University of Pennsylvania
ScholarlyCommons

Technical Reports (CIS)   Department of Computer & Information Science

August 1990

Statistical Decision Theory for Sensor Fusion

Raymond McKendall
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/cis_reports

Statistical Decision Theory for Sensor Fusion

Abstract

This article is a brief introduction to statistical decision theory. It provides background for understanding the research problems in decision theory motivated by the sensor-fusion problem.

Comments

Statistical Decision Theory
For Sensor Fusion

MS-CIS-90-55
GRASP LAB 227

Raymond McKendall

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

August 1990

Statistical Decision Theory for Sensor Fusion

Raymond McKendall*
GRASP Laboratory
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104-6389

Abstract

This article is a brief introduction to statistical decision theory. It provides background for understanding the research problems in decision theory motivated by the sensor-fusion problem.

1 Introduction

This article is a brief introduction to statistical decision theory. It provides background for understanding the research problems in decision theory motivated by the sensor fusion problem. In particular, this article is an introduction for the articles *Robust Multi-Sensor Fusion: A Decision Theoretic Approach* [Kamberova and Mintz, 1990] and *Non-Monotonic Decision Rules for Sensor Fusion* [McKendall and Mintz, 1990] of these Proceedings. The principal references for this review are [Ferguson, 1967], [Berger, 1985], and [DeGroot, 1970]. The appendices of [McKendall, 1990] give an expanded discussion of sensor fusion and an expanded introduction to decision theory.

Section 2 states the general research problem in statistical estimation. Section 3 describes the sensor fusion problem. Section 4 introduces statistical decision theory and formulates the research problem as a decision problem. Section 5 gives some examples of the research problems studied.

2 Research Problem

The problem of this research to estimate the location parameter $\theta \in \Theta$ of a single observation Z in the model

$$Z = \theta + V.$$

Thus, the random variable Z is a measurement of the location parameter θ in continuous, additive noise V. The goal of this research is to estimate θ. The tool for analysis is statistical decision theory.

There are two versions of this problem, standard estimation and robust estimation. In a *standard-estimation* problem, the distribution function F 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 $F \sim \mathcal{N}(0, 1)$. (The notation $\mathcal{N}(\mu, \sigma^2)$ indicates a normal or Gaussian distribution with location μ and scale σ.) In a *robust-estimation* problem, the distribution F 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 $\sigma \in (0, 1]$ is unknown; in this case $F \in \mathcal{F}$ where $\mathcal{F}$ is the set of $\mathcal{N}(0, \sigma^2)$ distribution functions with $\sigma \in (0, 1]$. Robust estimation accounts for inexact characterizations of the noise. Many problems in robust estimation reduce to problems in standard estimation.

The statement of the problem in terms of mathematical statistics is to estimate the location parameter $\theta \in \Theta$ of the random variable Z, where

$$Z \sim F_Z(\cdot|\theta)$$

and

$$F_Z(z|\theta) = F(z - \theta), \qquad \forall z \in \Re.$$

The distribution $F_Z(\cdot|\theta)$ of Z is the *sampling distribution*, and the distribution F of V is the *nominal distribution*. Similarly, the density of Z, $f_Z(\cdot|\theta)$, is the *sampling density*, and the density f of V is the *nominal density*. The density functions are related by the equation

$$f_Z(z|\theta) = f(z - \theta), \qquad \forall z \in \Re.$$

3 Motivation

The location-estimation model of this research is fundamental to research in robust fusion of location data. *Location data* are sensors' measurements of the position of an object. *Fusion* is the combination of location data from different sensors. *Robust fusion* accounts for uncertainty in the description of the underlying system. The goals of the research in sensor fusion 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.

Example Figure 1 illustrates a sensor-fusion problem with three sensors. The sensors S_1, S_2, and S_3 may be different kinds of sensors. For example, S_1 may be a laser sensor, S_2 may be a sonar sensor, and S_3 may be a camera. The output of each sensor S_i is a measurement Z_i of the distance θ of the object T from the horizontal axis. The dashed box around each sensor represents
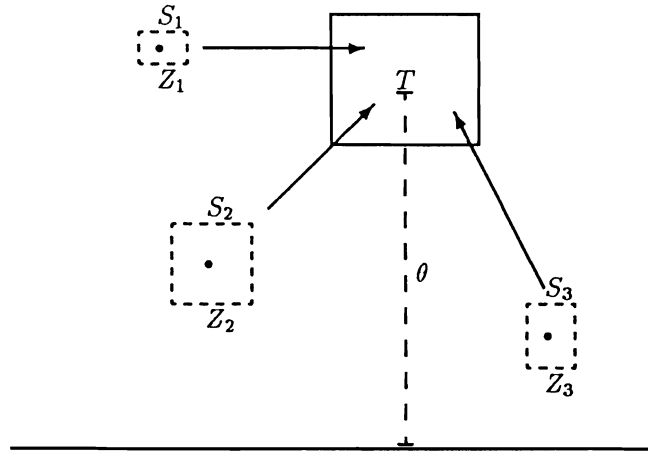
Figure 1: Sensor-fusion paradigm

the noise associated with the sensor's measurement. For example, there may be uncertainty in the exact position of each sensor. The box around the object T represents the prior information about the location of the object. For example, the object may be in a room with known dimensions.

The fusion problem is to combine the three measurements Z_1, Z_2, and Z_3 of the distance θ into a single estimate. Fusion of the data requires that the data are consistent: The consistency problem is to verify that Z_1, Z_2, and Z_3 are measurements of the same parameter. □

The location-data paradigm consists of a measurement Z of an unknown parameter θ in statistical uncertainty, noise due to the environment or to the sensor itself. 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 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.

The sensor-fusion problem has multiple measurements Z_1, ..., Z_n of the location θ in additive noise. These measurements originate from different sensors. The fusion problem is to combine these data into a single value for the location θ. 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. The goal of fusion is to find an estimator that minimizes the probability of unacceptable error. The fusion problem subsumes the problem of consistency, which is to verify that the data Z_1, ..., Z_n are in fact measurements of the same location.

4 Decision Theory

The tool for the analysis of the location-estimation problem is statistical decision theory. This section introduces

this theory and formulates the location-estimation model as a decision problem.

The Decision Problem

Figure 2 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 *parameter space* is the set of possible values for the unknown statistical parameters. For standard estimation, the parameter space is $\Omega = \Theta$. For robust estimation, the parameter space is $\Omega = \Theta \times \mathcal{F}$. 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:

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

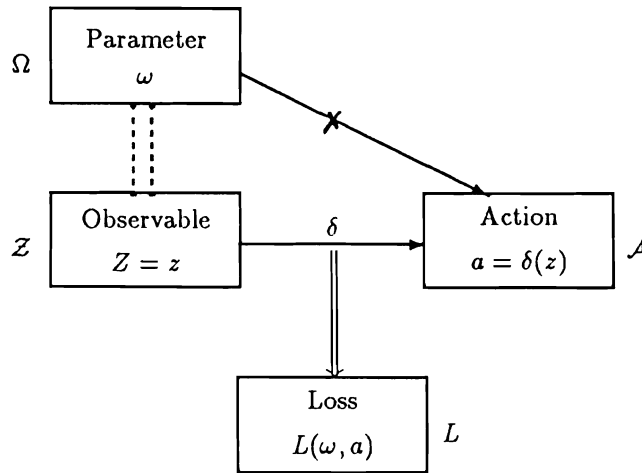The *observable* is a random variable whose distribution depends on the unknown parameters and thus contains

Figure 2: A statistical decision problem

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

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

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

Optimal 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. Thus, the first step in evaluating the performance of a decision rule δ is to find its average loss or *risk* $R(\omega, \delta)$:

$$R(\omega, \delta) := \mathrm{E}[L(\omega, \delta(Z)]$$
$$= \int_Z L(\omega, \delta(z))\, dF_Z(z|\theta)$$

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

$$R(\omega, \delta) = \int_Z L_e(\theta, \delta(z))\, dF_Z(z|\theta)$$
$$= \int_{\{z:|\delta(z)-\theta|>e\}} dF_Z(z|\theta)$$
$$= P_\omega\{|\delta(Z) - \theta| > e\}$$

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

Comparison of risk gives a weak optimality criterion. A decision rule δ_1 is preferable to a decision rule δ_2 if the risk of δ_1 is smaller than the risk of δ_2 uniformly in ω. A decision rule is *admissible* if there is no other rule preferable to it. Comparison of risk, however, is an incomplete criterion since the risk varies in the unknown parameter ω. (See figure 3.) Thus, 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.

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.

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 π on the parameter space Ω through which the risks are averaged. This distribution is the *prior distribution* on Ω. A decision rule δ^* is *Bayes* against π if its weighted-average risk under π is the smallest possible weighted-average risk:

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

Thus, a Bayes rule guards against the worst-possible weighted-average risk.

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

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

The goal of this research is to find a minimax rule for the location parameter θ of the measurement $Z = \theta + V$,
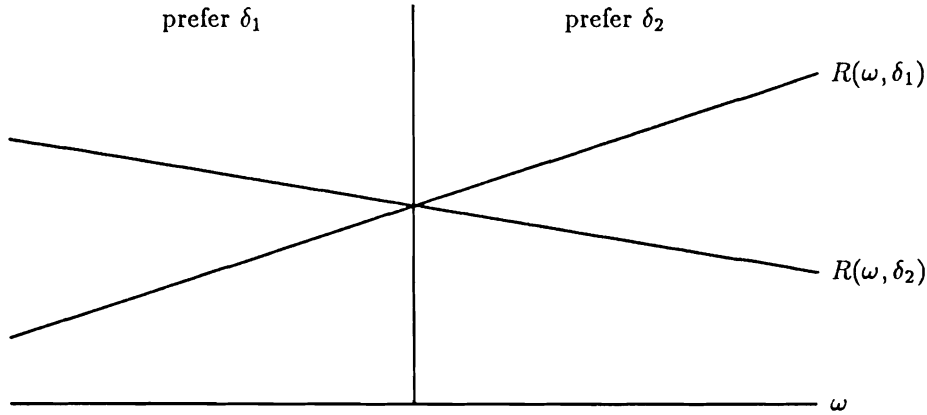
Figure 3: Incomplete comparison of decision rules through risk

but direct computation of a minimax rule from the definition is usually not possible. Instead, the Bayes and equalizer approaches provide an indirect strategy for finding minimax rules. A standard result from statistical decision theory states that a Bayes equalizer rule is minimax:

Theorem 1 *Let π be a distribution on Ω, and suppose that the decision rule δ is Bayes against π. If δ is an equalizer rule, then δ is minimax.*

Proof See [Ferguson, 1967, p. 90] or [McKendall, 1990, p. 271].

Thus, the strategy for finding a minimax rule is first to construct an equalizer rule and second to show that it is Bayes against some probability distribution on Ω. Theorem 2 gives an extension of this strategy:

Theorem 2 *Let π be a distribution on Ω, and suppose that the decision rule δ is Bayes against π. Suppose that there is a constant C such that the following two conditions are met:*

1. $R(\omega, \delta) \leq C$ for all $\omega \in \Omega$.

2. $\mathrm{P}\{\omega : R(\omega, \delta) = C\} = 1$.

Then δ is minimax.

Proof See [Ferguson, 1967, p. 90] or [McKendall, 1990, p. 272].

The probability distribution of these theorems is a mathematical tool; it has no interpretation for application. It is a least-favorable distribution. A distribution π_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.)

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 π be a distribution on Ω, and let $\pi(\cdot|z)$ 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 π.

Proof See [Ferguson, 1967, pp. 43–45].

The conditional distribution $\pi(\cdot|z)$ on Ω is the *posterior distribution* on Ω. The expected value under $\pi(\cdot|z)$ of the loss $L(\omega, a)$ is the *posterior expected loss* of an action a. Thus, a strategy for finding a Bayes rule against a prior distribution is to minimize the posterior expected loss under the corresponding posterior distribution.

Utility 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 $\mathcal{F}$. 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

$$C_\delta(Z) := [\delta(Z) - e, \delta(Z) + e].$$

The confidence coefficient is $\inf_\omega \mathrm{P}_\omega\{C_\delta(Z) \ni \theta\}$, where $\mathrm{P}_\omega\{C_\delta(Z) \ni \theta\}$ is the probability under ω that the con-

fidence interval covers θ. If δ^* is a minimax rule, then

$$\inf_\omega P_\omega\{C_{\delta^*}(Z) \ni \theta\} = \sup_\delta \inf_\omega P_\omega\{C_\delta(Z) \ni \theta\}.$$

This confidence procedure provides a test of hypothesis that two measurements Z_1 and Z_2 are consistent.

5 Examples

Example 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 $F \sim \mathcal{N}(0,1)$. The possible values of θ are the elements of $\Theta = \{-1, 0, 1\}$. This example is a standard-estimation problem since the nominal distribution F is known. Thus $\Omega = \Theta$ or $\omega = \theta$. 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:

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

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 $Z \ge 0.803$ is $\hat\theta = 1$.

The risk function of δ^* is this:

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

This decision rule is an equalizer rule with risk 0.422.

Furthermore, the rule δ^* is Bayes against the distribution on Θ that assigns these probabilities:

$$\begin{aligned} p(-1) &= 0.2982 \\ p(0) &= 0.4036 \\ p(1) &= 0.2982 \end{aligned}$$

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

Example 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:

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

The minimax decision rule δ^* is this:

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

Here $a = 0.3992$. (See figure 4.) This rule has $|\delta^*(z)| \le 0.2$ since the error tolerance is 0.1.

The risk function of δ^* is this:

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

This decision rule has constant risk (0.6176) except for the points $\theta = \pm 0.1$, which have smaller risk. Since these points together have zero probability under any continuous distribution, this rule is essentially an equalizer rule. In particular, theorem 2 applies to this rule.

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

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

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

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

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

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

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 the previous example.

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 the last example with the nominal distribution replaced by $\mathcal{N}(0, 0.25^2)$. In particular, the minimax rule is given by definition 1 with $a = 0.0808$.

(See [Zeytinoglu and Mintz, 1988] for the analysis underlying this example. See [Martin, 1987] and [McKendall, 1990] for other problems in robust estimation.) □

References

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

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

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

[Kamberova and Mintz, 1990] G. Kamberova and M. Mintz. Robust multi-sensor fusion: a decision theoretic approach. In these Proceedings, October 1990.
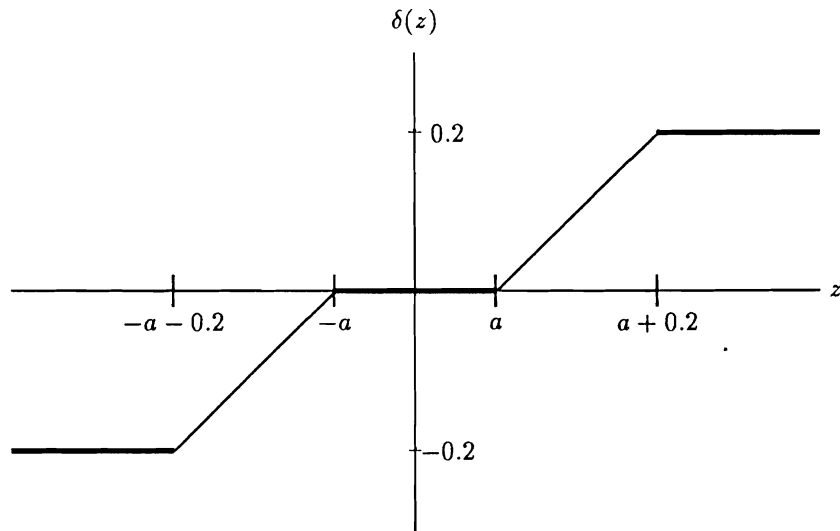
$\delta(z)$

0.2

$-a - 0.2$ $-a$ a $a + 0.2$ z

-0.2

Figure 4: A minimax rule when $F \sim \mathcal{N}(0,1)$, $\Theta = [-0.3, 0.3]$, and $c = 0.1$.

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

[Martin and Mintz, 1984] K.E. Martin and M. Mintz. Randomized robust confidence procedures. In *Proceedings of the Twenty-Second Annual Allerton Conference on Communication, Control, and Computing*, pages 309–317, University of Illinois, October 1984.

[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, Department of Computer and Information Science, University of Pennsylvania, May 1990.

[McKendall and Mintz, 1990] R. McKendall and M. Mintz. Non-monotonic decision rules for sensor fusion. In these Proceedings, October 1990.

[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.