Probabilistic Fault Diagnosis
with Automotive Applications

Anna Pernestål
Probabilistic Fault Diagnosis
with Automotive Applications

© 2009 Anna Pernestål

annap@isy.liu.se
http://www.vehicular.isy.liu.se
Department of Electrical Engineering,
Linköping University,
SE–581 83 Linköping,
Sweden.

ISBN 978-91-7393-493-0
ISSN 0345-7524

Printed by LiU-Tryck, Linköping, Sweden 2009
To my parents
Kjell and Eva
Abstract

The aim of this thesis is to contribute to improved diagnosis of automotive vehicles. The work is driven by case studies, where problems and challenges are identified. To solve these problems, theoretically sound and general methods are developed. The methods are then applied to the real world systems.

To fulfill performance requirements automotive vehicles are becoming increasingly complex products. This makes them more difficult to diagnose. At the same time, the requirements on the diagnosis itself are steadily increasing. Environmental legislation requires that smaller deviations from specified operation must be detected earlier. More accurate diagnostic methods can be used to reduce maintenance costs and increase uptime. Improved diagnosis can also reduce safety risks related to vehicle operation.

Fault diagnosis is the task of identifying possible faults given current observations from the systems. To do this, the internal relations between observations and faults must be identified. In complex systems, such as automotive vehicles, finding these relations is a most challenging problem due to several sources of uncertainty. Observations from the system are often hidden in considerable levels of noise. The systems are complicated to model both since they are complex and since they are operated in continuously changing surroundings. Furthermore, since faults typically are rare, and sometimes never described, it is often difficult to get hold of enough data to learn the relations from.

Due to the several sources of uncertainty in fault diagnosis of automotive systems, a probabilistic approach is used, both to find the internal relations, and to identify the faults possibly present in the system given the current observations. To do this successfully, all available information is integrated in the computations.

Both on-board and off-board diagnosis are considered. The two tasks may seem different in nature: on-board diagnosis is performed without human integration, while the off-board diagnosis is mainly based on the interactivity with a mechanic. On the other hand, both tasks regard the same vehicle, and information from the on-board diagnosis system may be useful also for off-board diagnosis. The probabilistic methods are general, and it is natural to consider both tasks.

The thesis contributes in three main areas. First, in Paper 1 and 2, methods are developed for combining training data and expert knowledge of different kinds to compute probabilities for faults. These methods are primarily developed with on-board diagnosis in mind, but are also applicable to off-board diagnosis. The methods are general, and can be used not only in diagnosis of technical system, but also in many other applications, including medical diagnosis and econometrics, where both data and expert knowledge are present.

The second area concerns inference in off-board diagnosis and troubleshooting, and the contribution consists in the methods developed in Paper 3 and 4.
The methods handle probability computations in systems subject to external interventions, and in particular systems that include both instantaneous and non-instantaneous dependencies. They are based on the theory of Bayesian networks, and include event-driven non-stationary dynamic Bayesian networks (nsDBN) and an efficient inference algorithm for troubleshooting based on static Bayesian networks. The framework of nsDBN event-driven nsDBN is applicable to all kinds of problems concerning inference under external interventions.

The third contribution area is Bayesian learning from data in the diagnosis application. The contribution is the comparison and evaluation of five Bayesian methods for learning in fault diagnosis in Paper 5. The special challenges in diagnosis related to learning from data are considered. It is shown how the five methods should be tailored to be applicable to fault diagnosis problems.

To summarize, the five papers in the thesis have shown how several challenges in automotive diagnosis can be handled by using probabilistic methods. Handling such challenges with probabilistic methods has a great potential. The probabilistic methods provide a framework for utilizing all information available, also if it is in different forms and. The probabilities computed can be combined with decision theoretic methods to determine the appropriate action after the discovery of reduced system functionality due to faults.
Sannolikhetsbaserad Diagnos med Fordonstillämpningar


Diagnos handlar om att hitta fel som är närvarande i ett system genom att använda ett flertal observationer från systemet och relationer mellan dessa. I dagens och morgondagens moderna fordon innebör detta många utmaningar, i synnerhet eftersom de flesta relationer innehåller osäkerheter. Det är utmanande att konstruera noggranna och tillförlitliga fysikaliska modeller av systemen, då de är mycket komplexa och verkar i en omgivning som ständigt förändras när fordonet kör på vägen. Vidare är det ofta svårt att samla data från fordonen för att lära relationer mellan observationer, i synnerhet från feltillstånd, eftersom fel typiskt är ovanliga och ibland till och med har okänd effekt på observationerna. Dessutom är beräkningskapaciteten, åtminstone för diagnos som ska utföras ombord på fordonet, ofta begränsad. Detta beror på att de processorer som klarar den utsatta miljön ombord har betydligt sämre prestanda än processorer till exempel i en PC. På verkstaden möts man av svårigheten att felen
i fordonet inte nödvändigtvis är synligt när fordonet är stilla. Till exempel är det svårt att upptäcka problem med bromsarna när inte bromsarna används.

Flera av utmaningarna inom fordonsdiagnos är relaterade till osäkerheter och otillräcklig information. Därför antas ett sannolikhetsbaserat förhållningssätt i den här avhandlingen, både när det gäller att hitta relationerna mellan observationerna, och för att detektera fel. Målet är att beräkna sannolikheterna att lika fel är närvarande. För att lyckas med detta är det viktigt att all tillgänglig information används i beräkningarna.

I avhandlingen betraktas både diagnos utförd ombord på fordonet och diagnos gjord på verkstad. Diagnos ombord och verkstadsdiagnos kan förefalla vara två helt olika problem. Ombord görs diagnosen automatiskt i styrsystemet och (i de flesta fall) helt utan inblandning av människor, till skillnad från diagnos på verkstäder som i första hand utförs av mekanikern, stöttad av ett felsökningsverktyg. Än andra sidan gäller diagnosen samma fordon, och information från diagnosen i styrsystemet ombord kan vara till stor hjälp under felsökningen på verkstaden. Inom det ramverk för diagnos, baserat på sannolikhetsteori, som används och utvecklas i den här avhandlingen, är metoderna generella och kan appliceras på diagnos både ombord och på verkstaden. Därför blir det naturligt att betrakta båda typerna av diagnos.

Den här avhandlingen bidrar i första hand inom tre områden. Det första området är metoder för att kombinera olika typer av information i sannolikhetsberäkningar. I artiklarna 1 och 2 har metoder utvecklats för att kombinera träningsdata och expertkunskap av olika typer. Metoderna är generella och kan inte bara användas inom diagnos, utan även inom många fält, till exempel medicinsk diagnos och ekonomisk modellering. Metoderna i artiklarna 1 och 2 har i första hand utvecklats med avseende på diagnos ombord, men kan självklart även användas inom verkstadsdiagnos.


Det tredje bidraget, presenterat i artikel 5, är en jämförelse och utvärdering av olika metoder lära relationer mellan observationer och fel från träningsdata.
Att lära från data för diagnos ställer särskilda krav på algoritmerna som används, och i artikel 5 har ett fem olika metoder anpassats till diagnos-problemet och deras prestanda har jämförts.

Genom hela avhandlingen har arbetet drivits av fallstudier av delsystem i en modern lastbil, där olika problem och svårigheter har identifierats. Teoretiskt sunda och generella metoder har utvecklats för att lösa dessa problem. Metoderna har sedan applicerats på de riktiga systemen i lastbilen.
I believe searching faults is like a detective’s work. We observe the system, discuss the hidden relations, using whatever we know about the system, and draw conclusions about whether there are faults present and, if so, which faults. Therefore, searching faults and doing diagnostic work is about understanding relations between observations and different faults, and to distinguish the relevant information in the observations. To design a diagnosis system, we have to find the relations. To perform diagnostic work, we have to reason using the relations and the current observations.

There are several different methods for learning the hidden relations in systems to diagnose: building models, using data, applying expert systems, and so on. However, digging deeper into the problem designing a diagnosis system, we notice that the available information is (often) not sufficient to exactly determine if there are faults present, nor to distinguish between them. We are left with a bunch of possible explanations.

This fact leads into the field of probability theory. When dealing with probabilities, and in particular probabilities about “real-world” events, such as “what is the probability that this truck is fault free?”, one need to know what “probability” is.

So, what is probability? Before beginning the work with this thesis, I would have said something like “Well, the probability is the relative frequency. I suppose.” However, I must confess, I had some problems with this interpretation. First, even if fault $F$ is present in 1 out of 100 trucks, i.e. has relative frequency 0.01, what is the probability that the fault is present in this particular truck?
Second, if a person I trust tells me that this truck is fault free, what is the probability that this the truck is fault free then? It is reasonable that it depends on how much I trust the person?

My problems with the interpretation of probability are, at least philosophically, solved through inspiring and interesting discussions with Mikael Sternard and Mathias Johansson at the Signals and Systems group at Uppsala University five years ago. They introduced me to E. T. Jaynes’ book *Probability – the Logic of Science* on probability as an extension to logic. According to Jaynes, probability is a property of the spectator and his state of knowledge rather than a “physical” property of the object. This gave me an understanding of probability as a measure of belief that has made this thesis possible. Without Mikael and Mathias it is highly probable that this thesis had been something completely different.

One of the most important persons during the work with this thesis has been my supervisor Dr. Mattias Nyberg. He has supported me through this work by pushing my ideas further, and efficiently puncturing my bad ideas. He has always new questions coming up, and new ideas about how the world and the work is. It has been an intellectual challenge to work with Mattias - and I love challenges.

This thesis has been performed as a collaborative industrial research project between Scania CV AB in Södertälje and the division of Vehicular Systems, Department of Electrical Engineering, Linköping University. I thank my managers at Scania for supporting this work and making it financially possible. Thanks to Prof. Lars Nielsen, for letting me join the Vehicular Systems group in Linköping, and to the people at the group, and in particular at the diagnosis group of Vehicular Systems, for the interesting discussions and for broadening my perspective on diagnosis (and many other things).

Other persons that have been more important for this work are my co-supervisor Dr. Jose M. Peña, with his knowledge on Bayesian networks; Dr. Nils-Gunnar Vågstedt and Hans Ivendahl at Scania, with their encouragement, and “real-world related questions” that have helped me to focus on the real problems; and Prof. Petri Myllymäki and Hannes Wettig at the CoSCo group at Helsinki University for hosting me and introducing me to learning methods.

Carl Svärd, Håkan Warnquist, and Dr. Tony Lindgren have proof-read parts of this thesis. Your comments have been invaluable.

A special thank to Dr. Erik Frisk for his support on LATEX, his never-ending interest, and his clever comments and questions.

To Support(er) Petter Lindh, for his infectious harmony, and his thoughtful comments, always given to me with excellent timing and content.

Many people know that I am addicted to long-distance running, and I think that the work with this thesis has been much like running a marathon race. A marathon is a challenge that, during the race, is sometimes simply fun, some-
times painful and heavy, often exhausting – but, the whole way through, a great
pleasure! I will end this marathon with thanking my supporters that have helped
me, encouraged me, and supported me through this marathon: my friends, my
grandparents, and my wonderful family Karin, Kjell, Eva, and Johan.

Anna Pernestål
Linköping 2009
Contents

I Introduction 1

1 Introduction 3
  1.1 Background 3
  1.1.1 Why Automotive Diagnosis? 3
  1.1.2 Diagnosis is a Challenge 4
  1.1.3 Approaches to Diagnosis 5
  1.2 Problem Formulation 6

2 Contributions 9
  2.1 Thesis Overview 9
  2.2 Appended Papers – Summary and Contributions 11
    2.2.1 Paper 1 - Data and Process Knowledge 11
    2.2.2 Paper 2 - Data and Likelihood Constraints 12
    2.2.3 Paper 3 - Non-Stationary Dynamic Bayesian Networks 13
    2.2.4 Paper 4 - Modeling and Inference for Troubleshooting 14
    2.2.5 Paper 5 - Comparing Methods for Learning 16
  2.3 List of Publications 17

References 19
II  Probability Theory in Diagnosis

3  Bayesian Probability Theory
   3.1  Dealing With Uncertainty .......................... 23
   3.2  Interpretations of Probability ....................... 25
   3.3  The Interpretation of Probability Used in the Thesis 26

4  A Brief Survey of Probability Based Diagnosis 29
   4.1  Model-Based Diagnosis ............................... 29
   4.1.1  Diagnosis Methods ............................. 29
   4.1.2  Logical Models ................................ 30
   4.1.3  Black Box Models ............................... 30
   4.1.4  Physical Models ................................ 31
   4.1.5  Discrete Event Systems ......................... 32
   4.2  Probabilistic Methods for Diagnosis ................ 32
   4.2.1  An Example: the Car Start Problem .......... 32
   4.2.2  What is Probabilistic Diagnosis? ............ 32
   4.3  Methods For Probabilistic Diagnosis .............. 33
   4.3.1  Dynamic Physical Models ..................... 34
   4.3.2  Data-Driven Black Box Models ................. 35
   4.3.3  Bayesian Networks ............................. 36

References 39

III  Papers

1  Bayesian Fault Diagnosis for Automotive Engines by Combin-
ing Data and Process Knowledge 47
   1  Introduction ..................................... 50
   2  The Automotive Diagnosis Problem .................. 51
      2.1  Motivating Application ......................... 52
      2.2  Requirements on the Solution .................. 54
      2.3  Related Work ................................ 55
      2.4  Example Application: The Diesel Engine ........ 56
   3  Problem Formulation ................................ 57
      3.1  Notation ..................................... 59
      3.2  Formal Problem Formulation ................... 60
   4  Two Types of Knowledge .............................. 60
      4.1  Training Data ................................ 61
      4.2  Process Knowledge .............................. 61
   5  Diagnosis Using Training Data ....................... 62
      5.1  One Observation ................................ 63
Bayesian Inference by Combining Training Data and Background Knowledge Expressed as Likelihood Constraints

1 Introduction

2 Preliminaries
2.1 Notation
2.2 Background Knowledge

3 Inference Using Data Only

4 Inference Using Data and Background Knowledge
4.1 Background Knowledge as Constraints
4.2 Computing the Probability of $Z$ under constraints
4.3 Parameter Transformation

5 Computing the Integrals
5.1 Characteristics of the Integral

6 Examples
6.1 Analytical Solution vs. Laplace Approximation
6.2 Diagnosis Example

7 Related Work

8 Conclusions
3 Non-stationary Dynamic Bayesian Networks in Modeling of Troubleshooting Processes

1 Introduction ............................................. 124
2 Related Work ........................................... 125
3 The Troubleshooting Scenario ............................ 126
  3.1 The OPG System .................................... 126
  3.2 Variables ........................................... 127
  3.3 Troubleshooting Actions ............................ 130
  3.4 Actions, Evidence, and Events ..................... 130
4 Dynamic Bayesian Networks ............................. 130
  4.1 Definitions of BN and DBN ......................... 130
  4.2 Characterizing an nsDBN ........................... 132
5 Building Non-stationary DBN Driven by Events ......... 133
  5.1 Initial BN ......................................... 133
  5.2 Nominal Transition BN ............................. 133
  5.3 Effects of Events ................................ 134
6 Inference in Event Driven non-stationary DBN ......... 135
  6.1 A Recursive Inference Algorithm .................. 135
  6.2 Frontier and Interface Algorithms ................ 137
7 Application to Troubleshooting ........................ 138
  7.1 Preparation: Building nsDBN for Troubleshooting .. 138
  7.2 Inference: Computing Probabilities ............... 142
8 Conclusions ............................................ 145
References .................................................. 146

4 Modeling and Efficient Inference for Troubleshooting Automotive Systems

1 Introduction ............................................. 152
2 Preliminaries .......................................... 154
  2.1 Notation .......................................... 154
  2.2 Bayesian Networks ................................ 154
3 The Troubleshooting Scenario and System ............... 155
  3.1 Motivating Application - the Retarder .............. 155
  3.2 The Troubleshooting Scenario ....................... 156
  3.3 The Troubleshooting System ......................... 156
  3.4 Variables ......................................... 158
4 Planner ................................................ 160
  4.1 Optimal Expected Cost of Repair ................... 160
  4.2 Search Graph ...................................... 161
5 Modeling for Troubleshooting .......................... 163
  5.1 Practical Issues when Building BN for Troubleshooting . 164
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Repairs, Operations, and Interventions</td>
<td>165</td>
</tr>
<tr>
<td>5.3 Event-Driven Non-stationary DBN</td>
<td>166</td>
</tr>
<tr>
<td>6 Diagnoser: Belief State Updating</td>
<td>169</td>
</tr>
<tr>
<td>6.1 Observation Actions</td>
<td>170</td>
</tr>
<tr>
<td>6.2 Repair Actions</td>
<td>170</td>
</tr>
<tr>
<td>6.3 Operation Actions</td>
<td>171</td>
</tr>
<tr>
<td>7 Diagnoser: BN Updating</td>
<td>171</td>
</tr>
<tr>
<td>7.1 BN Updating Example</td>
<td>173</td>
</tr>
<tr>
<td>7.2 BN Updating Algorithm</td>
<td>176</td>
</tr>
<tr>
<td>8 Modeling Application</td>
<td>183</td>
</tr>
<tr>
<td>9 Conclusion and Future Work</td>
<td>185</td>
</tr>
<tr>
<td>9.1 Conclusion</td>
<td>185</td>
</tr>
<tr>
<td>9.2 Future Work</td>
<td>186</td>
</tr>
<tr>
<td>References</td>
<td>193</td>
</tr>
</tbody>
</table>

5 A Comparison of Bayesian Approaches to Learning in Fault Isolation 195

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>198</td>
</tr>
<tr>
<td>2 Preliminaries</td>
<td>199</td>
</tr>
<tr>
<td>2.1 Notation</td>
<td>200</td>
</tr>
<tr>
<td>2.2 Fundamentals of Bayesian Networks</td>
<td>200</td>
</tr>
<tr>
<td>3 Bayesian Fault Isolation</td>
<td>200</td>
</tr>
<tr>
<td>3.1 Problem Formulation</td>
<td>201</td>
</tr>
<tr>
<td>3.2 Performance Measures</td>
<td>202</td>
</tr>
<tr>
<td>4 Modeling Methods</td>
<td>203</td>
</tr>
<tr>
<td>4.1 Modeling Assumptions</td>
<td>203</td>
</tr>
<tr>
<td>4.2 Direct Inference</td>
<td>206</td>
</tr>
<tr>
<td>4.3 Bayesian Network Methods</td>
<td>206</td>
</tr>
<tr>
<td>4.4 Regression</td>
<td>208</td>
</tr>
<tr>
<td>5 Experiments</td>
<td>210</td>
</tr>
<tr>
<td>5.1 Experimental Setup</td>
<td>210</td>
</tr>
<tr>
<td>5.2 Results</td>
<td>211</td>
</tr>
<tr>
<td>6 Conclusions</td>
<td>213</td>
</tr>
<tr>
<td>References</td>
<td>215</td>
</tr>
</tbody>
</table>

IV Concluding Remarks 219

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Concluding Remarks</td>
<td>221</td>
</tr>
<tr>
<td>1 Conclusions</td>
<td>221</td>
</tr>
<tr>
<td>2 Future Research</td>
<td>223</td>
</tr>
<tr>
<td>References</td>
<td>225</td>
</tr>
</tbody>
</table>
A Interpretations of Probability

A.1 Dealing With Uncertainty

A.2 Interpretations of Probability
   A.2.1 Bayesians and Frequentists
   A.2.2 Switching Between Interpretations

A.3 The Bayesian View: Probability as an Extension to Logic
   A.3.1 Consistency and Common Sense
   A.3.2 The Statements Behind the | -sign

A.4 Assigning Numbers
   A.4.1 Principle of Indifference
   A.4.2 Jeffreys Prior
   A.4.3 Maximum Entropy
   A.4.4 Reference Priors
   A.4.5 Betting Game

References
Part I

Introduction
1

Introduction

You insist that there is something a machine cannot do. If you tell me precisely what it is that a machine cannot do, then I can always make a machine that does just that!

J. von Neumann, 1948

1.1 Background

1.1.1 Why Automotive Diagnosis?

To meet steadily increasing requirements on performance, safety, and decreased environmental impact, modern automotive vehicles are becoming increasingly complex products. For example, functions are developed for active safety systems, for exhaust gas after-treatment, and to optimize fuel economy. The functions typically integrate mechanics, chemical processes, hydraulics, and electric components, as well as electronic control units (ECUs) and software. The number of ECUs is steadily increasing to satisfy requirements on increased functionality. As an example, during the last fifteen years the number of ECUs in an Scania heavy truck has increased from about five fifteen years ago to about 35-40 in modern trucks of today.

The complexity and increased functionality of modern vehicles make them more challenging to monitor, diagnose, and troubleshoot. At the same time the requirements on the diagnosis system itself are increasing As depicted in
Chapter 1. Introduction

Figure 1.1: There are several interesents with requirements on the diagnosis system of an automotive vehicle.

Figure 1.1, there are several interesents having requirements on the diagnosis system in a heavy truck. At the workshop, the mechanic needs support to be able to perform fast and efficient troubleshooting and repair of the complex automotive vehicles. To fulfill demanding environmental legislation, faults that increase exhaust emissions must be detected within specified times, and safety legislation regulates faults related to safety issues. The manufacturer needs a diagnosis system that is easily configured during development of new products, and that can be used also in the early phases of product testing. A powerful diagnosis system is also an important factor for manufacturers of automotive vehicles in the competition for customers, and will continue to be so in the future. For the driver, the diagnosis system should reduce safety risks without producing any unexpected behavior of the truck, nor any annoying false alarms. For haulage contractors, increased uptime and reduced service and maintenance costs are important. This can be achieved with an accurate and efficient diagnosis system.

1.1.2 Diagnosis is a Challenge

Fault diagnosis is about finding faults that possibly are present in the system by using numerous observations and their internal relations. The internal relations can be described by different types of models of the system and the faults. However, in complex systems, such as automotive vehicles of today and tomorrow, finding the internal relations and building models is a most challenging task, since the relations often are hidden and may include uncertainty. Building accurate physical models of the automotive systems is complicated, both due to the complexity of the systems and since the systems are operated
1.1. Background

In continuously changing surroundings. In addition, in particular considering heavy trucks and buses, many vehicles are rebuilt or reconfigured after leaving the factory to satisfy customers' specific needs. Reconfigurations can for example include containers with refrigerators for food transport, changed in-take air systems for trucks operating in deserts, external systems for handling timber, or changed rear axle gear ratio. These reconfigurations lead to that the knowledge about the actual configuration of the vehicle in the on-board diagnosis system in the ECU is limited. Uncertainty is further increased by measurement noise in sensors, and by the dispersion in quality in the sensor populations.

In automotive diagnosis, collecting data to learn from is often difficult, mainly since faults are rare. One alternative is to implement faults and collect data. However, since there are many different faults, and some are difficult, or even impossible, to implement, there will most often only be a limited data available from a small subset of faults that should be diagnosed. In particular, there will typically only be data from single faults, but the diagnosis system should also handle multiple faults. Moreover, there may be faults that causes abnormal behavior but that are previously unseen.

On-board the vehicle, diagnosis is performed in ECUs, where the hardware capacity in terms of CPU power and data storage is limited. Off-board, at the workshop, the hardware capacity is less limited. On the other hand, there can be faults that are present in the vehicle but that are not excited while the vehicle is at the workshop.

1.1.3 Approaches to Diagnosis

One common and efficient approach to diagnosis is to use models of the system and apply model based diagnosis (MBD). The models can be of different types. Each of the model types have different advantages and drawbacks in the automotive application.

One important class of models used for diagnosis is physical models, as for example in [Peischl and Wotawa, 2003, Lucas, 2001, Hamscher et al., 1992, Korbicz et al., 2004, Gertler, 1998]. However, as described in the previous section, accurate modeling of automotive vehicles is difficult due to several sources of uncertainty. Other diagnosis methods are based on models learned from data, as for example the ones in [Gustafsson, 2001, Russell et al., 2000, Basseville and Nikiforov, 1993]. The collection of data for diagnosing automotive vehicles is associated with two main problems. First, to distinguish faults with data driven techniques, data from both the fault free case as well as from fault situations is needed. Data from the fault free case can typically be collected by running and observing the system. Data from faults is, on the other hand, difficult to collect from observing the system since faults are rare. Second, the diagnosis system should work when the product is newly released to the market, but at this point the amount of data is often limited. In addition, each new release
of a product needs a new set of data, even when the differences from previous releases only are small.

The uncertainties described in the section above makes it difficult, or even impossible, to determine exactly which fault that is present in the vehicle. Many diagnosis algorithms, for example those based on the General Diagnosis Engine (GDE) [de Kleer, 1992] and its extension Sherlock [de Kleer and Williams, 1992] or Reiter’s method based on first principles logic [Reiter, 1992], determines a list of all faults that can possibly explain the current behavior of the system. With uncertainty in models and measurements, this list may be very long, since faults can not be excluded with certainty. In GDE, Sherlock, and Reiter’s methods these lists are focused on more probable faults, mainly in the sense that explanations with a small number of faulty components are preferred before explanations with a larger number of faulty components. In this thesis, we handle the uncertainties by taking a probabilistic approach, and compute the probabilities for the faults and combination of faults, given all available information. In the probabilistic approach, faults that are impossible are assigned probability zero, and possible faults are ranked after their probability. In addition, having computed the probabilities for faults, we can apply a decision-theoretic approach, where the probabilities are combined with a loss function to determine the counter action to perform. The concept of combining probabilities with loss functions can be used both for on-board and off-board diagnosis, but with different loss functions.

1.2 Problem Formulation

The main objective in this thesis is to contribute to improved diagnosis of automotive vehicles. We let the work be driven by case studies of real applications, where challenges and problems are identified. Methods for solving the identified problems are developed, and applied to the real systems. Fault diagnosis is a challenging and complicated task, and although the tasks of diagnosing different systems or subsystems are similar, there are also differences, for example in the type of background knowledge available. We strive for making the diagnosis methods theoretically sound and general. The soundness of the methods makes it easier to track and understand the meaning of their output and to guarantee their performance. Moreover, development engineers can tailor the general methods to suite their particular application.

We consider both on-board and off-board diagnosis of automotive vehicles. The two tasks may seem different in nature. On-board diagnosis is performed in the automotive on-board control system during operation of the vehicle, mostly without human integration. Off-board diagnosis is performed by a mechanic supported by a troubleshooting tool. In the troubleshooting tool, diagnosis is based on the possibility of human interaction with the system. On the other
1.2. Problem Formulation

hand, both on- and off-board diagnosis regard the same vehicle, and models used in diagnosis rely on the same internal relations in, or models of, the system.

Within the probabilistic framework used in this thesis, the main objective is to compute the probability distribution for faults, or system status, given all information available:

\[ p(\text{system status}|\text{all available information}) . \tag{1.1} \]

This probability can then be combined with decision theory to determine the appropriate action; for example the best on-board control strategy, the best troubleshooting action, whether to set of an alarm or not, etc. The probability (1.1) is used both on-board and off-board. “All available information” can be divided into three main parts: expert knowledge about the system, data, and current observations. The expert knowledge and the data are the same in both on- and off-board diagnosis of the same vehicle, and therefore it is natural to consider both tasks in the thesis. However, since different kinds of observations are available for on-board diagnosis during operation and at off-board at the workshop, different subparts of expert knowledge and data may play different roles. This also means that information stored from the on-board diagnosis may contribute to improved off-board diagnosis, and vice versa.

The computation of the probability (1.1) is central in this thesis. We consider different kinds of information, or knowledge, and use different computation approaches. In particular, we focus on the following questions:

- How do standard methods for learning from data perform in the computation of (1.1)?
- Which are the main issues regarding the training data available for diagnosis?
- In the computation of (1.1), the different pieces of information are to be combined. The different information pieces can be of widely different types, and include for example dynamical physical models, state machines, fault models, structural knowledge about fault effects, experimental and observational data, function specifications. How should these different kinds of information be integrated in the computations?
- To compute (1.1), dependency relations between different subparts of the diagnosed system are used. In probabilistic terms, the dependency relations represent information flow. However, the physical relation that caused the dependency may not be present at the time the relation is used. For example, at the workshop it can be observed that oil has leaked out during operation of the system, although there is no oil leaking out when the system is at rest. In particular, during off-board diagnosis, what are the effects of these different kinds of dependencies?
During off-board diagnosis and troubleshooting at workshops, “all information available” includes knowledge about that parts of the system have been repaired. The repairs are external interventions that change the dependency structure of the system. Therefore, one important question is: how should external interventions be handled in the computation of (1.1)?

In on-board diagnosis, hardware capacities are limited, and in off-board diagnosis fast computations are crucial to reduce troubleshooting and repair time. Therefore, one important question is thus: how to compute the probability (1.1) as efficiently as possible?
Contributions

We balance probabilities and choose the most likely. It is the scientific use of the imagination.

Sherlock Holmes, in “The Hound of the Baskervilles”, 1902

2.1 Thesis Overview

Besides this introductory part, Part I, this thesis consists of three parts: an introduction and brief survey of probabilistic methods for diagnosis, five appended papers, and conclusions. An overview of the three parts, and relations between the papers and chapters is shown in Figure 2.1.

Part II is an introductory survey of probability, diagnosis, and in particular probabilistic methods for diagnosis. It constitutes, together with the current Part I, the bottom layers in Figure 2.1. In Chapter 3, a brief introduction to Bayesian probability is given. Rather than a being a reference on probability theory presenting computation rules, it is intended as a discussion of interpretations of probability. In particular the interpretation used in this thesis is presented. In Chapter 4 a brief survey of previous works on model-based diagnosis, and in particular probabilistic diagnosis.

Part III is the main part of this thesis, and consists of the five appended papers. In all five papers there are both application-related and theoretical contributions. The theoretical contributions are in the fields of learning, modeling and inference. As depicted in Figure 2.1, Papers 1, 2, and 5 contribute to the
Chapter 2. Contributions

Figure 2.1: Overview of the thesis.
theory of learning, while Papers 3 and 4 consider learning. All five papers have theoretical contributions in the field of inference. In the application-related view, Papers 1 and 2 have a clear focus on on-board diagnosis, while Papers 3 and 4 focus on off-board troubleshooting. In Paper 5 theoretical methods are handled that are applicable to both on- and off-board diagnosis.

In Part IV, a conclusion of the work and the results in the thesis are presented. Moreover, an outlook is provided, discussing future challenges and applications of probabilistic diagnosis in automotive systems.

2.2 Appended Papers – Summary and Contributions

In this section we give an overview of the appended papers, together with a brief summary of each of the papers. For each paper, we also present the contributions, both the theoretical contributions related to development of new methods, and the application-related contributions related to diagnosis of real automotive vehicles.

2.2.1 Paper 1 - Data and Process Knowledge


Paper 1 is based on the publication:


Summary

The objective is to develop a diagnosis method that computes probabilities of faults, and that is applicable to real automotive systems. A careful application study is performed, and requirements on the diagnosis system are listed.

The diagnosis method should compute the probabilities for faults, using all available information. The case study has shown that the available information comprise several types of information: training data; different kinds of monitoring functions, such as diagnostic tests or residuals; and sensor readings. The training data available is typically limited in amount. Furthermore, the training data is often experimental, i.e. collected after first actively implementing faults, instead of simply observing the system and wait for faults to appear. For many automotive systems there are physical models available of the system, but they are typically not detailed enough to rely on alone in fault diagnosis. Finally,
the computational burden should be kept small to meet hardware capacity limitations of on-board ECU processors.

A method for computing the probabilities of faults given both the physical models and the (limited amount of) training data is developed. The method is a combination of two previous types of methods – consistency based methods using the Fault Signature Matrix (FSM) such as Sherlock [de Kleer and Williams, 1992] and structured hypothesis testing [Nyberg, 2000], and standard probabilistic methods using training data only, see for example [Heckerman et al., 1995]. In an application to the task of diagnosing the gas flow of a heavy truck diesel engine, the new method is illustrated on real world data.

In the paper it is also discussed how the new, combined method relates to these previous methods for diagnosis, and that the diagnosis result always is at least as good as using one of the previous methods.

**Contributions**

- The detailed investigation of the automotive diagnosis problem.
- The translation of physical characteristics of the diagnosed process to assumptions in the probability computations.
- The method for combining training data and expert knowledge in terms of an FSM in computations of probabilities of faults.
- The application of the new method to the diagnosis of a real world automotive diesel engine.
- The investigation of the new method’s relation to previous works such as Sherlock [de Kleer and Williams, 1992], structured hypotheses testing [Nyberg, 2000], model-based probabilistic methods, and Bayesian networks.

**2.2.2 Paper 2 - Data and Likelihood Constraints**


Paper 2 is based on the publication:

Summary

A new method is developed for learning the posterior probability distribution of a class variable $C$ given an observation vector $\mathbf{x} = (x_1, \ldots, x_n)$ and background information $i$ consisting of a combination of training data and expert knowledge in terms of likelihood constraints. Likelihood constraints are constraints on linear combinations on the parameters in the distributions $p(x_i|C, i)$. The likelihood constraints are very general, and can be used to express several types of expert knowledge, such as explicit knowledge about certain values of the parameters in the probability computations, or knowledge about the values of linear combinations of parameters. Also, constraints such as “variable $X_i$ has the same, but unknown, distribution given $C = c_1$ and $C = c_2$” can be expressed using likelihood constraints.

Likelihood constraints appear naturally in many different kinds of applications, such as medical and technical diagnosis and econometrics. In particular, the constraints in probability computations considered in the previous papers [Boutilier et al., 1996] and [Jaeger, 2004] are special cases of the likelihood constraints considered here.

In the paper, the derivation of the new method is shown in detail. The method leads to multidimensional integrals that do not have any closed form solutions in general. In the paper, an approximate solution method based on Laplace approximation is proposed. All the computations are illustrated in detail on two examples, of which one is a diagnosis task.

Contributions

- The method for integration of expert knowledge in terms of likelihood constraints and training data in probability computations.
- The translation of constraints in general terms into likelihood constraints.
- The application of the new method to the diagnosis problem.

2.2.3 Paper 3 - Non-Stationary Dynamic Bayesian Networks


Paper 3 is partly based on the publication:

Summary

The task of troubleshooting automotive vehicles is considered, and in particular the computation of probabilities of faults in a process that is subject to external interventions. The task is further complicated by the fact that there is a mixture of two kinds of dependencies that are used in troubleshooting: instantaneous and non-instantaneous. For example, during operation of a vehicle there may be oil leaking out from a pipe through a worn out gasket. When the system is at rest, the oil on the outside of the pipe can be used to identify the leakage, although the oil is not leaking out at rest. If the oil is cleaned up, the system must be operated again in order to verify whether the leakage is still present.

The external interventions changes the dependency structure of in the model: we say that they cause events. To to model processes with both instant and non-instant dependencies and events, the framework of event-driven non-stationary dynamic Bayesian networks (nsDBN) is developed. The framework is general, not only to troubleshooting, but to modeling of all kinds processes where there are events. It is also shown how an event-driven nsDBN is efficiently characterized by an initial Bayesian network (BN), a nominal transition BN, and three sets used to define the events.

Modeling is an artwork, and in the paper we provide guidelines for development engineers to simplify the task. We also describe the troubleshooting problem in the framework of event-driven nsDBN, and illustrate the computations on a typical subsystem of an automotive vehicle.

Contributions

- The general framework of event-driven nsDBN, that facilitates probability computations in systems that are subject to external interventions that affects the dependency structure.

- The formulation of the troubleshooting problem within this framework. This opens for solving troubleshooting problems in the automotive field, where it is important to handle general dependency structures, multiple faults, and without any simple function verification.

- The illustration of the use of event-driven nsDBN on an automotive example.

2.2.4 Paper 4 - Modeling and Inference for Troubleshooting

Paper 4 is partly based on the publications:


**Summary**

The objective in this paper is to propose a troubleshooting system that applies to real automotive applications. To do this, a case study of a mechatronic system of an automotive heavy truck, an auxiliary braking system, is performed. Three main issues are identified as important to account for in the troubleshooting system: the need for assembling/disassembling the vehicle during troubleshooting, the difficulty to verify whether the system is fault free, and the need for time efficient inference to reduce waiting time for the mechanic. The first two issues leads to that probabilities need to be computed in a system that is subject to external interventions.

A decision-theoretic approach is used to design a troubleshooting system consisting of two parts: a planner, that suggests the next troubleshooting action; and a diagnoser that supports the planner with probability computations. To compute the probabilities in the diagnoser the framework of event-driven nsDBNs presented in Paper 3 can be used. In the nsDBN probabilities for all ingoing variables can be used, but the diagnoser it is shown to be sufficient to compute conditional probabilities for observations. Therefore, we take off in the nsDBNs, and develop a new method for computing the necessary probabilities in the diagnoser. The method is based on an algorithm that through simple manipulations updates a static BN as events occur. The algorithm is carefully derived and proved in the paper. In the paper we also discuss practical issues related to modeling for troubleshooting.

**Contributions**

- The development of a troubleshooting system that is applicable to real automotive systems. In particular, assembling/disassembling of the system is possible, and no specific function verification is presumed.

- The detailed case study, and the extensive discussion of practical issues related modeling for troubleshooting.
Chapter 2. Contributions

- The new efficient inference algorithm for troubleshooting, based on an algorithm that updates a static Bayesian network as external interventions occur. In particular, it is proved that the algorithm provides the same probabilities as an nsDBN.

2.2.5 Paper 5 - Comparing Methods for Learning


Paper 5 is based on the publications:


Summary

In this paper, five approaches for learning from data are compared and evaluated on the problem of fault diagnosis and isolation. Based on the five approaches are previously presented in the literature, eight methods where derived. The compared methods are: Direct Inference [Pernestål and Nyberg, 2007], two versions of naive Bayesian networks [Jensen and Nielsen, 2007] with discrete and binary observations respectively, two versions of general Bayesian networks [Jensen and Nielsen, 2007, Silander and Myllymäki, 2006] with discrete and binary observations respectively, linear regression [Bishop, 2005], logistic regression [Roos et al., 2005], and weighted logistic regression, a version of logistic regression that is developed to handle experimental training data. The methods are tailored to suite the fault diagnosis and isolation problem, and to handle issues in fault diagnosis, such the experimental data and that there are faults from which there is not data.

To evaluate the methods, relevant performance measures are discussed. Finally the methods are compared on data from a real-world automotive diesel engine. Among the compared methods, logistic regression is shown to perform best on this

Contributions

- The application and comparison of eight different Bayesian methods for learning from data, applied to the fault diagnosis problem.

- The investigation of special characteristics of training data in diagnosis, for example that the amount of data often is limited, and that data typically is experimental.
• The tailoring of these methods to suite the fault diagnosis problem, and
  in particular the unseen fault patterns and the experimental data.

2.3 List of Publications

Here follows a list of publications that are not appended to the thesis, but that
constitute an important background work to the appended papers. They are
listed in order of publication.

• Anna Pernestål, Mattias Nyberg, and Bo Wahlberg. (2006). A Bayesian
  Approach to Fault Isolation with Application to Diesel Engine Diagnosis.
  In Proceedings of 17th International Workshop on Principles of Diagnosis
  (DX’06), Peñaranda, Spain.

• Anna Pernestål, Mattias Nyberg, and Bo Wahlberg. (2006). A Bayesian
  Approach to Fault Isolation Structure Estimation and Inference. In Pro-
  ceedings of IFAC Symposium on Fault Detection, Supervision and Safety
  of Technical Processes (SAFEPROCESS 2006), Beijing, China.

  discussion on the Assignment of Priors. In Reglermöte 2006, Stockholm,
  Sweden.

• Anna Pernestål. (2007). A Bayesian Approach to Fault Isolation with Ap-
  plication To Diesel Engine Diagnosis. Licentiate Thesis. Royal Institute
  of Technology, Stockholm, Sweden.

• Anna Pernestål and Mattias Nyberg. (2007). Using Data and Prior
  Information in Bayesian Classification. Tech. Report LiTH-ISY-R-2811.
  Linköping University, Linköping, Sweden.

• Anna Pernestål and Mattias Nyberg. (2007). Probabilistic Fault Diag-
  nosis Based on Incomplete Data. In Proceedings of the European Control
  Conference (ECC 2007), Kos, Greece.

• Anna Pernestål and Mattias Nyberg. (2007). Using Prior Information in
  Bayesian Classification - with Application to Fault Diagnosis. In 27th
  International Workshop on Bayesian Inference and Maximum Entropy
  Methods in Science and Engineering (MaxEnt 2007), Albany, USA.

• Anna Pernestål and Mattias Nyberg. (2007). Experimental and Observa-
  tional Data in Learning for Bayesian Inference. Tech. Report LiTH-ISY-
  R-2834. Linköping University, Linköping, Sweden.


References


Part II

Probability Theory in Diagnosis
Bayesian Probability Theory

Probability is nothing but common sense reduced to calculation.

Laplace, 1812

In automotive diagnosis, there are several sources of uncertainty: noise, model errors, lack of training data, etc. In this thesis we use probability theory to handle these uncertainties, and to determine faults that are possibly present in the monitored system. Rules for manipulating and updating probabilities are described for example in [Blom, 1994, Durrett, 2004, Casella and Berger, 2001]. However, one problem that remains when using probabilities to infer about the real world is to assign numbers to the probabilities. To do this, it is necessary to understand the word “probability”. In this chapter, we briefly discuss different interpretations of probability and, in particular the interpretation of probability used in this thesis. This chapter is a shorter version of Appendix A.

3.1 Dealing With Uncertainty

Human life is to a great deal a life lived under uncertainty. Every day we make decisions under uncertainty, both in professional life and in private. For example: will the stock market raise or fall today? My car does not start, which part has caused the failure? Should I bring an umbrella tonight? How much should I bet on my favorite soccer team in the next game? Should I fold in the poker game? What conclusions can be drawn from the laboratory experiment? There is no upper limit on the number of such situations.
The situations listed above are very different in their nature. Sometimes the probability calculation relies on data, as in laboratory experiments. In other cases the probability calculations are based on known facts, for example, the number of spades in a deck of card is well known and thus the probability of drawing a spade can be computed. In yet other cases, it seems like probabilities are more or less based on personal feelings, for example in sports betting.

In each situation, the human brain deals with uncertainty. It considers the available information, for example: yesterday’s stock market trend or the observation that the headlights of my car does not light. The brain weighs factors speaking fore and against an event, and makes decisions (which may be more or less clever).

In the problem considered in this thesis, diagnosis of automotive vehicles, we deal with uncertainty in a formal way. Given observations of different kinds from a system, the aim is to construct an algorithm that, just like the human brain, considers the available information and evaluates the probabilities that different faults are present. The available information can for example comprise data, different kinds of models with unknown model errors, drawings, and functionality specification documents. To be able to transform these fundamentally different types of information and construct the diagnosis algorithm that computes probabilities for faults, one might ask oneself questions as: What is this “uncertainty”? What is “probability”? What does the “probability that it will rain tonight” mean? Is it unique? Can we put a number on it?

In reference literature on probability theory, for example [Blom, 1994, Durrett, 2004, Casella and Berger, 2001, O’Hagan and Forster, 2004], formulas and tools for manipulating probabilities are presented, as in the following toy example.

**Example 3.1.1 (Was it the Sprinkler?).**

Sanna wakes up a morning and wants to know whether it has rained during the night. She knows that the prior probability for rain is \( p(\text{rain}) = 0.3 \). Moreover, she knows that, if it has rained, the lawn will be wet, i.e. that \( p(\text{wet lawn}|\text{rain}) = 1 \). She also knows that, if there is no rain, there is a sprinkler that cause the lawn to be wet with probability \( p(\text{wet lawn}|\text{no rain}) = 0.2 \).

After waking up, Sanna notices that the lawn is wet. She can then compute the probability that it has rained by using Bayes’ rule and marginalization [Blom, 1994] as follows:

\[
p(\text{rain}|\text{wet lawn}) = \frac{p(\text{wet lawn}|\text{rain})p(\text{rain})}{p(\text{wet lawn})} = \frac{1 \cdot 0.3}{1 \cdot 0.3 + 0.2 \cdot 0.7} = 0.68\ldots
\]
These computations are perfectly fine as long as the numbers, such as “the probability for rain is 0.3”, are known. In the example above, the numbers where simply stated, but how are they found? To assign numbers in the probability distributions to use in computations, it is necessary to know what “probability” means.

3.2 Interpretations of Probability

The discussion about the definition of the word “probability” has been going on for more than 200 years [Hacking, 1976]. Depending on the background of the researchers, there where several different interpretations during these years. Among the different interpretations of probability, there are two main paths [Hacking, 1976, O’Hagan and Forster, 2004, Jaynes, 2001]: the idea of probability as a frequency in an ensemble, often called the frequentist view or frequency-type, on the one hand; and the idea of probability as the degree of belief in a proposition, often referred to as the Bayesian view or belief-type, on the other hand. In frequentist view, probability is defined by the relative frequency of an event, and is a property of the object. Consider for example the statement:

*This coin is biased towards heads. The probability of getting heads is about 0.6.*

This statement expresses probability in the frequency-type meaning, and is true depending on “how the world is”. This statement can (at least hypothetically) be tested by tossing the coin (infinitely) many times. If the relative frequency for heads is 0.6, the statement is true, if the relative frequency for heads is something else, the statement is false. In the Bayesian view, probability is the degree of belief, given some evidence. Consider now this sentence about the same coin:

*Taking all the evidence into consideration, the probability of getting a head in the next roll is about 0.6.*

This statement is true depending on how well evidence supports the particular probability assignment. The probability is subjective in the sense that it depends on the evidence. This statement can be true, depending on the evidence, even if the relative frequency turns out to be something else than 0.6.

These two views, the frequency-type and the belief-type, are different in a philosophical sense, and a natural question is why the same word, “probability”, is used for both of them. Hacking [Hacking, 1976] gives one explanation: in
daily life, we (humans) switch back and forth between the two perspectives. Consider the following example.

**Example 3.2.2 (Switching Between Frequency and Belief).**

A truck of model R arrives to a mechanic at a workshop. The mechanic knows that among all model R trucks, one out of ten of the trucks that arrives to the workshop has fault $F$ present. The mechanic concludes that choosing a random model R truck of those that has been (or are) at the workshop, the probability that fault $F$ is found is 0.1. This probability is of frequency-type.

Consider now the particular truck that just arrived to the workshop. What is the probability that this truck is has fault $F$? The truck is either faulty or fault free, so there is no randomness, but still the mechanic would (probably) say that the probability is 0.1. He reasons as follows. Out of all model R trucks that has visited the workshop, fault $F$ was present in 1 out of 10. This truck is a model R and has arrived to the workshop. Taking those three pieces of information into account, the probability that this particular truck has fault $F$ is 0.1.

The two interpretations of probability, as well as methods for assigning probabilities is further discussed in Appendix A. Instead, we now concentrate on the interpretation of probability used in this thesis.

### 3.3 The Interpretation of Probability Used in the Thesis

In this thesis, as in Example 3.2.2, we consider a specific vehicle. The vehicle is either fault-free or faulty, but since we, in general, not have enough information about the vehicle to determine its fault status, we use probabilities.

Although not being dogmatic, we will in this thesis mainly take a Bayesian, or belief-type, view on probability. We let the probability be determined by the evidence, or background information, given. To denote this, if $i$ denotes all information given, we write the probability for an event $A$ as $p(A|i)$. We let the probability be defined by is given behind the $|$-sign, i.e. by the evidence of background information. In this interpretation, the probability is subjective in the sense that different evidence give different probabilities. On the other hand, the probability is objective in the sense that we assume that it is uniquely determined about what is given behind the $|$-sign. This implies that we, to be formal, require enough information behind the $|$-sign to uniquely determine the probability. For example, if $D$ denotes the number of eyes coming up when rolling a dice, the probability for getting six eyes in a certain trial is written

\[ p(D = 6|S) = \frac{1}{6}. \]
where

\[ S = \text{The dice is unbiased. The dice has six sides. We apply principle of indifference, that says that if there are } n \text{ possible events and there is no reason for favoring any of the events over the others, each event should be assigned probability } 1/n. \]

The example above show a quite lengthy and intricate way of writing something that is implicitly understood. Furthermore, in many situations, it is uninteresting and/or extremely complicated to explicitly state every piece of information that is behind the \( \mid \) -sign. Therefore, we often simply denote this knowledge “background knowledge” (background information) and write \( i \). When the background knowledge is clear from the context we sometimes omit \( i \) as well.

We have, in this thesis, adopted the Bayesian interpretation of probability since it is appealing and natural for the reasoning in the problems related to diagnosis that we are faced to, or, as O’Hagan [O’Hagan and Forster, 2004] expresses it: “the Bayesian interpretation is fundamentally sound, very flexible, produce clear and direct inferences, and make use of all information”\(^1\).

However, we are not dogmatic, and there are cases where the frequentist view is similar or equal. Technically, the rules of probabilities and the computations are the same in both interpretations of probabilities [Hacking, 1976]. This means that the methods presented in this thesis are valid and make sense regardless of the probability interpretation of the user.

\(^1\)In contrast to classical methods that have “philosophical flaws”, limited range, indirect interpretation of the inference, and not utilize prior information [O’Hagan and Forster, 2004].
4

A Brief Survey of Probability Based Diagnosis

4.1 Model-Based Diagnosis

4.1.1 Diagnosis Methods

During the last two decades, fault diagnosis of technical systems has become a steadily increasing field of research. One important reason is the introduction of more complex and capable computers and electronic control units (ECU), that mitigates improved system functionality that make the systems more difficult to diagnose. At the same time, the better ECUs provide a platform for improved diagnosis algorithms.

There is a huge number of different methods for doing diagnosis. In its most general form, diagnosis is to, based on knowledge about the system, study observations from the system and then draw conclusions about the state of the system. Different diagnosis methods are based on different “knowledge about the system” and consider observations in different ways.

In model-based diagnosis (MBD), models of the system under diagnosis are used to describe the relations between observations and faults, see Figure 4.1. The model typically describes how possible faults affect the observations. During diagnosis, these relations are inverted and the observations are used to draw conclusions about which faults that are present. There is a wide variety in model-types that can be used in diagnosis, and in Figure 4.2 an overview is given. This is by no means the only way of characterizing model-based diagnosis methods, and it is not complete, but gives an idea of some model-types that
Figure 4.1: A diagnosis model describing how faults affect observations of the system. During diagnosis, observations are made and the inverted relations are used to make inference about which faults that may be present.

appears in the literature. In the remainder of this section we present four types of models and a selection of works based on each of the model type.

### 4.1.2 Logical Models

Among the first modern methods for MBD we find Reiter’s method based on first order logic [Reiter, 1992]. The system under diagnosis is described by using logical statements. In Reiter’s method, the diagnoses are assignments of component states to all components in the system that are consistent with the observations made of the system. During the same time period as Reiter’s method was developed, the General Diagnostic Engine (GDE) [de Kleer, 1992] and its descendant Sherlock [de Kleer and Williams, 1992] based on similar ideas were presented.

### 4.1.3 Black Box Models

Black box models, or *data driven models*, are learned from training data, and can for example be various classification methods [Duda et al., 2001, Devroye et al., 1996, Bishop, 2005, Russell et al., 2000, Chiang et al., 2001, Sorsa et al., 1991], among which we find for example Support Vector Machines (SVM) [Lee et al., 2007, Ge et al., 2004, Saunders et al., 2000], methods for Case Based Reasoning (CBR) [Bregon et al., 2007], and Bayesian networks learned from data [Verron et al., 2007, Pernestål et al., 2008]. Since data driven models are learned from data, they require no explicit knowledge about the process
under diagnosis. The main drawback with the data driven models in diagnosis is that they, in their general form, require data from all fault cases that are to be diagnosed – a situation that is rarely fulfilled in fault diagnosis applications since faults are rare.

4.1.4 Physical Models

Examples of physical model types are for example state space models and Differential Algebraic Equations (DAE). Physical models are used in diagnosis in several ways [Blanke et al., 2003, Patton et al., 2000, Isermann and Balle, 2007, Isermann, 2006, Cordier et al., 2004, Staroswiecki and Comtet-Varga, 2001]. Among the diagnosis methods based on physical models we find for example parity space [Basseville and Nikiforov, 1993, Gertler, 1998, Zhang et al., 2006], structural analysis [Krysander, 2006], structural hypothesis testing [Nyberg, 2000], Bayesian network methods learned from physical principles [Roychoudhury et al., 2006, Schwall, 2005], and qualitative models [Daigle et al., 2007, Mosterman and Biswas, 1999]. In diagnosis using physical models, data is sometimes needed to tune the model, but the diagnosis result depend to a larger extent on the accuracy of the model than on the data. For automotive systems, the operation conditions and surroundings are continuously changing and it is typically difficult to build a model that is sufficiently accurate in all
operating conditions.

4.1.5 Discrete Event Systems

One large branch of diagnosis concerns diagnosis of Discrete Event Systems (DES), see for example the workshop series DCDS [Dotoli and Larizza, 2009]. When considering DES, the system is modeled by a set of states and transitions between these states [Kurien and Nayak, 2000]. Some states represent that the system is faulty. Diagnosis then becomes the task of tracking the sequence of states that system has been in, given observations from the system. Two commonly used model approaches are Petri nets [Murata, 1989, Aghasaryan et al., 1998] and state automata [Lunze and Supavatanakul, 2002, Supavatanakul et al., 2006]..

4.2 Probabilistic Methods for Diagnosis

4.2.1 An Example: the Car Start Problem

In this thesis, we apply probabilistic methods for diagnosis. Basically, this means that we compute probabilities for faults. The idea of using probabilistic methods for diagnosis is not new. In fact, diagnosis is one of the most common applications in introductory courses on probability theory. One example is the “Car start problem” [Jensen and Nielsen, 2007], where the task is to determine why a car does not start. A simple version of the car start problem is shown in Figure 4.3, where variables are shown as circles and dependencies between the variables are given by directed edges, point in the direction of causal influence. For example, the amount of fuel (Fuel?) and whether the starter rolls (Starter Roll?) have causal impact on the whether the car starts (Car Start?), and the state Fuel?. So, if probabilistic diagnosis problems can be solved in the basic course on probability, what is the problem? In the example above, the model, i.e. the causal dependencies between variables, is assumed to be known. This is typically not the case in real applications. Furthermore, dependencies need to be quantified. Finally, we need methods for inference, for example, to determine the probability that the fuel tank is empty, given that the car does not start and that the battery is fully charged. These three tasks are often challenging. In next section, we give a more precise formulation of the challenges in probabilistic diagnosis.

4.2.2 What is Probabilistic Diagnosis?

As stated in Chapter 1, the aim is to compute the probability distribution

\[ p(\text{system status} | \text{all available information}) \],

(4.1)
Figure 4.3: A basic example of diagnosis: the car start problem. The probability that the car starts is dependent on the fuel tank level ($Fuel$?), the battery status ($Battery$), and the status of the starter motor ($Starter\ Motor$).

where “all available information” may include current observations, training data, and other kinds of knowledge about the system. This distribution (4.1) is sometimes referred to as the belief state.

The knowledge about the system is often represented with some kind of model, for example one of those described in Section 4.1. Regardless of which kind of model that is used, it is often impossible determine exactly which faults that are present in the system. Reasons may for example be that the number of observation points is limited, that there are noise and model errors present, or the unknown and changing operating environment. These factors cause us to reason under uncertainty.

We divided the probabilistic diagnosis problem into two subproblems:

1. Learning. To construct, or learn, an adequate model of the system under diagnosis, including dependency structures and strength of dependencies.

2. Inference. To use the model to make inference, and compute the probability distribution for the system state, or for faults.

Depending on the model type used, these two steps will be more or less difficult.

**4.3 Methods For Probabilistic Diagnosis**

There are numerous methods for probabilistic diagnosis in the literature, based on different kinds of models. In this section, we review methods based on
three model-types that are most closely related to the methods presented in the appended papers in Part III. We discuss the kind of dependency relations and uncertainties that are modeled within each model type. We also consider the complexity of the two steps Learning and Inference, and summarize advantages and drawbacks.

4.3.1 Dynamic Physical Models

**Model Type.** Dynamic physical systems, such as combustion engines, automotive robots, chemical plants and many others, are often described by a state space model or by differential algebraic equations (DAEs) [Wahlström, 2009, Verma et al., 2004, Patton et al., 2000]. In a probabilistic setting, a discrete time state space model can for example written as

\[
\begin{align*}
    z_t &\sim p(z_t|z_{0:t-1}, x_{0:t-1}, y_{1:t-1}, u_{1:t-1}) \\
    x_t &= f(z_t, x_{t-1}, w_t, u_t), \quad w_t \sim p(w_t) \\
    y_t &= g(z_t, x_t, v_t, u_t), \quad v_t \sim p(v_t)
\end{align*}
\]

where \(y_t\) are sensor readings, \(u_t\) known control signals, \(x_t\) continuous internal states, \(z_t\) discrete internal states, \(w_t\) and \(v_t\) are process and measurement noise respectively. In this model, \(y_t\) and \(u_t\) comprise the observations, and the faults states is a (subset) of \(z_t\). All variables may be scalar or vector-valued.

**Learning.** Learning a dynamic physical model consists in determining the functions \(f\) and \(g\), and distribution of the internal state \(z_t\), and the distributions \(p(w_t)\) and \(p(v_t)\) of the noise \(w_t\) and \(v_t\). The functions \(f\) and \(g\) are often equations representing the physical behavior of the system, and known by domain experts. The distribution \(p(z_t|z_{0:t-1}, x_{0:t-1}, y_{1:t-1}, u_{1:t-1})\) describes transitions between discrete states in the system. The discrete variable \(z_t\) represents faults, and the probability for transitions is often assumed to be known. The distributions \(p(w_t)\) and \(p(v_t)\) are generally assumed to be known, and often considered to be Gaussian.

**Inference.** With this type of model, the belief state (4.1) that we search is the probability \(p(z_t|y_{1:t-1}, u_{1:t-1})\). If the functions \(f\) and \(g\) are linear (or linearized), and \(v_t\) and \(w_t\) are (assumed to be) Gaussian the Kalman Filter can be used to determine the belief state, see for example [Gustafsson, 2001].

A more general approach, that applies to non-linear \(f\) and \(g\), and non-Gaussian \(v_t\) and \(w_t\), is the Particle Filter [Doucet et al., 2001], where the relevant distributions are approximated using a swarm of “particles”, or realizations of \(p(x_t|x_{0:t-1}, y_{1:t-1}, z_{0:t-1})\) and \(p(y_t|x_{0:t}, y_{0:t-1}, z_{t-1})\). There are several

**Advantages.** If the state space description is known and can be linearized, the Kalman Filter method is straightforward and computational efficient. State space models often exists for control, and these models can be reuse for diagnosis.

**Drawbacks.** The Kalman and Particle filters can often be used straightforwardly to detect abnormal behavior of the system. However, to isolate the particular fault that is present is often more challenging. Methods for diagnosis typically require multiple copies of the model and a bank of filters. This increases the computational burden. Furthermore, to isolate faults, models describing the effects of the faults on the process are needed.

### 4.3.2 Data-Driven Black Box Models

**Model Type.** Black Box models are learned from training data. The structure of the model does not aim to represent any physical relations between inputs and outputs. Examples of model types are given in Section 4.1.3. Sometimes, learning black box models is referred to as machine learning.

**Learning.** If no explicit information is known about the system under diagnosis, but there is a lot of *training data*, i.e. tuples of observations and corresponding fault statuses, from the system under diagnosis, there are methods for learning the black-box models presented in literature [Duda et al., 2001, Devroye et al., 1996, Bishop, 2005, Russell et al., 2000]. The methods are generally based on optimization of a performance measure by tuning parameters in the models. For a Bayesian approach, data can be used to learn a Bayesian network (BN) [Silander and Myllymäki, 2006], where the nodes in the BN represent observations and faults.

**Inference.** Depending on the type of black box model used, inference may be simple or complicated. However, in most of the methods, the learning part is the most time consuming, and designed to provide straight-forward inference. This is particularly true for regression methods and neural networks.

**Advantages.** No explicit knowledge about the process is needed.
Drawbacks. The main drawback with the data-driven black box probabilistic methods is that a large amount of data from all faults considered is needed. Often it is difficult to obtain data from the faulty cases, since faults are rare. The black box methods may also be difficult to interpret, and therefore they may also be difficult to verify. Even if there is knowledge of the process available, the existing methods for learning data driven probabilistic models can typically not integrate this information with the data.

4.3.3 Bayesian Networks

Model Type. A Bayesian network (BN) is a representation of a factorization of a joint distribution of a set of variables $X_1, \ldots, X_n$. A BN is a directed, acyclic graph, where nodes represent variables and edges between nodes represent dependency relations. To each node there is a conditional probability distribution (CPD) for the corresponding variable given its parents associated. Introductions to Bayesian networks are given for example in [Jensen and Nielsen, 2007] and [Russell and Norvig, 2003].

Learning. In literature several ways of learning BNs for diagnosis are presented. The most common are: to learn from data, see Section 4.3.2; to use BNs set up by experts as in [Lerner et al., 2000, Schwall, 2005]; or to systematically derive the BNs from sets of physical equations by using a bond graph [Roychoudhury et al., 2006]. Also, for a given dependency, the CPDs can be learned from data.

The structures of the BNs used for diagnosis in the literature are different. Some of the most common are: two-layer BNs, where the nodes are either observations (in terms of sensor signals, residuals or diagnostic tests) or components as in [Schwall, 2005, Verron et al., 2009], multilayer BNs including internal variables and capturing the structure of the system as in [Schwall and Gerdes, 2002], and dynamic Bayesian networks (DBN) capturing the dynamics of systems [Murphy, 2002, Roychoudhury et al., 2006].

Inference. When the BN is known, standard methods can be used for inference. The most common are variable elimination and join tree. For large BNs with many nodes and many dependencies the inference methods may become computationally intractable and approximation methods must be applied [Jensen and Nielsen, 2007]. Methods for learning DBNs are presented in [Murphy, 2002].

Advantages. BNs representing physical structures are usually easy to interpret and validate. Also, they can be easily updated with local changes if the system under diagnosis is changed [Russell and Norvig, 2003].
**Drawbacks.** In some systems there may be several unknown and hidden effects. These may be difficult to learn and model in the BN, but may be important for the diagnosis result [Pernestål et al., 2006]. Furthermore, even if dependency structures of BNs for diagnosis by experts, learning the numbers in the CPDs is often more difficult since standard methods for parameter learning, as for example in [Heckerman et al., 1995] require data from all faults to be detected.
Chapter 4. A Brief Survey of Probability Based Diagnosis
References


Part IV

Concluding Remarks
Concluding Remarks

It is a capital mistake to theorize before you have all the evidence. It biases the judgment.

*Sherlock Holmes, 1888. In “A Study in Scarlet”*

## 1 Conclusions

The main objective with this thesis has been to contribute to improved diagnosis of automotive vehicles. The work been driven by case studies of real applications, such as automotive engines and breaking systems. We have studied both on-board diagnosis, in Paper 1, 2, and 5, and off-board diagnosis for troubleshooting, in Paper 3 and 4. In the case studies, challenges and problems have been identified. In both on- and off-board diagnosis, the limited amount of training data and the uncertainties in models of the system are two of the most important challenges. To face these challenges we have chosen a probabilistic approach, and compute the *probabilities* that faults are present in the system under diagnosis.

Considering on-board diagnosis, two of the most important issues are to handle the experimental training data, need for integration of different kinds of knowledge of the diagnosed system, and the hardware capacity limitations. In Paper 1 a method for combining expert knowledge in terms of a Fault Signature Matrix (FSM) with experimental training data has been developed, and in Paper 2 a method that combine likelihood constraints with data has been
developed. Both these methods are generic, and applicable to several different fields of applications.

In Paper 5 five approaches resulting in eight methods for learning fault diagnosis and isolation has been compared. The comparison is made with on-board diagnosis in mind, but they are applicable also to off-board diagnosis. In the survey, the methods based on logistic regression have proved to have the best performance, in particular in relation to the small number of parameters needed.

In off-board diagnosis for troubleshooting, we have identified three main issues: in models for troubleshooting there are both instant and non-instant edges, the need for computing probabilities of variables in a system that is subject to interventions, and the need for time efficient probability computations. This has led to the development of the framework of event-driven non-stationary dynamic Bayesian networks (nsDBN) in Paper 3, and its further development in Paper 4 to the algorithm updateBN that is optimized for probability computations in troubleshooting. The framework of event-driven nsDBNs is a general framework for modeling processes with external interventions, and is applicable not only to troubleshooting.

In Chapter 1 we formulated the problem to be solved in the thesis as five questions. We are now ready to answer these.

- How do standard methods for learning from data perform in the computation of (1.1)? For eight methods from five different approaches, including different types of Bayesian networks and regression, this question is answered in Paper 5. Of course, the results are dependent on the particular diagnosis situation, but one conclusion can be drawn: it is important that the method handles experimental data. Furthermore, methods with a smaller number of parameters perform better than those with more parameters.

- Which are the main issues regarding the training data available for diagnosis? Training data is used in Paper 1, 2, and 5, and in these papers we have identified two main challenges: (a) the lack of data from faults and fault combinations that are to be diagnosed, and (b) the fact that data is experimental. The handle (a), methods for combining data and knowledge are crucial. In particular, in Paper 1 and 2 experiments have shown that combining both data and expert knowledge improves the inference, compared to using data alone. The fact (b), that data is experimental, means that no information about the prior distribution of faults (before observations are made) can be learned from the data. In all Paper 1, 2, and 5 the experimental training data is handled in different ways, depending on the over-all strategy in each of the papers. However, all three methods allows for integration of prior probabilities.
2. Future Research

- **How should these different kinds of information be integrated in the computations?** In Paper 1 and 2, it has been shown that two of the most common types of expert knowledge in diagnosis, namely in form of an FSM and in form of relations between conditional distributions of single observations, appears as different kinds of constraints in the computations. In particular, the relations between single observations are translated to likelihood constraints, and it is shown that the likelihood constraints can be used to represent a broad class of information.

- **What are the effects of the different kinds of dependencies?** In probability computations for troubleshooting with interventions this has been shown to be a very important question to get the probabilities right. In Paper 3 and 4, the concepts of instant and non-instant dependencies and persistent and non-persistent variables are introduced to handle this task.

- **How should external interventions be handled in the computation of (1.1)?** In Paper 3 and 4 the event-driven non-stationary nsDBNs and the algorithm updateBN been developed as an answer to this question.

- **How to compute the probability (1.1) as efficiently as possible?** This question is very difficult, or even impossible, to answer, since it depends on the available information and the requirements on the accuracy of the computed result. However, in all five papers in the thesis it has been one main issue to limit computation time and, in particular when considering on-board diagnosis, to optimize storage requirements.

2 Future Research

In this thesis, steps have been taken towards the use of probabilistic methods for diagnosis in automotive applications. Although answering several questions, including the five listed in Section 1, many new have appeared during the work. In this section we make an outlook on future work and research using a broad and holistic view. Detailed suggestions on future work are presented in each of the five appended papers.

**Other Background Knowledge.** In the thesis, we have considered background knowledge in terms of a Fault Signature Matrix in Paper 1, and in terms of likelihood constraints in Paper 2. These two types of background knowledge are general and can describe many types of expert knowledge. It is shown in the papers that the same kind of background knowledge appears in many different areas of applications. A natural next step is to investigate which other kinds of background knowledge that exist, and how they can be combined with data in probability computations. Furthermore, to increase the possibility...
of diagnosing and isolating faults, it would be interesting to combine different kinds of background knowledge with each other.

**Finding Dependencies and Numbers.** In probabilistic models, both the structure of dependencies between variables and the underlying conditional probability distributions need to be determined. Data could be used to learn the models, but as discussed in the thesis, the amount of data is often limited. In particular this is true when systems are under development or freshly released to the market. In addition, engineers that develop an automotive system possess a large amount of knowledge and intuition about it. To use their knowledge in the diagnosis, it must be translated to a form that can be used in the probability computations. To get the most out of the probability computations, future research concerning the translation of experts’ knowledge to probability distributions is interesting. This is particularly important in modeling for troubleshooting as in Paper 3 and 4.

**Fault Tolerant Control.** In Paper 4 we have discussed troubleshooting from a decision-theoretic view-point, and combined probabilities for faults with loss functions to compute the best action for a workshop mechanic to perform. Similarly, for on-board diagnosis, it would be interesting future work to combine probability computations with loss-functions. Interesting future work would be to apply this approach also in Fault Tolerant Control (FTC), where the objective is to control the control-systems in the vehicle to avoid damaging consequences of faults.

**Performance Measures.** In order to compare and evaluate diagnosis methods, performance measures are necessary. In the literature there are several performance measures, such as percentage of correct classification, log-loss-scoring function, or mean-square error, see for example [Devroye et al., 1996, Gustafsson, 2001]. However, these are general performance measures, and not developed for diagnosis. Is it the case that a fault diagnosis system with good score in these performance measures performs well in diagnosis? Furthermore, what is a desired behavior of a fault diagnosis algorithm? The answer depends, of course, on how the output from the diagnosis system is supposed to be used. Indeed, the probability for a fault itself is rather uninteresting, as long as no reaction on the fault is suggested or performed. Therefore, one attractive alternative is to combine the fault diagnosis algorithm with a loss function and compute the expected loss, or the risk. For example, in troubleshooting the Expected Cost of Repair, defined in Paper 4 could be a suitable performance measure. Future work in this area includes for example finding proper loss functions and evaluating diagnosis systems to understand which properties that gives high scores.
Scaling. In all five appended papers, as in many research areas, the problems related to scaling of the methods to larger problems are identified as important and interesting future work. A related question is whether the method can be applied to subsystems and the results from each subsystem combined, instead of scaling the methods to larger systems?

Other Data Driven Methods and Training Data. In Papers 1, 2, and 5 we have focused on learning from data, and focused on probabilistic methods in general and Bayesian methods in particular. In the literature, there are also other methods for retrieving knowledge from data, such as Support Vector Machines, Neural Networks, and Nearest Neighbor-methods, see for example [Duda et al., 2001, Bishop, 2005]. Work has been performed on applying such methods to the diagnosis task, see for example [Russell et al., 2000, Verron et al., 2007, Lee et al., 2007]. However, these methods are based on data only, and expert knowledge of the kinds used in Papers 1 and 2 are not used. Interesting future work include applying these methods to fault diagnosis, and investigate how expert knowledge can be integrated in these methods.

References


Chapter 5. Concluding Remarks
Interpretations of Probability

Life’s most important questions are, for the most part, nothing but probability problems

Laplace, 1814

Computations with probabilities follow well defined rules, such as the Sum Rule, the Product Rule, and Bayes’ Rule [Blom, 1994, Durrett, 2004, Casella and Berger, 2001]. However, to use these tools for computing probabilities, it is necessary to find the numbers on conditional probabilities and prior probabilities. To determine these numbers, it is necessary to know what the “probability” really is.

A.1 Dealing With Uncertainty

Human life is to a great deal a life lived under uncertainty. Every day we make decisions under uncertainty, both in professional life and in private. For example: will the stock market raise or fall today? My car does not start, which part has caused the failure? Should I bring an umbrella tonight? How much should I bet on my favorite soccer team in the next game? Should I fold in the poker game? What conclusions can be drawn from the laboratory experiment? There is no upper limit on the number of such situations.

The situations listed above are very different in their nature. Sometimes the probability calculation relies on data, as in laboratory experiments. In other cases the probability calculations are based on known facts, for example, the
number of spades in a deck of card is well known and thus the probability of
drawing a spade can be computed. In yet other cases, it seems like probabilities
are more or less based on personal feelings, for example in sports betting.

In each situation, the human brain deals with uncertainty. It considers
the available information, for example: yesterday’s stock market trend or the
observation that the headlights of my car does not light. The brain weighs
factors speaking fore and against an event, and makes decisions (which may be
more or less clever).

In the problem considered in this thesis, diagnosis of automotive vehicles,
we deal with uncertainty in a formal way. Given observations of different kinds
from a system, the aim is to construct an algorithm that, just like the human
brain, considers the available information and evaluates the probabilities that
different faults are present. The available information can for example comprise
data, different kinds of models with unknown model errors, drawings, and func-
tionality specification documents. To be able to transform these fundamentally
different types of information and construct the diagnosis algorithm that com-
putes probabilities for faults, one might ask oneself questions as: What is this
“uncertainty”? What is “probability”? What does the “probability that it will
rain tonight” mean? Is it unique? Can we put a number on it? In reference liter-
ature on probability theory, for example [Blom, 1994, Durrett, 2004, Casella and
Berger, 2001, O’Hagan and Forster, 2004], formulas and tools for manipulating
probabilities are presented, as in the following toy example.

Example A.1.1 (Was it the Sprinkler?).
Sanna wakes up a morning and wants to know whether it has rained during the
night. She knows that the prior probability for rain is $p(\text{rain}) = 0.3$. Moreover,
she knows that, if it has rained, the lawn will be wet, i.e. that $p(\text{wet lawn}|\text{rain}) = 1$. She also knows that, if there is no rain, there is a sprinkler that cause the
lawn to be wet with probability $p(\text{wet lawn}|\text{no rain}) = 0.2$.

After waking up, Sanna notices that the lawn is wet. She can then compute the
probability that it has rained by using Bayes’ rule and marginalization [Blom,
1994] as follows:

$$p(\text{rain}|\text{wet lawn}) = \frac{p(\text{wet lawn}|\text{rain})p(\text{rain})}{p(\text{wet lawn})} =$$

$$= \frac{p(\text{wet lawn}|\text{rain})p(\text{rain})}{p(\text{wet lawn}|\text{rain})p(\text{rain}) + p(\text{wet lawn}|\text{no rain})p(\text{no rain})} =$$

$$= \frac{1 \cdot 0.3}{1 \cdot 0.3 + 0.2 \cdot 0.7} = 0.6818\ldots$$

These computations are perfectly fine as long as the numbers, such as “the
probability for rain is 0.3”, are known. In the example above, the numbers where
simply stated, but how are they found? To assign numbers in the probability
distributions to use in computations, it is necessary to know what “probability” means.

### A.2 Interpretations of Probability

The discussion about the definition of the word “probability” has been going on for more than 200 years [Hacking, 1976]. Depending on the background of the researchers, there were several different interpretations during these years. The first rigorous description of probability is often considered as the one given by Pierre-Simon Laplace [Laplace, 1951] in 1814:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

Since Laplace’s definition the reactions and discussion about the meaning of the word “probability” has been numerous. No consistent definition of the word exists, instead interpretations are considered. The clash of opinions was commented by Savage [Savage, 1954] in 1954:

> As to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel.

#### A.2.1 Bayesians and Frequentists

The discussion about the definition of the word “probability” has been going on for more than 200 years [Hacking, 1976]. Depending on the background of the researchers, there where several different interpretations during these years. Among the different interpretations of probability, there are two main paths [Hacking, 1976, O’Hagan and Forster, 2004, Jaynes, 2001]: the idea of probability as a frequency in an ensemble, often called the frequentist view or frequency-type, on the one hand; and the idea of probability as the degree of belief in a proposition, often referred to as the Bayesian view or belief-type, on the other hand. There are several labels on the two interpretations of probability, such as subjective/objective, epistemic/aleatory, belief-type/frequency-type, Number 1/Number 2 [Hacking, 1976].
In frequentist view, probability is defined by the relative frequency of an event, and is a property of the object. Consider for example the statement:

This coin is biased towards heads. The probability of getting heads is about 0.6.

This statement expresses probability in the frequency-type meaning, and is true depending on “how the world is”. This statement can (at least hypothetically) be tested by tossing the coin (infinitely) many times. If the relative frequency for heads is 0.6, the statement is true, if the relative frequency for heads is something else, the statement is false. In the Bayesian view, probability is the degree of belief, given some evidence. Consider now this sentence about the same coin:

Taking all the evidence into consideration, the probability of getting a head in the next roll is about 0.6.

This statement is true depending on how well evidence supports the particular probability assignment. The probability is subjective in the sense that it depends on the evidence. This statement can be true, depending on the evidence, even if the relative frequency turns out to be something else than 0.6.

For a dogmatic frequentist, probabilities exists only when dealing with experiments that are random and well-defined. The probability of random event is defined as the relative frequency of occurrence of the outcome of the experiment, when repeating the experiment infinitely many times [Hacking, 1976]. Famous frequentists are Jerzy Neyman, Egon Pearson, and Ronald Aylmer Fisher.

In the frequentist interpretation, the probability of an event is a property of the event, and it is well defined only for events that can be repeated infinitely many times. Thus, questions such as “what is the probability for rain tomorrow?” are not defined, because there is only one today and one tomorrow, and it is impossible to construct repeated experiments to investigate the relative frequency of rainy days the day after today\(^1\). However, asking the weather office the answer would be something like “It is mid december, and it was rain yesterday. During the last thirty years there has been rain 50% of the days, and of those days, there has been rain the following day for about 50%”. Thus, in a frequentistic view, the probability of rain tomorrow is the probability of rain a “general day in mid December, where it has been rain the day before”, rather than tomorrow. This is a different interpretation from the Bayesian view.

In the Bayesian view, probabilities can be assigned to any statement, regardless of whether there is any random process involved. The probability of an

---

\(^1\)In his book [Jaynes, 2001], Jaynes takes this argument even further and claims that there are (almost) no experiments that can be controlled so perfectly that it is guaranteed that they are repetitions of the same event.
A.2. Interpretations of Probability

event represents an individual's degree of belief in that event, given all information that the individual has at hand. In the Bayesian view, the probability is a property of the spectator and in particular the information the spectator has at hand, and not a property of the event. Famous Bayesians are for example Bruno de Finetti, Frank Ramsey, L. J. Savage, and Edwin T. Jaynes. The difference between the frequentist and Bayesian view is illustrated in the following example.

Example A.2.2 (Urn Experiment - Frequentists vs. Bayesians).
Statement $S$: “There is an urn with equally many white and black balls.” For a frequentist $F_1$ the probability of drawing a white ball is 0.5, since if balls where drawn from the urn infinitely many times half of them would be white. For a Bayesian $B_1$ with information $S$, the probability of drawing a white ball is 0.5, since there is no reason for the Bayesian to favor white or black\(^2\). For a Bayesian $B_2$ with information $S$ together with the statement $S_1$: “the black balls where put into the urn before the white balls”, would have a higher probability for drawing a white ball than Bayesian $B_1$.

The urn example above illustrates two important things. First, and proven in [O'Hagan and Forster, 2004], for repeatable and independent random events, such as drawing a ball from an urn, the Bayesian and frequentist views coincide. Also, all computational rules of probabilities, such as the product rule, the sum rule, and Bayes’ rule can be used with both frequentist and Bayesian definitions of probabilities.

Second, it is clear that for Bayesian $B_2$ with information $S_1$ in addition to $S$ has another probability for drawing a white ball than Bayesian $B_1$ has. However, the exact value of the probability for Bayesian $B_2$ is not easily determined. For a frequentist, the probability of an event is defined as its relative frequency. It is a property of the object, and is sometimes said to be objective. For a Bayesian the probability for an event is subjective in the sense that it is determined by the information the person has at hand. However, as discussed in the next section, the probability for an event is not arbitrary.

A.2.2 Switching Between Interpretations

These two views, the frequency-type and the (Bayesian) belief-type, are different in a philosophical sense, and a natural question is why the same word, “probability”, is used for both of them. Hacking [Hacking, 1976] gives one explanation: in daily life, we (humans) switch back and forth between the two perspectives. Consider the following example.

Example A.2.3 (Switching Between Frequency and Belief).
A truck of model R arrives to a mechanic at a workshop. The mechanic knows

\(^2\)This is often called the Principle of Indifference.
that among all model R trucks, one out of ten of the trucks that arrives to the
workshop has fault $F$ present. The mechanic concludes that choosing a random
model R truck of those that has been (or are) at the workshop, the probability
that fault $F$ is found is 0.1. This probability is of frequency-type.
Consider now the particular truck that just arrived to the workshop. What is
the probability that this truck is has fault $F$? The truck is either faulty or fault
free, so there is no randomness, but still the mechanic would (probably) say
that the probability is 0.1. He reasons as follows. Out of all model R trucks
that has visited the workshop, fault $F$ was present in 1 out of 10. This truck
is a model R and has arrived to the workshop. Taking those three pieces of
information into account, the probability that this particular truck has fault $F$
is 0.1.

\section{A.3 The Bayesian View: Probability as an Extension to Logic}

In this section, we follow the reasoning by Jaynes in [Jaynes, 2001], and show
how the belief-type, (or Bayesian) interpretation of probability can be subjective
without being arbitrary. To do this, we use the language of logic, and extend
it to also consider uncertain events. For example, assume that it is known that
$A \Rightarrow B$, and that we know that the event $A$ is true. We can then draw the
conclusion that also $B$ is true. On the other hand, if $B$ is known to be true we
can not say anything about $A$ with certainty. However, our common sense says
that if $B$ is known to be true, $A$ is more likely to be true.

\subsection{A.3.1 Consistency and Common Sense}

We will now formalize this reasoning, but first we recall the traditional definition
of probability, by Kolmogorov’s axioms [Blom, 1994, Jaynes, 2001]:

- For every event $A$ it holds that $p(A) \in [0, 1]$.
- For the whole sample space $\Omega$ it holds that $p(\Omega) = 1$.
- If $A$ and $B$ are mutually exclusive, it holds that $p(A \cup B) = p(A) + p(B)$
(“Sum Rule”).

Furthermore, the conditional probability of $A$ given $B$ is defined by

$$p(A|B) = \frac{p(AB)}{p(B)} \text{ "Product Rule".}$$
Jaynes [Jaynes, 2001], based on Cox [Cox, 1946], takes another approach. Starting from three fundamental desiderata, including requirements on consistency and common sense, they show that probability must fulfill the sum and product rules. The three desiderata are:

I Degrees of plausibility are represented by real numbers.

II Qualitative agreement with common sense.

III Consistency:

(a) If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.

(b) All evidence relevant to the question should be taken into account. Some of the information can not be arbitrary ignored and the conclusions drawn on what remains.

(c) Equivalent states of knowledge should always be represented by equivalent plausibility assignments. That is, if in two problems the state of knowledge is the same, then the same plausibilities must be assigned in both.

In the desiderata above, uncertainty is expressed in terms of plausibility. In [Jaynes, 2001] Jaynes states that probability is a monotonic function \( p \) of plausibility. Adding the requirement that probability should be described by a real number between 0 and 1, and adopting the convention that 1 represents that an event is true with certainty, and 0 that an event is certainly false, Jaynes [Jaynes, 2001] shows how the rules for probability computations can be computed from the three desiderata given above. In particular, this holds for the sum rule and the product rule.

The results in [Jaynes, 2001] and [Cox, 1946] are criticized and debated for example in [Halpern, 1999] and [Arnborg and Sjödin, 2000]. However, as remarked by Arnborg and Sjödin in [Arnborg and Sjödin, 2000], the “authors advocating standard Bayesianism have not been strengthened or weakened” by their analysis.

A.3.2 The Statements Behind the \(|\cdot|\)-sign

In the Bayesian view, the probability of an event is determined uniquely by the information behind the \(|\cdot|\)-sign. In [Jaynes, 2001], Jaynes argues that it is nonsense to talk about the probability of an event \( A \) without expressing the information \( i \) which it is based on. Even if there are no other explicit
events available, i includes general information, for example about how prior probabilities are assigned.

A.4 Assigning Numbers

In the discussion above we have seen that there are two main interpretations of probabilities, the frequentist and the Bayesian view, and as discussed in Chapter 3 we switch between these interpretations in ever-day life. We have discussed that also when the relative frequency, i.e. the probability according to the frequentist view, is not defined or relevant, we can use the Bayesian, belief-type, view. However, in the Bayesian case, there is still one main challenge left: How to assign numbers to the probabilities?

In order to obtain a non-arbitrary theory for probabilities, we need objective ways for determining the numbers. There are two cases to consider: (i) assigning probability distribution for a variable $X$, given a certain state of knowledge $i^*$; and (b) assigning probabilities of an event $A$ given a certain state of knowledge $i^*$.

The “state of knowledge” may be a defined background knowledge, for example “I rolled this dice yesterday, and it showed five eyes” if $A$ is the event “roll the dice and obtain six eyes”. However, $i^*$ often represent the “prior knowledge” about $A$ (or $X$). In many situations, the prior knowledge is used to express ignorance, i.e. “knowing nothing”. In this case, the prior probability distribution, the probability distribution conditioned on $i^*$ only, should be non-informative.

In the following sections we present four commonly used approaches for assigning prior probability distributions, followed by a method for assigning probabilities. Methods for assigning priors is further discussed for example in [O’Hagan and Forster, 2004].

A.4.1 Principle of Indifference

Suppose that there are $n > 1$ possible events, the principle of indifference then says that if there is no reason for favoring any of the events over the others, each event should be assigned probability $1/n$ (see [Jaynes, 2001, O’Hagan and Forster, 2004]). The Principle of Indifference is sometimes called the Principle of Insufficient Reason.

A.4.2 Jeffreys Prior

Jeffreys Prior for a real-valued variable $X$ is given by $p(x) = 1/x$. It is an improper prior, i.e. it does not integrate to one. However, since prior probability distributions are always used together with likelihoods $p(x|y)$ to obtain a posterior probability $p(y|x) \propto p(x|y)p(x)$, they can be safely used [O’Hagan and Forster, 2004].
Jeffreys prior has two interesting properties. First, it is invariant to scaling of \( x \), and, second, it is uniform in the logarithm of \( x \). With Jeffreys prior, the probability of obtaining a number in the interval \([1, 10]\) is equal to the probability of obtaining a number in the interval \([10, 100]\).

### A.4.3 Maximum Entropy

A more general approach to assigning prior probabilities, is to use the concept of entropy [Jaynes, 2001],

\[
H_p(x) = -\sum_i p(x_i|x^*) \log p(x_i|x^*),
\]

where the sum is replaced by an integral sign in the continuous case. The idea is to use the distribution \( p^* \) that is consistent with the available information \( i^* \) and that maximizes \( H_p \).

### A.4.4 Reference Priors

A method, related to the maximum entropy method, for assigning priors is the concept of reference priors introduced by Bernardo in [Bernardo, 1979]. Bernardo considers the problem of probability updating, i.e. the computation of the probability of \( x \) after learning \( y \), given by

\[
p(x|y, i^*) = \propto p(y|x, i^*)p(x|i^*).
\]

The likelihood \( p(y|x, i^*) \) in the equation above is assumed to be known, and \( p(x|i^*) \) is the prior to be assigned.

The reference prior is the “least informative” prior in the sense that as much as possible is learned about \( X \) through the likelihood. This means that the difference in information (or knowledge) about \( X \) in the posterior distribution \( p(x|y, i^*) \) relative to the prior \( p(x|i^*) \) is maximized. The reference prior is obtained by maximizing the expected Kullback-Leibler divergence of the posterior distribution relative to the prior. Technically, the reference prior is defined in the asymptotic limit, i.e., the limit of the priors obtained by maximizing the expected Kullback-Leibler divergence to the posterior as the number of data points goes to infinity.

### A.4.5 Betting Game

On possibility for assigning probabilities to events that are not possible to repeat several times is to use a betting exercise as described in [Jensen and Nielsen, 2007, Jeffrey, 2004]. For example, what is the probability that there will be snow in Linköping on December 18 2010? Anna, based on her background and
experience, estimate the probability to $p_A$. Bill, with other background knowledge and other experience, may estimate the probability to another value, say $p_B$. In this sense, the probability for snow in Linköping 2010 is subjective. One way to assign numbers to the subjective probabilities is the following. Assume that there is a ticket that is worth €100 each if there is snow in Linköping on December 2010. Anna thinks that €10 is the right price for this ticket, and thus $p_A = 0.1$. Bill, on the other hand, may think that €1 is just the right price, and thus $p_B = 0.01$. Betting games as the one described above are used to predict markets for commercial means [Hanson, 2007].

References


References


<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>updateBN</td>
<td>176</td>
</tr>
<tr>
<td>0/1-loss</td>
<td>203</td>
</tr>
<tr>
<td>action</td>
<td>156</td>
</tr>
<tr>
<td>observation</td>
<td>130</td>
</tr>
<tr>
<td>operation</td>
<td>130</td>
</tr>
<tr>
<td>repair</td>
<td>130</td>
</tr>
<tr>
<td>action request</td>
<td>157</td>
</tr>
<tr>
<td>action result</td>
<td>157</td>
</tr>
<tr>
<td>AO*</td>
<td>162</td>
</tr>
<tr>
<td>assembly state</td>
<td>156</td>
</tr>
<tr>
<td>Automotive diagnosis</td>
<td>55</td>
</tr>
<tr>
<td>automotive process</td>
<td>52</td>
</tr>
<tr>
<td>background information</td>
<td>26</td>
</tr>
<tr>
<td>background knowledge</td>
<td>100</td>
</tr>
<tr>
<td>Bayesian network</td>
<td>36, 83, 131, 154, 200, 208</td>
</tr>
<tr>
<td>Bayesian view</td>
<td>25, 26</td>
</tr>
<tr>
<td>BDe score</td>
<td>208</td>
</tr>
<tr>
<td>belief</td>
<td>25, 26</td>
</tr>
<tr>
<td>belief state</td>
<td>33, 157, 169</td>
</tr>
<tr>
<td>Binary Diagnostic Matrix</td>
<td>62</td>
</tr>
<tr>
<td>car start problem</td>
<td>32</td>
</tr>
<tr>
<td>component</td>
<td>127, 158, 164</td>
</tr>
<tr>
<td>d-separate</td>
<td>137, 155</td>
</tr>
<tr>
<td>DBN</td>
<td>131, 155</td>
</tr>
<tr>
<td>non-stationary</td>
<td>166</td>
</tr>
<tr>
<td>degree of belief</td>
<td>25</td>
</tr>
<tr>
<td>diagnosed mode</td>
<td>77</td>
</tr>
<tr>
<td>diagnoser</td>
<td>157, 169</td>
</tr>
<tr>
<td>diagnosis</td>
<td>54, 158</td>
</tr>
<tr>
<td>diagnostic trouble code</td>
<td>159</td>
</tr>
<tr>
<td>diesel engine</td>
<td>56</td>
</tr>
<tr>
<td>Dirichlet parameter</td>
<td>206</td>
</tr>
<tr>
<td>dynamic Bayesian network</td>
<td>131, 155</td>
</tr>
<tr>
<td>ECR</td>
<td>160</td>
</tr>
<tr>
<td>ECU</td>
<td>155</td>
</tr>
<tr>
<td>electronic control unit</td>
<td>3</td>
</tr>
<tr>
<td>empty event</td>
<td>133, 167</td>
</tr>
<tr>
<td>epoch</td>
<td>133, 166</td>
</tr>
<tr>
<td>event</td>
<td>130, 134, 166, 171</td>
</tr>
<tr>
<td>event-driven nsDBN</td>
<td>166, 171</td>
</tr>
<tr>
<td>evidence</td>
<td>130, 171</td>
</tr>
<tr>
<td>expected cost of repair</td>
<td>160, 224</td>
</tr>
</tbody>
</table>
INDEX

experimental data, 61, 201
expert knowledge, 7, 164
explaining away effect, 204

family of structure classes, 179
Fault Information System, 62, 97
Fault Signature Matrix, 62, 97, 206
fault tolerant control, 224
feature selection, 79
frequency, 25
frequentist view, 25
frontier, 137
FSM, 97

goal state, 158

inference, 33
initial BN, 132, 133
instant edge, 134, 167
interpretations of probability, 25
intervention, 125, 165

Kalman Filter, 34

Laplace approximation, 107
learning, 33
logistic score, 202

macrofault, 77
mechanic, 164
minimal repairable unit, 164
mode, 59
mode variable, 59
model
  black box, 30
  data driven, 30
  logical, 30
model based diagnosis, 5, 29
monitoring functions, 52
Multivariate Unimodal, 106

Naive Bayes, 207
nominal transition BN, 133
non-stationary DBN, 131

NOx emissions, 52
observable symptom, 129, 159
observational, 61
observational data, 69, 201
off-board diagnosis, 6
oil-pipe-gasket system, 127
on-board diagnosis, 6
On-board processor, 54
OPG system, 127
outgoing interface, 134

parameter constraints, 102
Particle Filter, 34
percentage of correct classification, 203
performance measure, 74
persistent variable, 134, 168
planner, 157, 160
proper score, 203

regression
  linear, 208
  logistic, 209
relative frequency, 25
repair-influenced BN, 178
residual structure, 62
response information, 61, 71
retarder, 155, 164

selective naive Bayes, 207
Sherlock algorithm, 80
structure class, 178
structured hypothesis testing, 81
structured residuals, 81
successor function, 162

symptom
  observable, 159
system status, 7

temporal edges, 155
temporal link, 131
time slice, 131
training data, 53
transition BN, 132
troubleshooting action, 130, 156
troubleshooting BN, 159
troubleshooting process, 133
troubleshooting session, 166
troubleshooting strategy, 157, 160
A DAE formulation for Multi-Zone Thermodynamic Models and its Application to CVCP Engines
Per Öberg, Dissertation No. 1257, 2009.

Control of EGR and VGT for Emission Control and Pumping Work Minimization in Diesel Engines

Efficient Simulation and Optimal Control for Vehicle Propulsion

Single-Zone Cylinder Pressure Modeling and Estimation for Heat Release Analysis of SI Engines

Modeling for Fuel Optimal Control of a Variable Compression Engine

Fault Isolation in Distributed Embedded Systems

Design and Analysis of Diagnosis Systems using Structural Methods

Air Charge Estimation in Turbocharged Spark Ignition Engines
Per Andersson, Dissertation No. 989, 2005.

Residual Generation for Fault Diagnosis

Model Based Diagnosis: Methods, Theory, and Automotive Engine Applications

Spark Advance Modeling and Control

Driveline Modeling and Control