Handling of uncertainty in petroleum applications

Lars Holden and Petter F. Mostad

February 17, 1997

Preface

This note reports the work done by NCC during 1996 for PSTI as part of the PSTI Juniper project. The Juniper project is concerned with desicion making under uncertainty in the petroleum industry. The NCC part of the project focuses on Geostatistical models, Bayesian networks, and comparing Bayesian modelling with Interval Probability Theory. Lars Holden and Petter Mostad have worked on this project at NCC, in cooperation with the Juniper team in Bristol.

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

Contents

Chapter 1

Assumptions in Geostatistical models

1.1 Introduction

In order to make precise predictions for reservoir performance, it is necessary to have a detailed reservoir description. The reservoir is unknown and may only be described in a stochastic model. This model is in fact a model of the knowledge or beliefs about the reservoir, not a model of the reservoir itself, as the reservoir itself is fixed, and not stochastic.

A stochastic model of the reservoir is based on many assumptions. Some of the assumptions are controversial and there may be several assumptions where the probability that the assumption is not correct is not negligible. With many such assumptions, the probability that all the assumptions are correct may be small. However, many of the assumptions are only critical in order to make a stochastic model of the reservoir but not critical for the actual estimate. This may be illustrated by the hydro carbon volume of a reservoir which depends on the average saturation, not the spatial distribution of the saturation of the fluids. Which assumptions that are critical for the reservoir performance may be difficult to evaluate and may depend on other reservoir properties.

The objective of this chapter is to discuss the assumptions in geostatistical models, the consequences of these assumptions on the prediction of the reservoir performance and propose a simplified non-geostatistical approach which may use information from geostatistical models.

1.2 A simple model

The purpose of this section is to illustrate in more detail the assumptions in a small example. In the example the oil in place in a reservoir is estimated. The example is based on a simple model of the reservoir. Well and seismic data may be used.

We will make the following assumptions:

The reservoir volume is treated as a stochastic variable

This is a philosophical assumption and it is not critical for the estimate. The reservoir volume is not stochastic, but we do not know the value. The stochastic model describes our knowledge about the reservoir. This is similar to the result of a dice which is thrown but where we are not able to observe which side is up. The uncertainty is connected to our knowledge of the reservoir. Hence, the uncertainty decreases as our knowledge increases. The stochastic properties are connected to repeatability, e.g. the actual reservoir is only one of many possible reservoirs which all are possible according to our present knowledge.

Geological interpretation

The reservoir is assumed bounded by a water/oil contact (WOC) and a top surface with depth $Z(x)$ where $x = (x_1, x_2)$ is the horizontal coordinate. This is a critical assumption for the estimate and it is quite often not obvious.

Porosity ϕ, saturation S(x), velocity v(x) and boundaries are well-defined

This is more a philosophical question. In the engineering approach this is just a matter of scale.

The seismic reflection time $t(x)$ is interpreted correctly

This is also a critical assumption for the estimate and quite often interpretation errors are made. The consequences of an error decreases as the number of wells increases. Measurement error is not included in this assumption.

We set the depth to the top structure $Z(x) = v(x)t(x) + \epsilon(x)$. Since all $v(x), Z(x)$ and $v(x)$ are well-defined this is not an assumption. The properties of $\epsilon(x)$ may however be discussed.

Next assumptions are:

The random functions $v(x), \phi(x), \epsilon(x)$ and $S(x)$ are Gaussian random fields

A detailed study will probably show that this assumption is not correct. In order to simulate the random fields it is necessary to assume that the random fields are from one of the very few types it is possible to simulate from. This may be considered as an approximation. For volume calculations it is probably not critical as long as the expectation and trends are correct.

Semivariogram and expectation of the Gaussian random fields

It is assumed that the expectation of the velocity depends linearly on the reflection time, i.e., $E\{v(x)\} = a + bt(x)$. It is easier to assume that the velocity depends on reflection time than depth. For the porosity, saturation, and ϵ, we

assume constant expectation. This may be considered as approximations. The approximation may be poor if there are other trends that are left out. The effect of the assumptions decreases as the number of wells increases.

The user should specify the distributions for each of the parameters in the Gaussian fields. These distributions are not assumptions, they just reflect the knowledge of the user about the reservoir. The hydrocarbon volume is defined as an area integral which is based on a vertical averaging in ϕ and S by:

$$H = \int (Z(x) - Z_{woc})\phi(x)S(x)dx$$

The final assumptions are :

Simulation algorithm simulates from the model

Also here there are approximations both due to the simulation algorithm and grid resolution. This is probably not critical.

The number of simulations is sufficiently large to reflect the uncertainty in the estimate of the production

In this simple model there is not expected to be a complicated relationship between different variables, and it is cheap to generate a sufficient number of realizations. This assumption is therefore not critical for the estimate

There are no errors in the implementation, and the program is correctly used.

There will always be bugs in large programs and possible to misuse it. If satisfactory quality control is used, the probability for a significant error is very small, particularly in this simple example.

Summary of small example.

Based on the computer program several realizations for the surfaces are generated conditioned on the available data and the corresponding volumes H are calculated. The result of the study is that given all the assumptions, the estimate on H is given by the distribution of the calculated values of H. The only assumptions which are critical for the estimate are that the refection time is interpreted correctly and that there are no trends which are not included. If these assumptions are considered controversial, it is possible to expand the model and include alternative versions.

1.3 The GRUS study and a simplified model

GRUS is a large uncertainty study of the production of oil from the Veslefrikk field. The study was performed by the Norwegian Computing Center in 1991-1994. It was based on interpretations performed by the operator Statoil. See (Lia, Omre, Tjelmeland, Holden, MacDonald, Hustad & Qi 1994), (Egeland,

Omre, Lia, Tjelmeland, Holden & Andersen 1994), (Lia, Omre, Tjelmeland & Holden 1995).

In the study, well data, seismic and geological knowledge were used. The model consisted of stochastic models for depth conversion, facies, petrophysics, fluids and seismic faults. Apriori distributions for more than 100 parameters were specified and then estimated from the data. Each realization was upscaled and a grid was calculated depending on the realization. About 300 full field reservoir realizations were generated and analyzed using more than 2 month of CPU time on a Unix work station. Some of the conclusions, given that all the assumptions are correct, are:

- The 95% prediction interval for the production after 20 years was $\pm 15\%$.
- The 95% prediction interval for original oil in place was $\pm 14\%$.
- The 95% prediction interval for the production after 20 years given the oil in place was $\pm 8\%$ Oil in place and recovery seemed to be quite independent variables.
- The best estimate for the recovery of the mobile oil is 57% with a 95% prediction interval [51%,62%].
- About 75% of the uncertainty in the production is due to parameters, the rest is due to heterogeneities.
- The parameters which contributed most to the uncertainty were fault transmissibilities, saturations and seismic velocity.

In this section, we will discuss the assumptions in more detail and propose a technique which may make it possible to quantify the production also under much weaker assumptions. In order to illustrate the approach, quantification of the effects of the different assumptions is tried. This quantification is just an illustration and not based on a thorough evaluation.

The order in which the assumptions are made is critical for the evaluation of the importance of each assumption. F.ex. the stochastic model neglects the fine scale variability. If the stochastic model assumption is made first, this makes the assumption regarding the reservoir simulation less critical.

The assumptions are divided into four different classes depending on how they may be treated in the quantification.

1.3.1 Philosophical assumptions

Typical assumptions are:

The reservoir is stochastic

The different variables are well-defined

These assumptions will probably have no effect on the quantitative estimates and are not discussed further.

1.3.2 Assumption which may be handled in the model

Distribution of the parameters

There is a large number of parameters in the model. Distributions are specified

for more than 100 parameters. Parameters considered to be of less importance are kept constant. One may argue that the specified distributions reflect the knowledge of the user and hence are not assumptions. With a large number of parameters and limited time available the judgment will not always be the best possible. However, it is possible, by using the model, to find the consequences of a particular value/distribution.

It may be reasonable to argue that this assumption is not critical for the quantification.

1.3.3 Assumptions which may be handled in a simplified model

Let us use the much simpler model

$$P_n = HR_n = \phi SVR_n$$

where

- P_n is production after n years
- H is the hydrocarbon in place in the reservoir.
- ϕ is average porosity in a specified region
- S is average saturation in the specified region
- V is the volume of the region
- R_n is the recovery in the region after n years

This model may be made a little more detailed by summing over different layers (cross flow must be treated) and recovery may be split in sweep efficiency and micro scale recovery. From the GRUS study it is possible to find joint distributions for these variables given all the assumptions. The advantage of using such a simple model is that fewer assumptions are required. The disadvantage is of course that the precision is much poorer. It is possible to base the quantification of the variables in such a model on a detailed study like the GRUS model and then modify the distributions in order to compensate for the assumptions in the GRUS model.

The following assumptions are from this class:

The stochastic model represents the uncertainty in the model

There are many simplifications and approximations in the stochastic model compared to the reservoir. In the GRUS study subseismic faults and variability on a small scale were neglected, facies and surface geometry were approximated. The stochastic model also depends strongly on the geological interpretation. It is believed that the uncertainty of the estimate of oil in place is underestimated and that the uncertainty in the recovery is probably underestimated more severely.

Upscaling does not change the properties

It is possible to test the effects of this assumption by performing sensitivity runs on a finer scale. The uncertainty in the recovery is underestimated slightly due to this assumption.

The reservoir simulator model of the flow in the reservoir

It is known that the reservoir simulator uses a simplified model of the flow. It is difficult to quantify the size of the error due to this simplication. Most simulators are based on the same mathematical model such that different simulators does not give independent estimates of the flow in the reservoir. Laboratory experiments and performance of other fields may be used in addition to more theoretical studies in a quantification of this error. The Veslefrikk field used in the GRUS study is a quite simple oil/water reservoir, so the effect of this assumption should not be very large.

Depletion strategy is fixed in the reservoir

During the development of the field there will be considerable increase in knowledge og the reservoir and technical improvements. An optimalization of the recovery during the development would probably increase the recovery drastically.

Simulation algorithm simulates from the model

Also here there are approximations both due to the simulation algorithm and grid resolution. This is probably not critical.

No errors in the implementation and correct usage

It is believed that the probability for a significant error is negligible.

1.3.4 Assumptions where the effect of failure is difficult to quantify

Complete misunderstanding of reservoir
This is very difficult to quantify. Veslefrikk is not considered as a complicated field such that the probability for a complete misunderstanding is small.

1.3.5 Summary on a simplication of the GRUS model

In the GRUS study many assumptions were made and some of these are controversial. Based on the assumptions in the model it is possible to quantify the performance and find all possible joint distributions between different input and output variables.

It is also possible to build a simplified model with much less assumptions. Also in the simplified model it is necessary to specify distributions for the different variables. This specification may be based on experience from the GRUS study and an evaluation of the effect of the different assumptions made.

We will try illustrate how this may be done in a simplified model based on the GRUS study. It is reasonable to argue for a slight increase in the uncertainty in the oil in place due to limitations in the GRUS stochastic model. The expected value of the oil in place should probably be the same in the simplified model as in the GRUS study. The uncertainty in the recovery could be increased due to the uncertainty in small scale variation and flow properties. By optimizing the depletion strategy it should be possible in increase the recovery significantly, particularly for the reservoirs with small recovery. This effect may be so large

that the uncertainty in the recovery in the simplified model could be less that in the GRUS model In the GRUS model the oil in place and recovery seemed as almost independent variables. This property should probably be used also in the simplified model.

Chapter 2

Mathematics for uncertainty modelling

This chapter describes and compares some mathematical models that may be used for uncertainty modelling. After an introduction, we go on to describe the mathematics in Section 2.2, and then Interval Probability Theory is compared with Bayesian modelling in Section 2.3.

2.1 Introduction

In our everyday life, we frequently encounter information we may describe as uncertain, unspecific, vague, ambiguous, or in other ways not precise and certain. At the same time, we must often base desicions on such information. The Juniper project has focused on such desicion making in the petroleum industry. In this section, we will look at what kind of mathematical models we may use in this context.

If we want to make a mathematical model incorporating some or all of these phenomena, we need to be very careful about what is "uncertain" or "fuzzy", and what is not. A mathematical model is by definition a completely described set of possibilities, or states, or variables, and relations between them. Our basic scientific method is to build precise models (in many sciences mathematical models), conjecture that they represent reality in some sense, (i.e., by connecting their states to concepts in the real world), and then test whether predictions from the model describe or predict well the real world. A mathematical model cannot in itself be "uncertain" or "fuzzy". The term "open world model" must thus be interpreted as a model with precisely defined, and well known, states, but where some of these states may be used to *model* unforeseen possibilities in the real world. It follows that an "open world model" is really no different from any other model: We must still describe all its states presiciely, and the relationship between all states, whether or not these states are used to represent "unforeseen events" or "foreseen events".

We must thus divide our work in two: First, the description and understanding of a model used for uncertainty modelling, and then, we must evaluate its appropriateness for modelling real world events. The last task may (and indeed must) be done by evaluating how well the model predicts events, compared to

other models. This is the standard way of testing models in any science, and should not be different for models dealing with uncertainty.

The traditional way of modelling uncertainty has been with probability theory and stochastic theory. Several researchers, some of them with background in Artificial Intelligence, have set up other mathematical theories for this purpose, (Krause & Clark 1993). In this note, we will focus on Bayesian modelling and Interval Probability Theory, and in addition mention some other models.

2.2 Models

Probability theory has traditionally been the way to model uncertain knowledge. Mathematically, such a model consists of a set, and a non-negative function on that set whose sum (or integral) is one. The set is then meant to represent the set of "possible worlds" considered, while the function represents their probability (or probability density). Note that the set must be fully described in the model, but that of course some subset may be interpreted as "unforeseen worlds", i.e., situations we at this point cannot fully describe.

The main difference between the traditional frequentist probability modelling and Bayesian modelling, is that Bayesian modelling assumes there is a completely specified probability distribution at *any* given time in the modelling, also before what is considered data in this modelling step is introduced. In other words, in a frequentist approach, a parameter of the model may be completely unspecified at some point, while in the Bayesian approach, there will always exist some prior information about it.

We will return to Bayesian models in the next chapter; let us now look at Interval Probability Theory. The basic idea of this theory is to use a numerical interval instead of a single value for probabilities. The most direct interpretation of this is that the probability of an event is known only to be in an interval $[a, b]$; we do not want to model where in the interval the probability might be. Let us phrase this in the same language as the Bayesian model: We have a set (representing "all possible worlds"), and instead of modelling a probability for all subsets, we say that some subsets have a probability in a given interval. In this way, an IPT model may in fact be seen as a set of stochastic models: The set of stochastic models consistent with the given intervals. Using this interpretation, a calculus for IPT may be developed. We then write $P(A) = [a, b]$, although we should, according to the above interpretation, formally write $P(A) \in [a, b]$. For example if $A = A1 \cup A2$, and if

$$P(A1) = [N_1, P_1]$$

and

$$P(A2) = [N_2, P_2],$$

we get that

$$P(A) = [N_1 \vee N_2, 1 \wedge (P_1 + P_2)], \tag{2.1}$$

where $\vee$ and $\wedge$ mean taking the maximum and the minimum, respectively.

However, the interpretation above may in fact be too simplistic. In (Cui & Blockley 1990), Section 6, the following information is set up:

$$\begin{aligned} P(A|B1) &= [N_1, P_1], && (2.2)\\ P(A|\overline{B1}) &= [N_2, P_2], && (2.3)\\ P(A|B2) &= [N_3, P_3], && (2.4)\\ P(A|\overline{B2}) &= [N_4, P_4], && (2.5)\\ P(B1) &= [N_5, P_5], && (2.6)\\ P(B2) &= [N_6, P_6]. && (2.7) \end{aligned}$$

If we interpret A, $B1$ and $B2$ as being subsets of a general set (of "possible worlds"), the overbar to mean set complement, and $P(A|B)$ to mean $P(A \cap B)/P(B)$, it is possible to compute the resulting information about the probability of A: Equations 2.2, 2.3 and 2.6 give

$$\begin{aligned} P(A) = &[(N_1N_5 + N_2(1-N_5)) \wedge (N_1P_5 + N_2(1-P_5)),\\ &(P_1N_5 + P_2(1-N_5)) \vee (P_1P_5 + P_2(1-P_5))], \end{aligned} \quad (2.8)$$

while Equations 2.4, 2.5 and 2.7 give

$$\begin{aligned} P(A) = &[(N_3N_6 + N_4(1-N_6)) \wedge (N_3P_6 + N_4(1-P_6)),\\ &(P_3N_6 + P_4(1-N_6)) \vee (P_3P_6 + P_4(1-P_6))]. \end{aligned} \quad (2.9)$$

Combining the information in Equations 2.8 and 2.9, we get that

$$\begin{aligned} P(A) = &[((N_1N_5 + N_2(1-N_5)) \wedge (N_1P_5 + N_2(1-P_5))) \vee\\ &((N_3N_6 + N_4(1-N_6)) \wedge (N_3P_6 + N_4(1-P_6))),\\ &((P_1N_5 + P_2(1-N_5)) \vee (P_1P_5 + P_2(1-P_5))) \wedge\\ &((P_3N_6 + P_4(1-N_6)) \vee (P_3P_6 + P_4(1-P_6)))]. \end{aligned} \quad (2.10)$$

However, in their publication, Cui and Blockley compute $P(A)$ by viewing A as the set-theoretic union $A = A1 \cup A2$, where $A1$ and $A2$ represent different "proof paths" for A[1]. Presumably substituting Equations 2.2 to 2.5 with

$$\begin{aligned} P(A1|B1) &= [N_1, P_1], && (2.11)\\ P(A1|\overline{B1}) &= [N_2, P_2], && (2.12)\\ P(A2|B2) &= [N_3, P_3], && (2.13)\\ P(A2|\overline{B2}) &= [N_4, P_4], && (2.14) \end{aligned}$$

one now gets that

$$\begin{aligned} P(A) = &[((N_1N_5 + N_2(1-N_5)) \wedge (N_1P_5 + N_2(1-P_5))) \vee\\ &((N_3N_6 + N_4(1-N_6)) \wedge (N_3P_6 + N_4(1-P_6))),\\ &1 \wedge (((P_1N_5 + P_2(1-N_5)) \vee (P_1P_5 + P_2(1-P_5))) +\\ &((P_3N_6 + P_4(1-N_6)) \vee (P_3P_6 + P_4(1-P_6))))], \end{aligned} \quad (2.15)$$

which is different from Equation 2.10. An example of the difference may be seen by setting all the N's and P's to 0.5; then Equation 2.10 gives $P(A) = [0.5, 0.5]$, while Equation 2.15 gives $P(A) = [0.5, 1]$.

[1] David Blockley has confirmed this as the correct approach in a letter to me.

Thus it may be necessary to substitute the straightforward interpretation of IPT above with a more complex axiomatic development to derive the calculus presented in (Cui & Blockley 1990).

Indeed, in order to get an "open world model", one has tried in IPT to avoid talking about an underlying set of "possible worlds". In an axiomatic development, the A's and B's above could simply be the elements of the model (representing "proof paths" or even "processes"), rather than subsets of some set of "possible worlds". One must then define operators corresponding to logical "and", "or", and "not" on the set of such A's and B's, and a function P assigning to each element an interval: $P(A) = [a, b]$. One must then establish a consistent set of axioms for this theory. Developments along this line may be found in (Krause & Clark 1993) and (Blockley 1985). However, we will not go further into this here, and in what follows, we will go back to the initial interpretation of IPT above.

Finally, we should mention that a number of other mathematical models have been developed to represent uncertainty. Maybe one of the most interesting (at least from a mathematical point of view) is Dempster-Shafer theory (Shafer 1975). In that theory, there *is* an underlying "set of worlds" Ω, but instead of a probability function on its subsets, there is a "belief function" Bel such that $Bel(\Omega) = 1$, but such that $Bel(A) + Bel(\overline{A})$ is not necessarily 1. It is not entirely obvious how to best compare this theory with IPT, but its "rule of combination" is clearly more restrictive than in IPT.

2.3 Comparing IPT with Bayesian modelling

Having described the mathematical theories, the question remains which theory is best suited to model actual uncertain knowledge. The purpose of mathematical modelling is after all that computations within the model should yield insights (or "predictions") that are true, and non-obvious. But, to compare IPT with Bayesian modelling, let us first compare them on a mathematical level.

Let us return to viewing IPT as a set of possible stochastic models on the set Ω of possible worlds. Assume our information is given as a set of statements of interval probabilities, and conditional interval probabilities, for a set of subsets $\{A_1, \ldots, A_n\}$. Then Ω may be partitioned into a set of disjoint subsets $B = \{B_1, \ldots, B_m\}$ such that all A_i are unions of sets in B. For each stochastic model on Ω which is possible according to the information, we may associate to it the point in $\mathbb{R}^m$ corresponding to the vector $(P(B_1), \ldots, P(B_m))$. The set of possible stochastic models is now exactly the set associated in this way to points in a certain subset $S \subset \mathbb{R}^m$. It is straightforward to see that S must be a convex set, as all interval probabilities and conditional interval proabilities impose convex restrictions in $\mathbb{R}^m$. This may in fact be utilized to develop IPT calculus rules.

In Bayesian modelling, S would of course only be a point. Having S larger than a point may be regarded as the difference between the two models. So one way to compare the theories is to compare IPT with a double-stochastic model where we use a probability distribution on S. For example, the statement $P(A) = [a, b]$ of IPT could be compared with a model where we use an underlying stochastic variable p, with $P(A \mid p) = p$ and p having some specific probability distribution within the interval $[a, b]$. Comparing the two theories,

it turns out that modelling a probability distribution on S may be crucial when combining information.

2.3.1 Combining information

To be concrete, let us look at a specific example from oil exploration. Let A be the statement "There is oil at location L", where L is in an undrilled area. From general experience, one might produce an educated guess that the chance of A being true is 0.5; it would be a vague guess.

Assume that, before drilling at L, 50 other wells are drilled in the area, of which 25 turn out to be dry, the others have oil. Assuming (unrealistically) that there is no spatial pattern in the wells with oil, we could now say that the chance of A being true is 0.5, but that this is a much "firmer" guess. We may express this difference by using a double-stochastic model.

For example, the vague guess might be expressed by the following model: Let p be a parameter with a uniform probability density function between 0.1 and 0.9, and say that A given the parameter p has probability p. Simple probability theory tells us that to find the expectation of A, we must first compute the expectation of A given p, which is p, and then compute the expectation of the distribution for p, which is 0.5. The model for A given the observations in the 50 other wells would then be the same except that the probability density function for p would have a spike around 0.5. The expectation would still be 0.5.

In fact, we may compute what this last probability distribution would look like, assuming that all the 51 well locations have the same initial probability for oil, and that the correlation between them is indeed expressed by an underlying parameter p as above (i.e., they are independent given the value of p). If the probability density function for p is $f(p)$ after a certain number of wells have been drilled, then it will be proportional to $f(p)p$ after including the information that a well with oil has been found, and $f(p)(1-p)$ after including the information from a dry well. Thus including information from the 50 wells will give you the probability density function

$$f(p) = \begin{cases} \frac{1}{C}p^{25}(1-p)^{25} & \text{if } 0.1 \le p \le 0.9 \\ 0 & \text{otherwise} \end{cases}$$

where

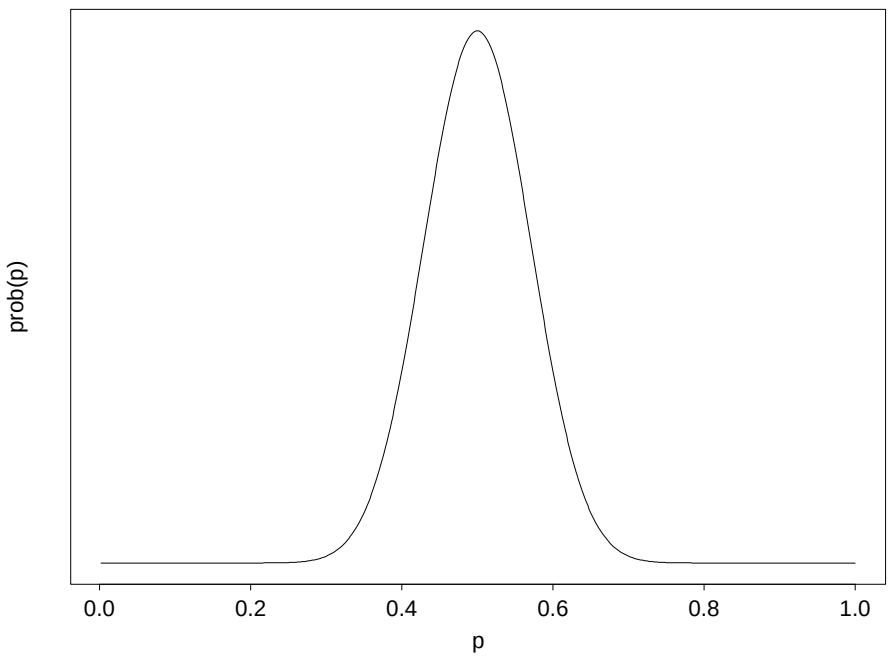
$$C = \int_{0.1}^{0.9} p^{25}(1-p)^{25}dp.$$

This function has a spike at 0.5, see Figure 2.1.

What would be the difference between A after no observations and A after the 50 wells? If we only wanted to make our best possible guess at whether A was true, there would be no difference: In both cases, the best guess (the expectation) would be 0.5. However, there would be a difference when combining with new information. Assume we drilled another well at yet another location X, and found oil there. If we had included the information from the 50 other wells, the new expectation for A would now be 0.509, while if we had not included this information, the new expectation would be 0.607.

Note that the possible *range* of p after these updates is still $[0.1, 0.9]$. If we had neglected to consider the probability distribution for p within this interval,

Figure 2.1: Probability density for p



we whould still only know that p was in the interval $[0.1, 0.9]$, even after drilling 50 wells.

In the example above, we have combined knowledge from several different drillings, and geological (or "prior") knowledge, to derive at a "posterior" model that seems to represent numerically something similar to what we would have concluded without uncertainty modelling. Thus it seems that, in this case, Bayesian updating models quite well the process of increasing experience. How can we use IPT in this situation?

The IPT interval for A given no other drillings could be $[0.1, 0.9]$ in analogy with the model above. If we assume, as above, that this interval may be used for all the wells when we are not considering information from the other wells, and if we further assume that the dependencies between the wells is via an underlying probability p, as above, then we may in fact not infer anything about the IPT interval for A from drilling the 50 wells. This was shown mathematically above, and may be understood as follows: "Testing" the underlying probability p by drilling wells that have the probability p to contain oil can only give uncertain information about p. Drilling 25 wells with oil makes it less likely that p is close to 0.1, but it can never *prove* that it is above a number greater than 0.1: The next 225 wells drilled *could* be dry.

To get something more useful out of IPT in this situation, we need to consider other dependencies between the different pieces of information. The question is then how to define these dependencies. Presumably, we would want them to produce a narrower interval $[c, d]$ for the probability of A given the results

from the 50 first wells. We would also want to get the same interval $[c, d]$ no matter how we divided the set of wells into subsets, and included the information stepwise. It is not obvious how such dependencies should be set up. Requireing a user of an automated system to specify these dependencies would leave a heavy burden on the user, and it could be close to impossible to ensure consistensy in the sense that splitting up the well information in different subsets, as above, should give the same result.

2.3.2 Modelling the unknown

Above, we compared IPT with Bayesian modelling by putting a probability distribution on the set S associated with IPT. However, the comparison could of course be done in many ways.

Saying $P(A) = [a, b]$ could be interpreted as: There is a chance a that we are in a situation where A is known to be true, there is a chance $1 - b$ that we are in a "world" where A is known to be false, and there is a chance $b - a$ that we are in a "world" where we chose not to model whether A is true or not. This interpretation would be in line with the ideas of an "open world model". We would still have a specific set of "possible worlds" Ω, and a single probability distribution on it. But for each statement A for which we care to model an interval probability $P(A) = [a, b]$, we would have associated a subset $A_+ \subset \Omega$, with $P(A_+) = a$, where we would know A to be true, a subset $A_- \subset \Omega$, with $P(A_-) = 1 - b$, where A would be known to be false, and a subset $A_0 \subset \Omega$, with $P(A_0) = b - a$ where we would not know whether A was true or false.

Let us now, as above, divide Ω into disjoint subsets $B_1, \ldots, B_m$ such that all the sets above are uions of the B_i's. If we specify all the necessary dependencies, we get a vector $(P(B_1), \ldots, P(B_m))$ of probabilities. In this situation, we may make all necessary computations using standard probability theory. Note however that we need to specify *all* dependencies. If the variables we use are specified in a sequence $A, B, C, \ldots$, we would for example need to specify the probabilities for B being true, false, or unknown, given that A is true, false, or unknown. If we always specified that B was unknown given that A was unknown, we could very soon end up with variables that are mostly unknown. If a variable we were interested in was expressed as being dependent on a large subtree of variables, we would very likely end up with a computation showing that our variable of interest is mostly unknown. This is unlikely to be a useful result.

Thus we need to routinely specify that a variable B may be known to be true or false even if an underlying variable A is unknown. Our model will then be no different from a Bayesian model: For each variable A, we use (when necessary) an underlying variable signifying whether we explicitly model A or not. If we do not model A, we must use other information in the modelling of dependent variables. An example of this kind of modelling is given in the next chapter in the petroleum exploration section.

2.4 Practical uncertainty modelling

In the previous sections, different mathematical models were presented and compared. Assuming that the different models have been specified mathematically,

the question remains of how well they are adapted to model real world concepts. Testing this depends on how the mathematical concepts of each theory is mapped to concepts in the real world. Several interpretations of IPT may be given, and so we cannot now conclude whether IPT is a "good" theory or not.

In a practical sense, what matters is whether a theory can help you with new insights in real examples. Thus to investigate the theories further, we must look at examples, where the interpretation of each theory is fixed, and compare the results of the different modelling teqhniques this way.

Chapter 3

Bayesian Networks

Modelling uncertainties may be done in several ways, as seen in the last chapter. In this chapter, we will focus on the Bayesian approach. We will start with some of the philosophy behind this this kind of modelling. After a short introduction in Bayesian networks, we will look at two examples in some depth. We will end with some general comments about using Bayesian nets for our kind of applications.

3.1 A Bayesian World View

We all have ideas and opinions about how the real world, or certain specific parts of it, works. In any situation, we use our powers of association to classify and describe it, and based on these associations, we will always make some kind of internal model of the situation, whether this model is very fixed and certain, or quite preliminary and uncertain. For example, letting go of a pencil one meter above the floor, we are pretty sure it will fall down, although this theory has only been tested a finite number of times. On the other hand, seeing a UFO will probably make most of us quite unsure about what happens next. But when pressed for a prediction, we would probably all have some predictions: Some would expect to reveal the thing as an illusion or a hoax immediately, others would expect to be taken in for "experiments".

The central assumption of Bayesian modelling is that the knowledge and beliefs of a person at any given time may be represented as a stochastic model, i.e., a probability distribution over the set of possible worlds. In any practical modelling, this set is of course a particular simplification of the world, focusing on the attributes we need to model. The second assumption of Bayesian modelling is that when a person observes "new data", his or her knowledge after these observations corresponds to the probability distribution conditioned on these data (in the mathematical sense of conditioning). We will not go further into the appropriateness of these assumptions, but we will just note the following points:

- We assume that our model of a persons knowledge is, at all times, a *fully specified* stochastic model. This contrasts to approaches where one uses a partially defined stochastic model to model the knowledge. Both

non-Bayesian statistics and the interpretation of IPT used in the previous chapter fall in this category.

Many people react to this feature of Bayesian modelling. And it *is* often difficult to specify good prior distributions for all parameters in a complex model. However, the alternative, to leave some parts of the model completely unspecified, is probably also unsatisfactory. As we tried to demonstrate above: As long as we understand a stochastic model, and manage to connect it to our own internal model of our knowledge, we will always have *some* idea about what we think its parameters should be, and this knowledge should be utilized.

- A strength of Bayesian modelling is that the updating of the model by new data may of course be done in several steps. Thus as long as the original stochastic model has been formuated realistically and as a true representation of a persons diverse knowledge, the effect of all additional information the person receives may be computed by simply mathematically conditioning the model on the data. In practice, of course, the conditional model will not always be a realistic representation of the new knowledge of the person. This is because the stochastic model is necessarily a simplification of the persons knowledge. We cannot in practice model a persons reaction to any possible data set that may appear. However, we may try to incorporate into the model that some "unexpected" data may give some "unforeseen" posterior model. It is then important to classify to what extent the "unforeseen" posterior model differs from the foreseen ones. In many cases, the failing of one model assumption will *not* throw the whole model into disarray, it will simply widen the uncertainty about certain aspects of it. It is important to include also this kind of information into the model; otherwise, it may be overwhelmed by the possibility of "unforeseen" and completely unknown posterior models.

3.2 Bayesian Networks

We will here focus on Bayesian models which may be represented as a set of one-dimensional variables. Our goal is to find a common framework to treat such models in. We would like to have a tool that could do the neccessary mathematics of finding the posterior model, once the prior model has been specified.

Assume a stochastic model consists of the stochastic parameters $a_1, \ldots, a_n$, with some dependencies between them. The total probability distribution may be written $\pi(a_1, \ldots, a_n)$, where π denotes the generic function giving the probability density. We may always split this expression into conditional distributions

$$\pi(a_1, \ldots, a_n) = \pi(a_1)\pi(a_2 \mid a_1)\pi(a_3 \mid a_1, a_2) \ldots \pi(a_n \mid a_1, \ldots, a_{n-1})$$

In many stochastic models in practice, it turns out that a variable is not dependent on all the previous variables, only some of them. In a concrete example, we could have, instead of the above expression,

$$\pi(a_1, a_2, a_3, a_4, a_5) = \pi(a_1)\pi(a_2)\pi(a_3 \mid a_1, a_2)\pi(a_4 \mid a_2)\pi(a_5 \mid a_3, a_4).$$

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

This interdependency is illustrated graphically in Figure 3.1, where each stochastic variable is represented by a box, and each dependency by an arrow.

Clearly, for any stochastic model, there could be several different graphs representing it, depending on the ordering of the variables. In particular, the graph may be transformed so that any variable is the first one, and thus dependent on no other variables.

Once a stochastic model has been formulated, there are several things one might want to do with it. First of all, if one of the variables represents a "response" you are interested in, you might want to find its distribution. The net may help you here: There are several algorithms to find the expectation or distribution of a variable in the net, using analytical computations, numerical computations, or simulation.

One of the most common changes in a stochastic model, is that previously unknown variables are observed, and may thus be substituted with a fixed value. After that has been done, one generally would like to find the effect on the whole model (i.e., find the posterior model). This updating may be done by algorithms connected to the Network formulation. Conceptually, the Network may be transformed so that the "data-node" becomes one which depends on no other variables. Then its distribution may simply be changed to a fixed value. This process may be called Bayesian inversion.

Using a Bayesian network, it is easy to include in the model the uncertainty in the parameters of your model: One may introduce underlying stochastic variables expressing this uncertainty. Of course, the parameters of the underlying stochastic variable may again be thought to be uncertain, and so one may need another underlying variable to specify this, and so on. The issue is sometimes raised that, as this process must be terminated at some point, some uncertainty is somehow lost.

It turns out, however, that if one specifies exactly what kind of answers one wants from the Bayesian net, the process above may be terminated after a specific number of steps without changing these answers. This proves that no uncertainty is "lost". Let us be more precise: Assume we have a Bayesian net, with a subset of nodes which we call "interesting", i.e., which have clear interpretations, and then a nubmer of underlying nodes which represent for example uncertainty in parameters. Let us assume we want to use the network to answer two types of questions: The marginal distribution of each of the "interesting" nodes, and the marginal distribution of "interesting" nodes, given new distributions for some other "interesting" nodes.

One may then show that it is only necessary to model uncertainty in parameterers to a certain depth. More precisely, all variables that do not appear in "minimal paths" between "interesting nodes", may in fact be removed from the network (by integrating out their effect). A "minimal path" is here a path in the network that cannot be made shorter by deforming it inside the network.

The above is of course a very incomplete introduction to Bayesian Networks. This field has been developing for a number of years, and a lot of litterature, and network implementations, exist. One reference is (Spiegelhalter, Dawid, , Lauritzen & Cowell 1993); other references may be found there. In the litterature, one may find algorithms for how Bayesian inversion of networks may be done. Most algorithms focus on first transforming the network to a simply connected graph by collecting together variables, and then inverting that graph one link at a time.

Instead of going deeper into this theory, we will consider below how Bayesian Nets may be applied in two concrete examples.

3.3 Example: Prospecting

To make the abstractions above more concrete, lets look at a specific stochastic model and formulate it as a Bayesian Network.

Assume we are in an early phase of exploration. We have identified a probable kitchen, two traps that are possible reservoirs, and possible migration paths. The situation may be illustrated by figure 3.2. Let us denote by K the binary stochastic variable specifying whether the kitchen has produced "adequate" amounts of petroleum. Based on experience, we may for example say that $P(K) = 0.5$. Now let $P1$ and $P2$ denote the variables for the migration paths to be OK, and let $S1$ and $S2$ denote the variables for the corresponding traps to be adequately sealing. Finally, let $R1$ and $R2$ be the variables signifying whether there is petroleum in the two traps. We may now say that $R1$ has a probability distribution dependent on K, $P1$ and $S1$. In the simplest of models, we may say that $R1$ is true if and only if K, $P1$ and $S1$ are true, i.e., $R1$ is *completely determined* if the three underlying variables are given. We could also say that only the probability for $R1$ is determined, given the underlying variables, to include some uncertainties not captured by our model. But for now, we stick to the simple model. Describing a similar relationship for $R2$, we get the pattern of dependencies given in figure 3.2.

Having set up a model like this, we should have some information about the prior probabilities for $P1$, $P2$, $S1$, and $S2$. For the sake of an example, let's say that $P(P1) = P(P2) = 0.3$ and $P(S1) = P(S2) = 0.8$. We may then trivially compute the probabilities for $R1$ and $R2$:

$$P(R1) = P(R2) = 0.5 \cdot 0.3 \cdot 0.8 = 0.12.$$

In more complex and realistic networks, it may not be so straightforward to compute the probabilities, or probability distributions, for the interesting variables. Several algorithms are available to do this, depending on the network. One may always use simulation, so that a value for each variable is simulated based on simulated values further down in the network.

More interesting than the computations above are the ones where we introduce new data about, say, $R2$. Assume we drill in $R2$, and petroleum is discovered there. Then we may compute the conditional probability of $R1$, given that $R2$ is either true or false:

$$P(R1 \mid R2) = 1.0 \cdot 0.3 \cdot 0.8 = 0.24$$

and

$$P(R1 \mid \overline{R2}) = 0.104.$$

We get the expected result, i.e., that the probability for $R1$ increases when $R2$ is true, and decreases when it is false. However, it is clear that our model so far is too simplistic. For example, the chance of $S1$ (the sealing of reservoir 1) being OK is probably not independent of $S2$ being OK. We could model this dependence by saying that there is an underlying parameter s, with a uniform

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

distribution $s \sim \mathrm{U}[0.6, 1]$ such that $P(S1 \mid s) = P(S2 \mid s) = s$. This parameter would express the probability of good sealing in this particular situation, and the information that there is a good seal in one trap could influence the probability for a seal in the other. Let us similarily say that there is an underlying parameter p expressing the dependence between $P1$ and $P2$: $P(P1 \mid p) = P(P2 \mid p) = p$, while $p \sim \mathrm{U}[0, 0.6]$. We then get the model shown at the top of Figure 3.3

Note that the expectation of all variables is the same as in the previous model, so that $P(R1) = P(R2) = 0.12$. However, the conditional probabilities are now different:

$$P(R1 \mid R2) = 0.327$$

and

$$P(R1 \mid \overline{R2}) = 0.092$$

Note that the probability for $R1$ now depends more on $R2$ than in the previous model. This is because we have specified more dependence between the two variables. In this way, we could stepwise refine the model to include more and more knowledge. At some point, we would become quite dependent on some assumed geological model, and so we might want to express that even the assumption of this model is uncertain. The two reservoirs might for example not have the same kitchen. One may handle the uncertainty around the choice of model by adding a variable M signifying whether the given model is correct. If it fails, we must resort to simpler methods to predict $R1$ and $R2$: We could let G be a variable describing the probability of finding petroleum in a borehole in the area, given no information about the subsurface geology. If we assume that the model fails, G could be our best guess for $R1$ and $R2$. We then get a network of variables as depicted at the bottom of Figure 3.3.

Let us, for the sake of computations, assume that $P(M) = 0.9$ and that $P(R1 \mid G) = P(R2 \mid G = G = 0.1$. From this, we compute that

$$P(R1) = P(R2) = 0.118$$

while

$$P(R1 \mid R2) = 0.307$$

and

$$P(R1 \mid \overline{R2}) = 0.106.$$

We see how the expectation for oil is a weighed mean between what it would be in the two different models. Similarily, the conditional probabilities are between the first and the second models, as we are now unsure about wether the wells are really dependent.

We have seen how we may stepwise refine our model, and how the model at each step may provide reasonable answers. Instead of going further with this example, we will look at another one, where the complexity of the model is more apparent.

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

3.4 Example: Oil/water contacts

We turn to an example involving computation of oil/water contacts based on saturation- and pressure-logs. The example is based on (Abrahamsen 1995), and uses that notation. From the oil/water contacts, the computations in the article go on to also use the data to find which compartments of the reservoir that are in communication. We will use the exposition and notation of the article.

Note that the model presented in (Abrahamsen 1995) falls within our framework in that one starts with a full specification of a prior model. The posterior model is then deduced, and the posterior distributions for some interesting quantities are numerically estimated. An important feature of the model is that all stochastic variables are either discrete or one-dimensional. Thus we may have some hope of fitting this model into the general framework we try to build.

Let us try to make diagrams of how the stochastic variables in the model are related. We look at the compartment named A1. Figure 3.4 shows part of how the variables are related: The distributions for a, b, $\lambda(=l)$ and $\sigma_{S_w}(=s_{sw})$ are given in (Abrahamsen 1995). z_{owc}^{A1} denotes the oil/water contact in A1. When a, b, λ, σ_{S_w} and z_{owc}^{A1} are given, the distribution for an observation S_{wi}^{A1} of water saturation at depth z_i in A1 is given by a normal distribution.

Figure 3.4 similarily shows the situation for the pressure measurements P_i^{A1}. Putting the two diagrams together, we have described the model in compartment A1. If we, for the moment, assume a prior distribution for z_{owc}^{A1} which is uniform in some interval, we have a fully specified model. Given the actual observations of pressures and water saturations, we may find the posterior model for z_{owc}^{A1}. In (Abrahamsen 1995), the posterior pdf for z_{owc}^{A1} is computed, up to a constant.

Our general goal is to make a program that can do the kind of computations done in this example automatically, after the prior has been specified. This is not easy, but as this application of Bayesian nets is clearly non-trivial, it demonstrates that such a general tool would be very useful.

3.5 Applied Bayesian networks

We have shown that problems in several petroleum applications of probability theory may fit into the framework of Bayesian nets, and non-trivial solutions to these problems may be found by using general tools on such networks.

Our goal should be to have a general tool that may:

- Help a user formulate problems like the above in terms of a network.
- Be general enough for use in many different areas.
- Give estimates for the probabilities or expected values for variables that are of interest.
- Update the model with new data.
- Update part of the model in terms of increased complexity, when that is necessary.
- Perform sensitivity analysis of the effect of aquireing certain data on the distributions of other variables.

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

- Do all of the above in a user friendly way.

We have in this project studied the possibilities for such a tool. We have made a simple inplementation of a Bayesian net to study Bayesian inversion algorithms, and above all, we have studied how easy it is to formulate and solve commonplace problems in terms of Bayesian nets. The conclusion so far must be that such reformulations can be done. It also seems that the necessary techniques for treating Bayesian nets are available. To evaluate further the potential for this tool, more realistic and larger case studies must be performed.

List of Figures

Bibliography

Abrahamsen, P. (1995), Probability density functions for hydrocarbon contacts, NR-note SAND/11/95, Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway.

Blockley, D. I. (1985), 'Fuzziness and probability: a discussion of gaines' axioms', *Civil Engineering Syst.* **2**, 195–200.

Cui, W. & Blockley, D. I. (1990), 'Interval probability theory for evidential support', *International Journal of Iintelligent Systems* **5**, 183–192.

Egeland, T., Omre, H., Lia, O., Tjelmeland, H., Holden, L. & Andersen, T. (1994), GRUS data analysis, NR-note SAND/17/1994, Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway.

Krause, P. & Clark, D. (1993), *Representing Uncertain Knowledge*, Kluwer Academic Publishers.

Lia, O., Omre, H., Tjelmeland, H. & Holden, L. (1995), The great reservoir uncertainty study - grus, *in* J. Olsen, S. Olaussen, T. Jensen, G. H. Landa & L. Hinderaker, eds, 'Profit, project summary reports', Norwegian Petroleum Directorate, pp. 191–206.

Lia, O., Omre, H., Tjelmeland, H., Holden, L., MacDonald, A., Hustad, O. S. & Qi, Y. (1994), GRUS — great reservoir uncertainty study. summary report, NR-note SAND/11/94, Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway.

Shafer, G. (1975), *A mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J.

Spiegelhalter, D., Dawid, A. P., , Lauritzen, S. L. & Cowell, R. G. (1993), 'Bayesian analysis in expert systems', *Statistical Science* **8**, 219–283.