September 1990

Sensor-Fusion With Statistical Decision Theory: A Prospectus of Research in the GRASP Lab

Raymond McKendall
University of Pennsylvania

Max L. Mintz
University of Pennsylvania, mintz@cis.upenn.edu

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-90-68.

Sensor-Fusion With Statistical Decision Theory: A Prospectus of Research in the GRASP Lab

Abstract

The purpose of this report is to describe research in sensor fusion with statistical decision theory in the GRASP Lab, Department of Computer and Information Science, University of Pennsylvania. This report is thus a tutorial overview of the general research problem, the mathematical framework for the analysis, the results of specific research problems, and directions of future research. The intended audience for this report includes readers seeking a self-contained summary of the research as well as students considering study in this area. The prerequisite for understanding this report is familiarity with basic mathematical statistics.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-90-68.

Sensor-Fusion With Statistical Decision Theory: A Prospectus Of Research In The GRASP Lab

MS-CIS-90-68
GRASP LAB 234

Raymond McKendall
Max Mintz

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

September 1990

Sensor-Fusion with Statistical Decision Theory: A Prospectus of Research in the GRASP Lab

Raymond McKendall and Max Mintz*

GRASP Lab
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104-6389

September 1990

Abstract

The purpose of this report is to describe research in sensor fusion with statistical decision theory in the GRASP Lab, Department of Computer and Information Science, University of Pennsylvania. This report is thus a tutorial overview of the general research problem, the mathematical framework for the analysis, the results of specific research problems, and directions of future research. The intended audience for this report includes readers seeking a self-contained summary of the research as well as students considering study in this area. The prerequisite for understanding this report is familiarity with basic mathematical statistics.

*Acknowledgments: Navy Contract N0014-88-K-0630. AFOSR Grants 88-0244, 88-0296. Army/DAAL 03-89-C-0031PRI. NSF Grants CISE/CDA 88-22719, IRI 89-06770. The Dupont Corporation.

Contents

List of Figures

1 Introduction

The purpose of this report is to describe research in sensor fusion with statistical decision theory in the GRASP[1] Lab, Department of Computer and Information Science, University of Pennsylvania. This report is thus a tutorial overview of the general research area, the mathematical framework for the analysis, the results of specific research problems, and directions for future research. The intended audience for this report includes readers seeking a self-contained summary of the research as well as students considering study in this area. The prerequisite for understanding this report is familiarity with basic mathematical statistics.

This report has nine sections. Section 2 gives a heuristic description of the sensor fusion problem. Section 3 describes the mathematical model of sensor fusion adopted in the GRASP Lab. Section 4 discusses uncertainty classes used in modeling sensor noise. Section 5 is an introduction to statistical decision theory, the framework for the mathematical analysis. Section 6 gives some examples of research problems solved. Section 7 is a summary of the main research papers. Section 8 poses other directions of research. Section 9 lists publications and presentations by researchers in the GRASP Lab.

2 Sensor Fusion and Consistency

The *sensor-fusion problem* is the problem of combining multiple measurements from sensors into a single measurement of the sensed object or attribute. There may be several different sensors or multiple measurements from a single sensor. The fusion problem arises when there are discrepancies in the data: In most applications, the measurements do not coincide. The variations in the data are due to the uncertainty associated with a sensor. The uncertainty may be due to the environment or to the sensor itself. Possible sources of uncertainty include temperature variation, electronic interference, weather conditions, worn or defective components, miscalibration, and quantization. Despite the variation, however, the desired output of the entire system is a single measurement based on the measurements from the sub-systems. Fusion is combination of these discrepant data. The goal of fusion, in particular, is optimal combination of multiple measurements. The criterion of optimality depends on the system and its model.

Example 2.1 Figure 1 (p. 2) illustrates a sensor-fusion problem with three different sensors, labeled S_1, S_2, and S_3. The output of each sensor is a measurement Z_i of the distance of the airplane from a common origin. In general, the measurements Z_1, Z_2, and Z_3 will differ from each other because of uncertainty in each sensor. For example, the assumed position of each sensor may inexact, or the accuracies of the

[1] General Robotics and Active Sensory Perception

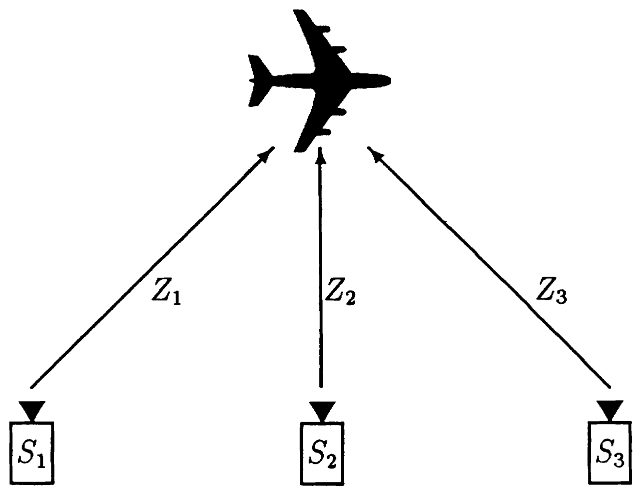

Figure 1: A sensor-fusion problem

sensors may vary at different ranges. The fusion problem is to combine these three measurements into a single measurement of the airplane's distance. Simple examples of fusion schemes include the average or median of the data. An optimal scheme may maximize the average performance of the system or minimize the worst-possible performance of the system, for example. □

The sensor-fusion problem induces the consistency problem. A requirement of fusion is that the combined data are measurements of the same object. The *consistency problem* is to verify within the uncertainty of the system that the data are in fact measurements of the same object. Thus, the first stage of a sensor-fusion problem is a test of consistency. Only consistent data are combined. Moreover, consistency verifies only the precision of the system, but not its accuracy. The mathematical framework alone can neither resolve inconsistencies in the data nor verify accuracy of consistent data: These questions must be addressed through some criterion, belief, or experience external to the framework.

Example 2.2 Figure 2 (p. 2) illustrates a consistency problem. The goal of this system of five sensors is to measure the range of a single airplane. The system comprises five sub-systems corresponding to the five sensors. The output is a combination of the five measurements Z_1, Z_2, ..., Z_5 — provided that each sensor is observing the same airplane. The consistency problem is to verify this stipulation. In this example, a consistency test partitions the data set into two subsets of consistent measurements: $\{Z_1, Z_2, Z_4\}$ and $\{Z_3, Z_5\}$. The test cannot, however, determine which subset if either is correct. Overall, the measurements are not consistent, and so they are not combined. The system must be evaluated with some external assumptions in order to proceed. For example, prior experience may indicate that sensors S_3 and S_5 are faulty,

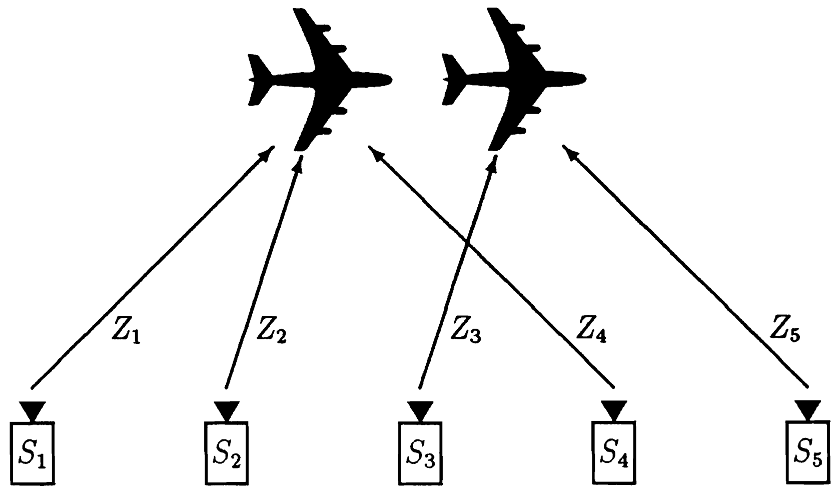

Figure 2: A consistency problem in sensor fusion

and so only the consistent measurements Z_1, Z_2, Z_3 are combined. Alternatively, the inference from the system may be that there are two airplanes in the scope of the system, that they are too close to distinguish one from the other, and consequently that they are too close for safety. □

In summary, this research in sensor fusion addresses the problem of combining multiple measurements of a single attribute or parameter into a single estimate of that parameter. The approach taken by this research to sensor fusion has some underlying assumptions about the sensor system and its application. First, it assumes that sensor measurements contain uncertainty and that multiple measurements can help to reduce the uncertainty. Second, it assumes that statistical models of the noise can be derived. The source of these models may be either physical or empirical. Derivation of these models is an important component of the analysis. Third, this approach assumes that the possible values of the parameter are known, and it uses this prior information in the fusion process. It neither assumes nor uses probabilistic prior information about the parameter. Finally, this approach assumes that there is a tolerance or threshold for error: An estimate within the tolerance is good enough for the application, but an estimate outside of this tolerance is not good enough for the application. Furthermore, there is no distinction among estimates within this threshold and no distinction among estimates outside of the threshold.

This research addresses a low-level aspect of sensor fusion amenable to statistical analysis. Its scope is a single estimate of a parameter through multiple measurements of the parameter. This aspect of sensor fusion is low-level in the sense that its scope is limited to finding statistical information about a single attribute. This research does not address other questions of sensor fusion, such as how to use a measurement, what measurements to take, or how to integrate measurements of different attributes.

Research in the GRASP Lab considering other aspects of sensor fusion is included in [Durrant-Whyte, 1986], [Durrant-Whyte, 1988], [Hager, 1988], and [Hager, 1990].

3 Mathematical Model

The research on sensor fusion in the GRASP Lab concentrates on statistical analysis of this problem. In particular, this research studies fusion of location data. The location-data model consists of measurements Z_i of an unknown parameter θ in statistical uncertainty. (These may be multi-dimensional.) A location model of a measurement assumes that the parameter governs only the location of the noise but not its shape; the model assumes that the shape of the noise is independent of the parameter. (Such noise is called additive.) For example, a measurement Z of a parameter θ may be modeled as a Gaussian or normally distributed random variable with mean θ: $Z \sim \mathcal{N}(\theta, \sigma^2)$. Then the shape of the noise is $\mathcal{N}(0, \sigma^2)$ regardless of the location θ of the mean.

Statistical Formulation of Problem

The underlying statistical model of sensor fusion in the GRASP Lab has multiple measurements Z_i consisting of a location parameter θ_i in additive noise V_i:

$$Z_i = \theta_i + V_i, \qquad i = 1, \ldots, n$$

The random variable V_i is the statistical uncertainty in the measurement Z_i. The scalar θ is a location parameter for the distribution function F_{Z_i} of Z_i: For all $\theta_i \in \Theta_i$,

$$F_{Z_i}(z|\theta_i) = F_{V_i}(z - \theta_i), \qquad \forall z \in \Re.$$

The set Θ consists of the possible locations and thus represents prior information about the location parameter. The consistency problem is to verify for all i and j that $\theta_i = \theta_j$ within statistical uncertainty. If the data are consistent, the model is simplified:

$$Z_i = \theta + V_i, \qquad i = 1, \ldots, n$$

The fusion problem is to find an optimal estimator $\delta_n(Z_1, \ldots, Z_n)$ of the common location θ from consistent data $Z_1, \ldots, Z_n$. For constructing an optimal estimator, the model assumes a tolerance e for error: An estimate $\hat{\theta}$ for θ is acceptable if the absolute error of estimation $||\hat{\theta} - \theta||$ is at most e; otherwise, the error is unacceptable. An *optimal estimator* minimizes the (maximum) probability of unacceptable error.

This formulation of sensor-fusion permits flexibility in modeling the noise. For example, the distribution of the noise may be asymmetric, multi-modal, or uncertain. In particular, there is no restriction to Gaussian noise. Moreover, the model allows measurements from sensors with different noise distributions: There is no assumption that the random variables $V_1, \ldots, V_n$ are identically distributed. Also, there may be statistical dependence among these variates.

Example 3.1 The location model of the sensor-fusion problem of example 2.1 is this:

$$\begin{aligned} Z_1 &= \theta_1 + V_1 \\ Z_2 &= \theta_2 + V_2 \\ Z_3 &= \theta_3 + V_3 \end{aligned}$$

Each measurement Z_i consists of an unknown location parameter $\theta_i \in \Theta$ in additive noise V_i. The set Θ represents prior knowledge about the range of the airplane. The noise V_i models the uncertainty associated with sensor S_i. The problem of consistency is to verify that $\theta_1 = \theta_2$, that $\theta_1 = \theta_3$, and that $\theta_2 = \theta_3$ within statistical uncertainty. If the measurements are consistent, then the fusion problem is to find an optimal estimator $\delta_n(Z_1, Z_2, Z_3)$ of the common location θ. □

Scalar Problem

Currently, the main research problem motivated by this paradigm is to estimate the scalar location parameter $\theta \in \Theta$ of a single observation Z in the model

$$Z = \theta + V.$$

There are two versions of this problem, standard estimation and robust estimation. In a *standard-estimation* problem, the distribution function F_V of the additive noise V is known. An example is to estimate the mean θ of $Z \sim \mathcal{N}(\theta, 1)$; in this case $F_V \sim \mathcal{N}(0, 1)$. In a *robust-estimation* problem, the distribution F_V is uncertain: It is an unknown member of a given class $\mathcal{F}$ of distribution functions, an uncertainty class. An example is to estimate the mean θ of $Z \sim \mathcal{N}(\theta, \sigma^2)$ when the scale $\sigma \in (0, 1]$ is unknown; in this case $F_V \in \mathcal{F}$ where $\mathcal{F}$ is the set of $\mathcal{N}(0, \sigma^2)$ distribution functions with $\sigma \in (0, 1]$. Robust estimation is important because it accounts for inexact characterizations of the noise. Standard estimation is important both because it is a starting point for the analysis of robust estimation and because many problems in robust estimation reduce to problems in standard estimation.

Example 3.2 Figure 3 (p. 6) illustrates the structures of standard estimation and robust estimation. In a standard-estimation model, the statistical path from the location parameter θ to the measurement Z is known: The noise distribution F_V represents this path. In a robust-estimation model, however, only the *possible* statistical paths are known: The noise distribution is known only to be either F_1, F_2, F_3, or F_4. The *exact* path is uncertain. The set of possible paths is the uncertainty class $\mathcal{F} = \{F_1, F_2, F_3, F_4\}$. Standard estimation is thus the special case of robust estimation in which $\mathcal{F}$ consists of a single distribution function. □

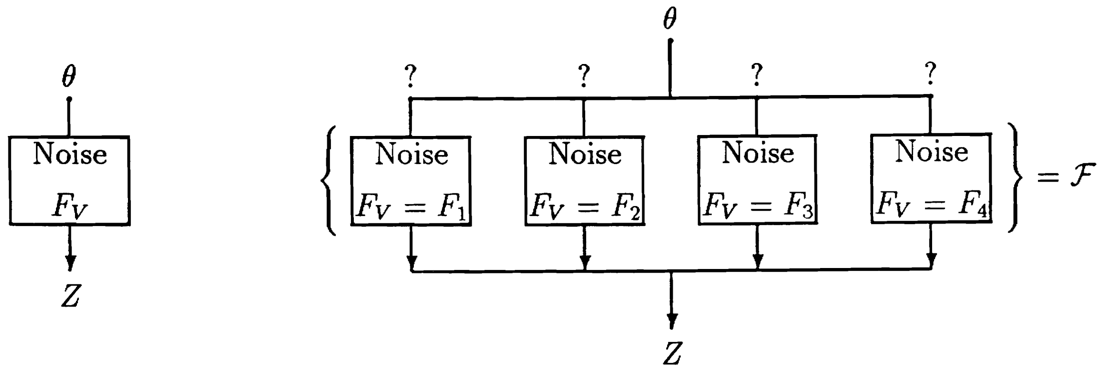

Figure 3: Standard estimation and robust estimation

Analysis of the single-observation model $Z = \theta + V$ is the first step in the analysis of the multiple-observation model. Furthermore, preliminary research suggests that in some instances an optimal estimator $\delta_n(Z_1, \ldots, Z_n)$ is the composition of an optimal estimator $\delta(Z)$ from the single-observation model with a scalar statistic $T_n(Z_1, \ldots, Z_n)$ of the data: $\delta_n(Z_1, \ldots, Z_n) = \delta(T_n(Z_1, \ldots, Z_n))$.

The location-data model $Z = \theta + V$ is also a first step in the analysis of the general model $Z = h(\theta, V)$, in which h is some transformation of location and noise. It is the simple case in which $h(\theta, v) = \theta + v$. Example transformations h in the general case include truncation and saturation functions. The most general situation is a robust model in which the transformation h is an uncertain member of a set $\mathcal{H}$ of possible transformations. For example, the transformation h may be a saturation function for which the point of saturation is not exactly known.

Finally, analysis of the single-observation model provides a test of hypothesis that two measurements $Z_1 = \theta_1 + V_1$ and $Z_2 = \theta_2 + V_2$ are consistent. With $Z := Z_1 - Z_2$, or

$$Z := (\theta_1 - \theta_2) + (V_1 - V_2) =: \theta + V,$$

the data are consistent if $\theta = 0$ within statistical uncertainty. A test that $\theta = 0$ accepts this hypothesis if the confidence interval $[\delta(Z) - e, \delta(Z) + e]$ contains zero. Here $\delta(Z)$ is an estimator of θ in the model $Z = \theta + V$. Analysis of this hypothesis test is an important research problem in consistency.

Figure 4 (p. 7) illustrates the components of the sensor-fusion problem. The goal is fusion of consistent measurements from a sensor system. The scalar estimation problem is the starting point for the statistical analysis.

Remark Multiple-observation models $Z_i = \theta_i + V_i$ also are classified as standard or robust. The adjectives *standard* and *robust* refer to the model of the noise distribution. □

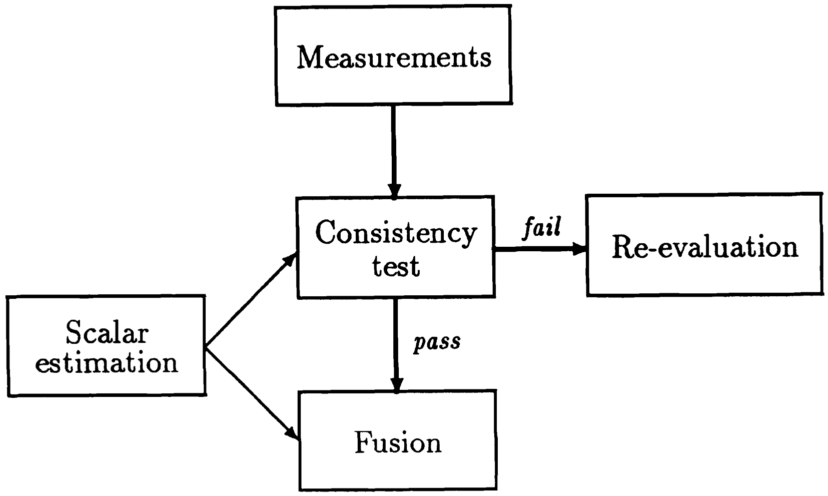

Figure 4: Components of research in sensor fusion

4 Uncertainty Classes

Uncertainty classes model a variety of problems in which the noise distribution is not exactly known. Some examples are uncertainty in scale, uncertainty in origin, and ϵ-contamination.

Example 4.1 A model for uncertainty in the origin of a sensor is

$$Z = \theta + V' + \eta, \qquad \eta \in [\eta_1, \eta_2].$$

Here, the distribution $F_{V'}$ of V' is known, but the shift η is unknown. This parameter represents the error in positioning the sensor. The model assumes that this error has known bounds η_1 and η_2. With $V := V' + \eta$, this problem becomes a robust-estimation problem with

$$F_V \in \{F : F(x) = F_{V'}(x - \eta),\ \eta_1 \leq \eta \leq \eta_2\}.$$

(See figure 5, p. 9.) This set extends to an uncertainty class $\mathcal{F}$ such that

$$\mathcal{F} \subseteq \{F : F_{V'}(x - \eta_2) \leq F(x) \leq F_{V'}(x - \eta_1)\}.$$

A further extension replaces the lower and upper boundary distributions $F_{V'}(\cdot - \eta_2)$ and $F_{V'}(\cdot - \eta_1)$ derived from translations of a single distribution $F_{V'}$ with boundary distributions F_{b_l} and F_{b_u} that are not necessarily related through translation. □

Example 4.2 A model for uncertainty in the scale of a sensor's precision is

$$Z = \theta + \sigma V', \qquad \sigma \in (0, \sigma_m].$$

Here, too, the distribution $F_{V'}$ of V' is known, but the scale σ is unknown. The scale represents the sensor's precision through its reciprocal $1/\sigma$. The model assumes a known upper bound σ_m for the scale or, equivalently, a known lower bound $1/\sigma_m$ for the precision. With $V := \sigma V'$, this problem becomes a robust-estimation problem with

$$F_V \in \{F : F(x) = F_{V'}(x/\sigma),\ 0 < \sigma \leq \sigma_m\}.$$

(See figure 6, p. 9.) This set extends to an uncertainty class $\mathcal{F}$ such that

$$\mathcal{F} \subseteq \{F : F(x) \leq F_{V'}(x/\sigma_m) \text{ for } x < 0 \text{ and } F(x) \geq F_{V'}(x/\sigma_m) \text{ for } x \geq 0\}.$$

A further extension replaces the boundary $F_{V'}(\cdot/\sigma_m)$ based on scale with some other boundary F_b not necessarily related to scale. □

Remark Mathematical analysis motivates the extended uncertainty classes of examples 4.1 and 4.2: Many problems in robust estimation reduce to problems in standard estimation in which the noise distribution is taken from the uncertainty class. For uncertainty in scale, the appropriate noise distribution is the boundary $F_{V'}(\cdot/\sigma_m)$. For uncertainty in origin, initial research suggests that the appropriate noise distribution is a convex combination of the boundaries $F_{V'}(\cdot - \eta_1)$ and $F_{V'}(\cdot - \eta_2)$. Thus, the boundaries of these uncertainty classes often contain the essential information for robust estimation. An intuitive interpretation of this reduction of robust estimation to standard estimation based on boundary distributions is that the boundaries represent the worst-possible noise. □

Example 4.3 A model for uncertainty in both the origin of a sensor and the scale of its precision is

$$Z = \theta + \sigma V' + \eta, \qquad \sigma \in (0, \sigma_m],\ \eta \in [\eta_1, \eta_2].$$

With $V := \sigma V' + \eta$,

$$F_V \in \{F : F(x) = F_{V'}((x - \eta)/\sigma),\ 0 < \sigma \leq \sigma_m,\ \eta_1 \leq \eta \leq \eta_2\}. \ \square$$

Example 4.4 A model for sporadic interference or noise is ϵ-contamination. The distribution function of ϵ-contaminated noise V is $F_V = (1 - \epsilon)\Phi + \epsilon\Psi$, where Φ is a known distribution function, Ψ is an unknown distribution function, and small $\epsilon \in (0, 1)$ is known: With high probability $1 - \epsilon$ the distribution of V is the known distribution Φ, but with low probability ϵ the distribution of V is contaminated by an unknown distribution Ψ and is thus uncertain. The corresponding uncertainty class for F_V is

$$\mathcal{F} = \{F : (1 - \epsilon)\Phi(x) < F(x) < (1 - \epsilon)\Phi(x) + \epsilon\}.$$

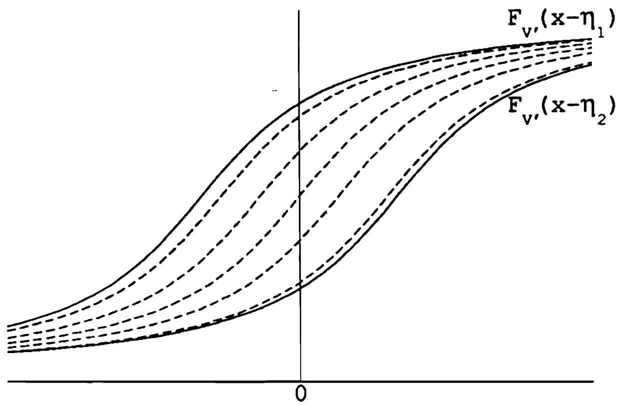

Figure 5: An uncertainty class for origin

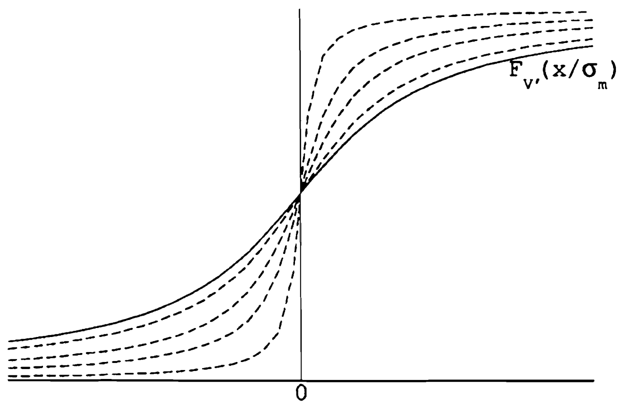

Figure 6: An uncertainty class for scale

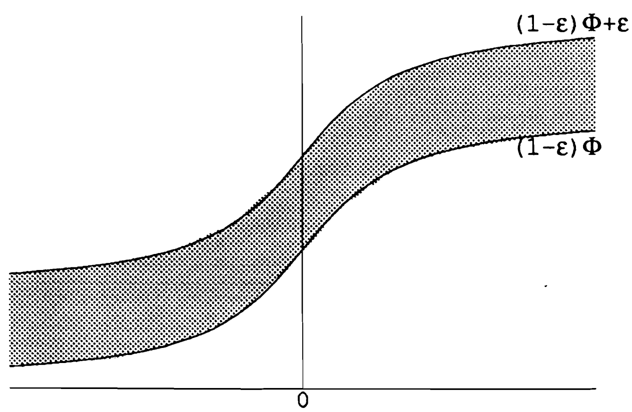

Figure 7: An uncertainty class for ϵ-contamination

(See figure 7, p. 9.) The lower boundary $(1-\epsilon)\Phi(\cdot)$ corresponds to contamination concentrated at infinity, and the upper boundary $(1-\epsilon)\Phi(\cdot)+\epsilon$ corresponds to contamination concentrated at negative infinity:

$$(1-\epsilon)\Phi(x) = (1-\epsilon)\Phi(x) + \epsilon \lim_{y\to\infty} \Psi_y(x)$$

$$(1-\epsilon)\Phi(x) + \epsilon = (1-\epsilon)\Phi(x) + \epsilon \lim_{y\to-\infty} \Psi_y(x)$$

$$\Psi_y(x) := \begin{cases} 0 & \text{if} \quad x < y \\ 1 & \text{if} \quad x \geq y \end{cases} \quad \Box$$

Remark In some cases, an uncertainty class for ϵ-contamination is a generalized uncertainty class for origin or for scale if substochastic distributions are permitted. A *substochastic distribution function* F has the following properties:

1. F is non-decreasing.
2. F is right continuous.
3. $0 \leq F(x) \leq 1$ for all x.
4. $\int_{\Re} dF(x) < 1$
5. $\lim_{x\to-\infty} F(x) > 0$ or $\lim_{x\to\infty} F(x) < 1$.

A substochastic distribution assigns less than unit probability to the real line $\Re$. An interpretation is that such a distribution has probability mass at negative or positive infinity. Also, a substochastic distribution formalizes the notion of infinitely bad outliers.[2]

Example 4.5 The uncertainty class for ϵ-contamination of example 4.4 can be expressed as a generalized uncertainty class for origin provided that substochastic distributions are allowed. In particular, the upper and lower boundary distributions are

$$F_{b_l}(x) := (1-\epsilon)\Phi(x) \quad \text{and} \quad F_{b_u}(x) := (1-\epsilon)\Phi(x) + \epsilon.$$

Consequently, $\mathcal{F}$ is a generalized uncertainty class for origin:

$$\mathcal{F} \subset \{F : F_{b_l}(x) \leq F(x) \leq F_{b_u}\}$$

The boundary distributions are substochastic:

$$\int_{\Re} dF_{b_l}(x) = \int_{\Re} dF_{b_u}(x) = 1 - \epsilon.$$

Also, $F_{b_l}(\infty) = 1 - \epsilon$ and $F_{b_u}(-\infty) = \epsilon$. $\Box$

[2] P.J. Huber, *Robust Statistical Procedures*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1977, p. 30.

Example 4.6 The uncertainty class for ϵ-contamination of example 4.4 is a generalized uncertainty class for scale if the contamination is restricted to symmetric distributions Ψ: $\Psi(-x^+)+\Psi(x^+) = 1$ for all x. In this case, the extremes of contamination correspond to concentration of the probability mass at zero or to concentration of half of the mass at negative infinity and half of the mass at infinity. Any distribution $F \in \mathcal{F}$ satisfies these bounds:

$$\forall x < 0, \qquad (1-\epsilon)\Phi(x) < F(x) < (1-\epsilon)\Phi + \epsilon/2$$

$$\forall x \geq 0, \ (1-\epsilon)\Phi(x) + \epsilon > F(x) > (1-\epsilon)\Phi + \epsilon/2$$

Thus if $F_b(x) := (1-\epsilon)\Phi(x) + \epsilon/2$, which represents concentration of the mass at positive and negative infinities, then

$$\mathcal{F} \subset \{F : F(x) \leq F_b(x) \text{ for } x < 0 \text{ and } F(x) \geq F_b(x) \text{ for } x \geq 0\}.$$

The boundary distribution F_b is substochastic. □

5 Statistical Decision Theory

Statistical decision theory is the mathematical framework for analysis of the location-estimation problem in the GRASP Lab. This section introduces this theory and formulates the location-estimation model as a decision problem. The principal references for this discussion are [Ferguson, 1967] and [Berger, 1985]. Secondary references include [DeGroot, 1970], [Wald, 1971], and [Bickel and Doksum, 1977, Chapter 10].

The Decision Problem

Figure 8 (p. 12) illustrates the structure of a statistical decision problem. The task is to make a decision or perform some action a from a set $\mathcal{A}$ of allowable actions. The parameter ω determines the correct action to take, but the value of this parameter is not known. There are, however, two types of information about ω. First, the possible values are known. These are the elements of the set Ω. Second, there is an observable random variable Z whose distribution depends on ω and thus contains statistical information about ω. The goal of a decision problem is to choose an action from $\mathcal{A}$ by using the observable to gain information about the unknown parameter. The objective is to find a *decision rule* δ that maps the sample space $\mathcal{Z}$ of the observable Z to the action space $\mathcal{A}$: The decision or action for an observation $Z = z$ is $\delta(z) \in \mathcal{A}$. Because the action taken is based on a random variable, the decision process has error. The loss function L gives the penalty for this error: The loss incurred by action a for the parameter ω is $L(\omega, a)$.

In summary, a decision problem is a quadruple $(\Omega, \mathcal{A}, L, Z)$ consisting of a parameter space Ω, an action space $\mathcal{A}$, a loss function L, and an observable Z. The

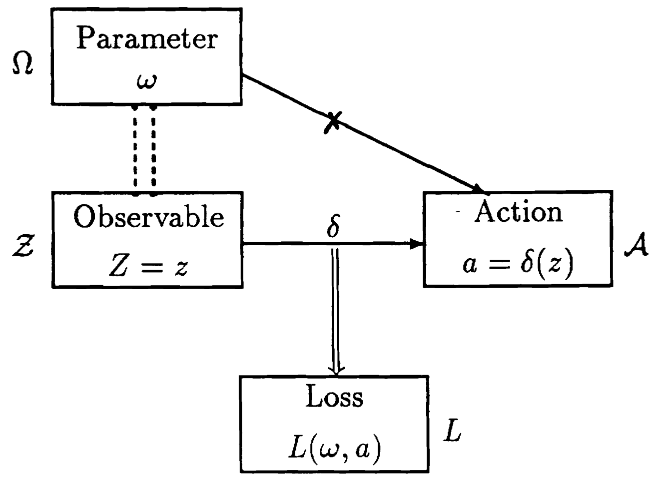

Figure 8: A statistical decision problem

parameter space is the set of possible values for the unknown statistical parameters. For standard estimation, the parameter space is $\Omega = \Theta$ and $\omega = \theta$. For robust estimation, the parameter space is $\Omega = \Theta \times \mathcal{F}$ and $\omega = (\theta, F_V)$. The *action space* is the set of available decisions. The action space of the location-estimation problem is $\mathcal{A} = \Theta$; an action $a \in \mathcal{A}$ is an estimate of θ. The *loss function* is a scalar function on $\Omega \times \mathcal{A}$. The loss $L(\omega, a)$ for $\omega \in \Omega$ is the cost of the estimate a of θ. This research uses the *zero-one (e) loss* function, L_e, to model error tolerance:

$$L_e(\omega, a) := \begin{cases} 0 & \text{if} \quad ||\theta - a|| \leq e \\ 1 & \text{if} \quad ||\theta - a|| > e \end{cases}$$

The *observable* is a random variable whose distribution depends on the unknown parameters and thus contains information about them. For the location-estimation problem, the observable is $Z = \theta + V$.

A decision rule $\delta(Z)$ in an estimation problem is an estimator of θ. The decision rule is chosen according to an optimality criterion. This research constructs minimax decision rules: Under zero-one (e) loss, an estimator $\delta^*(Z)$ of the location parameter θ is minimax if

$$\sup_{\omega} \mathrm{P}_\omega\{||\delta^*(Z) - \theta|| > e\} = \inf_{\delta} \sup_{\omega} \mathrm{P}_\omega\{||\delta(Z) - \theta|| > e\}.$$

Thus, a minimax estimator based on zero-one (e) loss minimizes the maximum probability that the absolute error of estimation is greater than the error tolerance e. Equivalently, a minimax estimator minimizes the maximum probability of unacceptable error.

Remark The zero-one (e) loss function models the notion of acceptable *versus* unacceptable estimates. An estimate a of θ has acceptable error if $||\theta - a|| \leq e$

($L_e(\theta, a) = 0$); otherwise, the estimate has unacceptable error ($L_e(\theta, a) = 1$). The minimax criterion under zero-one loss models the goal of minimizing the maximum probability of unacceptable error. □

Optimal Decision Rules

An optimal decision rule is in some sense better than or preferable to other decision rules. A decision rule δ_1 is preferable to a decision rule δ_2 if the loss under δ_1 is smaller than the loss under δ_2. The loss function alone, however, is not enough to choose between two decision rules since $L(\omega, \delta(Z))$ is a random variable. Consequently, the first step in evaluating the performance of a decision rule δ is to find its average loss or *risk* $R(\omega, \delta)$:

$$\begin{aligned} R(\omega, \delta) &:= \mathrm{E}[L(\omega, \delta(Z))] \\ &= \int_{Z} L(\omega, \delta(z))\, dF_Z(z|\theta) \end{aligned}$$

The risk $R(\omega, \delta)$ is the weighted-average loss of δ, where the weight is given by the distribution $F_Z(\cdot|\theta)$.

Example 5.1 When the loss is zero-one (e), the risk of a rule δ is the probability under ω that the absolute error exceeds e:

$$\begin{aligned} R(\omega, \delta) &= \int_{Z} L_e(\theta, \delta(z))\, dF_Z(z|\theta) \\ &= \int_{\{z:||\delta(z)-\theta||>e\}} dF_Z(z|\theta) \\ &= \mathrm{P}_\omega\{||\delta(Z) - \theta|| > e\} \end{aligned}$$

Thus, small risk implies small probability of unacceptable error of estimation. □

Comparison of risk gives the weak optimality criterion of admissibility. A decision rule δ_1 is better than a decision rule δ_2 if the risk of δ_1 is smaller than the risk of δ_2 uniformly in ω:

$$\text{For all } \omega \in \Omega,\ R(\omega, \delta_1) \leq R(\omega, \delta_2).$$

$$\text{For some } \hat{\omega} \in \Omega,\ R(\hat{\omega}, \delta_1) < R(\hat{\omega}, \delta_2).$$

A decision rule is *inadmissible* if there is another rule that is better than it. A decision rule is *admissible* if there is no better rule. Admissibility, however, is an incomplete criterion since the risk varies in the unknown parameter ω. (See figure 9.) Consequently, the second step in finding a decision rule is to remove the dependence of a choice on the unknown parameter. This step leads to three types of decision rules: minimax, Bayes, and equalizer.

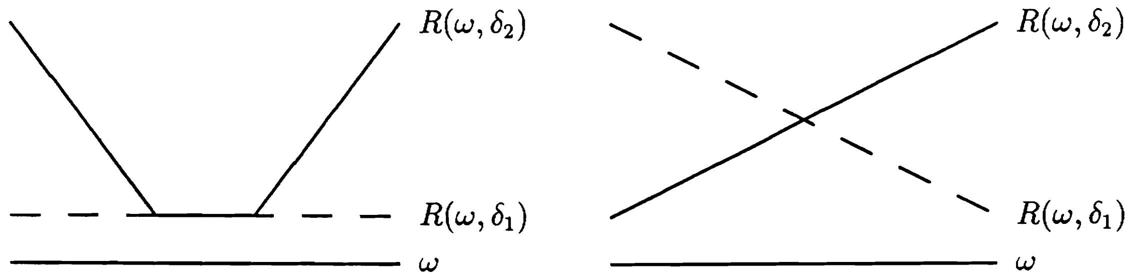

Figure 9: Complete and incomplete comparison of decision rules through risk

The minimax approach eliminates the unknown parameter ω from the risk by comparing the maximum risks of two decision rules. A decision rule δ^* is a *minimax* rule if its maximum risk is the smallest possible maximum risk:

$$\sup_\omega R(\omega, \delta^*) = \inf_\delta \sup_\omega R(\omega, \delta)$$

Thus, a minimax rule guards against the worst-possible risk.

Example 5.2 A decision rule δ^* in a standard-estimation problem is minimax if for all rules δ,

$$\sup_\theta R(\theta, \delta^*) \leq \sup_\theta R(\theta, \delta).$$

A decision rule δ^* in a robust-estimation problem is minimax if for all rules δ,

$$\sup_{(\theta, F_V)} R((\theta, F_V), \delta^*) \leq \sup_{(\theta, F_V)} R((\theta, F_V), \delta). \square$$

Example 5.3 Example 5.1 shows that under zero-one (e) loss the risk of a decision rule δ is

$$R(\omega, \delta) = \mathrm{P}_\omega\{||\delta(Z) - \theta|| > e\}.$$

Consequently, a decision rule is minimax under zero-one (e) loss if and only if

$$\sup_\omega \mathrm{P}_\omega\{||\delta^*(Z) - \theta|| > e\} = \inf_\delta \sup_\omega \mathrm{P}_\omega\{||\delta(Z) - \theta|| > e\}. \square$$

The Bayes approach eliminates ω by comparing the weighted-average risks of two decision rules. This approach assumes that there is a known probability distribution $\boldsymbol{\pi}$ on the parameter space Ω through which the risks are averaged. This distribution is a *prior distribution* on Ω. A decision rule δ_π^* is *Bayes* against a prior $\boldsymbol{\pi}$ if its weighted-average risk under $\boldsymbol{\pi}$ is the smallest possible weighted-average risk under $\boldsymbol{\pi}$:

$$\mathrm{E}[R(\omega, \delta_\pi^*)] = \inf_\delta \mathrm{E}[R(\omega, \delta)]$$

Thus, a Bayes rule has best average performance.

The equalizer approach eliminates ω by choosing a decision rule with constant risk. A decision rule δ is an equalizer rule if for all $\omega \in \Omega$,

$$R(\omega, \delta) = \text{constant}.$$

Constant risk is not by itself an optimality criterion, but minimax rules are often equalizer rules or almost equalizer rules.

Remark A minimax or Bayes rule is optimal with respect to the class $\mathcal{D}$ of all decision rules $\delta : \mathcal{Z} \to \mathcal{A}$. (The definitions take the infimum over all $\delta \in \mathcal{D}$.) It is sometimes useful, however, to find a decision rule which is optimal only with respect to a restricted class $\mathcal{D}_r \subset \mathcal{D}$ of decision rules. In particular, a decision rule $\delta_r^* \in \mathcal{D}_r$ is $\mathcal{D}_r$-minimax if for all decision rules $\delta_r \in \mathcal{D}_r$,

$$\sup_\omega R(\omega, \delta_r^*) \leq \sup_\omega R(\omega, \delta_r).$$

Similarly, a decision rule δ_r^* is $\mathcal{D}_r$-Bayes against a prior distribution $\boldsymbol{\pi}$ if for all decision rules $\delta_r \in \mathcal{D}_r$,

$$\mathrm{E}[R(\omega, \delta_r^*)] \leq \mathrm{E}[R(\omega, \delta_r)]. \square$$

The goal of this research is to find a minimax rule for the location parameter θ of the measurement $Z = \theta + V$, but direct computation of a minimax rule from the definition is usually not possible. Instead, the Bayes and equalizer approaches provide an indirect route to minimax rules. A standard result from statistical decision theory states that a Bayes equalizer rule is minimax:

Theorem 1 *Let $\boldsymbol{\pi}$ be a distribution on Ω, and suppose that the decision rule δ is Bayes against $\boldsymbol{\pi}$. If δ is an equalizer rule, then δ is minimax.*

Proof See [Ferguson, 1967, p. 90] or [Berger, 1985, p. 350]. □

Theorem 1 gives a strategy for finding a minimax rule. The first step is to hypothesize a class of decision rules likely to contain a minimax rule. The structure of the decision problem usually suggests several candidates. The second step, if possible, is to find an equalizer rule δ in this class. The final step, if possible, is to construct a distribution $\boldsymbol{\pi}$ on Ω such that δ is Bayes against $\boldsymbol{\pi}$. Theorem 2 gives an extension of this strategy.

Theorem 2 *Let $\boldsymbol{\pi}$ be a distribution on Ω, and suppose that the decision rule δ is Bayes against $\boldsymbol{\pi}$. Let $C = \sup_\omega R(\omega, \delta)$. If $\mathrm{P}\{\omega : R(\omega, \delta) = C\} = 1$, then δ is minimax.*

Proof See [Ferguson, 1967, p. 90] or [Berger, 1985, p. 350]. □

Theorem 2 relaxes the constraint of constant risk in theorem 1 by permitting some points to have risk less than the maximum risk. It requires instead that there are not too many such points in the sense that the probability of the totality of these points is zero under the prior distribution on Ω. For example, if Ω is a continuum of points, then the theorem permits the risk to vary from the maximum risk for any countable number of points because the probability of a countable set is zero under a continuous distribution. A decision rule is *almost an equalizer rule* if it satisfies the hypotheses of theorem 2 and it is not an equalizer rule.

The probability distribution of these theorems is a mathematical tool; it does not necessarily have an interpretation as prior knowledge about the parameter. It is a least-favorable distribution: A distribution $\boldsymbol{\pi}_0$ on Ω is *least favorable* if

$$\inf_{\delta} \mathrm{E}^{\pi_0}[R(\omega,\delta)] = \sup_{\pi} \inf_{\delta} \mathrm{E}^{\pi}[R(\omega,\delta)].$$

(The superscripts indicates the distribution on Ω.) Thus, the Bayes risk against a least-favorable distribution is the worst-possible Bayes risk against any distribution.

Computation of a Bayes rule is usually easier than computation of a minimax rule from the definition. Theorem 3 outlines a strategy for finding a Bayes rule:

Theorem 3 *Let $\boldsymbol{\pi}$ be a distribution on Ω, and let $\boldsymbol{\pi}(\cdot|z)$ be the conditional distribution on Ω given the observation $Z = z$. If for all z,*

$$\mathrm{E}^{\pi(\cdot|z)}[L(\omega,\delta(z))] = \inf_{a} \mathrm{E}^{\pi(\cdot|z)}[L(\omega,a)],$$

then δ is Bayes against $\boldsymbol{\pi}$.

Proof See [Ferguson, 1967, pp. 43–45] or [Berger, 1985, p. 159]. □

The conditional distribution $\boldsymbol{\pi}(\cdot|z)$ on Ω is the *posterior distribution* on Ω. The expected value under $\boldsymbol{\pi}(\cdot|z)$ of the loss $L(\omega,a)$ is the *posterior expected loss* of an action a. Thus, theorem 3 states that a Bayes rule minimizes the posterior expected loss under the corresponding posterior distribution.

Example 5.4 For a standard-estimation problem with a prior density p on a continuous parameter space $\Omega = \Theta$, the conditional density on Θ given $Z = z$ is

$$p(\theta|z) = \frac{f_Z(z|\theta)p(\theta)}{g_Z(z)} \quad \text{where} \quad g_Z(z) := \int_{\Theta} f_Z(z|\theta)p(\theta)\,d\theta,$$

provided that $g_Z(z)$ is positive. Thus the posterior expected loss of an action a is

$$\int_{\Theta} L_e(\theta,a)\,\frac{f_Z(z|\theta)p(\theta)}{g_Z(z)}\,d\theta.$$

Accordingly, theorem 3 implies that a decision rule δ is Bayes against p if for all z,

$$\int_\Theta L_e(\theta, \delta(z))\, f_Z(z|\theta)p(\theta)\, d\theta = \inf_a \int_\Theta L_e(\theta, a)\, f_Z(z|\theta)p(\theta)\, d\theta.$$

Thus in practice, the strategy for finding a Bayes action corresponding to an observation $Z = z$ is to minimize

$$\int_\Theta L_e(\theta, a)\, f_Z(z|\theta)p(\theta)\, d\theta$$

over all possible actions a. □

Features of Decision Theory

This decision-theoretic formulation of the location problem has several features. First, standard estimation and robust estimation coincide within the framework of statistical decision theory. The only difference is the specification of the parameter space: $\Omega = \Theta$ or $\Omega = \Theta \times \mathcal{F}$. The tools of statistical decision theory, however, apply to either specification. Second, decision theory incorporates prior information about the unknown parameters through the minimax criterion by optimizing over $\omega \in \Omega$. Third, a decision problem accounts for the consequences of the estimate through the loss function. Zero-one (e) loss, in particular, models error tolerance: An estimate within e of θ is sufficiently close and so incurs no penalty, and an estimate greater than e from θ is too far and thus incurs full penalty. Also, zero-one loss is independent of the distribution F_V. Finally, a minimax estimator $\delta^*(Z)$ based on zero-one (e) loss induces an optimal fixed-size ($2e$) confidence procedure that maximizes the confidence coefficient among all fixed-size ($2e$) confidence procedures. This fixed-size confidence procedure induced by an estimator δ of θ is

$$\mathrm{C}_\delta(Z) := [\delta(Z) - e, \delta(Z) + e].$$

The confidence coefficient is $\inf_\omega \mathrm{P}_\omega\{\mathrm{C}_\delta(Z) \ni \theta\}$, where $\mathrm{P}_\omega\{\mathrm{C}_\delta(Z) \ni \theta\}$ is the probability under ω that the confidence interval covers θ. If δ^* is a minimax rule, then

$$\inf_\omega \mathrm{P}_\omega\{\mathrm{C}_{\delta^*}(Z) \ni \theta\} = \sup_\delta \inf_\omega \mathrm{P}_\omega\{\mathrm{C}_\delta(Z) \ni \theta\}.$$

Thus, a confidence procedure based on a minimax estimator under zero-one loss maximizes the minimum probability of covering θ. This confidence procedure also provides a test of hypothesis that two measurements are consistent.

6 Examples of Research Problems

This section gives examples of research problems in statistical decision theory motivated by the sensor-fusion problem. These examples are specific instances of the general results summarized in section 7.

Example 6.1 This example gives a minimax rule for the location or mean θ of a measurement $Z \sim \mathcal{N}(\theta, 1)$ with $\theta \in \{-1, 0, 1\}$ when the error tolerance e is 0.

The random variable Z has the structure $Z = \theta + V$ where $F_V \sim \mathcal{N}(0,1)$. The possible values of θ are the elements of $\Theta = \{-1, 0, 1\}$. This example is a standard-estimation problem since the noise distribution F_V is known. Thus $\Omega = \Theta$ and $\omega = \theta$. Also, the action space $\mathcal{A}$ is Θ. The loss function is the zero-one (0) loss function:

$$L_0(\theta, a) := \begin{cases} 0 & \text{if} \quad a = \theta \\ 1 & \text{if} \quad a \neq \theta \end{cases}$$

The minimax decision rule δ^* is this:

$$\delta^*(z) = \begin{cases} -1 & \text{if} \quad z \leq -0.803 \\ 0 & \text{if} \quad -0.803 < z < 0.803 \\ 1 & \text{if} \quad 0.803 \leq z \end{cases}$$

The point $x = 0.803$ is the unique solution of the equation

$$2F_V(-x) = F_V(x-1).$$

This rule implies, for example, that the estimate corresponding to the observation $Z = 0.5$ is $\hat{\theta} = 0$. Similarly, the estimate corresponding to any observation $Z \geq 0.803$ is $\hat{\theta} = 1$.

The risk function of δ^* is this:

$$\begin{aligned} R(-1, \delta^*) &= 1 - F_V(-0.803 + 1) \\ R(0, \delta^*) &= 2F_V(-0.803) \\ R(1, \delta^*) &= F_V(0.803 - 1) \end{aligned}$$

This decision rule is an equalizer rule with risk 0.422.

Furthermore, the rule δ^* is Bayes against the distribution on Θ with probabilities $p(0) = 0.4036$ and $p(\pm 1) = 0.2982$. This distribution is a least-favorable distribution.

(See [McKendall, 1990] for the analysis underlying this example and for similar problems in standard estimation.) □

Example 6.2 This example gives a minimax rule for the location θ of a Cauchy measurement $Z \sim \mathcal{C}(\theta, 1)$ with $\theta \in \{-1, 0, 1\}$ when the error tolerance e is 0. The distribution and density of a Cauchy random variable $V \sim \mathcal{C}(0,1)$ are

$$F_V(v) = \frac{1}{\pi}\arctan(v) + \frac{1}{2} \quad \text{and} \quad f_V(v) = \frac{1}{\pi(1+v^2)}.$$

The difference between this example and example 6.1 is the noise distribution.

The minimax decision rule δ^* is this:

$$\delta^*(z) = \begin{cases} -\delta^*(-z) & \text{if} \quad z < 0 \\ 0 & \text{if} \quad 0 \leq z < 0.739 \\ 1 & \text{if} \quad 0.739 \leq z < 5.733 \\ 0 & \text{if} \quad 5.733 \leq z \end{cases}$$

Thus the minimax estimate is $\hat{\theta} = 0$ for small positive observations ($0 < z < 0.739$) and $\hat{\theta} = 1$ for large observations ($0.739 \leq z < 5.733$). For very large observations ($z \geq 5.733$), however, the minimax estimate is again $\hat{\theta} = 0$.

The risk function of δ^* is this:

$$\begin{aligned} R(0, \delta^*) &= 2F_V(\mu(x^*)) - 2F_V(x^*) \\ R(\pm 1, \delta^*) &= F_V(x^* - 1) + F_V(1 - \mu(x^*)) \end{aligned}$$

$$\mu(x) := (x-2)/(2x-1), \qquad x \neq 1/2$$

Here $x^* := 0.739$ satisfies the equation

$$2F_V(\mu(x)) - 2F_V(x) = F_V(x-1) + F_V(1 - \mu(x)).$$

Thus, this decision rule is an equalizer rule with risk 0.485.

The rule δ^* is Bayes against the distribution on Θ that has probabilities $p(0) = 0.4198$ and $p(\pm 1) = 0.2901$. This distribution is least favorable.

(See [McKendall, 1990] for the analysis underlying this example.) □

Example 6.3 This example gives a minimax rule for the location θ of a measurement $Z \sim \mathcal{N}(\theta, 1)$ with $\theta \in [-0.3, 0.3]$ when the error tolerance e is 0.1.

This example is also a standard-estimation problem. The parameter space and action space both are the interval $[-0.3, 0.3]$. The zero-one (0.1) loss function is this:

$$L_{0.1}(\theta, a) := \begin{cases} 0 & \text{if} \quad |\theta - a| \leq 0.1 \\ 1 & \text{if} \quad |\theta - a| > 0.1 \end{cases}$$

The minimax decision rule δ^* is this:

$$\delta^*(z) = \begin{cases} -\delta^*(-z) & \text{if} \quad z < 0 \\ 0 & \text{if} \quad 0 \leq z < a \\ z - a & \text{if} \quad a \leq z < a + 0.2 \\ 0.2 & \text{if} \quad a + 0.2 \leq z \end{cases} \tag{1}$$

Here $a = 0.3992$ is the unique solution of the equation

$$2F_V(-a - e) = F_V(a - e).$$

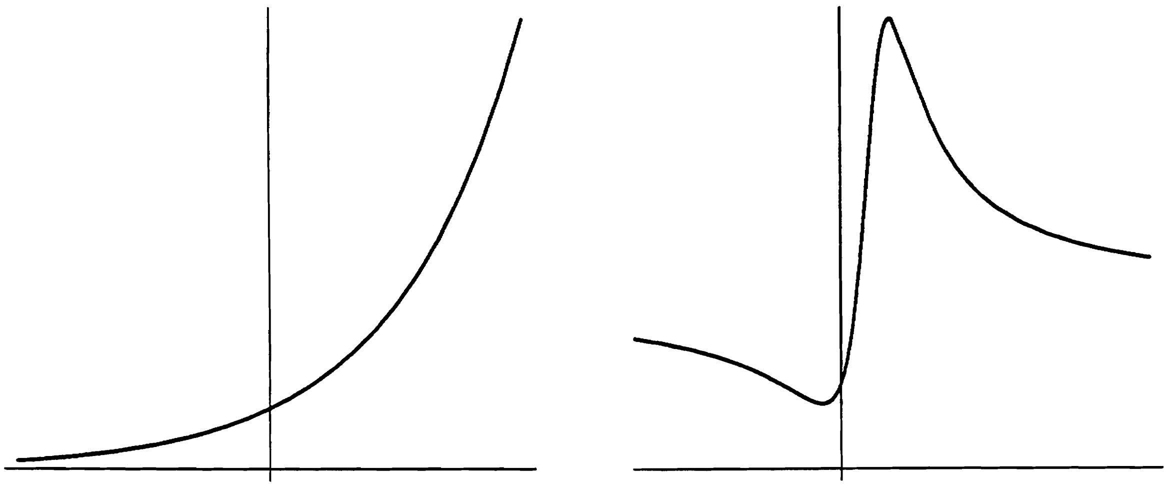

Figure 10: Likelihood ratios $f_Z(\cdot|1)/f_Z(\cdot|0)$ for the $\mathcal{N}(0,1)$ and $\mathcal{C}(0,1)$ distributions

This rule has $|\delta^*(z)| \leq 0.2$ since the error tolerance is 0.1.

The risk function of δ^* is this:

$$R(\theta,\delta^*) = \begin{cases} R(-\theta,\delta^*) & \text{if} \quad \theta < 0 \\ 2F_V(-a-0.1) & \text{if} \quad 0 \leq \theta < 0.1 \\ F_V(-a-0.1) & \text{if} \quad \theta = 0.1 \\ F_V(a-0.1) & \text{if} \quad 0.1 < \theta \leq 0.3 \end{cases}$$

This decision rule has constant risk (0.6176) except for the points $\theta = \pm 0.1$, which have smaller risk. Thus, this rule is almost an equalizer rule in the sense of theorem 2 (p. 15).

The rule δ^* is Bayes against the distribution on Θ that has this density function:

$$p(\theta) = \begin{cases} 1.62 & \text{if} \quad -0.3 \leq \theta \leq -0.1 \\ 1.76 & \text{if} \quad -.01 < \theta < 0.1 \\ 1.62 & \text{if} \quad 0.1 \leq \theta \leq 0.3 \end{cases}$$

This distribution is least favorable.

(See [Zeytinoglu and Mintz, 1984] for the analysis underlying this example.) □

Remark The minimax rules of examples 6.1, 6.2, and 6.3 are admissible. □

Remark The minimax rules of examples 6.1 and 6.3 are non-decreasing. Their monotonicity reflects the monotone likelihood ratio[3] of the $\mathcal{N}(0,1)$ noise distribution. The minimax rule of example 6.2 is not monotonic. Its behavior reflects the non-monotonic behavior of the $\mathcal{C}(0,1)$ likelihood ratio. (See figure 10, p. 20.) □

[3] A random variable Z with a density function $f_Z(\cdot|\theta)$, for $\theta \in \Theta$, has a *monotone likelihood ratio* if the ratio $f_Z(\cdot|\theta_1)/f_Z(\cdot|\theta_2)$ is non-decreasing for all $\theta_1 > \theta_2$.

Example 6.4 This example extends example 6.3 to multiple measurements. It gives a minimax estimator $\delta_3^*(Z_1, Z_2, Z_3)$ for θ when there are three independent and consistent measurements Z_1, Z_2, Z_3 with $Z_i \sim \mathcal{N}(\theta, 1)$.

The solution is based on the sample mean $(Z_1 + Z_2 + Z_3)/3$ and its centered distribution $F' \sim \mathcal{N}(0, 1/3)$. With $\boldsymbol{z} := (z_1, z_2, z_3)$ and $T(\boldsymbol{z}) := (z_1 + z_2 + z_3)/3$, a minimax decision rule δ_3^* is this:

$$\delta_3^*(\boldsymbol{z}) = \begin{cases} -\delta^*(-T(\boldsymbol{z})) & \text{if} \quad T(\boldsymbol{z}) < 0 \\ 0 & \text{if} \quad 0 \leq T(\boldsymbol{z}) < a', \\ T(\boldsymbol{z}) - a' & \text{if} \quad a' \leq T(\boldsymbol{z}) < a' + 0.2 \\ 0.2 & \text{if} \quad a' + 0.2 \leq T(\boldsymbol{z}) \end{cases}$$

Here a' is the unique solution of the equation $2F'(-a' - e) = F'(a' - e)$. Thus, this rule is the composition $\delta^* \circ T$ of the minimax rule δ^* for the single-sample problem with noise $V \sim F'$ and the sample-mean statistic T.

(The assumption of Gaussian noise is essential to this example because the sample mean of independent $\mathcal{N}(\theta, 1)$ random variables is a sufficient statistic for θ. There are no similar results supporting composition of a single-sample minimax rule with the sample mean when sample mean is not sufficient for the location parameter. See [Zeytinoglu and Mintz, 1988] for the analysis and for other results composing a single-sample rule with a scalar statistic in multi-sample problems.) □

Example 6.5 This example gives a minimax rule for the location $\theta \in [-0.3, 0.3]$ of a measurement $Z \sim \mathcal{N}(\theta, \sigma^2)$ with some uncertain scale $\sigma \leq 0.25$ when the error tolerance is 0.1.

This example is a robust-estimation problem since the scale σ and hence the noise distribution $F_V \sim \mathcal{N}(0, \sigma^2)$ are uncertain. The uncertainty class is

$$\mathcal{F} = \{\mathcal{N}(0, \sigma^2),\ \sigma \leq 0.25\}.$$

The parameter space Ω is $\Theta \times \mathcal{F}$ or, equivalently, $[-0.3, 0.3] \times (0, 0.25]$. The action space and loss function are the same as those of example 6.3.

This problem reduces to a standard-estimation problem since the largest possible scale is sufficiently small relative to the error tolerance. The minimax rule for this example is the minimax rule for the standard-estimation problem of example 6.3 with the noise distribution replaced by $\mathcal{N}(0, 0.25^2)$. In particular, the minimax rule is given by definition 1 (p. 19) with $a = 0.0808$.

(See [Zeytinoglu and Mintz, 1988] for the analysis. See [Martin, 1987] for a similar problem in which the largest possible scale is too large for the problem to reduce to standard estimation.) □

Example 6.6 This example gives a minimax rule for the location $\theta \in \{-u, 0, u\}$ of a measurement $Z = \theta + V$ when the noise distribution F_V is uncertain. The uncertainty class $\mathcal{F}$ is a generalized uncertainty class for scale with boundary $\Phi \sim \mathcal{N}(0,1)$:

$$\mathcal{F} = \{F : F(v) \leq \Phi(v) \text{ for } v < 0 \text{ and } F(v) \geq \Phi(v) \text{ for } v \geq 0\}.$$

(See example 4.2, p. 7.) The error tolerance is $e = 0$.

This problem also reduces to standard estimation provided that the unit u is sufficiently large ($u \geq -2\Phi^{-1}(1/4)$). If $u = 1.5$, for example, then solution of this problem is similar to solution of the example 6.1:

$$\delta^*(z) = \begin{cases} -1 & \text{if } z \leq -1.011 \\ 0 & \text{if } -1.011 < z < 1.011 \\ 1 & \text{if } 1.011 \leq z \end{cases}$$

The point $x = 1.011$ is the unique solution to the equation $2\Phi(-x) = \Phi(x-u)$. This decision rule is not an equalizer rule, but it is almost an equalizer rule in the sense of theorem 2. (See [McKendall, 1990] for the analysis.) □

Example 6.7 This example is a robust-estimation problem which does not reduce to standard estimation. The possible locations are $\{-1, 0, 1\}$ and the error tolerance is zero. The uncertainty class is

$$\mathcal{F} := \{\mathcal{N}(0,1), \mathcal{N}(0, 2.5^2)\}.$$

The largest possible scale (2.5) is too large for the problem to reduce to standard estimation.

A minimax rule δ^* is the following:

$$\delta^*(z) := \begin{cases} -\delta^*(-z) & \text{if } z < 0. \\ 0 & \text{if } 0 \leq z < x_1 \\ 1 & \text{if } x_1 \leq z < x_2 \\ 0 & \text{if } x_2 \leq z < x_3 \\ 1 & \text{if } x_3 \leq z \end{cases}$$

(See figure 15, p. 26.) Here $x_1 := 1.09833$, $x_2 := 2.59355$, and $x_3 := 3.095$.

This decision rule is Bayes against the prior distribution on the parameter space $\{-1, 0, 1\} \times \{1, 2.5\}$ with these probabilities:

$$\begin{aligned} p(0,1) &:= 0 \\ p(0,2.5) &:= 0.406 \\ p(\pm 1, 1) &:= 0.048 \\ p(\pm 1, 2.5) &:= 0.249 \end{aligned}$$

This distribution is least-favorable.

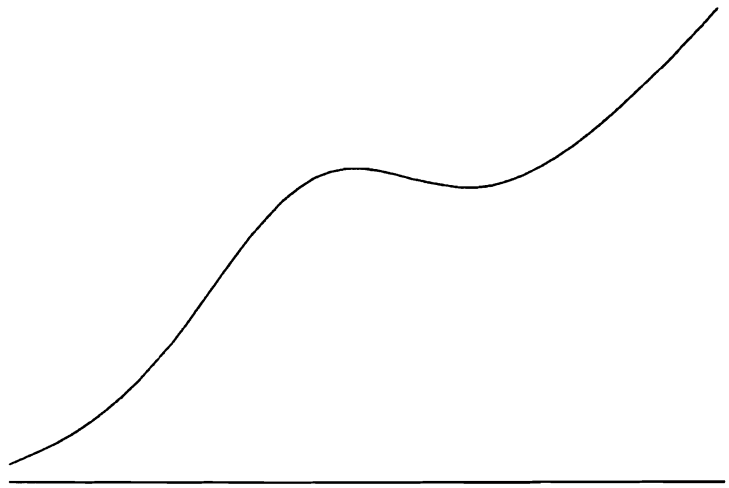

Figure 11: Likelihood ratio $\hat{f}_Z(\cdot|1)/\hat{f}_Z(\cdot|0)$

The risk function of δ^* is this:

$$R((0,1),\delta^*) = 0.26453$$

$$R((0,2.5),\delta^*) = R((\pm 1,1),\delta^*) = R((\pm 1,2.5),\delta^*) = 0.576597$$

The rule δ^* is almost an equalizer rule: Although the risk for the parameter $(0,1)$ is less than the equalized risk for the other points, the probability mass for $(0,1)$ is zero under the least-favorable distribution. (See [McKendall, 1990] for the analysis.) □

Remark In the standard-estimation problems of examples 6.1, 6.2, and 6.3, the shape of the minimax rule reflects the shape of the likelihood ratio corresponding to the density of Z. In the robust-estimation problem of example 6.7, however, the non-monotonic shape of the minimax rule mimics the shape of the likelihood ratio corresponding to the *marginal* density of Z given θ under the least-favorable probability function p:

$$\hat{f}_Z(z|\theta) := f_Z(z|(\theta,1))\,p(\theta,1) + f_Z(z|(\theta,2.5))\,p(\theta,2,5), \qquad z \in \Re$$

(See figure 11 (p. 23).) □

7 Summary of Results

This section summarizes research in the GRASP Lab on decision problems relevant to the sensor-fusion paradigm. It outlines the main results of [Zeytinoglu and Mintz,

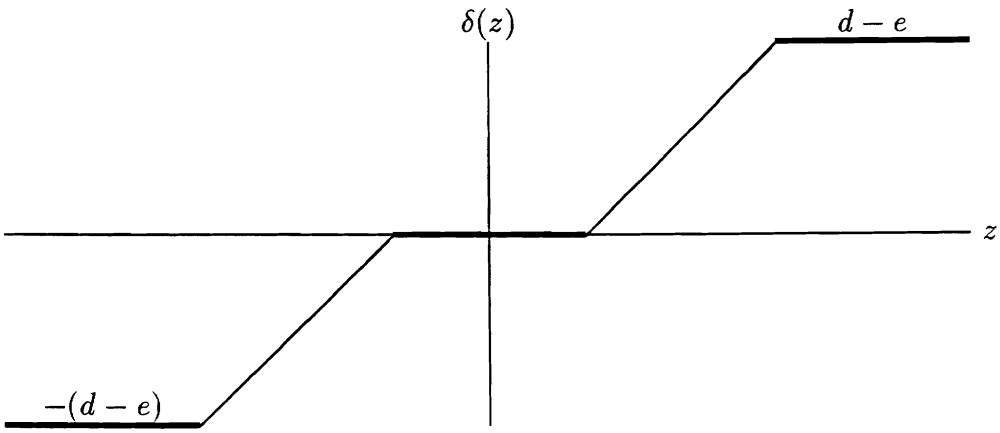

Figure 12: A minimax rule of [Zeytinoglu and Mintz, 1984]

1984], [Zeytinoglu and Mintz, 1988], [Martin, 1987], [McKendall, 1990], and [Kamberova and Mintz, 1990]. The research problems are decision problems in which Θ is a finite interval $[-d, d]$ or its discrete analog $\{0, \pm u, \ldots, \pm Nu\}$. In standard-estimation problems, the minimal assumptions on the noise are that V has a continuous, increasing distribution function and a positive density function. The uncertainty classes $\mathcal{F}$ in robust-estimation problems are uncertainty classes for origin, scale, or ϵ-contamination. Different assumptions about the parameters e, d or N, and F_V or $\mathcal{F}$ define the specific research problems studied.

Zeytinoglu and Mintz, 1984 The standard-estimation problem of [Zeytinoglu and Mintz, 1984] has parameter space $[-d, d]$. The noise distribution has a unimodal, symmetric density function and a monotone likelihood ratio. The minimax rule δ is an odd, continuous, non-decreasing, piecewise-linear function. Its linear elements alternate between segments of unit slope and segments of zero slope. (See figure 12, p. 24.) The changes occur at points determined by the solution of a nonlinear system of equations in the noise distribution F_V. Without the assumption that F_V has a monotone likelihood ratio, the decision rule δ is a restricted minimax rule in the class of odd, non-decreasing, non-randomized decision rules.

Zeytinoglu and Mintz, 1988 In [Zeytinoglu and Mintz, 1988], the standard-estimation problem of [Zeytinoglu and Mintz, 1984] is extended to robust estimation. The uncertainty class is a generalized uncertainty class for scale in which the boundary distribution Φ has a symmetric, unimodal density with a monotone likelihood ratio. If e exceeds a bound e^* depending on d/e and Φ, then the decision rule δ of [Zeytinoglu and Mintz, 1984] with $F_V = \Phi$ is minimax. Also, the rule δ is restricted minimax in the sense of [Zeytinoglu and Mintz, 1984].

Martin, 1987 In [Martin, 1987], the robust-estimation problem of [Zeytinoglu and Mintz, 1988] with $e < e^*$ is studied through a specific example in which $d/e = 3$ and Φ is the standard normal distribution, $\mathcal{N}(0,1)$. The uncertainty class is

$$\mathcal{F} := \{F : F \sim \mathcal{N}(0,\sigma_1^2) \text{ or } F \sim \mathcal{N}(0,\sigma_2^2),\ \sigma_1 < \sigma_2\}.$$

The minimax rule for this problem is a randomized decision rule. A *randomized decision rule* is a probability distribution on the space $\mathcal{D}$ of all possible decision rules. (See [Ferguson, 1967] or [Berger, 1985] for further discussion of randomization in statistical decision problems.) The randomized rule of this result confines its mass to a few translations of the minimax rule obtained in [Zeytinoglu and Mintz, 1984] with $F_V = \Phi$.

McKendall, 1990 The decision problems of [McKendall, 1990] extend the problems of [Zeytinoglu and Mintz, 1984] and [Zeytinoglu and Mintz, 1988] to the discrete parameter space $\{0, \pm u, \ldots, \pm Nu\}$, where N is a positive integer and u is a positive unit. There are two standard-estimation problems, in which the distribution function F_V is continuous and increasing on $\Re$ and has a continuous density function. In the first F_V has a monotone likelihood ratio. The corresponding minimax rules are non-decreasing step functions whose jumps have magnitude u. (See figure 13, p. 26.) In contrast, the second problem assumes that F_V is the standard Cauchy distribution, which does not have a monotone likelihood ratio. The minimax rules for this problem are also step functions with unit jumps, but they are not monotonic. (See figure 14, p. 26.) In both cases, the steps in the decision rules occur at points determined by the solution of nonlinear systems of equations in the noise distribution F_V.

There are also two robust-estimation problems, which are similar to the problems of [Zeytinoglu and Mintz, 1988] and [Martin, 1987]. The uncertainty class is the generalized uncertainty class for scale in which Φ has a monotone likelihood ratio. If u exceeds a bound u^* depending on N, e and Φ, then the decision rule δ of the first standard-estimation problem with $F_V = \Phi$ is minimax. The case $u < u^*$ is studied through specific examples in which $N = 1$, $e = 0$, and $\mathcal{F}$ is the uncertainty class of [Martin, 1987]. The minimax rules for these examples are step functions with unit jumps, but they are not monotonic. (See figure 15, p. 26.)

Kamberova and Mintz, 1990 The results of [Kamberova and Mintz, 1990] extend conclusions of [Zeytinoglu and Mintz, 1984] and [Zeytinoglu and Mintz, 1988] to problems with fewer assumptions about the noise distribution. In particular, these results do not assume that the density of F_V is symmetric or unimodal, and they do not assume that the support of F_V is the real line. These results also show how to reduce some problems in robust estimation to standard estimation for uncertainty in origin and uncertainty in origin and scale combined.

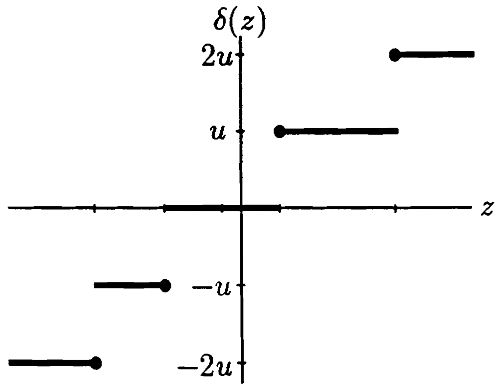

Figure 13: A minimax rule of [McKendall, 1990] when F_V has a monotone likelihood ratio

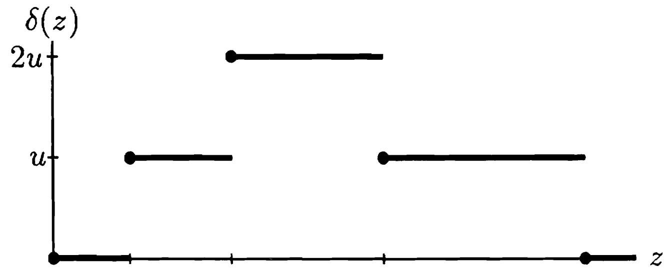

Figure 14: A minimax rule of [McKendall, 1990] when F_V is standard Cauchy ($z \geq 0$)

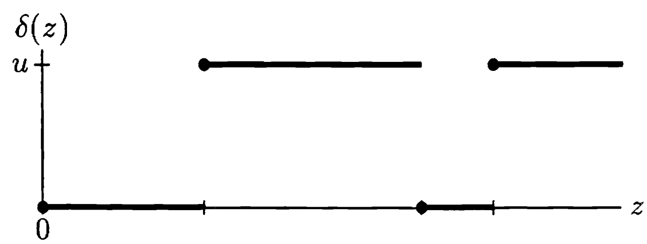

Figure 15: A minimax rule of [McKendall, 1990] when $u < u^*$ ($z \geq 0$)

8 Research Directions

The goals of this research are to model sensor fusion as a statistical problem, to analyze the model with statistical decision theory, and to develop mathematical statistics for the analysis. The past and current emphasis of this research is the third goal — finding minimax rules for decision problems motivated by sensor fusion. In particular, the current emphasis is to extend the results of section 7 to more general structures. This section outlines these directions for research.

Non-monotone likelihood ratio In practical applications, likelihood ratios are not monotone. For example, ϵ-contamination models and scale-mixture models of Gaussian distributions are both widely accepted as realistic models for representing uncertainty in sampling distributions. The likelihood ratios in these models, however, are generally not monotone. Thus an important extension is to allow noise distributions that do not have monotone likelihood ratios. Research issues include identifying classes of likelihood ratios and corresponding classes of decision rules. This extension is important in robust estimation that does not reduce to standard estimation as well as in both single-sample and multi-sample standard estimation. Initial results in this direction are included in [McKendall, 1990].

Multi-dimensional location Problems in which the space Θ of possible locations is multi-dimensional are also important to applications: $\Theta \subset \Re^2$ or $\Theta \subset \Re^3$, for example. A related problem is to jointly estimate the location parameter and scale parameter of a bivariate distribution by transforming the scale into a second location parameter through the logarithm. Research issues include both stochastic dependence among the components of an observation and algebraic dependence in a non-rectangular space of locations, such as $\{(\theta_1, \theta_2) : \theta_1^2 + \theta_2^2 \leq r^2\}$. Another issue is approximation with a minimax rule from a similar problem that does not have dependencies.

Multiple samples The multi-sample problem is an important extension of the single-sample results. It is fundamental to any application of sensor-fusion. One approach to multi-sample problems is direct computation of decision rules on the measurements. Another approach is to reduce the multi-sample problem to a single-sample problem by composing a single-sample rule with a scalar statistic on the measurements: Here, the problem is to identify an appropriate noise distribution and statistic. A third approach is to approximate a global minimax rule with a restricted minimax rule: Here, the problem is to choose the restricted class so that the restricted minimax rule is a good approximation that is easy to find. Some initial efforts using the last two approaches are included in [Zeytinoglu and Mintz, 1984].

Robust estimation One direction in robust estimation is to develop the problems of [McKendall, 1990] and [Martin, 1987] that do not reduce to standard estimation.

Another direction is to extend the results of [McKendall, 1990] to uncertainty in origin and ϵ-contamination. A general research issue is to identify and analyze other uncertainty classes.

Sensor models The first and second goals of this research — to model sensor fusion as a statistical problem and to analyze the model with statistical decision theory — provide broad directions for research. Development and testing of models for specific sensors include acquisition and analysis of data. Preliminary work analyzes statistical variation in gray-value measurements from both monochrome and color CCD cameras used in computer-vision systems. Possible additional sources of data are laser, tactile, and ultrasound sensors. Another direction is to consider different models of a sensor's measurements. An extension to the linear location-data model $Z = \theta + V$ is the non-linear model $Z = h(\theta) + V$. Both of these models are special cases of the the general structure $Z = h(\theta, V)$. Possible transformations h are truncation, saturation, and quantization Also, the function h on Θ may be uncertain, and so robust estimation must guard against this uncertainty, too.

Research tools The research philosophy embraces computational investigation in addition to traditional mathematical analysis: Numerical and symbolic computation are used heavily to gain insight into the solution of the decision problems. The computing tools of the GRASP Lab provide rich resources for this philosophy. The computing environment consists primarily of a Sun 4/280 and a VaxStation 3500 under UNIX. These are accessed through many Sun, Vax, IBM, and Hewlett-Packard workstations running X11. There is also a Connection Machine with two front ends, one for C/Paris and another for *LISP. The general-purpose programming language in the lab is C, and the main editor is GNU-Emacs. The main symbolic package is *Mathematica*, although MACSYMA is available. The main numerical packages are *Numerical Recipes in* C and IMSL. The document processor is LaTeX/TeX with PostScript.

9 Publications and Presentations

This section lists publications and presentations about sensor fusion by researchers in the GRASP Lab.

1. *Robust Multi-Sensor Fusion: A Decision-Theoretic Approach.*
   G. Kamberova and M. Mintz.
   In *Proceedings of the 1990 DARPA Image Understanding Workshop,* pp. 867–873, Morgan Kaufmann Publishers, Inc., San Mateo, CA.
   (GRASP Lab technical report MIS-CIS-90-57 (229).)
   September 1990.

2. *Statistical Decision Theory for Sensor Fusion.*
   R. McKendall.

In *Proceedings of the 1990 DARPA Image Understanding Workshop,* pp. 861–866, Morgan Kaufmann Publishers, Inc., San Mateo, CA. (GRASP Lab technical report MIS-CIS-90-55 (227).) September 1990.

3. *Non-Monotonic Decision Rules for Sensor Fusion.* R. McKendall and M. Mintz. In *Proceedings of the 1990 DARPA Image Understanding Workshop,* pp. 874–880, Morgan Kaufmann Publishers, Inc., San Mateo, CA. (GRASP Lab technical report MIS-CIS-90-56 (228).) September 1990.

4. *Minimax Estimation of a Discrete Location Parameter for a Continuous Distribution.* R. McKendall. PhD dissertation, Systems Engineering, University of Pennsylvania. (GRASP Lab technical report MIS-CIS-90-28 (214).) May 1990.

5. *Robust Fusion of Sensor Information: A Decision-Theoretic Approach.* R. McKendall. Presented to *Army Research Office Short Course on Multi-Sensor Fusion,* GRASP Lab. April 1990.

6. *Robust Sensor Fusion.* M. Mintz. Presented to *Army Research Office Short Course on Multi-Sensor Fusion,* GRASP Lab. April 1990.

7. *Robust Multi-Sensor Fusion: A Decision-Theoretic Approach.* G. Kamberova and M. Mintz. In *SPIE Conference Proceedings, Sensor Fusion II: Human and Machine Strategies*, 1198:192–201. November 1989.

8. *Robust Fusion of Sensor Information: A Decision-Theoretic Approach.* R. McKendall. Presented to *Workshop on Robust Estimation Techniques for Computer Vision,* Center for Automation Research, Department of Computer Science, University of Maryland, College Park, MD. July 1989.

Also in *GRASP News* 6(1):3–12, GRASP Lab technical report MS-CIS-90-15, Fall 1989.

9. *Confidence Regions for Restricted Location and Log-Scale Parameters for a Normal Distribution.*
   G. Kamberova and M. Mintz.
   GRASP Lab research notes.
   March 1989.

10. *Statistical Decision Theory for Sensor Integration.*
    R. McKendall.
    Presented to *Army Research Office Short Course on Multisensory Integration,* TACOM (Tank-Automotive Command), Warren, MI.
    November 1988.

11. *Robust Optimal Fixed Sized Confidence Procedures for a Restricted Parameter Space.*
    M. Zeytinoglu and M. Mintz.
    In *The Annals of Statistics*, 16(3):1241–1253.
    September 1988.

12. *Robust Fusion of Location Information.*
    R. McKendall and M. Mintz.
    In *Proceedings of the 1988 IEEE International Conference on Robotics and Automation*, pp. 1239–1255.
    April 1988.
    Also in *Technical Proceedings of the 1988 Tri-Service Data Fusion Symposium,* 1:201–207, Naval Air Development Center, Warminster, PA.
    May 1988.

13. *Randomized Robust Confidence Procedures.*
    K. Martin.
    PhD dissertation, Systems Engineering, University of Pennsylvania.
    December 1987.

14. *Models of Sensor Noise and Optimal Algorithms for Estimation and Quantization in Vision Systems.*
    R. McKendall and M. Mintz.
    GRASP Lab technical report MS-CIS-86-80.
    December 1986.

15. *Robust Minimax Location Parameter Estimation Under Zero-One Loss With a Compact Parameter Space.*
    M. Zeytinoglu.

PhD dissertation, Systems Engineering, University of Pennsylvania.
May 1985.

16. *Randomized Robust Confidence Procedures.*
    K. Martin and M. Mintz.
    In *Proceedings of the Twenty-Second Annual Allerton Conference on Communication, Control, and Computing,* pp 309–317, Department of Electrical Engineering, University of Illinois, Urbana, Il.
    October, 1984.

17. *Optimal Fixed Sized Confidence Procedures for a Restricted Parameter Space.*
    M. Zeytinoglu and M. Mintz.
    In *The Annals of Statistics*, 12(3):945–957.
    September 1984.

References

[Berger, 1985] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York, second edition, 1985.

[Bickel and Doksum, 1977] P.J. Bickel and K.A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, Inc., Oakland, CA, 1977.

[DeGroot, 1970] M.H. DeGroot. *Optimal Statistical Decisions.* McGraw-Hill Book Company, New York, 1970.

[Durrant-Whyte, 1986] H.F. Durrant-Whyte. *Integration, Coordination and Control of Multi-Sensor Robot Systems.* PhD dissertation. GRASP Lab technical report MIS-CS-86-67 (73), Department of Computer and Information Science, University of Pennsylvania, August 1986.

[Durrant-Whyte, 1988] H.F. Durrant-Whyte. *Integration, Coordination and Control of Multi-Sensor Robot Systems.* Kluwer Academic Publishers, Boston, MA, 1988.

[Ferguson, 1967] T.S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, Inc., Orlando, FL, 1967.

[Hager, 1988] G.D. Hager. *Active Reduction of Uncertainty in Multi-sensor Systems.* PhD dissertation. GRASP Lab technical report MIS-CS-88-47 (147), Department of Computer and Information Science, University of Pennsylvania, July 1988.

[Hager, 1990] G.D. Hager. *Task-Directed Sensor Fusion and Planning, A Computational Approach.* Kluwer Academic Publishers, Boston, MA, 1990.

[Kamberova and Mintz, 1990] G. Kamberova and M. Mintz. Robust multi-sensor fusion: a decision theoretic approach. In *Proceedings of the 1990 DARPA Image Understanding Workshop*, pages 867–873, Morgan Kaufmann Publishers, Inc., San Mateo, CA, September 1990. Technical report MIS-CIS-90-57 (229), Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA.

[Martin, 1987] K.E. Martin. *Randomized Robust Confidence Procedures.* PhD thesis, Systems Engineering, University of Pennsylvania, December 1987.

[McKendall, 1990] R. McKendall. *Minimax Estimation of a Discrete Location Parameter for a Continuous Distribution.* PhD dissertation. GRASP Lab technical report MIS-CS-90-28 (214), Department of Computer and Information Science, University of Pennsylvania, May 1990.

[Wald, 1971] A. Wald. *Statistical Decision Functions.* Chelsea Publishing Company, Bronx, New York, second edition, 1971.

[Zeytinoglu and Mintz, 1984] M. Zeytinoglu and M. Mintz. Optimal fixed sized confidence procedures for a restricted parameter space. *The Annals of Statistics*, 12(3):945–957, September 1984.

[Zeytinoglu and Mintz, 1988] M. Zeytinoglu and M. Mintz. Robust optimal fixed sized confidence procedures for a restricted parameter space. *The Annals of Statistics*, 16(3):1241–1253, September 1988.