Mechanical Systems and Signal Processing, vol. 48 (2014) pp.1–14

Failure Detection with Likelihood Ratio Tests and Uncertain Probabilities: An Info-Gap Application

Yakov Ben-Haim[1] and Miriam Zacksenhouse[2]
Technion—Israel Institute of Technology
Ronit Eshel,[3] Raphael Levi,[4] Avi Fuerst[5] and Wayne Bentley[6]
Intel Corporation

Contents

Abstract The likelihood ratio test (LRT) has attractive failure-detection properties. However, evaluating the likelihood ratio and implementing the LRT require knowledge of the underlying probability distributions. Data or knowledge, especially about future failures, is often quite limited. In this paper we employ the info-gap robustness function in specifying the parameters of the LRT when the probability distributions are imperfectly known. We develop an info-gap analog of the probabilistic detection-error trade off curve, and demonstrate the results by application to pressure measurements on an industrial production device.

Keywords failure detection, likelihood ratio test, info-gap robustness, uncertain probability, detection-vs-error trade off.

---

[1]Yitzhak Moda'i Chair in Technology and Economics, Faculty of Mechanical Engineering, Technion — Israel Institute of Technology, Haifa 32000 Israel, yakov@technion.ac.il

[2]Faculty of Mechanical Engineering, Technion — Israel Institute of Technology, Haifa 32000 Israel, yakov@technion.ac.il

[3]Intel Israel, Kiriat Gat, ronit.eshel@intel.com

[4]Intel Israel, Kiriat Gat, raphael.levi@intel.com

[5]Intel USA, avi.fuerst@intel.com

[6]Intel USA, Chandler, Arizona, wayne.bentley@intel.com

1 Introduction

Failure detection is the discovery of the occurrence of anomalous behavior, as distinguished from failure diagnosis that characterizes the nature and origin of the anomaly. We employ info-gap theory in formulating a failure detection algorithm based on the likelihood ratio test (LRT), given imperfect knowledge of the probability distribution of future failures.

Failure detection has been studied systematically and in depth at least since the early decades of the 20th century, and is the subject of numerous books and articles studying both theory and application. This literature applies a wide range of tools from statistics and signal processing. For instance, Gertler (1998, chap. 11) applies statistical tests on residuals (the difference between measurements and model predictions) to detect the presence of failure. Patton, Frank and Clark (1989) discuss a range of filter and observer techniques for failure detection. Williams (1998) summarizes model-based failure detection. Willsky (1976, 1984) discusses likelihood ratio methods for detection of abrupt changes. Pau (1981, chap. 4) explores time-minimization statistical tests. Basseville and Nikiforov (1993) provide a masterful and very accessible discussion of statistical tools—especially based on the likelihood ratio concept—for detecting abrupt changes in dynamical systems. Campbell and Nikoukhah (2004) explore interrogative methods—applying external signals—for failure detection. We mention just a few of the myriad recent specific applications. Tibaduiza *et al.* (2013) build statistical models of dynamic behavior for detecting and classifying damage in structural health monitoring. Hwang, Lee and Hwang (2013) develop a condition monitoring method based on likelihood change of a stochastic model of the system in normal operation. Earls (2013) uses non-contact and very sparse contact inspection methods for non-destructive evaluation and testing for hidden corrosion in steel bridge connections.

Much effort has been invested in studying the relations among these various approaches to failure detection. Quite often seemingly distinct methods—such as likelihood ratio tests and residual parity checks—are equivalent in the sense that either method can be represented by the other (Basseville and Nikiforov, 1993, p.241). Nonetheless, representationally equivalent methods may be very different in computational difficulty or in their requirements for prior information. The choice of a failure detection strategy depends on the type of system, the type of failure, and the available information.

The LRT has an optimality property that make it very attractive: minimal probability of missed detection (type II error), for given probability of false alarm (type I error). However, evaluating the likelihood ratio and implementing the LRT requires knowing the underlying probability distributions. Data or knowledge, especially about future failures, is often quite limited. In this paper we employ the info-gap robustness function to specify the parameters of the LRT when the underlying probability distributions are imperfectly known.

Robustness to uncertainty is a central concept in many approaches to failure detection. Chow and Willsky (1984), Gertler (1988), Frank (1990) and many others discuss methods of "analytical redundancy" and "residual generation" for exploiting redundant information in system models to enhance robustness against modeling error. Ding *et al.* (2000) continue this direction and discuss observer-based methods for robustly detecting unknown inputs to linear time-invariant systems.

'Robustness' has many meanings. The concept of robustness used in this paper derives from a prior concept of non-probabilistic uncertainty. Knight (1921) distinguished between 'risk' based on known probability distributions and 'true uncertainty' for which probability distributions are not known. Similarly, Ben-Tal and Nemirovski (1999) are concerned with uncertain data within a prescribed uncertainty set, without any probabilistic information. Likewise Hites *et al.* (2006, p.323) view "robustness as an aptitude to resist to 'approximations' or 'zones of ignorance'", an attitude adopted also by Roy (2010). We also are concerned with robustness against Knightian uncertainty. We consider uncertainty in probability distributions but we do not pursue an explicitly statistical approach to robustness as studied by Huber (1981) and many others.

Our approach is in the tradition of Wald. Wald (1945) studied Knightian uncertainty in the problem of statistical hypothesis testing based on a random sample whose probability distribution is not known, but whose distribution is known to belong to a given class of distribution functions.

Wald states that "in most of the applications not even the existence of ... an a priori probability distribution [on the class of distribution functions] ... can be postulated, and in those few cases where the existence of an a priori probability distribution ... may be assumed this distribution is usually unknown." (p.267).

In this paper we quantify Knightian uncertainty using info-gap theory (Ben-Haim, 2006; info-gap.com). Info-gap theory has been applied to various problems of statistical inference where probabilistic properties or distributions are incompletely known. Pierce, Worden and Manson (2006) use info-gap theory in assessing the reliability of artificial neural nets for damage detection. Zacksenhouse *et al.* (2009) use info-gap theory in managing data uncertainty in linear regression for neural decoding of brain-machine interfaces. Ben-Haim (2010) employs info-gap theory in regression of economic data and for confidence interval estimation given uncertain probabilities. Mirer and Ben-Haim (2010) use info-gap theory in the design and analysis of "penalty tests" in which excess stresses are applied to an explosive material in assessing its safety against accidental actuation and its reliability of operational actuation. Ben-Haim (2011) applies info-gap theory in statistical evaluation of null results—not detecting the presence of a prenicious agent—when the degree of statistical correlation between observations is uncertain.

The basic properties of the LRT are discussed and illustrated in sections 2 and 3. The robustness function is formulated in section 4, and demonstrated on industrial data in section 5. Specification of the parameters needed for implementation of the LRT algorithm, and use of the robustness function in assessing confidence in the decision, are illustrated in section 6. We also demonstrate an info-gap analog of the probabilistic detection-error trade off curve (also known as the Receiver Operating Characteristic, or ROC).

2 Statistical Concepts and Hypothesis Tests

In this section we define the requisite statistical concepts and notation.

Probability distributions and innovations. Our task is to detect the occurrence of change in the system. Let θ be a vector of parameters that specifies the system, where θ_0 and θ_1 are the values before and after the onset of failure, respectively. We will assume that θ_0 is known from system identification in failure-free conditions, while θ_1 is unknown or highly uncertain.

We assume that we have access to the inputs and outputs of the system. A set of consecutive control inputs is denoted $Z_i^j = \{z_i, \ldots, z_j\}$ at time steps $i, i+1, \ldots, j$. A set of consecutive output measurements is denoted $X_i^j = \{x_i, \ldots, x_j\}$.

The measured output vector x_i is a random variable whose probability density function (pdf) is denoted $p(x|\theta)$. Thus x is distributed according to $p(x|\theta_0)$ before failure, and according to $p(x|\theta_1)$ after failure. We will assume that $p(x|\theta_0)$ is known or reliably estimated. In contrast, we are uncertain about some aspects of $p(x|\theta_1)$.

Given measured inputs Z_1^i and outputs X_1^{i-1} and a specification θ of the system we can construct the conditional pdf of the next output, x_i, denoted $p(x_i|Z_1^i, X_1^{i-1}, \theta)$. This may employ system identification, or state-estimators such as the Kalman filter. If $i = 1$, so that x_1 is the first output, we define $p(x_1|Z_1^1, X_1^0, \theta)$ as $p(x_1|z_1, \theta)$. We can write the joint pdf of the outputs X_1^m as:

$$p(X_1^m|\theta) = \prod_{i=1}^{m} p(x_i|Z_1^i, X_1^{i-1}, \theta) \tag{1}$$

We can then evaluate the conditional expectation of x_{i+1}, denoted $\mathrm{E}(x_{i+1}|Z_1^{i+1}, X_1^i, \theta)$. When x_{i+1} is observed we can calculate the *innovation:*

$$y_{i+1} = x_{i+1} - \mathrm{E}(x_{i+1}|Z_1^{i+1}, X_1^i, \theta) \tag{2}$$

A consecutive sequence of innovations is denoted $Y_i^j = \{y_i, \ldots, y_j\}$. If θ is a correct specification of the system, then, in many situations, the innovations Y_i^j are zero-mean and uncorrelated, or sometimes even statistically independent.

Hypotheses. Given inputs Z_1^{m+1} and outputs X_1^m (indexed with respect to the start of a given data window), we wish to decide between the following two hypotheses. Hypothesis H_0 states that no change has occurred in the system. Hypothesis H_1 states that change in the system has occurred, or begun, at time step i_c. These hypotheses are stated more formally as follows.

$$H_0: \qquad \theta = \theta_0, \qquad 1 \le i \le m \tag{3}$$

$$H_1: \begin{cases} \theta = \theta_0, & 1 \le i < i_\mathrm{c} \\ \theta = \theta_1, & i_\mathrm{c} \le i \le m \end{cases} \tag{4}$$

i_c denotes the time step at which change began. Thus, according to hypothesis H_1, measurements $X_1^{i_\mathrm{c}-1}$ are governed by θ_0, and measurements $X_{i_\mathrm{c}}^m$ are governed by θ_1. Beginning in section 4 we will deal with the uncertainty in θ_1.

Decision function. Given inputs Z_1^{m+1} and outputs X_1^m, D denotes the decision algorithm for choosing between the hypotheses:

$$D = \begin{cases} 0: & \text{Accept } H_0 \\ 1: & \text{Reject } H_0 \end{cases} \tag{5}$$

Our task is to choose a specific structure for decision algorithm D.

Error probabilities. A decision algorithm can err in either of two ways. The algorithm could falsely reject H_0, which is called a type I error or "false alarm", or it could falsely reject H_1, called a type II error or "missed detection". The probabilities of these errors depend on the choice of the decision function, D. Hence evaluation of these error probabilities can be used to evaluate and choose the decision function.

Given inputs Z_1^{m+1} and outputs X_1^m, the probabilities of type I and type II errors are formally defined as:

$$\alpha_0 = \mathrm{Prob}\,(D = 1 |\, H_0) \tag{6}$$

$$\alpha_1 = \mathrm{Prob}\,(D = 0 |\, H_1) \tag{7}$$

α_0 is the probability of type I error: the probability that the algorithm will decide in favor of H_1 when in fact H_0 holds. α_0 is the probability of false alarm. α_1 is the probability of type II error: the probability that the algorithm will decide in favor of H_0 when in fact H_1 holds. α_1 is the probability of missed detection. It is well known that these two detection-error probabilities trade off against one another as the specification of the decision algorithm is changed.

Likelihood ratio. Given an innovation, y_i, at time t_i, the log of the likelihood ratio is:

$$s_i = \ln \frac{p(y_i | Z_1^i, X_1^{i-1}, \theta_1)}{p(y_i | Z_1^i, X_1^{i-1}, \theta_0)} \tag{8}$$

Given inputs Z_1^{m+1} and outputs X_1^m, we define the log likelihood ratio of the innovations as:

$$S_1^m = \ln \frac{\prod_{i=1}^{i_\mathrm{c}-1} p(y_i | Z_1^i, X_1^{i-1}, \theta_0) \prod_{i=i_\mathrm{c}}^m p(y_i | Z_1^i, X_1^{i-1}, \theta_1)}{\prod_{i=1}^m p(y_i | Z_1^i, X_1^{i-1}, \theta_0)} \tag{9}$$

$$= \sum_{i=i_\mathrm{c}}^m s_i \tag{10}$$

S_1^m is sometimes called the cumulative sum of the log likelihood ratios.

Likelihood ratio test. The likelihood ratio test (LRT) is defined as follows. For any fixed value of λ, decide between hypotheses H_0 and H_1 according to:

$$D = \begin{cases} 0 \text{ if } & S_1^m < \lambda \quad (\text{Accept } H_0) \\ 1 \text{ if } & S_1^m \ge \lambda \quad (\text{Reject } H_0) \end{cases} \tag{11}$$

The Neyman-Pearson lemma specifies widely occurring conditions under which the likelihood ratio provides an optimal test. The lemma also provides a statistical meaning of optimality. Roughly speaking, the LRT has maximum power from among all tests of the same significance (Basseville and Nikiforov, 1993, p.129).

3 Nominal Implementation of the LRT: Known Normal Distributions

We first briefly consider the implementation of the LRT given full probabilistic information.

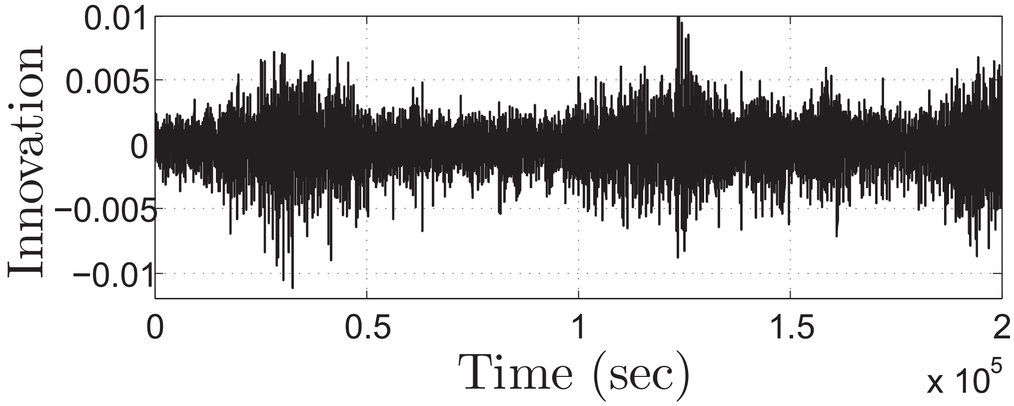

Figure 1: Innovations from an industrial production device.

Fig. 1 shows innovations calculated from pressure measurements on an industrial production device that is supposedly operating normally and for which the conditional expectation in eq.(2) is effectively zero. The innovations fluctuate erratically around zero as we expect for normal operation. The vast majority of the innovations are less than 0.004 in absolute magnitude; some values exceed 0.005 and rarely reach 0.01. The number of observations is 99,998, taken at 2-second intervals and thus cover about 55 hours.

We base the LRT on the assumption of normal distributions of the innovations under both hypotheses, and knowledge of the moments of these distributions. That is, the hypotheses in eqs.(3) and (4) are:

$$H_0: \qquad y_i \sim \mathcal{N}(\mu_0, \sigma_0^2), \qquad 1 \le i \le m \tag{12}$$

$$H_1: \begin{cases} y_i \sim \mathcal{N}(\mu_0, \sigma_0^2), & 1 \le i < i_{\rm c} \\ y_i \sim \mathcal{N}(\mu_1, \sigma_1^2), & i_{\rm c} \le i \le m \end{cases} \tag{13}$$

where μ_i and σ_i^2 are specified for $i = 0, 1$.

The log likelihood ratio of a sequence of innovations, eq.(10), is:

$$S_1^m = (m - i_{\rm c} + 1) \ln \frac{\sigma_0}{\sigma_1} + \frac{1}{2\sigma_0^2} \sum_{i=i_{\rm c}}^{m} (y_i - \mu_0)^2 - \frac{1}{2\sigma_1^2} \sum_{i=i_{\rm c}}^{m} (y_i - \mu_1)^2 \tag{14}$$

We note that evaluation of the likelihood ratio depends on the innovations and on the probability distributions under both hypotheses. H_0 is accepted or rejected by comparing S_1^m against the threshold, λ, as stated in eq.(11).

Fig. 2 shows the log likelihood ratio, S_1^m in eq.(14), with $\mu_0 = 0$ and $\sigma_0 = 0.0015$, which are very close to the sample mean and standard deviation of the entire data set, and $\mu_1 = 0$ and $\sigma_1 = 0.005$ which represents a possible altered condition. The entire data set is divided into short test windows, each containing 600 data points, so $m = 600$. Failure is hypothesized to start at the 450th datum in each window, so $i_{\rm c} = 450$.

Suppose, as an example, that the decision threshold is $\lambda = 100$. From eq.(11) and fig. 2 we see that H_0 is accepted throughout most of the data set, except for repeated rejections of H_0 in the time interval from 25,000 to 35,000 seconds, around 120,000 seconds, and around 190,000 seconds.

We now explore the robustness of decisions such as these, in the face of uncertainty in the statistical properties of the data after onset of failure.

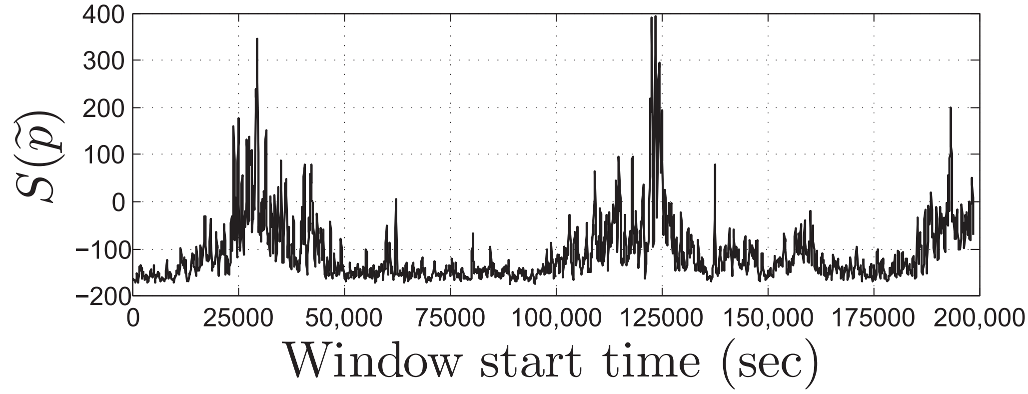

Figure 2: Nominal cumulative-sum log likelihood ratio.

4 Robustness of the Likelihood Ratio Test

In section 4.1 we define and discuss the info-gap model for uncertainty in the pdf under failure hypothesis H_1. In section 4.2 we formulate two info-gap robustness functions.

4.1 Info-Gap Models of Uncertainty

We distinguish between the uncertain pdf's of the innovations $p(y|\theta_i)$ under hypotheses H_i ($i = 0$ or 1) and their known estimates, $\widetilde{p}(y|\theta_i)$, which will be referred to as the 'nominal' pdfs. The uncertainty in these distributions is quantified using an info-gap model. We will shortly consider a specific example, but for now it is sufficient to define a generic info-gap model for uncertainty in the pdf's.

In this context, an info-gap model is a unbounded family of nested sets of pdf's $(p(y|\theta_0),\, p(y|\theta_1))$. All info-gap models, denoted $\mathcal{U}(h)$ for $h \geq 0$, have two properties:

$$\text{Contraction:} \qquad \mathcal{U}(0) = \{(\widetilde{p}(y|\theta_0),\, \widetilde{p}(y|\theta_1))\} \tag{15}$$

$$\text{Nesting:} \qquad h < h' \quad \Longrightarrow \quad \mathcal{U}(h) \subseteq \mathcal{U}(h') \tag{16}$$

'Contraction' states that the uncertainty set $\mathcal{U}(0)$ is a singleton, containing only the estimated pdf's. 'Nesting' asserts that the uncertainty sets $\mathcal{U}(h)$ become more inclusive as h increases. These two properties endow h with its meaning as an horizon of uncertainty. The value of h is unknown so an info-gap model is an unbounded family of nested sets of uncertain entities.

The info-gap model represents severe uncertainty about the estimated pdf's in the sense that there is no known worst case and that there is no known probability measure on the space of uncertain entities. This sort of non-probabilistic uncertainty is sometimes called "true uncertainty" (Knight, 1921) or "deep uncertainty" (Lempert, *et al.,* 2003). The robustness functions to be discussed in section 4.2 quantify the impact of this uncertainty on the interpretation of the LRT decision. The generic results derived from these properties are illustrated in a specific industrial example. The uncertainty is also severe because of the large potential impact of an uncertain failure in an industrial context. Even seemingly innocuous system changes can have huge impact on production processes and product quality.

Consider a specific example, to which we will return later. First suppose that $p(y|\theta_0)$, the pdf of the innovations under the no-failure hypothesis H_0, is known and normal. There is no uncertainty in $p(y|\theta_0)$ so it need not be distinguished from its estimate, $\widetilde{p}(y|\theta_0)$, and hence it is not included in the info-gap model.

Furthermore suppose that strong evidence exists that $p(y|\theta_1)$, the pdf of the innovations under failure hypothesis H_1, is normal. However, the mean and standard deviation of $p(y|\theta_1)$ are highly uncertain because the origin and nature of the failure—if there is one—is unknown.

We have estimates of the mean and standard deviation of $p(y|\theta_1)$, denoted $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$, but these estimates are highly uncertain (and may be little more than guesses) because the failure is unknown.

We have approximate error estimates, denoted s_μ and s_σ, but the actual errors in the estimated moments under H_1 may be greater. In fact, we have no realistic or meaningful knowledge of the maximum error in these estimated moments. This information supports the following fractional-error info-gap model for $p(y|\theta_1)$ (Ben-Haim, 2006):

$$\mathcal{U}(h) = \left\{ p(y|\theta_1) \sim \mathcal{N}(\mu_1, \sigma_1^2) : \left| \frac{\mu_1 - \widetilde{\mu}_1}{s_\mu} \right| \leq h, \ \left| \frac{\sigma_1 - \widetilde{\sigma}_1}{s_\sigma} \right| \leq h, \ \sigma_1 \geq 0 \right\}, \quad h \geq 0 \tag{17}$$

Each set, $\mathcal{U}(h)$, contains all normal pdfs whose mean or standard deviation deviates fractionally from the estimates by no more than h, the horizon of uncertainty. The value of h is unknown. Hence this is an unbounded family of nested sets of normal distributions and obeys the properties of contraction and nesting. That is, there is no known worst case, and there is no known probability distribution on the uncertain moments.

4.2 Definition of Robustness

In this section we define the robustness to uncertainties in the pdf's in the context of a generic info-gap model, $\mathcal{U}(h)$, which is restricted only by the properties of contraction and nesting, eqs.(15) and (16). We will develop two different robustness functions, whose properties are discussed in section 5. Robustness will provide a tool for specifying the decision threshold of the LRT and for interpreting the test decisions, discussed in section 6.

Nominal acceptance of H_0. Suppose that, given a sequence of innovations Y_1^m, the likelihood ratio $S_1^m(\widetilde{p})$, based on the nominal pdf's $\widetilde{p}(y|\theta_0)$ and $\widetilde{p}(y|\theta_1)$, **accepts** H_0 at threshold λ:

$$S_1^m(\widetilde{p}) < \lambda \tag{18}$$

We ask: how robust is this acceptance of H_0 to uncertainty in the pdf's? The robustness is the greatest horizon of uncertainty up to which H_0 is accepted for the given sequence of innovations Y_1^m:

$$\widehat{h}_0(\lambda) = \max \left\{ h : \left(\max_{p(y|\theta_i) \in \mathcal{U}(h)} S_1^m(p) \right) < \lambda \right\} \tag{19}$$

$\widehat{h}_0(\lambda)$ is the greatest horizon of uncertainty, h, up to which a nominal decision to accept H_0 is unchanged.

Suppose $\widehat{h}_0(\lambda)$ is large. This means that hypothesis H_0 is accepted, and failure is not declared, for any realization of the pdf's of the innovations up to a large horizon of uncertainty. If failure has actually not occurred, and if the correct distributions $p(y|\theta_0)$ and $p(y|\theta_1)$ are actually within the horizon of uncertainty $\widehat{h}_0(\lambda)$, then failure will, correctly, not be declared. In this sense, $\widehat{h}_0(\lambda)$ can be thought of as the robustness against type I error (false alarm) resulting from error in the pdf's: When $\widehat{h}_0(\lambda)$ is large, failure will not be erroneously declared unless $p(y|\theta_0)$ and $p(y|\theta_1)$ are far from the nominal distributions, $\widetilde{p}(y|\theta_0)$ and $\widetilde{p}(y|\theta_1)$.

Nominal rejection of H_0. Suppose that, given a sequence of innovations Y_1^m, the likelihood ratio $S_1^m(\widetilde{p})$, based on the nominal pdf's $\widetilde{p}(y|\theta_0)$ and $\widetilde{p}(y|\theta_1)$, **rejects** H_0 at threshold λ:

$$S_1^m(\widetilde{p}) \geq \lambda \tag{20}$$

We ask: how robust is this rejection of H_0 to uncertainty in the pdf's? The robustness is the greatest horizon of uncertainty up to which a nominal decision to reject H_0 is unchanged, for the given sequence of innovations Y_1^m:

$$\widehat{h}_1(\lambda) = \max \left\{ h : \left(\min_{p(y|\theta_i) \in \mathcal{U}(h)} S_1^m(p) \right) \geq \lambda \right\} \tag{21}$$

Suppose $\widehat{h}_1(\lambda)$ is large. This means that hypothesis H_1 is accepted, and failure is declared, for any realization of the pdf's of the innovations up to a large horizon of uncertainty. If failure has actually occurred, and if the correct distributions $p(y|\theta_0)$ and $p(y|\theta_1)$ are actually within the horizon of uncertainty $\widehat{h}_1(\lambda)$, then this failure will be detected. In this sense, $\widehat{h}_1(\lambda)$ can be thought of as the robustness against type II error (missed detection) resulting from error in the pdf's: When $\widehat{h}_1(\lambda)$ is large, failures will not be missed unless they occur when $p(y|\theta_0)$ and $p(y|\theta_1)$ are far from the nominal distributions, $\widetilde{p}(y|\theta_0)$ and $\widetilde{p}(y|\theta_1)$.

As will be shown in section 5, in connection with eqs.(24) and (25), increasing λ has opposite effects on $\widehat{h}_0(\lambda)$ and $\widehat{h}_1(\lambda)$ so there is a trade off between them.

Let $M_i(h)$, for $i = 0$ or 1, denote the inner optimum in the definition of the robustness, eq.(19) or eq.(21) respectively. The nesting property of the info-gap model implies that $M_0(h)$, the inner maximum in eq.(19), is a non-decreasing function of h. Similarly, the nesting property implies that $M_1(h)$, the inner minimum in eq.(21), is a non-increasing function of h. In both cases the robustness is the greatest value of h at which $M_i(h) = \lambda$. This implies that $M_i(h)$ is the inverse of the robustness, $\widehat{h}_i(\lambda)$. Consequently, a plot of h vs. $M_i(h)$ is identical to a plot of the robustness, $\widehat{h}_i(\lambda)$ vs. λ, for either $i = 0$ or $i = 1$. The evaluation of $M_i(h)$ is explained in appendix A.

5 Properties of the Robustness Functions: Example

We now illustrate and discuss basic properties of the robustness functions, before demonstrating their application to the LRT in section 6. We use the info-gap model of eq.(17).

The data in fig. 1 are divided into windows, each with 600 data points ($m = 600$ in eqs.(12) and (13)) corresponding to 1,200 seconds. Failure is hypothesized, under H_1, to start at the 450th datum in each window, so $i_c = 450$. The pdfs are assumed to be normal, with known moments $\mu_0 = 0$ and $\sigma_0 = 0.0015$ under H_0, and estimated but uncertain moments $\widetilde{\mu}_1 = 0$ and $\widetilde{\sigma}_1 = 0.005$ under H_1. The error weights in the info-gap model, eq.(17), are $s_\mu = 0.001$ and $s_\sigma = 0.005$.

We look at 4 individual windows starting at 29,900sec, 79,900sec, 119,900sec and 179,900sec. The innovations for the 4 windows are shown in fig. 3. The innovations in the 1st and 3rd windows tend to be larger than in the other windows, as expected from fig. 2.

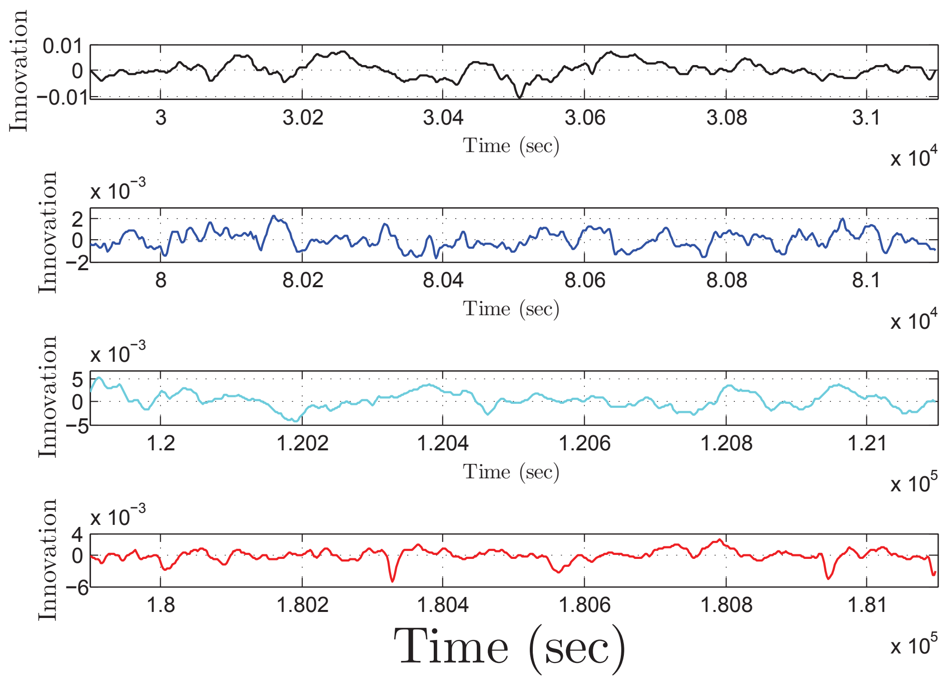

Figure 3: Innovations in 4 windows.

The robustness curves for these four time intervals are calculated by evaluating $M_i(h)$ vs. h as explained in appendix A. Plotting h vertically vs $M_i(h)$ horizontally is identical to plotting $\widehat{h}_i(\lambda)$ vertically vs. λ horizontally. The robustness curves for the 4 windows in fig. 3 are shown in figs. 4

and 5. The color codes are preserved between figs. 3, 4 and 5: each color refers to the same time interval.

The numerical results discussed subsequently are based on the info-gap model in eq.(17) with industrial data. However, the zeroing and trade off properties to be discussed hold for *any* info-gap model, and result from the properties of contraction and nesting, eqs.(15) and (16), as explained in the following discussion.

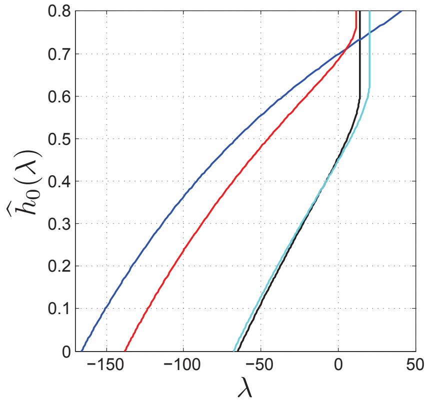

Figure 4: Robustness for nominal acceptance of H_0.

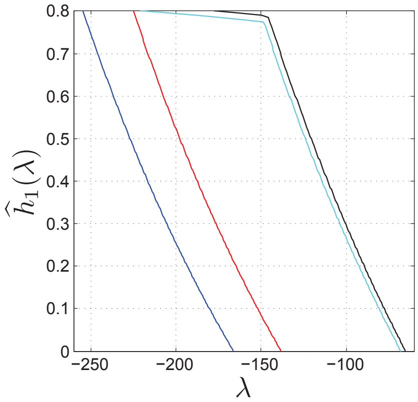

Figure 5: Robustness for nominal rejection of H_0.

Three points can be made from these figures: zeroing, trade off between robustness and threshold, and trade off between $\widehat{h}_0$ and $\widehat{h}_1$.

Zeroing. The inner extrema in the definitions of the robustness functions, eqs.(19) and (21), are $M_0(h)$ and $M_1(h)$, respectively. These two functions are equal when $h = 0$ because the info-gap uncertainty set, $\mathcal{U}(0)$, is a singleton and contains only the nominal pdf as a result of the contraction property of all info-gap models, eq.(15). Furthermore, these functions then precisely equal the nominal log likelihood ratio:

$$M_0(0) = S_1^m(\widetilde{p}) = M_1(0) \tag{22}$$

From eq.(22) and the monotonicity of $M_i(h)$ we conclude that, if $S_1^m(\widetilde{p})$ is adopted as the decision threshold, λ, then the robustness precisely equals zero:

$$\widehat{h}_i(\lambda) = 0 \quad \text{if} \quad \lambda = S_1^m(\widetilde{p}) \tag{23}$$

This is the zeroing property: attaining the estimated performance, $S_1^m(\widetilde{p})$, has zero robustness to uncertainty. $\widetilde{p}$ is our best—but highly uncertain—estimate of the pdf under failure. $S_1^m(\widetilde{p})$ is the resulting estimate of the likelihood ratio. However, using $S_1^m(\widetilde{p})$ as the decision threshold results in decisions that have no robustness to uncertainty in $\widetilde{p}$. The zeroing property is manifested in figs. 4 and 5 by the horizontal intercepts being the corresponding $S_1^m(\widetilde{p})$ values, which are the same in both figures as stated in eqs.(22) and (23).

Trade off between robustness and threshold. The robustness curves in fig. 4, $\widehat{h}_0(\lambda)$ vs. λ, all have positive slope: larger λ entails greater robustness to uncertainty. $\widehat{h}_0(\lambda)$ is the robustness for acceptance of H_0. Larger λ means that H_0 is accepted more easily. The robustness for acceptance of H_0 increases as the criterion for acceptance of H_0 is relaxed.

A similar interpretation applies to fig. 5. The robustness curves in fig. 5, $\widehat{h}_1(\lambda)$ vs. λ, all have negative slope: smaller λ entails greater robustness to uncertainty. $\widehat{h}_1(\lambda)$ is the robustness for rejection of H_0. Smaller λ means that H_0 is rejected more easily. The robustness for rejection of H_0 increases as the criterion for rejection of H_0 is relaxed.

These trade offs are completely general, and are not restricted to the info-gap model of eq.(17). The trade offs result from the nesting property of all info-gap models, eq.(16), and from the definitions of the robustness functions, eqs.(19) and (21). The inner maximum in eq.(19) is a *non-decreasing* function of the horizon of uncertainty, h, because it is the *maximum* on a family of nested sets: as h gets bigger the sets $\mathcal{U}(h)$ become more inclusive and the maximum increases. Likewise, the inner minimum in eq.(21) is a *non-increasing* function of the horizon of uncertainty, h, because it is the *minimum* on a family of nested sets: as h gets bigger the sets $\mathcal{U}(h)$ become more inclusive and the minimum decreases. Hence, as the threshold requirement, λ, increases, $\widehat{h}_0(\lambda)$ increases and $\widehat{h}_1(\lambda)$ decreases for any info-gap model.

We can summarize these trade offs as:

$$\frac{\partial \widehat{h}_0}{\partial \lambda} \geq 0 \qquad (24)$$

$$\frac{\partial \widehat{h}_1}{\partial \lambda} \leq 0 \qquad (25)$$

Note, however, that $\widehat{h}_0(\lambda)$ and $\widehat{h}_1(\lambda)$ are never positive at the same value of λ. From eqs.(19) and (21) we conclude that:

$$\widehat{h}_0(\lambda) = 0 \quad \text{for} \quad \lambda \leq S_1^m(\widetilde{p}) \qquad (26)$$

$$\widehat{h}_1(\lambda) = 0 \quad \text{for} \quad \lambda \geq S_1^m(\widetilde{p}) \qquad (27)$$

Trade off between $\widehat{h}_0$ and $\widehat{h}_1$. We note that the robustness-ranks of the four curves in fig. 4 are the reverse of the robustness-ranks of the four curves in fig. 5, throughout most of the λ-range. The most robust interval in fig. 4 (blue, 2nd interval), is the least robust in fig. 5. The 2nd-most robust interval in fig. 4 (red, 4th interval), is the 2nd-least robust in fig. 5. And so on. The 2nd and 4th intervals have the lowest innovations, so H_0 is most confidently accepted and the robustness for acceptance of H_0 is greatest. Equivalently, the robustness for rejection of H_0 is low in the 2nd and 4th intervals. Note however that the robustness curves in fig. 4 cross one another at high robustness, so this trade off is not universal.

6 Implementing and Assessing the LRT

6.1 Formulation

Implementing the LRT requires choice of the decision threshold, λ, and the estimated moments $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$ under failure hypothesis H_1. In this section we explain and illustrate the choice of these parameters, based on the robustness function. We also discuss the use of the robustness function for assessing the strength of rejection of H_0. This will lead to an info-gap analog of the probabilistic detection-error trade off curve. We use the info-gap model of eq.(17).

The mean and standard deviation, μ_0 and σ_0, under H_0 are calculated from a failure-free section of data, running from 50,000 to 100,000 [sec] in fig. 1. The data from 100,000 to 200,000 [sec] are tested for failure in overlapping test windows. The size of each test window is $m = 600$ steps (corresponding to 1,200 [sec]) and the index within each window, at which failure starts according to H_1, is $i_{\rm c} = 1$. A new test window is initiated each 50 steps. The moments $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$ under failure hypothesis H_1 are calculated from the innovations $y_{i_{\rm c}}, \ldots, y_m$ in each test window, eqs.(31) and (32) in appendix A. This maximizes the log likelihood ratio, S_1^m in eq.(14) (see appendix B), and thus maximizes the robustness for nominal rejection of H_0 in eq.(21). The error weights of these estimates in the info-gap model of eq.(17) are chosen as:

$$s_\mu = \max(0.001,\ \widetilde{\mu}_1), \quad s_\sigma = \max(0.005,\ \widetilde{\sigma}_1) \qquad (28)$$

In subsequent results, fig. 6, we will see that $s_\mu = 0.001$ and $s_\sigma = 0.005$.

Let $\alpha_{0\rm c}$ denote the largest probability of false alarm (type I error) that we are willing to accept. The choice of $\alpha_{0\rm c}$ is an operational judgment, for instance that a 5% false alarm rate is acceptable, in which case $\alpha_{0\rm c} = 0.05$.

For each test window we calculate the decision threshold, $\widehat{\lambda}$, as the smallest threshold for which the empirical false alarm probability in the failure-free section does not exceed the critical value $\alpha_{0\rm c}$. This is based on μ_0, σ_0 and $\widetilde{\mu}_1, \widetilde{\sigma}_1$ for the test window.

Specifically, consider a test window whose estimated moments under H_1 are $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$. Let $F(\lambda, \widetilde{\mu}_1, \widetilde{\sigma}_1)$ denote the fraction of non-overlapping windows of length m, in the failure-free section, for which H_0 is rejected. H_0 is rejected in a window if $S_1^m \geq \lambda$ for that window. For any window, S_1^m is evaluated with eq.(14), and with the innovations of that window and the values $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$. In summary, given $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$ estimated from the test window, the $\widehat{\lambda}$ for that test window is defined as:

$$\widehat{\lambda} = \min \{\lambda : \ F(\lambda, \widetilde{\mu}_1, \widetilde{\sigma}_1) \leq \alpha_{0c}\} \tag{29}$$

As in section 4.2, $M_1(h)$ denotes the inverse of the robustness function $\widehat{h}_1(\lambda)$. The evaluation of $M_1(h)$ is specified in appendix A. For each test window we evaluate the robustness to missed-detection associated with the decision threshold $\widehat{\lambda}$, namely $\widehat{h}_1(\widehat{\lambda})$, as the value of h satisfying:

$$M_1(h) = \widehat{\lambda} \tag{30}$$

6.2 Results

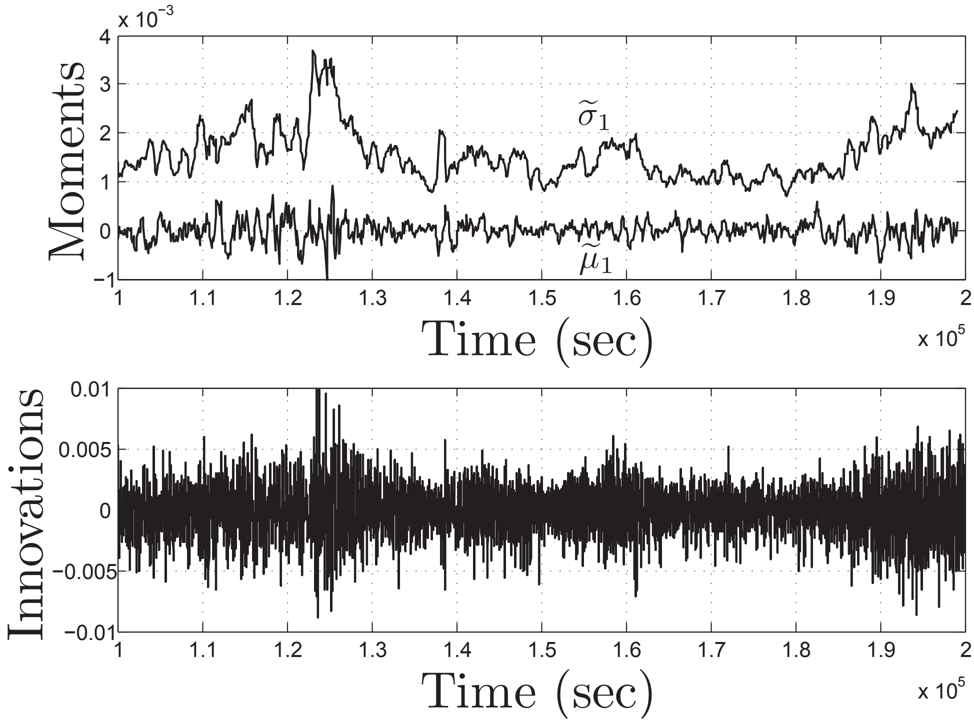

Figure 6: Innovations, means and standard deviations in the test section.

The upper frame of fig. 6 shows the estimated mean and standard deviation under H_1, $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$, for each test window. The lower frame, reproduced from fig. 1, shows the innovations. One sees that the moments reflect disturbances in the innovations, though it is also clear that the moments themselves are not clean or sensitive indications of anomalous innovations.

The 4 frames of fig. 7, from bottom to top, show the innovations in the test section, y_i, the decision threshold $\widehat{\lambda}$, the log likelihood ratio S_1^m, and the robustness for nominal rejection of H_0, $\widehat{h}_1(\widehat{\lambda})$, in each test window.

$\widehat{\lambda}$ and S_1^m in the 2nd and 3rd frames of fig. 7 show sharp deviations in regions of larger innovations, for instance around 125,000 and 195,000 [sec]. The robustness, $\widehat{h}_1$, in the top frame also deviates strongly in parallel to $\widehat{\lambda}$ and S_1^m.

The decision of the LRT for each test window is specified in eq.(11). From eq.(21) we see that $\widehat{h}_1$ is positive if and only if H_0 is nominally rejected in the corresponding test window. Thus from the top frame of fig. 7 we see that H_0 is rejected in most of the test windows before 162,000 and after 182,000 [sec].

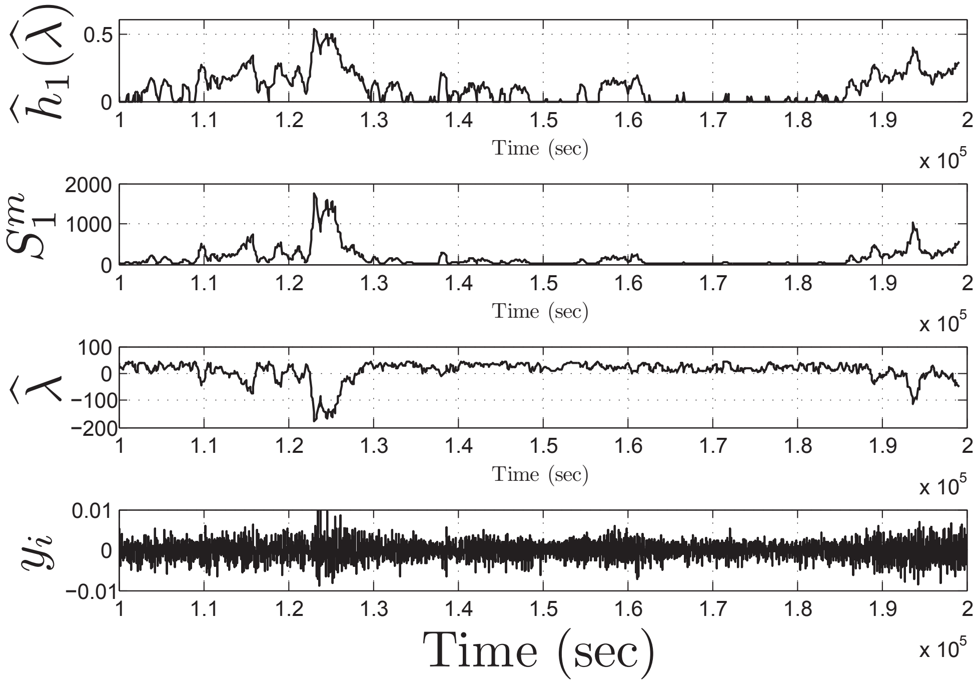

Figure 7: Innovations y_i, decision threshold $\widehat{\lambda}$, log likelihood ratio S_1^m, robustness $\widehat{h}_1(\widehat{\lambda})$, in the test section. $\alpha_{0c} = 0.05$.

However, the strength of rejection of H_0 is not uniform over the test section, as expressed by the variation of the robustness. $\widehat{h}_1$ is in the range of 0.4 to 0.5 in test windows around 125,000 [sec], and about 0.3 around 194,000 [sec], but much less over most of the remainder of the test section. From eq.(28) and the info-gap model of eq.(17) we understand that a robustness of 0.5 means that the estimated moments $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1$ under H_1, could err by as much as 0.5×0.001 and 0.5×0.005, respectively, without changing the decision. Referring to the moments in the upper frame of fig. 6, we see that a robustness of 0.5 implies quite large immunity against fluctuation in the moments. The corresponding decisions should be viewed confidently. A robustness of 0.3 is also fairly definitive. Substantally lower (but positive) robustness must be viewed cautiously as a tentative warning. The conventional LRT uses $\widehat{\lambda}$ and S_1^m to decide between H_0 and H_1 as stated in eq.(11), based on known or estimated moments. The added value of the robustness function, $\widehat{h}_1(\widehat{\lambda})$, is in assessing the confidence for rejecting H_0 in light of the severe uncertainty surrounding the moments of the pdf under failure.

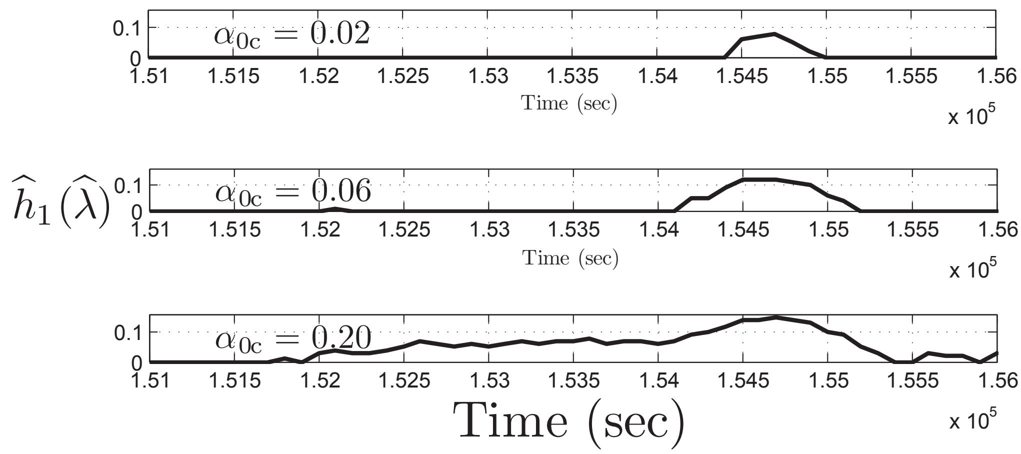

Figure 8: Robustness $\widehat{h}_1(\widehat{\lambda})$, in the test section. $\alpha_{0c} = 0.02$, 0.06, 0.20.

6.3 Detection-Error Trade Off

The 3 frames of fig. 8 show the robustness, $\widehat{h}_1$, in a short part of the test section around 155,000 [sec]. From top to bottom, the frames are evaluated at increasing α_{0c}, the maximum acceptable probability of false alarm. From the 3 frames of fig. 8 we see that the robustness increases as the maximum acceptable probability of false alarm increases: $\widehat{h}_1$ goes up as α_{0c} goes up. The LRT declaration of anomaly is more confident (larger $\widehat{h}_1$) as the likelihood of false alarm increases (larger α_{0c}). In other words, the confidence in declaring an anomaly trades off against the confidence in declaring no alarm.

This trade off suggests an info-gap robustness analog of the probabilistic detection-error trade off curve (a variation on what is sometimes called a Receiver Operating Characteristic). A traditional detection-error trade off curve shows α_1 vs. α_0, the probabilities of type II and type I errors in eqs.(6) and (7). We can estimate α_0 either with $p(y|\theta_0)$ or as $F(\lambda)$ based on failure-free data as explained earlier. In contrast, α_1 is unavailable (or highly unreliable) because $\widetilde{p}(y|\theta_1)$ is highly uncertain and data on future failures are scarce or non-existent. The info-gap analog of the probabilistic detection-error trade off replaces α_1 by $\widehat{h}_1$, and α_0 by α_{0c} as shown in fig. 9. A large value of $\widehat{h}_1$, and a small value of α_{0c}, are both desirable. The curve in fig. 9 is monotonic, expressing the trade off between the robustness against missed detection (type II error), $\widehat{h}_1$, and the probability of false alarm (type I error), α_{0c}.

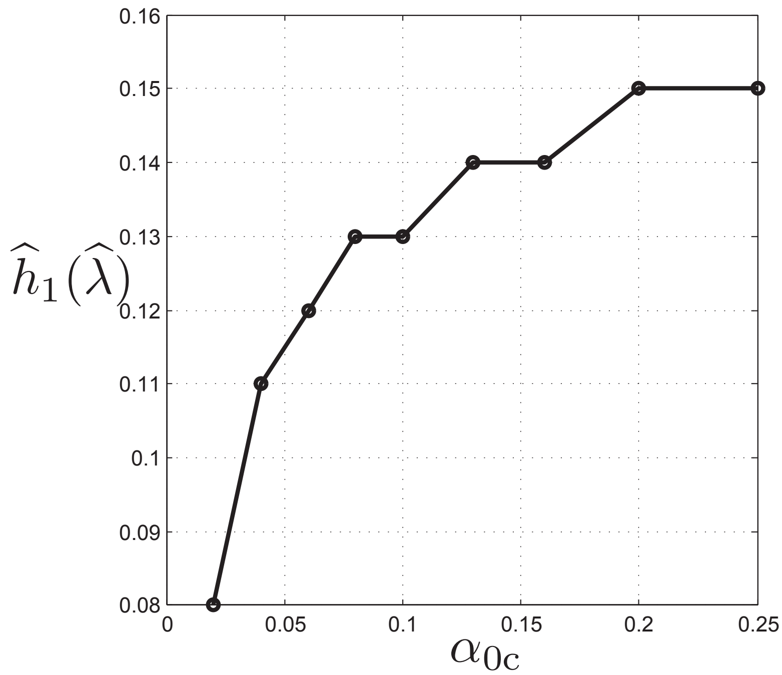

Figure 9: Robustness $\widehat{h}_1(\widehat{\lambda})$, evaluated at the test window at time 154,600 sec, as a function of the critical probability of falsely rejecting H_0, α_{0c}.

The monotonic trade off between $\widehat{h}_1(\widehat{\lambda})$ and α_{0c} is a generic property that does not depend on the specific info-gap model underlying fig. 9, as we now explain. The relation between $\widehat{\lambda}$ and α_{0c} is defined in eq.(29) which does not depend on the specific info-gap model used in the example. From eq.(29) we see that $\widehat{\lambda}$ is non-increasing as α_{0c} increases. From eq.(25) we see that $\widehat{h}_1(\widehat{\lambda})$ is non-increasing as $\widehat{\lambda}$ increases, which also is a generic property, independent of the specific info-gap model used in the numerical example (as explained prior to eq.(25)). Combining these observations we see that the trade off illustrated in fig. 9 is a general property that will be observed for any info-gap model.

7 Conclusion

The challenges of failure detection are very different from those of state estimation with additive noise. In state estimation—Kalman filtering and its derivatives for instance—one often presumes complete knowledge of the probability distribution of the noise. Failures, in contrast, are often unique and unprecedented. Stochastic characterization of failures, based on historical data, ignores the endless "creativity" of failure avenues and mechanisms. There is typically a substantial gap between what the analyst *does know* about failures before they occur, and what *needs to be known* for a full probabilistic analysis. This gap in the information can be modeled and managed with info-gap decision theory as demonstrated in this paper.

An info-gap robustness analysis does not presume that the analyst knows nothing about the uncertain failures. On the contrary, a great diversity of info-gap models of uncertainty are available for representing different degrees and types of knowledge and ignorance (Ben-Haim, 2006). All info-gap models share two properties—contraction and nesting—that reflect unbounded uncertainty in failure-dependent features and in how failures affect the discriminating features. However, one may exploit very specific and precise information, such as normality in our example, without requiring knowledge of other aspects of the failure.

The analyst has data, models, and some understanding about future failures. However, one cannot answer the question 'How wrong is this information?' without extensive—usually prohibitive—effort. One can, however, answer the question 'How much error can be tolerated?'. The info-gap robustness function provides a quantitative answer to this question. The robustness function is non-probabilistic and less informative than most stochastic assessments of risk. On the other hand, the robustness function, and the info-gap model of uncertainty upon which it is based, depend on less information than probabilistic models. The info-gap robustness analysis is attractive for designing failure detection algorithms because information is usually scarce.

These ideas are demonstrated in a simple yet challenging industrial application of failure detection from measurements. As is very common in failure detection, the industrial colleagues have extensive experience with the no-failure condition but only limited experience with the failed condition. Consequently it was possible to construct good models for the measured signals in the no-fail condition and to characterize well the pdf of the resulting prediction errors (the innovations). Hence the uncertainty associated with the innovations in the no-failure case is negligible compared with the uncertainty of the innovations under the failure condition. The latter is severe since experience with failure is limited to only a few examples that do not cover all the possible failure scenarios. Indeed, the system may fail in many other ways not covered by current experience, and thus the pdf of the no-fail innovations may differ from the estimated pdf under the failure condition. We make the simplifying yet reasonable assumption that the pdf of the innovations remains normal under failure, and we represent the uncertainty in the moments of the normal distribution (both mean and variance) by info-gap models centered on their estimates. Thus, there is no limit on the moments and no assumption about their distribution. The info-gap framework allows us to compute the robustness function of the nominal decision—failure or no-failure—to uncertainties in the moments: i.e., what is the maximum horizon of uncertainty for which the nominal decision remains unchanged. These robustness functions depend on the decision threshold for the LRT, and provide a method for selecting the threshold level. This method is demonstrated in specific examples, and generates an info-gap robustness analog of the probabilistic detection-error trade off curve (a variation on what is sometimes called a Receiver Operating Characteristic). In addition, the robustness functions for nominal acceptance or rejection of the null hypothesis can be interpreted as robustnesses against type I and type II errors that could result from erroneous probability distributions.

In summary, the likelihood ratio test is a powerful method for failure detection. However, it relies on knowledge of the pdf of the innovations under both the no-failure and the failure cases. We have described how to capture uncertainties in these pdfs, and in particular in the pdf under the failure condition, using info-gap models of uncertainty. We derive the resulting robustness curves of the nominal decision and characterize the underlying trade-offs. These ideas and features are demonstrated in the context of a simple yet challenging industrial application. By integrating info-gap theory with the likelihood ratio test we provide a tool for evaluating the robustness of the resulting failure detection.

8 References

1. Basseville, Michéle and Albert Benveniste, eds., 1984, *Detection of Abrupt Changes in Signals and Dynamical Systems,* Springer-Verlag, Berlin.

2. Basseville, Michéle and Igor V. Nikiforov, 1993, *Detection of Abrupt Changes: Theory and Application,* Prentice-Hall.

3. Ben-Haim, Yakov, 2006, *Info-Gap Decision Theory: Decisions Under Severe Uncertainty,* 2nd edition, Academic Press, London.

4. Ben-Haim, Yakov, 2010, *Info-Gap Economics: An Operational Introduction,* Palgrave-Macmillan, London.

5. Ben-Haim, Yakov, 2011, Interpreting null results from measurements with uncertain correlations: An info-gap approach, *Risk Analysis,* 31(1):78–85.

6. Ben-Tal, A. and A. Nemirovski, 1999, Robust solutions of uncertain linear programs, *Oper. Res. Lett.,* 25, 1–13.

7. Campbell, Stephen L. and Ramine Nikoukhah, 2004, *Auxiliary Signal Design for Failure Detection,* Princeton University Press, Princeton.

8. Chow, Edward Y. and Alan S. Willsky, 1984, Analytical redundancy and the design of robust failure detection systems, *IEEE Trans Automatic Control,* 29(7): 603–614.

9. Davies, A., ed., 1998, *Handbook of Condition Monitoring,* Chapman & Hall, London.

10. DeGroot, Morris H., 1986, *Probability and Statistics,* 2nd ed., Addison-Wesley, Reading, MA.

11. Ding, S.X., T. Jeinsch, P.M. Frank and E.L. Ding, 2000, A unified approach to the optimization of fault detection systems, *Int. J. Adapt. Control Signal Processing,* 14:725–745.

12. Earls, C.J., 2013, Bayesian inference of hidden corrosion in steel bridge connections: Non-contact and sparse contact approaches, *Mechanical Systems and Signal Processing,* 41: 420–432.

13. Frank, Paul M., 1990, Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-based Redundancy: A Survey and Some New Results, *Automatica,* 26(3): 459–474.

14. Gertler, Janos J., 1988, A survey of model based failure detection and isolation in complex plants, *IEEE Control Systems Magazine,* 8(6): 3–11.

15. Gertler, Janos J., 1998, *Fault Detection and Diagnosis in Engineering Systems,* Marcel Dekker Inc., New York.

16. Hites, R., Y. De Smet, N. Risse, M. Salazar-Neumann and P. Vincke, 2006, About the applicability of MCDA to some robustness problems, *Eur. J. Oper. Res.* 174: 322–332.

17. Huber, P.J., 1981, *Robust Statistics,* John Wiley, New York.

18. Hwang, Kyu Hwan, Jong Min Lee and Yoha Hwang, 2013, A new machine condition monitoring method based on likelihood change of a stochastic model, *Mechanical Systems and Signal Processing,* 41: 357–365.

19. Knight, F.H., 1921, *Risk, Uncertainty and Profit,* Hart, Schaffner and Marx. Re-issued by Harper Torchbooks, New York, 1965.

20. Lempert, Robert J., Steven W. Popper and Steven C. Bankes, 2003, *Shaping the Next 100 Years: New Methods for Quantitative, Long-Term Policy Analysis,* RAND Corp. Santa Monica, CA.

21. Mirer, Tanya and Yakov Ben-Haim, 2010, Reliability assessment of explosive material based on penalty tests: An info-gap approach, *J Risk and Reliability,* Proc. Inst. Mech. Engrs. Part O, 224(4): 346–355.

22. Patton, Ron, Paul Frank and Robert Clark, eds., 1989, *Fault Diagnosis in Dynamic Systems: Theory and Applications.* Prentice Hall, New York.

23. Pau, L.F., 1981, *Failure Diagnosis and Performance Monitoring,* Marcel Dekker, Inc., New York.

24. Pierce, S.G., K. Worden and G. Manson, 2006, A novel information-gap technique to assess reliability of neural network-based damage detection *J. Sound Vibration,* 293 (1–2): 96–111.

25. Roy, B., 2010, Robustness in operational research and decision aiding: A multi-faceted issue, *Eur. J. Oper. Res.* 200: 629–638.

26. Tibaduiza, D.-A., M.-A. Torres-Arredondo, L.E. Mujica, J. Rodellar and C.-P. Fritzen, 2013, A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring, *Mechanical Systems and Signal Processing,* 41: 467–484.

27. Wald, Abraham, 1945, Statistical decision functions which minimize the maximum risk, *Annals of Mathematics,* 46(2): 265–280.

28. Williams, J.H., 1998, System monitoring and the use of models, pp.219–239 in Davies, (1998).

29. Willsky, Alan S., 1976, A survey of design methods for failure detection in dynamic systems, *Automatica,* 12: 601–611.

30. Willsky, Alan S., 1984, Detection of abrupt changes in dynamic systems, pp.27–49 in Basseville and Benveniste (1984).

31. Zacksenhouse, Miriam, Simona Nemets, Mikhail A Lebedev and Miguel A Nicolelis, 2009, Robust satisficing linear regression: Performance/robustness trade-off and consistency criterion, *Mechanical Systems and Signal Processing,* 23: 1954–1964.

Appendix

A Derivation of the Inverse Robustness Functions

Let $M_i(h)$, for $i = 0$ or 1, denote the inner optimum in the definition of the robustness, eq.(19) or eq.(21) respectively. $M_i(h)$ is the inverse of $\widehat{h}_i(\lambda)$ as explained following eq.(21). $M_i(h)$ is evaluated as follows.

Given observed innovations $Y_1^m = (y_1, \ldots, y_m)$, denote the sample mean and sample variance from onset of failure as:

$$\overline{y} = \frac{1}{m - i_{\rm c} + 1} \sum_{i=i_{\rm c}}^{m} y_i \tag{31}$$

$$s_y^2 = \frac{1}{m - i_{\rm c} + 1} \sum_{i=i_{\rm c}}^{m} (y_i - \overline{y})^2 \tag{32}$$

Consider S_1^m in eq.(14) as a function of the moments μ_1 and σ_1. In appendix B we show that:

1. $S_1^m(\mu_1, \sigma_1)$ has a global maximum when:

$$\mu_1 = \overline{y} \quad \text{and} \quad \sigma_1^2 = s_y^2 \tag{33}$$

2. $S_1^m(\mu_1, \sigma_1)$ has no other local optima. That is, $S_1^m(\mu_1, \sigma_1)$ increases as μ_1 moves towards $\overline{y}$, or as σ_1 moves towards s_y.

These conclusions result from the fact that the sample mean and variance, $\overline{y}$ and s_y^2, are maximum likelihood estimates of the population moments (DeGroot, 1986, pp.342–343).

Nominal acceptance of H_0. We evaluate the inverse of the robustness, $\widehat{h}_0(\lambda)$, in eq.(19).

As we explained above, $S_1^m(\mu_1, \sigma_1)$ has a global maximum with μ_1 and σ_1 at the sample moments, $\overline{y}$ and s_y in eqs.(31) and (32). Also, $S_1^m(\mu_1, \sigma_1)$ increases as μ_1 or σ_1 move towards from $\overline{y}$ or s_y, respectively. We derive an explicit expression for $M_0(h)$, the maximum of $S_1^m(\mu_1, \sigma_1)$ at horizon of uncertainty h. $M_0(h)$ is the inverse of the robustness, $\widehat{h}_0(\lambda)$.

Let $(x)^+$ equal x if x is positive, and equal zero otherwise. Define:

$$\mu_- = \widetilde{\mu}_1 - s_\mu h \tag{34}$$

$$\mu_+ = \widetilde{\mu}_1 + s_\mu h \tag{35}$$

$$\sigma_- = (\widetilde{\sigma}_1 - s_\sigma h)^+ \tag{36}$$

$$\sigma_+ = \widetilde{\sigma}_1 + s_\sigma h \tag{37}$$

Now define:

$$\mu_1^\star = \begin{cases} \min(\mu_+,\ \overline{y}), & \text{if } \widetilde{\mu}_1 \le \overline{y} \\ \max(\mu_-,\ \overline{y}), & \text{else} \end{cases} \tag{38}$$

$$\sigma_1^\star = \begin{cases} \min(\sigma_+,\ s_y), & \text{if } \widetilde{\sigma}_1 \le s_y \\ \max(\sigma_-,\ s_y), & \text{else} \end{cases} \tag{39}$$

It is evident that:

$$M_0(h) = S_1^m(\mu_1^\star, \sigma_1^\star) \tag{40}$$

We now have an explicit expression for $M_0(h)$. Plotting $M_0(h)$ vs. h is identical to plotting λ vs. $\widehat{h}_0(\lambda)$.

Nominal rejection of H_0. We evaluate the inverse of the robustness, $\widehat{h}_1(\lambda)$, in eq.(21).

As we explained above, $S_1^m(\mu_1, \sigma_1)$ has a global maximum with μ_1 and σ_1 at the sample moments, $\overline{y}$ and s_y in eqs.(31) and (32). Also, $S_1^m(\mu_1, \sigma_1)$ decreases as μ_1 or σ_1 move away from $\overline{y}$ or s_y, respectively. We derive an explicit expression for $M_1(h)$, the minimum of $S_1^m(\mu_1, \sigma_1)$ at horizon of uncertainty h. $M_1(h)$ is the inverse of the robustness, $\widehat{h}_1(\lambda)$.

It is evident that $M_1(h)$ equals the value of $S_1^m(\mu_1, \sigma_1)$ with the minimizing combination of μ and σ from among eqs.(34)–(37):

$$M_1(h) = \min\left\{S_1^m(\mu_-, \sigma_-),\ S_1^m(\mu_+, \sigma_-),\ S_1^m(\mu_-, \sigma_+),\ S_1^m(\mu_+, \sigma_+)\right\} \tag{41}$$

We now have an explicit expression for $M_1(h)$. Plotting $M_1(h)$ vs. h is identical to plotting λ vs. $\widehat{h}_1(\lambda)$.

B Maximum of the Log-Likelihood Ratio

Lemma 1 *The sum of the log-likelihood ratios, $S_1^m(\mu_1, \sigma_1)$, has a single extremum.*
Given:

- *A sample $y_{i_c}, \ldots, y_m$.*
- *Sample estimates of μ_1 and σ_1^2, $\overline{y}$ and s_y^2 in eqs.(31) and (32).*
- *The sum of the log-likelihood ratios, $S_1^m(\mu_1, \sigma_1)$ in eq.(14), considered as a function of μ_1 and σ_1.*

Then:

$S_1^m(\mu_1, \sigma_1)$ has a single extremum—a maximum—at:

$$\mu_1 = \overline{y}, \quad \sigma_1^2 = s_y^2 \tag{42}$$

Proof of Lemma 1. We find this optimum from the first and second derivatives of $S_1^m(\mu_1, \sigma_1)$ in eq.(14) with respect to μ_1 and σ_1.

The first derivatives are:

$$\frac{\partial S}{\partial \mu_1} = \frac{1}{\sigma_1^2} \sum_{i=i_c}^{m} (y_i - \mu_1) \tag{43}$$

$$\frac{\partial S}{\partial \sigma_1} = -\frac{m - i_c + 1}{\sigma_1} + \frac{1}{\sigma_1^3} \sum_{i=i_c}^{m} (y_i - \mu_1)^2 \tag{44}$$

Note that these derivatives vanish at the sample moments in eqs.(31) and (32):

$$\mu_1 = \overline{y} \quad \text{and} \quad \sigma_1^2 = s_y^2 \tag{45}$$

These are the only values (for non-negative σ_1) at which the derivatives vanish. Thus the slopes change sign only once. Thus eq.(45) is a global extremum and there are no other local extrema. We will show that it is a maximum.

The elements of the Hessian matrix are:

$$\frac{\partial^2 S}{\partial \mu_1^2} = -\frac{m - i_c + 1}{\sigma_1^2} \tag{46}$$

$$\frac{\partial^2 S}{\partial \mu_1 \partial \sigma_1} = -\frac{2}{\sigma_1^3} \sum_{i=i_c}^{m} (y_i - \mu_1) \tag{47}$$

$$\frac{\partial^2 S}{\partial \sigma_1^2} = \frac{m - i_c + 1}{\sigma_1^2} - \frac{3}{\sigma_1^4} \sum_{i=i_c}^{m} (y_i - \mu_1)^2 \tag{48}$$

Evaluating these elements at $\overline{y}$ and s_y results in:

$$\frac{\partial^2 S}{\partial \mu_1^2} = -\frac{m - i_c + 1}{s_y^2} \tag{49}$$

$$\frac{\partial^2 S}{\partial \mu_1 \partial \sigma_1} = 0 \tag{50}$$

$$\frac{\partial^2 S}{\partial \sigma_1^2} = -2\frac{m - i_c + 1}{s_y^2} \tag{51}$$

The determinant of the Hessian matrix is positive and the first element is negative, so $\overline{y}$ and s_y are a local maximum. Since the derivatives change sign only once, this is a global maximum and there are no other local extrema. ■