsymbol{x}}) := \{\lambda \in \Lambda :$ for $\lambda \in N_\lambda \in \mathcal{B}_\Lambda$, we have $P_{\Lambda,\boldsymbol{x}}(N_\lambda) > 0\}$ where N_λ is any open neighborhood of λ. Thus, $P_{\Lambda,\boldsymbol{x}}(A \cap Q_{\boldsymbol{x}}^{-1}(D)) = P_{\Lambda,\boldsymbol{x}}(A \cap supp(P_{\Lambda,\boldsymbol{x}})) = P_{\Lambda,\boldsymbol{x}}(A)$ for $A \in \mathcal{B}_\Lambda$ and $\boldsymbol{x} \in \mathcal{X}$.

Then we show the second result in the following. We observe that $\bar{P}(\emptyset) = 0$ and $\bar{P}(\Lambda) = 1$ and $\bar{P}(A) \in [0,1]$ for any $A \in \mathcal{B}_\Lambda$, since $P_{\Lambda,\boldsymbol{x}}$ and $P_{\mathcal{X}}$ are probability measures on $(\Lambda, \mathcal{B}_\Lambda)$ and $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, respectively.

For all countable collections $\{A_i\}_{i=1}^{\infty}$ of pairwise disjoint sets,

$$\begin{aligned}
\bar{P}\left(\bigcup_i A_i\right) &= \int_{\boldsymbol{x}\in\mathcal{X}} P_{\Lambda,\boldsymbol{x}}\left(\bigcup_i A_i\right) dP_{\mathcal{X}}(\boldsymbol{x}), \\
&= \sum_i \int_{\boldsymbol{x}\in\mathcal{X}} P_{\Lambda,\boldsymbol{x}}(A_i) dP_{\mathcal{X}}(\boldsymbol{x}), \\
&= \sum_i \bar{P}(A_i),
\end{aligned}$$

since each $P_{\Lambda,\boldsymbol{x}}(A_i)$ is integrable and finite with respect to $\boldsymbol{x}$. This shows that $\bar{P}$ is a probability measure.

5.1.4 Proof of Theorem 3.3.2

Considering the probability distribution of $\boldsymbol{a}$, we have

$$\widetilde{Bel}_C(A) = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in Q_{\boldsymbol{x}}^{-1}(C)) I_{\{Q_{\boldsymbol{x}}^{-1}(C)\subset A\}}(\boldsymbol{x}) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in Q_{\boldsymbol{x}}^{-1}(C)) I_{\{Q_{\boldsymbol{x}}^{-1}(C)\neq\emptyset\}}(\boldsymbol{x}) dP_{\mathcal{X}}},$$

where P is a probability measure of $\boldsymbol{a}$ on $(\Lambda, \mathcal{B}_\Lambda)$ that can be disintegrated along generalized contours and $\widetilde{Bel}_C \in [0,1]$. We choose P to be any choice in the equivalence class of GFGDs or any absolutely continuous inverse distributions in $\mathbb{P}_{\Lambda,\boldsymbol{x}}$ for each $\boldsymbol{x} \in \mathcal{X}$. Note that $P(\boldsymbol{a} \in Q_{\boldsymbol{x}}^{-1}(C)) I_{\{Q_{\boldsymbol{x}}^{-1}(C)\neq\emptyset\}}(\boldsymbol{x}) = P(\boldsymbol{a} \in Q_{\boldsymbol{x}}^{-1}(C))$ as a measurable function of $\boldsymbol{x}$. Then, by the disinte-

gration of P as

$$P(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda,$$

we have

$$\begin{aligned}
\widetilde{Bel}_C(A) &= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)}} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} I_{\{Q^{-1}_{\boldsymbol{x}}(C) \subset A\}}(\boldsymbol{x}) dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)}} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}, \\
&= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)}} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} I_{\{Q^{-1}_{\boldsymbol{x}}(C) \subset A\}}(\boldsymbol{x}) dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{y \in C} dP_{\mathcal{D}_{\boldsymbol{x}}}(y) dP_{\mathcal{X}}},
\end{aligned}$$

since $P(Q^{-1}_{\boldsymbol{x}}(C)) = P_{\mathcal{D}_{\boldsymbol{x}}}(C)$. Similarly, we have

$$\begin{aligned}
\widetilde{Pl}_C(A) &= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in Q^{-1}_{\boldsymbol{x}}(C)) I_{\{Q^{-1}_{\boldsymbol{x}}(C) \cap A\}}(\boldsymbol{x}) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in Q^{-1}_{\boldsymbol{x}}(C)) I_{\{Q^{-1}_{\boldsymbol{x}}(C) \neq \emptyset\}}(\boldsymbol{x}) dP_{\mathcal{X}}}, \\
&= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)\cap A}} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} I_{\{Q^{-1}_{\boldsymbol{x}}(C) \cap A \neq \emptyset\}}(\boldsymbol{x}) dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)}} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)} I_{\{Q^{-1}_{\boldsymbol{x}}(C) \cap \Lambda \neq \emptyset\}}(\boldsymbol{x}) dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}, \\
&= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell \in \mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)\cap A}} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} I_{\{Q^{-1}_{\boldsymbol{x}}(C) \cap A \neq \emptyset\}}(\boldsymbol{x}) dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{y \in C} dP_{\mathcal{D}_{\boldsymbol{x}}}(y) dP_{\mathcal{X}}}.
\end{aligned}$$

Note that $I_{\{Q^{-1}_{\boldsymbol{x}}(C) \cap A \neq \emptyset\}}(\boldsymbol{x})$ is ignorable since $\mathcal{E}_{Q^{-1}_{\boldsymbol{x}}(C)\cap A}$ is an empty set when $Q^{-1}_{\boldsymbol{x}}(C) \cap A = \emptyset$. Then, the result follows by replacing $\{P_\ell\}$ with the uniform ansatz $\{P^u_\ell\}$ in the MEIDs.

5.1.5 Proof of Theorem 3.3.3

When $\boldsymbol{a}$ and $\boldsymbol{X}$ are considered independent in the SIP model, the SIP posterior by using the Bayes' theorem is computed as follows

$$\begin{aligned}
P(\boldsymbol{a} \in A | y \in C) &= \frac{P(\boldsymbol{a} \in A, y \in C)}{P(y \in C)} = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in A, y \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P(y \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}} \\
&= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in A, Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P(Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}} = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in A, Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P(Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C) dP_{\mathcal{X}}}.
\end{aligned}$$

The event $\{\boldsymbol{a} \in A, Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C\}$ is equivalent to the intersection of A and the contour event $Q^{-1}_{\boldsymbol{x}}(C)$, i.e. $\{\boldsymbol{a} \in A \cap Q^{-1}_{\boldsymbol{x}}(C)\}$. Now, the result depends on the choice of the probability distri-

bution of $\boldsymbol{a}$. In this case, we choose the generating distribution P_Λ such that $Q_{\boldsymbol{x}}(\boldsymbol{a})$ is distributed according to $P_{\mathcal{D}_{\boldsymbol{x}}}$ induced through

$$P_{\mathcal{D}_{\boldsymbol{x}}}(C) = P_\Lambda(Q_{\boldsymbol{x}}^{-1}(C)),$$

where $C \in \mathcal{D}_{\boldsymbol{x}}$. Thus,

$$P(\boldsymbol{a} \in A | y \in C) = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P_\Lambda(\boldsymbol{a} \in A \cap Q_{\boldsymbol{x}}^{-1}(C)) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P_{\mathcal{D}_{\boldsymbol{x}}}(C) dP_{\mathcal{X}}}.$$

By the disintegration theorem in (3.3), $P_\Lambda(\boldsymbol{a} \in A \cap Q_{\boldsymbol{x}}^{-1}(C))$ can be further disintegrated, and thus we have

$$P(\boldsymbol{a} \in A | y \in C) = \frac{\int_{\boldsymbol{x}\in\mathcal{X}} \int_{\ell\in\mathcal{E}_{A\cap Q_{\boldsymbol{x}}^{-1}(C)}} \int_{\lambda\in\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell)\cap A} dP_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P_{\mathcal{D}_{\boldsymbol{x}}}(C) dP_{\mathcal{X}}}.$$

Since P_Λ is unknown, we can replace the conditional probability measures $\{P_\ell\}$ by the uniform ansatz $\mathcal{P}_{\boldsymbol{x}}^u$ in practice for the approximation given the measurability of the map $\boldsymbol{x} \to P_{\Lambda,\boldsymbol{x}}(A)$ for any $A \in \mathcal{B}_\Lambda$.

The result also applies for the case in which $\boldsymbol{a}$ and $\boldsymbol{X}$ are dependent. In this case,

$$\begin{aligned} P(\boldsymbol{a} \in A | y \in C) &= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in A, y \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P(y \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}} \\ &= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in A, Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P(Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}} \\ &= \frac{\int_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{a} \in A, Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C | \boldsymbol{X} = \boldsymbol{x}) dP_{\mathcal{X}}}{\int_{\boldsymbol{x}\in\mathcal{X}} P_{\mathcal{D}_{\boldsymbol{x}}}(C) dP_{\mathcal{X}}}, \end{aligned}$$

where $P(\boldsymbol{a} \in A, Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C | \boldsymbol{X} = \boldsymbol{x})$ is the conditional probability following the argument of (Kolmogorov and Bharucha-Reid, 2018). In fact, the output is generated by the conditional probabilities in the SFP. To obtain the result, we further disintegrate $P(\boldsymbol{a} \in A, Q_{\boldsymbol{x}}(\boldsymbol{a}) \in C | \boldsymbol{X} = \boldsymbol{x})$ and replace $\{P_\ell\}$ with the uniform ansatz.

5.1.6 Proof of Theorem 3.4.1

We start with the following definitions and lemmas for the domain Λ and the probability measures on $(\Lambda, \mathcal{B}_\Lambda)$.

Definition 5.1.6.1. *A topological space is separable if it contains a countable dense subset.*

Definition 5.1.6.2. *A complete metric space is a metric space in which every Cauchy sequence is convergent.*

Definition 5.1.6.3. *A metric space (Λ, d) is totally bounded if and only if for every real number $\epsilon > 0$, there exists a finite collection of open ϵ-balls in Λ whose union contains Λ.*

Definition 5.1.6.4. *A family of probability measures $\mathbb{P}$ on $(\Lambda, \mathcal{B}_\Lambda)$ is tight if for every real number $\epsilon > 0$ there exists a compact set $K \subset \Lambda$ such that $P(K) > 1 - \epsilon$ for all $P \in \mathbb{P}$.*

Lemma 5.1.3. *Any countable collection of probability measures on $(\Lambda, \mathcal{B}_\Lambda)$ is tight.*

Proof. Let $\mathbb{P} = \{P_i\}_{i=1}^{\infty}$ be any countable collection of probability measures on $(\Lambda, \mathcal{B}_\Lambda)$. Since Λ is a compact metric space, it is separable and complete. Since Λ is separable, for each k there exists a sequence $\{A_{kj}\}_{j=1}^{\infty}$ of open $1/k$-balls covering Λ. For any real number $\epsilon > 0$, we choose k_i and large enough n_{k_i} such that $k_i \geq k_{i-1}$ and $P_i(B_{k_i}) > 1 - \epsilon 2^{-k_i}$ where $B_{k_i} = A_{k_i 1} \cup \cdots \cup A_{k_i n_{k_i}}$. Let $B = \cap_{i=1}^{\infty} B_{k_i}$. Since B is totally bounded and Λ is complete, the set B has a compact closure K. Thus $P_i(K) \leq \sum_{i=1}^{\infty} P_i(B_{k_i}^c) < \epsilon$ for any $i \geq 1$. □

Lemma 5.1.4 (Theorem 5.1, Billingsley (2013))**.** *If $\mathbb{P}$ is tight, then every sequence of probability measures in $\mathbb{P}$ has a subsequence converging weakly to a probability measure in the closure of $\mathbb{P}$.*

proof of Theorem 3.4.1. The result follows by the fact that each $\bar{P}_i$ for $i \geq 0$ is a probability measure on $(\Lambda, \mathcal{B}_\Lambda)$, and by Lemma 5.1.3 and Lemma 5.1.4. □

5.1.7 Proof of Theorem 3.4.2

By combining the disintegration of μ_Λ as

$$\mu_\Lambda(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda,$$

we have, for each $i \geq 0$ and $A \in \mathcal{B}_\Lambda$,

$$\begin{aligned} P^i_{\Lambda,\boldsymbol{x}}(A) &= \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} \rho^i_\ell(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\ell) d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\ &= \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} \rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) d\mu_\ell(\lambda) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\ &= \int_A \rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) d\mu_\Lambda. \end{aligned}$$

The density $\rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda))$ is unique almost everywhere in Λ. In Butler et al. (2014), the space $\mathcal{L}_{\boldsymbol{x}}$ can characterized as a transverse parameterization in Λ indexing the generalized contours. For instance, a line segment in Λ that crosses all generalized contours once and only once. In this case, the density function can be computed by using the transverse parameterization in Λ, since the map between the transverse parameterization and the range of the output $\mathcal{D}_{\boldsymbol{x}}$ is bijective.

5.1.8 Proof of Theorem 3.4.3

We first compute the density function for each EEI. By Fubini-Tonelli theorem, we have

$$\begin{aligned} \bar{P}^i(A) &= \int_{\boldsymbol{x} \in \mathcal{X}} \int_{\lambda \in A} \rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) \rho_{\boldsymbol{X}}(\boldsymbol{x}) d\mu_\Lambda d\mu_{\boldsymbol{X}}, \\ &= \int_{\lambda \in A} \int_{\boldsymbol{x} \in \mathcal{X}} \rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) \rho_{\boldsymbol{X}}(\boldsymbol{x}) d\mu_{\boldsymbol{X}} d\mu_\Lambda, \end{aligned}$$

for any $A \in \mathcal{B}_\Lambda$ and $i \geq 0$, since each $\rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) \rho_{\boldsymbol{X}}(\boldsymbol{x})$ is a measurable function on the product of σ finite measure spaces. Thus, we have the unique density function

$$\bar{\rho}^i = d\bar{P}^i / d\mu_\Lambda = \int_{\boldsymbol{x} \in \mathcal{X}} \rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)) \rho_{\boldsymbol{X}}(\boldsymbol{x}) d\mu_{\boldsymbol{X}},$$

almost everywhere in Λ for each $i \geq 0$. Consider the complete metric space $(\mathcal{L}^1_D, d)$ where the space $\mathcal{L}^1_D$ contains density functions on Λ and d is the L^1 metric defined as

$$d(f, g) = \int |f - g| d\mu_\Lambda.$$

Then, for any real number $\epsilon > 0$ and $i, j > N$,

$$\begin{aligned} d(\bar{\rho}^i, \bar{\rho}^j) &\leq \int\int_{\boldsymbol{x}\in\mathcal{X}} \left|\rho^i_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda) - \rho^j_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)\right| \rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda))\rho_{\boldsymbol{X}}(\boldsymbol{x})d\mu_{\boldsymbol{X}}d\mu_\Lambda, \\ &< \epsilon \int\int_{\boldsymbol{x}\in\mathcal{X}} \rho^0_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda)}(\lambda)\rho_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\lambda))\rho_{\boldsymbol{X}}(\boldsymbol{x})d\mu_{\boldsymbol{X}}d\mu_\Lambda, \\ &= \epsilon \int \bar{\rho}^0 d\mu_\Lambda = \epsilon. \end{aligned}$$

Thus, the sequence $\{\bar{\rho}^i\}_{i\geq 0}$ is Cauchy in $(\mathcal{L}^1_D, d)$, and $\bar{\rho}^i$ converges to a density function in $(\mathcal{L}^1_D, d)$ as $i \to \infty$.

5.1.9 Proof of Theorem 3.4.4

We start with the proof of the first direction. Since $\widetilde{P} \in \mathbb{P}_{\Lambda,\mathcal{X}}$, it can be disintegrated as

$$\widetilde{P}(A) = \int_{\ell\in\mathcal{E}_A} \int_{\lambda\in\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\cap A} d\widetilde{P}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \quad \forall A \in \mathcal{B}_\Lambda, \tag{5.2}$$

for $\boldsymbol{x} \in \mathcal{X}$. Thus, the inverse distribution $\widetilde{P}_{\Lambda,\boldsymbol{x}}$ using the ansatz $\{\widetilde{P}_\ell\}$ in (5.2) is the exact distribution $\widetilde{P}$. Then, we have

$$\bar{P}(A) = \int_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_{\Lambda,\boldsymbol{x}}(A) dP_{\mathcal{X}} = \int_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}(A) dP_{\mathcal{X}} = \widetilde{P}(A),$$

for any $A \in \mathcal{B}_\Lambda$. This concludes $\widetilde{P} = \mathcal{G}(\widetilde{P})$.

Conversely, for a fixed point $\widetilde{P}$ of $\mathcal{G}$ that is absolutely continuous with respect to μ_Λ, we have

$$\text{Disintegration of a measure: } \widetilde{P}(A) = \int_{\ell\in\mathcal{E}_A} \int_{\lambda\in\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\cap A} d\widetilde{P}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}},\widetilde{P}}(\ell), \tag{5.3}$$

$$\text{Compute inverse distributions: } \int_{\ell\in\mathcal{E}_A} \int_{\lambda\in\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell)\cap A} d\widetilde{P}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) = \widetilde{P}_{\Lambda,\boldsymbol{x}}(A), \tag{5.4}$$

$$\text{Compute the next EEI: } \widetilde{P}(A) = \int_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_{\Lambda,\boldsymbol{x}}(A) dP_{\mathcal{X}}, \tag{5.5}$$

for any $\boldsymbol{x} \in \mathcal{X}$ and $A \in \mathcal{B}_\Lambda$. Equation (5.5) implies $\widetilde{P} = \widetilde{P}_{\Lambda,\boldsymbol{x}}$ since $\widetilde{P}_{\Lambda,\boldsymbol{x}}$ is not conditioned on $\boldsymbol{x}$. Thus, by combining (5.4) and (5.5), we have

$$\widetilde{P}(A) = \int_{\ell \in \mathcal{E}_A} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \cap A} d\widetilde{P}_\ell(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \tag{5.6}$$

for any $A \in \mathcal{B}_\Lambda$, and $\widetilde{P} \in \mathbb{P}_{\Lambda,\mathcal{X}}$ since $\widetilde{P}_\ell$ is absolutely continuous with respect to μ_ℓ for $\ell \in \mathcal{L}_{\boldsymbol{x}}$ and $\boldsymbol{x} \in \mathcal{X}$. Now, we compare the results in equations (5.3) and (5.6). In fact, equation (5.6) implies that $\widetilde{P}$ induces $P_{\mathcal{L}_{\boldsymbol{x}}}$ through

$$P_{\mathcal{L}_{\boldsymbol{x}}}(E) = \widetilde{P}(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(E)),$$

for $E \in \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}}$. Recall that the definition of $P_{\mathcal{L}_{\boldsymbol{x}},\widetilde{P}}$ in (5.3) is defined as

$$\widetilde{P}_{\mathcal{D}_{\boldsymbol{x}}}(C) = \widetilde{P}(Q^{-1}_{\boldsymbol{x}}(C)), \quad C \in \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}},$$
$$P_{\mathcal{L}_{\boldsymbol{x}},\widetilde{P}}(\mathcal{E}_A) = \widetilde{P}_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A)), \quad A \in \mathcal{B}_\Lambda.$$

This is equivalent to the following computation

$$P_{\mathcal{L}_{\boldsymbol{x}},\widetilde{P}}(E) = \widetilde{P}(\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}}}(E)),$$

for any $E \in \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}}$. Thus, $P_{\mathcal{L}_{\boldsymbol{x}},\widetilde{P}} = P_{\mathcal{L}_{\boldsymbol{x}}}$ for $\boldsymbol{x} \in \mathcal{X}$.

5.2 Proofs of Lemmas and Theorems in Chapter 4

5.2.1 Proof of Theorem 4.1.1

Let the domain of $P_{\mathcal{D}_{\boldsymbol{x}}}$ be

$$S_{\boldsymbol{x}} = \{q \in \mathcal{D}_{\boldsymbol{x}} : P_{\mathcal{D}_{\boldsymbol{x}}}(N_q) > 0 \text{ for any open neighborhood } N_q \in \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}} \text{ of } q\}.$$

For $q \in S_{\boldsymbol{x}}$, we have $P_{\mathcal{D}_{\boldsymbol{x}}}(N_q) > 0$ for any open neighborhood $N_q \in \mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}}$. Thus, $P_\Lambda(Q_{\boldsymbol{x}}^{-1}(N_q)) = P_{\mathcal{D}_{\boldsymbol{x}}}(N_q) > 0$. If $q \in (Q_{\boldsymbol{x}}(\mathcal{K}_0))^c$ where $(Q_{\boldsymbol{x}}(\mathcal{K}_0))^c$ is the complement of the closed set $Q_{\boldsymbol{x}}(\mathcal{K}_0)$, then there exists an open neighborhood N_q of q in $\mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}}$ such that $Q_{\boldsymbol{x}}^{-1}(N_q) \cap \mathcal{K}_0 = \emptyset$. This implies $P_\Lambda(Q_{\boldsymbol{x}}^{-1}(N_q)) = P_{\mathcal{D}_{\boldsymbol{x}}}(N_q) = 0$, which is a contradiction. Thus, we have $S_{\boldsymbol{x}} \subset Q_{\boldsymbol{x}}(\mathcal{K}_0)$.

On the other hand, for $q \in Q_{\boldsymbol{x}}(\mathcal{K}_0)$, we have $Q_{\boldsymbol{x}}^{-1}(N_q) \cap \mathcal{K}_0 \neq \emptyset$ for any open neighborhood N_q of q in $\mathcal{B}_{\mathcal{D}_{\boldsymbol{x}}}$. Note that each $Q_{\boldsymbol{x}}^{-1}(N_q)$ is an open set in Λ. Then, by the definition of $\mathcal{K}_0$, there exists an open neighborhood $\tilde{N}_\lambda \in \mathcal{B}_\Lambda$ of a point $\lambda \in Q_{\boldsymbol{x}}^{-1}(N_q) \cap \mathcal{K}_0$ and $\tilde{N}_\lambda \subset Q_{\boldsymbol{x}}^{-1}(N_q)$ such that $P_\Lambda(\tilde{N}_\lambda) > 0$. This implies $P_{\mathcal{D}_{\boldsymbol{x}}}(N_q) = P_\Lambda(Q_{\boldsymbol{x}}^{-1}(N_q)) \geq P_\Lambda(\tilde{N}_\lambda) > 0$. Thus, we conclude $S_{\boldsymbol{x}} \supset Q_{\boldsymbol{x}}(\mathcal{K}_0)$.

5.2.2 Proof of Theorem 4.1.2

Let the support of $P_{\Lambda,\boldsymbol{x}}$ be $\tilde{S}_{\boldsymbol{x}} = \{\lambda \in \Lambda : \lambda \in N_\lambda \in \mathcal{B}_\Lambda \text{ s.t. } P_{\Lambda,\boldsymbol{x}}(N_\lambda) > 0\}$. For $\lambda \in \tilde{S}_{\boldsymbol{x}}$, we have $P_{\Lambda,\boldsymbol{x}}(N_\lambda) > 0$ for any open neighborhood N_λ of λ. However, if $\lambda \in (\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0)))^c$, then there exists an open neighborhood $\tilde{N} \in \mathcal{B}_\Lambda$ of λ such that $\tilde{N} \cap \mathcal{K}_0 = \emptyset$ and $P_{\Lambda,\boldsymbol{x}}(\tilde{N}) = 0$. This contradiction implies that $\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0))$ for $\lambda \in \tilde{S}_{\boldsymbol{x}}$. Note that this proof is independent of the choice of the ansatz, which shows the result of Corollary 4.1.2.1.

Then, for $\lambda \in \pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0))$ and any open neighborhood $N_\lambda \in \mathcal{B}_\Lambda$, we have $\pi_{\mathcal{L}_{\boldsymbol{x}}}(N_\lambda) \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0) \neq \emptyset$. Take $\tilde{\ell} \in \pi_{\mathcal{L}_{\boldsymbol{x}}}(N_\lambda) \cap \pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0)$. There exists an open neighborhood $N_{\tilde{\ell}} \in \mathcal{B}_{\mathcal{L}_{\boldsymbol{x}}}$ such that $N_{\tilde{\ell}} \subset \pi_{\mathcal{L}_{\boldsymbol{x}}}(N_\lambda)$, and $P_{\mathcal{L}_{\boldsymbol{x}}}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(N_{\tilde{\ell}})) > 0$ by the definition of $\pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0)$. Since $\mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap S) > 0$ in a $\mu_{\mathcal{L}_{\boldsymbol{x}}}$-almost everywhere sense where $S \in \mathcal{B}_\Lambda$ is any open subset of Λ, we have $\pi_{\mathcal{L}_{\boldsymbol{x}}}(N_{\tilde{\ell}}) \supset \cup_{n \geq 1} E_n$ where $E_n = \{\ell \in \pi_{\mathcal{L}_{\boldsymbol{x}}}(N_{\tilde{\ell}}) : \mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap N_\lambda) > 0\}$. Then there exists some n for which $P_{\mathcal{L}_{\boldsymbol{x}}}(E_n) > 0$. Hence,

$$P_{\Lambda,\boldsymbol{x}}(N_\lambda) \geq \int_{\pi_{\mathcal{L}_{\boldsymbol{x}}}(N_{\tilde{\ell}})} \mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap N_\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) \geq \int_{E_n} \mu_\ell(\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\ell) \cap N_\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}}(\ell) > \frac{P_{\mathcal{L}_{\boldsymbol{x}}}(E_n)}{n} > 0.$$

Consequently, $\lambda \in \tilde{S}_{\boldsymbol{x}}$. Then, $\pi_{\mathcal{L}_{\boldsymbol{x}}}^{-1}(\pi_{\mathcal{L}_{\boldsymbol{x}}}(\mathcal{K}_0)) = Q_{\boldsymbol{x}}^{-1}(Q_{\boldsymbol{x}}(\mathcal{K}_0))$ is true simply because the map between $\mathcal{L}_{\boldsymbol{x}}$ and $\mathcal{D}_{\boldsymbol{x}}$ is one-to-one and onto.

5.2.3 Proof of Theorem 4.2.2

We conclude $\mathcal{K}_{\mathcal{X}} \in \mathbb{B}_{\mathcal{X}}^{-1}$ from $Q_{\boldsymbol{x}}(\mathcal{K}_0) \subset Q_{\boldsymbol{x}}(\mathcal{K}_{\mathcal{X}}) \subset Q_{\boldsymbol{x}}(\Lambda_{\boldsymbol{x}}) = B_{\boldsymbol{x}}$ for any $\boldsymbol{x} \in \mathcal{X}$. Since $C \subset \Lambda_{\boldsymbol{x}}$ for any $C \in \mathbb{B}_{\mathcal{X}}^{-1}$ and $\boldsymbol{x} \in \mathcal{X}$, we conclude $\mathcal{K}_{\mathcal{X}}$ is the maximal element in $\mathbb{B}_{\mathcal{X}}^{-1}$.

5.2.4 Proof of Theorem 4.3.1

Because $\mathcal{K}_0 \subset \Lambda$ is compact, there exists $c_1, c_2 \in \mathbb{R}^1$ such that $\Lambda_{\boldsymbol{x}} = \{\lambda \in \mathbb{R}^d : Q_{\boldsymbol{x}}(\lambda) \in Q_{\boldsymbol{x}}(\mathcal{K}_0) = [c_1, c_2]\} \cap \Lambda = S_{c_1,c_2} \cap \Lambda$. S_{c_1,c_2} is a parallel slab when $Q_{\boldsymbol{x}}$ is linear since

$$S_{c_1,c_2} = \bigcup_{c \in [c_1,c_2]} \{\lambda \in \mathbb{R}^d : Q_{\boldsymbol{x}}(\lambda) = c\}$$

is a union of a family of hyperplanes in $\mathbb{R}^d$. The half space $S_1 = \{\lambda \in \mathbb{R}^d : Q_{\boldsymbol{x}}(\lambda) \geq c_1\}$ contains the support $\mathcal{K}_0$ and $S_1 \cap \mathcal{K}_0 \neq \emptyset$, and same for $S_2 = \{\lambda \in \mathbb{R}^d : Q_{\boldsymbol{x}}(\lambda) \leq c_2\}$. This coincides with the supporting hyperplanes of the support $\mathcal{K}_0$, namely, $H_1 := \partial S_1 = \{\lambda \in \mathbb{R}^d : Q_{\boldsymbol{x}}(\lambda) = c_1\}$ and $H_2 := \partial S_2 = \{\lambda \in \mathbb{R}^d : Q_{\boldsymbol{x}}(\lambda) = c_2\}$.

5.2.5 Proof of Theorem 4.3.2

By the disintegration theorem, it is equivalent to show

$$\int_{\pi_{\mathcal{L}_{\boldsymbol{x}},\Lambda_n}(A)} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}},\Lambda_n}(\ell) \cap A} dP_{\ell|\Lambda_n}(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}|\Lambda_n}(\ell) \rightarrow \int_{\pi_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(A)} \int_{\lambda \in \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(\ell) \cap A} dP_{\ell|\mathcal{K}_{conv}}(\lambda) dP_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}(\ell),$$

where $P_{\ell|\Lambda_n}, P_{\mathcal{L}_{\boldsymbol{x}}|\Lambda_n}$ are the distributions computed by using the domain Λ_n and $P_{\ell|\mathcal{K}_{conv}}, P_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}$ are the distributions computed by using the domain $\mathcal{K}_{conv}$. Note that $\pi_{\mathcal{L}_{\boldsymbol{x}},\Lambda_n} : \Lambda_n \rightarrow \mathcal{L}_{\boldsymbol{x}}$ is the equivalence map on domain Λ_n, $\pi_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}} : \mathcal{K}_{conv} \rightarrow \mathcal{L}_{\boldsymbol{x}}$ is the equivalence map on domain $\mathcal{K}_{conv}$, and $A \in \mathcal{B}_{\mathcal{K}_{conv}}$ implies that $\pi_{\mathcal{L}_{\boldsymbol{x}},\Lambda_n}(A) = \pi_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(A)$ and $\pi^{-1}_{\mathcal{L}_{\boldsymbol{x}},\Lambda_n}(\ell) \cap A = \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(\ell) \cap A$ for $\ell \in \pi_{\mathcal{L}_{\boldsymbol{x}},\Lambda_n}(A)$. Also note that $P_{\mathcal{L}_{\boldsymbol{x}}|\Lambda_n}(\pi_{\mathcal{L}_{\boldsymbol{x}},\Lambda_n}(A)) = P_{\mathcal{D}_{\boldsymbol{x}}}(Q_{\boldsymbol{x}}(A))$ for $A \in \mathcal{B}_{\mathcal{K}_{conv}}$, which implies $P_{\mathcal{L}_{\boldsymbol{x}}|\Lambda_n} = P_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}$ for $A \in \mathcal{B}_{\mathcal{K}_{conv}}$. Then it is sufficient to show that the following converges to

zero,

$$Res = \int_{\pi_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(A)} (P_{\ell|\Lambda_n}(R_\ell) - P_{\ell|\mathcal{K}_{conv}}(R_\ell)) dP_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}(\ell),$$

where $R_\ell = \pi^{-1}_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(\ell) \cap A$ and $P_{\ell|\cdot}$ is uniform distribution defined on $(\cdot)$ almost everywhere except on some zero-measure boundary points on $(\cdot)$ by disintegration theorem, namely, $P_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}$-null set I_0. Then,

$$\begin{aligned} Res &= \int_{\pi_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(A)(\backslash I_0)} (P_{\ell|\Lambda_n}(R_\ell) - P_{\ell|\mathcal{K}_{conv}}(R_\ell)) \frac{dP_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}}{d\mu_{\mathcal{L}_{\boldsymbol{x}}}}(\ell) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \\ &= \int_{\pi_{\mathcal{L}_{\boldsymbol{x}},\mathcal{K}_{conv}}(A)} (f_n(\ell) - f(\ell)) d\mu_{\mathcal{L}_{\boldsymbol{x}}}(\ell), \end{aligned}$$

where $f_n(\ell) = P_{\ell|\Lambda_n}(R_\ell) \frac{dP_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}}{d\mu_{\mathcal{L}_{\boldsymbol{x}}}}(\ell)$ and $f(\ell) = P_{\ell|\mathcal{K}_{conv}}(R_\ell) \frac{dP_{\mathcal{L}_{\boldsymbol{x}}|\mathcal{K}_{conv}}}{d\mu_{\mathcal{L}_{\boldsymbol{x}}}}(\ell)$. Note that $f_n \leq f$ and $f_n \to f$ pointwisely with f integrable. Then, by the dominated convergence theorem, $P_{\Lambda_n,\boldsymbol{x}}(A) \to P_{\mathcal{K}_{conv},\boldsymbol{x}}(A)$.

5.2.6 Proof of Theorem 4.3.3

For any $A \in \mathcal{B}_\Lambda$, we have

$$\begin{aligned} \bar{P}(A;\Lambda_n) - \bar{P}(A;\mathcal{K}_{conv}) &= \int (P_{\Lambda_n,\boldsymbol{x}}(A) - P_{\mathcal{K}_{conv},\boldsymbol{x}}(A)) f_{\mathcal{X}}(\boldsymbol{x}) d\mu_{\mathcal{X}}(\boldsymbol{x}), \\ &= \int (g_n(\boldsymbol{x}) - g(\boldsymbol{x})) d\mu_{\mathcal{X}}(\boldsymbol{x}), \end{aligned}$$

where $g_n(\boldsymbol{x}) = P_{\Lambda_n,\boldsymbol{x}}(A) f_{\mathcal{X}}(\boldsymbol{x})$ and $g(\boldsymbol{x}) = P_{\mathcal{K}_{conv},\boldsymbol{x}}(A) f_{\mathcal{X}}(\boldsymbol{x})$. We know that $g_n \leq g$ and $g_n \to g$ pointwisely where g is integrable by Theorem 4.3.2. Thus, the result is implied by the dominated convergence theorem.

5.2.7 Proof of Theorem 4.3.4

We know that $\bigcap_{i=1}^{\infty} \Lambda_{\boldsymbol{x}_i} \supset \mathcal{K}_{conv}$ since each Λ_θ is a generalized parallel slab and is convex.

If there exists a point $p \in \bigcap_{i=1}^{\infty} \Lambda_{\boldsymbol{x}_i}$ such that $p \notin \mathcal{K}_{conv}$, the hyperplane separation theorem implies that p can be separated from $\mathcal{K}_{conv}$ by a half-space. Let the half-space correspond to $\boldsymbol{x}$.

If $\boldsymbol{x} \in \{\boldsymbol{x}_i\}_{i=1}^{\infty}$, then $\bigcap_{i=1}^{\infty} \Lambda_{\boldsymbol{x}_i} \subset \mathcal{K}_{conv}$. Otherwise, the Euclidean distance of point p and $\mathcal{K}_{conv}$ is not zero and there exists a small neighborhood of p that is not contained in $\mathcal{K}_{conv}$ because $\mathbb{R}^d$ is a metric space that satisfies separation axioms in topology. Thus, there exists some $\boldsymbol{x}_i$ in the dense subset that is arbitrarily close to $\boldsymbol{x}$ such that $\Lambda_{\boldsymbol{x}_i}$ separates them and $\bigcap_{i=1}^{\infty} \Lambda_{\boldsymbol{x}_i} \subset \mathcal{K}_{conv}$.

5.2.8 Proof of Theorem 4.3.6

As for the first statement, it is easy to see that $\mathcal{K}_0 \subset \bigcap_{i=1}^{n} \Lambda_{\Theta_i}$. For each $\lambda \in \bigcap_{i=1}^{n} \Lambda_{\Theta_i}$ given $\Theta_1, \Theta_2, \ldots, \Theta_n$, $Q_{\Theta_i}(\lambda) \in Q_{\Theta_i}(\mathcal{K}_0)$ holds for $i = 1, \ldots, n$ by the construction of the support of inverse measure. Therefore, we rewrite it as $\bigcap_{i=1}^{n} \Lambda_{\Theta_i} \triangle \mathcal{K}_0 = \bigcap_{i=1}^{n} \Lambda_{\Theta_i} \backslash \mathcal{K}_0$. Then let $\Theta_i : (\Omega, \mathcal{F}, P) \to (I, \mathcal{B}_I, P_\Theta)$ and $Z_n = \mu_\Lambda(\bigcap_{i=1}^{n} \Lambda_{\Theta_i} \triangle \mathcal{K}_0)$, which is a monotone sequence. Let D_n be the length (Lebesgue measure) of the largest interval from $I = (-\frac{\pi}{2}, \frac{\pi}{2})$ not containing any of the points when first n choices. Then we can show that

$$\{\omega \in \Omega : \ D_n(\omega) \to 0 \ as \ n \to \infty\} \subset \{\omega \in \Omega : \ Z_n(\omega) \to 0 \ as \ n \to \infty\}$$

by the geometry and hyperplane separation theorem. This is shown in the proof of Theorem 4.3.4: Generally, each boundary point z of a convex set in $\mathbb{R}^n$ has a supporting hyperplane containing z. For a convex domain $\mathcal{K}_0 \subset \mathbb{R}^n$, this means that $\mathcal{K}_0$ is contained in some half-plane H with $z \in \partial H$. This H naturally separates $\mathcal{K}_0$ and $\bigcap_{i=1}^{n} \Lambda_{\Theta_i}$ and it is characterized by a point Θ in the interval I.

Therefore, we only need to show the convergence of $D_n \to 0$ as $n \to \infty$. Let $\epsilon > 0$ be given. We have

$$P([D_n > \epsilon] \ i.o.) = P(\limsup_{n\to\infty} [D_n > \epsilon]) = \lim_{N\to\infty} P(\bigcup_{n \geq N} [D_n > \epsilon]) = \lim_{n\to\infty} P(D_n > \epsilon)$$

by monotonicity of the D_n. We can bound this probability by breaking the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ into $\frac{2\pi}{\epsilon}$ disjoint intervals of length $\frac{\epsilon}{2}$. Then, $P(D_n > \epsilon)$ is no larger than the probability of no points in

one of the $\frac{2\pi}{\epsilon}$ disjoint intervals. Thus,

$$P([D_n > \epsilon] \; i.o.) = \lim_{n\to\infty} P(D_n > \epsilon) \leq \lim_{n\to\infty} \frac{2\pi}{\epsilon}(1 - \frac{\epsilon}{2\pi})^n = 0.$$

Statements 2-3 naturally follow the first statement. Under condition (PL), each supporting hyperplane has a unique point on the boundary either tangent to the point or not. As for those tangent points, they have tangent hyperplanes. In this case, each generalized parallel slab has two points on the boundary. Thus, we have a partition $G_1, G_2, \ldots, G_{2n}$ of $\bigcap_{i=1}^{n} \Lambda_{\Theta_i} \triangle \mathcal{K}_0 = \bigcap_{i=1}^{n} \Lambda_{\Theta_i} \backslash \mathcal{K}_0$ based on these $2n$ points. Each G_i is the area of an triangle minus the area of the lune bounded by the boundary arc, thus is bounded above from the area of the triangle. Let G_i denote the area of the triangle. Then

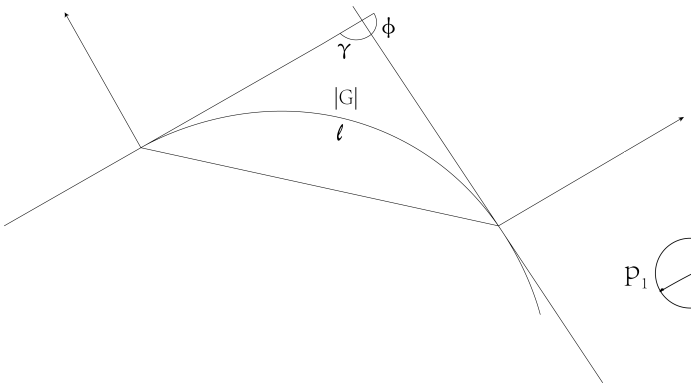
$$d_N\left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0\right) = \sum_{i=1}^{2n} G_i.$$

See Figure 5.1a. G is bounded by the area of an isosceles triangle with angle $\Delta\gamma > 0$ and the opposite side. Thus,

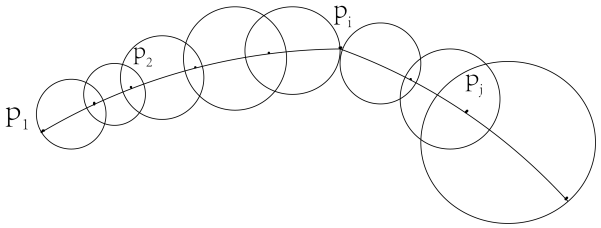
$$G \leq \frac{1}{4}\Delta l^2 \cot\frac{\Delta\gamma}{2} = \frac{1}{4}\Delta l^2 \tan\frac{\Delta\phi}{2},$$

where Δl is the boundary arc length for G, $\Delta\gamma = \pi - \Delta\phi$, and $\Delta\phi > 0$ denotes the difference of angles of slopes of supporting hyperplanes of two adjacent points, which means the distance between two θ_i which form the area G, where $\{\theta_i\}_{i=1}^{\infty}$ is some realization of $\{\Theta_i\}_{i=1}^{\infty}$.

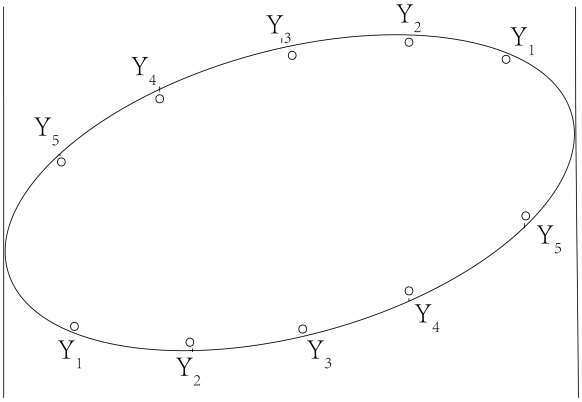
There exists a countable collection of points $\{z_i\}_{i=1}^{\infty}$ in $\mathbb{R}^1$ on the boundary such that the union of neighborhoods $\{B(z_i, r_i) \cap \partial\Lambda\}_{i=1}^{\infty}$ covers the arc segment Δl except two endpoints. The corresponding map of arc segment in each neighborhood is denoted $\{A_i\}_{i=1}^{\infty}$. We partition the line segment Δl as follows: If discontinuities of the curvature exist in this arc segment Δl, we partition the segment by the set of discontinuities and the set of points choosing from each intersection (overlap) of two neighborhoods $B(z_i, r_i) \cap B(z_{i+1}, r_{i+1})$ and the set of two endpoints. This parti-



(a) This is one part of the partition.



(b) Partition of the arc segment



(c) Partition of the curve by two parts

Figure 5.1: Partitions in Theorem 4.3.6.

tion is denoted as points $\{p_i\}_{i=1}^{\infty}$ satisfying

$$\Delta l = \sum_{i=1}^{\infty} \int_{p_i}^{p_{i+1}} ds,$$

where ds is the arc length measure, so the arc from p_i to p_{i+1} has the map A_i defined in *Condition* (PL), see Figure 5.1b. We know that there exist orthogonal coordinate systems for the maps A_i which are some rotations of the original coordinate system such that $A_i(x^i) = y^i$ where the superscript denotes the corresponding coordinate system of A_i. Map A_i defines points $p \in B(z_i, r_i) \cap \partial\Lambda$ after some rotation and $A_i'(x^i) = \tan(\phi^i)$ where ϕ^i is the angle of slope in the coordinate system of A_i. Then we have

$$\Delta l = \sum_{i=1}^{\infty} \int_{p_i}^{p_{i+1}} ds = \sum_{i=1}^{\infty} \int_{x_i^i}^{x_{i+1}^i} \sqrt{1 + A_i'^2(x^i)}\, dx^i \tag{5.7}$$

$$= \sum_{i=1}^{\infty} \int_{\phi_i^i}^{\phi_{i+1}^i} \sqrt{1 + \tan^2(\phi^i)}\, \frac{1 + \tan^2(\phi^i)}{A_i''(A_i'^{-1} \circ \tan^2(\phi^i))}\, d\phi^i \tag{5.8}$$

$$\leq \frac{(1+\delta)^{3/2}}{\tau} \sum_{i=1}^{\infty} \Delta\phi_i^i = \frac{(1+\delta)^{3/2}}{\tau} \sum_{i=1}^{\infty} \Delta\phi_i \leq \frac{(1+\delta)^{3/2}}{\tau} \Delta\phi \tag{5.9}$$

$$= c\Delta\phi, \tag{5.10}$$

where $\Delta\phi_i = \phi_{i+1} - \phi_i$ in the original coordinate system in contrast to $\Delta\phi_i^i = \phi_{i+1}^i - \phi_i^i$ in the ith coordinate system and $c = \frac{(1+\delta)^{3/2}}{\tau}$ is some positive constant. $\Delta\phi_i$ and $\Delta\phi_i^i$ are equal because of the invariance property. The last equality in (5.9) holds only if the supporting hyperplanes of the intersection of inverse supports at the two endpoints are tangent to these two points. Thus, for any

$\epsilon > 0$,

$$\begin{aligned} P(n^\beta d_N \left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0 \right) > \epsilon) &\leq P(n^\beta 2n \max G > \epsilon) \\ &\leq P(\frac{1}{2} n^{\beta+1} c^2 (\max \Delta\phi)^2 \tan \frac{\max \Delta\phi}{2} > \epsilon) \\ &\leq P(2n^{\beta+1} c^2 \tan^3 \frac{\max \Delta\phi}{2} > \epsilon) \\ &= P(\max \Delta\phi > C(\epsilon) \arctan^{1/3}(n^{-\beta-1})) \\ &\leq P(\max \Delta\phi > C(\epsilon) n^{-\frac{\beta+1}{3}}). \end{aligned}$$

Note that $\max \Delta\phi$ is exactly D_n as we defined in the first part of the proof since Θ_i's are uniformly distribution so that only the length of the arcs matters not the location. Hence,

$$P(n^\beta d_N \left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0 \right) > \epsilon) \leq c_1 n^{\frac{\beta+1}{3}} (1 - c_2 n^{-\frac{\beta+1}{3}})^n = c_1 n^{\frac{\beta+1}{3}} e^{-c_2 n^{\frac{2-\beta}{3}} - o(n^{\frac{2-\beta}{3}})},$$

where c_1, c_2 are positive constants. In this case, we only need to prove that

$$\sum_{i=1}^{\infty} n^{\frac{\beta+1}{3}} e^{-c_2 n^{\frac{2-\beta}{3}}} < \infty$$

when $\beta < 2$ and the result follows by Borel-Cantelli lemma. It can be proved that $n^{\frac{\beta+1}{3}} e^{-c_2 \frac{2-\beta}{3}}$ is positive monotone non-increasing after some large N. Then it is equivalent to prove

$$\sum_{i=1}^{\infty} 2^{n^{\frac{\beta+1}{3}}} e^{-c_2 2^{n^{\frac{2-\beta}{3}}}} < \infty$$

by Cauchy condensation test since its ratio limit is zero.

Similarly, statement 3 follows by

$$P(n^{\beta}/\log^{\gamma}(n)d_N\left(\bigcap_{i=1}^{n}\Lambda_{\Theta_i},\mathcal{K}_0\right)>\epsilon)\leq c_1\frac{n^{(\beta+1)/3}}{\log^{\gamma/3}(n)}(1-c_2n^{-(\beta+1)/3}\log^{\gamma/3}(n))^n$$
$$=c_1\frac{n^{(\beta+1)/3}}{\log^{\gamma/3}(n)}e^{-c_2n^{\frac{2-\beta}{3}}\log^{\gamma/3}(n)-o(n^{\frac{2-\beta}{3}}\log^{\gamma/3}(n))}.$$

In this case, $d_N\left(\bigcap_{i=1}^{n}\Lambda_{\Theta_i},\mathcal{K}_0\right)=o_p((\log(n))^{\gamma}/n^2)$ given $\gamma\geq 3$. This means $\beta=2$ when $\gamma\geq 3$. Likewise, when $\gamma>3$, we have

$$\sum_{i=1}^{\infty}\frac{n}{\log^{\gamma/3}(n)}e^{-\log^{\gamma/3}(n)}<\infty$$

since $\frac{n}{\log^{\gamma/3}(n)}e^{-\log^{\gamma/3}(n)}$ is a non-increasing sequence after some large $n>N$ and it is convergent by Cauchy condensation test since its ratio limit is zero. Note that In comparing the ratio, we can use the fact that

$$0\leq 2e^{(\log 2)^{\gamma/3}(k^{\gamma/3}-(k+1)^{\gamma/3})}\leq 2e^{(\log 2)^{\gamma/3}(-\gamma/3k^{\gamma/3-1})}\to 0.$$

Before turning to statement 4, we need a partition for the boundary $\partial\mathcal{K}_0$ such that on each part of the curve it is one-to-one and onto from θ to the tangent point. We choose the rightmost and leftmost points such that they partition the boundary $\partial\mathcal{K}_0$ into the upper curve and the lower curve. The rightmost and leftmost points are exactly the tangent points from the parallel slabs when $\theta=\pi/2$. On the upper curve, the angle that denotes the direction of the unit normal vector can be characterized by $\eta=\Theta+\pi/2$, while on the lower curve it is $\eta=\Theta+3\pi/2$ where Θ is defined above as the angle of slopes of tangent lines of parallel slabs. Define a map

$$\psi:x\to\int_{-\pi/2}^{x}f(t)dt.$$

Then $Y_i=\psi(\Theta_i)$ is uniformly distributed on $(0,1)$ and are independent. We also define $\Theta_j(n)$ by $\psi(\Theta_j(n))=Y_j(n)$ for $j=1,2,\ldots,n+1$ on both partitioned curves, which defines $\eta_j(n)=$

$\Theta_j(n) + \pi/2, j = 1, 2 \ldots, n$ on the upper curve and $\eta'_j(n) = \Theta_j(n) + 3\pi/2, j = 1, 2 \ldots, n$ on the lower curve. Since it is a closed curve as a boundary, $\eta_{n+1}(n) = \eta'_1(n) = \eta_1(n) + \pi$ and $\eta'_{n+1}(n) = \eta_1(n) + 2\pi$ which equivalently means $\Theta_{n+1}(n) = \Theta_1(n) + \pi$. Note that $Y_j(n), j = 1, 2 \ldots, n$ characterize unique points on each partitioned curve. On the upper curve, we can let $Y_{n+1}(n) = 1 + Y_1(n)$ since

$$\begin{aligned} Y_{n+1}(n) - Y_n(n) &= \int_{\Theta_n(n)}^{\Theta_{n+1}(n)} f(t)dt = \int_{\Theta_1(n)+\pi}^{\Theta_n(n)} f(t)dt = \left(\int_{\Theta_n(n)}^{\pi/2} + \int_{\pi/2}^{\Theta_1(n)+\pi} \right) f(t)dt \\ &= 1 - Y_n(n) + Y_1(n). \end{aligned}$$

See Figure 5.1c.

Then we use the same partition as in statement 2 (Figure 5.1a) to calculate $d_N \left(\bigcap_{i=1}^{n} \Lambda_{\Theta_i}, \mathcal{K}_0 \right)$ but indexed by the angles of normal vectors η. Let $\Delta(\eta_j(n), \eta_{j+1}(n))$ denote the Nikodym metric of the $(\eta_j(n), \eta_{j+1}(n))$ component of the partition. Then on the upper curve,

$$\Delta(\eta_j(n), \eta_{j+1}(n)) = \frac{1}{24} r^2_{\mathcal{K}_0}(\eta_j(n))(\eta_{j+1}(n) - \eta_j(n))(1 + o(1))$$

with $o(1) \to 0$ for $|\eta_{j+1}(n) - \eta_j(n)| \to 0$, uniformly in $\eta_j(n)$. Similarly on the lower curve,

$$\Delta(\eta'_j(n), \eta'_{j+1}(n)) = \frac{1}{24} r^2_{\mathcal{K}_0}(\eta'_j(n))(\eta'_{j+1}(n) - \eta'_j(n))(1 + o(1))$$

with $o(1) \to 0$ for $|\eta'_{j+1}(n) - \eta'_j(n)| \to 0$, uniformly in $\eta'_j(n)$, see [formulate 5.15] in McClure and Vitale (1975). By the uniform continuity of the function f, we have

$$Y_{j+1}(n) - Y_j(n) = f[\Theta_j(n)](\Theta_{j+1}(n) - \Theta_j(n))[1 + o(1)]$$

with $o(1) \to 0$ for $|\Theta_{j+1}(n) - \Theta_j(n)| \to 0$, uniformly in $\Theta_j(n)$. Hence, we have

$$\begin{aligned}
d_N\left(\bigcap_{i=1}^{n}\Lambda_{\Theta_i}, \mathcal{K}_0\right) &= \sum_{j=1}^{n}\frac{1}{24}r^2_{\mathcal{K}_0}(\eta_j(n))(\eta_{j+1}(n)-\eta_j(n))(1+o(1)) \\
&+ \sum_{j=1}^{n}\frac{1}{24}r^2_{\mathcal{K}_0}(\eta'_j(n))(\eta'_{j+1}(n)-\eta'_j(n))(1+o(1)) \\
&= \sum_{j=1}^{n}\frac{1}{24}r^2_{\mathcal{K}_0}(\Theta_j(n)+\pi/2)(\Theta_{j+1}(n)-\Theta_j(n))(1+o(1)) \\
&+ \sum_{j=1}^{n}\frac{1}{24}r^2_{\mathcal{K}_0}(\Theta_j(n)+3\pi/2)(\Theta_{j+1}(n)-\Theta_j(n))(1+o(1)) \\
&= \sum_{j=1}^{n}\frac{1}{24}\frac{r^2_{\mathcal{K}_0}[\psi^{-1}(Y_j(n))+\pi/2]+r^2_{\mathcal{K}_0}[\psi^{-1}(Y_j(n))+3\pi/2]}{f^3[\psi^{-1}(Y_j(n))]}U_j^3(n)(1+o(1))
\end{aligned}$$

with $o(1) \to 0$ for $\max_{1\le j\le n} U_j(n) \to 0$.

Now from Lemma 1, we have with probability 1

$$\lim_{n\to\infty} n^2 d_N\left(\bigcap_{i=1}^{n}\Lambda_{\Theta_i}, \mathcal{K}_0\right) = \frac{1}{4}\int_{(-\frac{\pi}{2},\frac{\pi}{2})}\frac{r^2_{\mathcal{K}_0}(\theta+\frac{\pi}{2})+r^2_{\mathcal{K}_0}(\theta+\frac{3\pi}{2})}{f^2(\theta)}d\theta.$$

Chapter 6

Summary and Future Work

In the dissertation, we study the aspects of both distributions and domains in the solution of the SIP. In Chapter 2, we introduce the established solution of the SIP in the discrete setting, with an extension by employing observable inputs in a model. In the SIP, it is crucial to consider the SFP, a data-generating process, in which the uncertainty in the output is propagated by a convolution of the uncertainty of all the inputs. We propose an approach that solves the SIP individually indexed by the observable inputs, which can be viewed as control variables that govern the experiments. This is especially advantageous to time-dependent problems in which the covariance of sequential experiments is hard to capture. We focus on finding a global solution, GFGD, that is not conditioned on any experiment in contrast to the established solution that is conditioned on an experiment due to the pre-specified ansatz. To obtain a GFGD, we propose an iterative approach that updates the ansatz information based on the solutions from each individual experiment. A GFGD is not associated with the observable inputs in its domain, which otherwise could cause bias in scientific inference, and reproduces all the output distributions. In addition, we connect the DS theory and the SIP approach by employing the technique of decomposition of a distribution. The DS functions give a bound for the degree of belief of the unobserved inputs through the information from the observable inputs only and ignore the information from the unobserved inputs themselves, while the SIP approach gives an exact approximation by integrating all the information. We also propose an extension of the classical Bayesian approach under the SIP setting for the case with limited sample sizes in practice, which can also be viewed as an extension of Bayesian meta-analysis.

In Chapter 3, we introduce the established solution of the SIP and the same iterative approach in the continuous setting, which can be applied in practice. We establish the equivalence class of GFGDs with absolute continuity dominated by the "volume" measure on the domain. This smoothness property helps to regularize the smoothness of the resulting distributions in the iterative

approach. In the discrete case, the property is automatic due to the fact that each finite distribution is absolutely continuous with respect to the counting measure on the domain. We show an example of three equivalent smooth GFGDs for the purposes of quantifying the uncertainty of an event and targeting a high-probability region. The distinct information they provide suggests that it is not appropriate to use a single distribution of the unobserved inputs to make inference, and thus we propose a bound among all possible equivalent solutions to for scientific inference. We demonstrate that our SIP approach is exactly an extension of the DS approach by disintegration, because the technique we employ, i.e. "averaging" the solutions from each individual experiment over the observable inputs, to initiate an iteration, is adopted from the DS framework.

The previous chapters highly rely on the input domain of the model which is often given as a compact space containing the "true" support of the inputs by scientists. The domain has a significant impact on the accuracy of the established solution because the solution is "spread out" on the domain according to the pre-specified ansatz along the contours that are determined by the domain. In addition, there might be multiple equivalent GFGDs that have different supports, and thus there is no one "true" support. We propose one unique support that is the minimal support containing all the supports of GFGDs. Thus, if we set the domain to be this unique support, the resulting solution of the SIP will have the least bias with respect to domains. We propose an approach that finds this unique support in linear models under some smoothness constraints of the support boundary.

Our methodologies in the dissertation are based on the physical processes modeled by Q and distributions of the output. In many problems in which the data-generating process is unknown, it is possible to approximate Q by setting up an appropriate expansion of Q, e.g. a Fourier series, to reduce the approximation error from using inappropriate models. It is of interest to explore a data version of the our methodology in which we observe data of the output according the distribution. In this case, we simply use data to approximate the output distribution, e.g. by a histogram or a Bayesian model, and thus the approximation error in the output propagates through to a potentially large error in the approximation of GFGDs. Similar version of the established solution of the SIP has been proposed in Butler et al. (2012, 2014). One of the important facts we have observed

in examples is that there might be multiple equivalent GFGDs that share some common features and also provide distinct information. In particular, the high-probability region might provide key information of the common mode of all possible GFGDs. However, there is no identifiability with which we can easily make inference. It is of interest to employ the possible GFGDs to find useful information and make appropriate inference in a similar way to fiducial inference and the DS theory in which the uncertainty of the unobservable inputs is considered. Furthermore, the computational cost enormously increases as the dimension of unobservable inputs increases. Adopting an approach in Bayesian updating, it is possible to consider experiments arriving sequentially. In this case, we can update the ansatz sequentially in the iteration and thus decrease the time complexity. In the discrete cases, the inverse distributions are essentially conditional distributions that are conditional on the Borel sigma algebra of the output distributions. It may be necessary to have smoothness and regularity on both the map Q and the domain Λ in the continuous cases. It is also of interest to explore how to sample the observable input efficiently and sufficiently when it is time or location. A dense sample in the space, e.g. some space-filling curve, is sufficient to provide all the information, or sometimes a smaller subset is useful enough to provide concentrated information. Different sampling methods might induce different results. We leave these topics for future research.

Bibliography

Aldrich, J. et al. (1997). Ra fisher and the making of maximum likelihood 1912-1922. *Statistical science*, 12(3):162–176.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Breidt, J., Butler, T., and Estep, D. (2011). A measure-theoretic computational method for inverse sensitivity problems i: Method and analysis. *SIAM journal on numerical analysis*, 49(5):1836–1859.

Butler, T., Estep, D., and Sandelin, J. (2012). A computational measure theoretic approach to inverse sensitivity problems ii: A posteriori error analysis. *SIAM journal on numerical analysis*, 50(1):22–45.

Butler, T., Estep, D., Tavener, S., Dawson, C., and Westerink, J. J. (2014). A measure-theoretic computational method for inverse sensitivity problems iii: Multiple quantities of interest. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):174–202.

Butler, T., Graham, L., Estep, D., Dawson, C., and Westerink, J. (2015). Definition and solution of a stochastic inverse problem for the manning'sn parameter field in hydrodynamic models. *Advances in water resources*, 78:60–79.

Chang, J. T. and Pollard, D. (1997). Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317.

De Vito, E., Rosasco, L., Caponnetto, A., De Giovannini, U., Odone, F., and Bartlett, P. (2005). Learning from examples as an inverse problem. *Journal of Machine Learning Research*, 6(5).

Dempster, A. P. (2008). Upper and lower probabilities induced by a multivalued mapping. In *Classic works of the Dempster-Shafer theory of belief functions*, pages 57–72. Springer.

Drobot, V. et al. (1982). Probabilistic version of a curvature formula. *The Annals of Probability*, 10(3):860–862.

Fisher, R. A. (1930). Inverse probability. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 26, pages 528–535. Cambridge University Press.

Guiasu, S. and Shenitzer, A. (1985). The principle of maximum entropy. *The mathematical intelligencer*, 7(1):42–48.

Higgins, J. P., Thompson, S. G., and Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1):137–159.

Kaipio, J. and Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Kolmogorov, A. N. and Bharucha-Reid, A. T. (2018). *Foundations of the theory of probability: Second English Edition*. Courier Dover Publications.

McClure, D. E. and Vitale, R. A. (1975). Polygonal approximation of plane convex bodies. *J. Math. Anal. Appl*, 51(2):326–358.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, pages 409–423.

Schneider, R. (1988). Random approximation of convex sets. *Journal of Microscopy*, 151(3):211–227.

Senn, S. (2000). The many modes of meta. *Drug Information Journal*, 34(2):535–549.

Senn, S. et al. (2011). You may believe you are a bayesian but you are probably wrong. *Rationality, Markets and Morals*, 2(42):48–66.

Smith, T. C., Spiegelhalter, D. J., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in medicine*, 14(24):2685–2699.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

Yager, R. R. and Liu, L. (2008). *Classic works of the Dempster-Shafer theory of belief functions*, volume 219. Springer.

Yang, L. (2018). Infinite dimensional stochastic inverse problems. *Ph.D. Dissertation In CSU*.

Zhang, J. and Liu, C. (2011). Dempster-shafer inference with weak beliefs. *Statistica Sinica*, pages 475–494.

Appendix A

Supplementary Material

```
% Sample code for results in Chapter 4
% Location parameters for Ushape
x1 = 0.5; x2 = 0.7; r = 0.1;
y1 = 0.2; y2 = y1 + 2*r; y3 = y2 + 2*r; y4 = y3+ 2*r;
% Plot Ushape support
imageSizeX = 1;
imageSizeY = 1;
x = 0:0.0001:1;
y = 0:0.0001:1;
circlePixels = ones(length(x), length(y));
for i = 1:length(x)
    for j = 1:length(y)
        if  x(i) >= x1 &&  x(i) <=x2 && y(j) >= y1 && y(j) <= y2
           || ...
                x(i) >= x1 &&  x(i) <=x2 && y(j) >= y3 && y(j) <=
                    y4 || ...
                x(i) >= x1-3*r &&  x(i) <=x1 && (x(i)-x1)^2+(y(j)
                   -(y2+y3)/2)^2 >= r^2 && (x(i)-x1)^2+(y(j)-(y2+
                   y3)/2)^2 <= (3*r)^2 || ...
                x(i) >= x2 &&  x(i) <=x2+r && y(j) >= y1 && y(j)
                   <= y2 && (x(i)-x2)^2+(y(j)-(y1+y2)/2)^2 <= r^2
                    || ...
                x(i) >= x2 &&  x(i) <=x2+r && y(j) >= y3 && y(j)
                   <= y4 && (x(i)-x2)^2+(y(j)-(y3+y4)/2)^2 <= r^2
```

```
                circlePixels(j,i) = 0;
            end
        end
    end
    imagesc(0:1,0:1,circlePixels);
    set(gca,'YDir','normal','FontSize',16)
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Generate sample on the Ushape
    n = 1000000;
    AB = rand(n,2);
    Ind = AB(:,1) >= x1 &  AB(:,1) <=x2 & AB(:,2) >= y1 & AB(:,2) <=
        y2 | ...
                AB(:,1) >= x1 &  AB(:,1) <=x2 & AB(:,2) >= y3 & AB
                    (:,2) <= y4 | ...
                AB(:,1) >= x1-3*r &  AB(:,1) <=x1 & (AB(:,1)-x1).^2+(
                    AB(:,2)-(y2+y3)/2).^2 >= r^2 & (AB(:,1)-x1).^2+(AB
                    (:,2)-(y2+y3)/2).^2 <= (3*r)^2 | ...
                AB(:,1) >= x2 &  AB(:,1) <=x2+r & AB(:,2) >= y1 & AB
                    (:,2) <= y2 & (AB(:,1)-x2).^2+(AB(:,2)-(y1+y2)/2)
                    .^2 <= r^2 | ...
                AB(:,1) >= x2 &  AB(:,1) <=x2+r & AB(:,2) >= y3 & AB
                    (:,2) <= y4 & (AB(:,1)-x2).^2+(AB(:,2)-(y3+y4)/2)
                    .^2 <= r^2;
    points = AB(Ind,:);
    [track,~] = size(points);
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Recover the convex hull of the Ushape
```

```
N = 500; K = 100;
% N: num of grid cells on one dimension
% K: num of experiments
x = linspace(0,1,N+1);
x(1) = [];
x = x-1/N/2;
y = linspace(0,1,N+1);
y(1)=[];
y = y-1/N/2;
theta = -pi/2 + 2*(pi/2)*rand(1,K);
zeroLab = ones(N);
[Xcol,Ycol] = meshgrid(y,x);
for k=1:K
    X = tan(theta(k));
    Y = points(:,1)*X+points(:,2); %% ax+b
    Ymin = min(Y);
    Ymax = max(Y);
    Ytrial = Xcol*X+Ycol;
    zeroLab = zeroLab.*(Ytrial>=Ymin & Ytrial<=Ymax);
end
imagesc(x,y,zeroLab)
colormap('summer')
colorbar
set(gca,'YDir','normal','FontSize',16)
% Plot the Ushape boundary
hold on
x_1 = x1:0.001:x2;    y_1 = y1*ones(1,length(x_1));
```

```
x_2 = x1:0.001:x2;    y_2 = y2*ones(1,length(x_2));
x_3 = x1:0.001:x2;    y_3 = y3*ones(1,length(x_3));
x_4 = x1:0.001:x2;    y_4 = y4*ones(1,length(x_4));
th_5 = pi/2:pi/50:3*pi/2;
x_5 = 3 * r * cos(th_5) + x1;
y_5 = 3 * r * sin(th_5) + (y2+y3)/2;
th_6 = pi/2:pi/50:3*pi/2;
x_6 = r * cos(th_6) + x1;
y_6 = r * sin(th_6) + (y2+y3)/2;
th_7 = -pi/2:pi/50:pi/2;
x_7 = r * cos(th_7) + x2;
y_7 = r * sin(th_7) + (y4+y3)/2;
th_8 = -pi/2:pi/50:pi/2;
x_8 = r * cos(th_8) + x2;
y_8 = r * sin(th_8) + (y1+y2)/2;
xx = [x_1 x_8 fliplr(x_2) fliplr(x_6) x_3 x_7 fliplr(x_4) x_5]';
yy = [y_1 y_8 fliplr(y_2) fliplr(y_6) y_3 y_7 fliplr(y_4) y_5]';
plot(xx,yy)
hold off
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Plot the EEI computed on the convex hull
N0 = 100;
AddLab = zeros(N);
IntLabel = zeros(N);
count = zeros(K,1);
[i0,j0] = find(zeroLab);
for k = 1:K
```

```
    Lab = zeros(N);
    X = tan(theta(k));
    Y = points(:,1)*X+points(:,2);
    Ymax = max(Y);
    Ymin = min(Y);
    edges = linspace(Ymin,Ymax,N0+1);
    [PY,~] = histcounts(Y, edges);
    Ytrial = Xcol*X+Ycol;
    for i = 1:length(i0)
        IntLabel(i0(i),j0(i)) = findInterval(edges,Ytrial(i0(i),
            j0(i)));
    end
    [i1,j1] = find(IntLabel);
    for i = 1:N0
        count(i) = sum(reshape(IntLabel==i,[],1));
    end
    for i = 1:length(i1)
        Lab(i1(i),j1(i)) = PY(IntLabel(i1(i),j1(i)))/track/count(
            IntLabel(i1(i),j1(i)));
    end
    AddLab = AddLab+Lab;
end
AddLab = AddLab/K;
s = surf(AddLab*N^2);
s.EdgeColor = 'none';
colormap('summer');
set(gca,'XTick',0:N/2:N)
```

```
set(gca,'XTickLabel',0:0.5:1)
set(gca,'YTick',0:N/2:N)
set(gca,'YTickLabel',0:0.5:1)
set(gca, 'FontSize',12);
```