Value of Information

Decision Analysis and the Value of Information

Jo Eidsvik

Jo.Eidsvik@ntnu.no

My background:

Professor of Statistics at NTNU in Trondheim, NORWAY.

Education:
- MSc in Applied Mathematics, Univ of Oslo
- PhD in Statistics, NTNU

Work experience:
- Norwegian Defense Research Establishment
- Statoil

Research interests:
- Spatial statistics, spatio-temporal statistics,
- Computational statistics, sampling methods, fast approximation techniques,
- Geoscience applications,
- Design of experiments,
- Decision analysis, value of information,

I like tennis, hiking, skiing, etc.

Plan for course

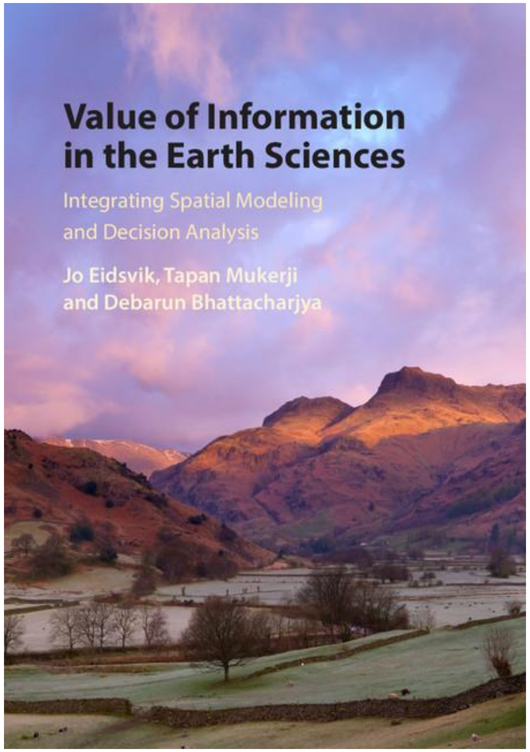
Time	**Topic**
Lecture 1	Introduction and motivating examples
	Elementary decision analysis and the value of information
	Multivariate statistical modeling, dependence, graphs
	Value of information analysis for dependent models
Lecture 2	Spatial statistics, spatial design of experiments
	Value of information analysis in spatial decision situations
	Examples of value of information analysis in Earth sciences
Lecture 3	Computational aspects
	Sequential decisions and sequential information gathering
	Examples from mining and oceanography

Every day: Small exercise s as we go, and computer project at the end.

Material:

Relevant background reading :

- Eidsvik, J., Mukerji, T. and Bhattacharjya, D., Value of information in the Earth sciences, Cambridge University Press, 2015.



- Howard R.A. and Abbas, A.E., Foundations of decision analysis, Pearson, 2015.
- Many spatial statistics books:
  - Cressie and Wikle (2011),
  Chiles and Delfiner (2012),
  Banerjee et al. (2014),
  Pyrcz and Deutsch (2014),
  etc.

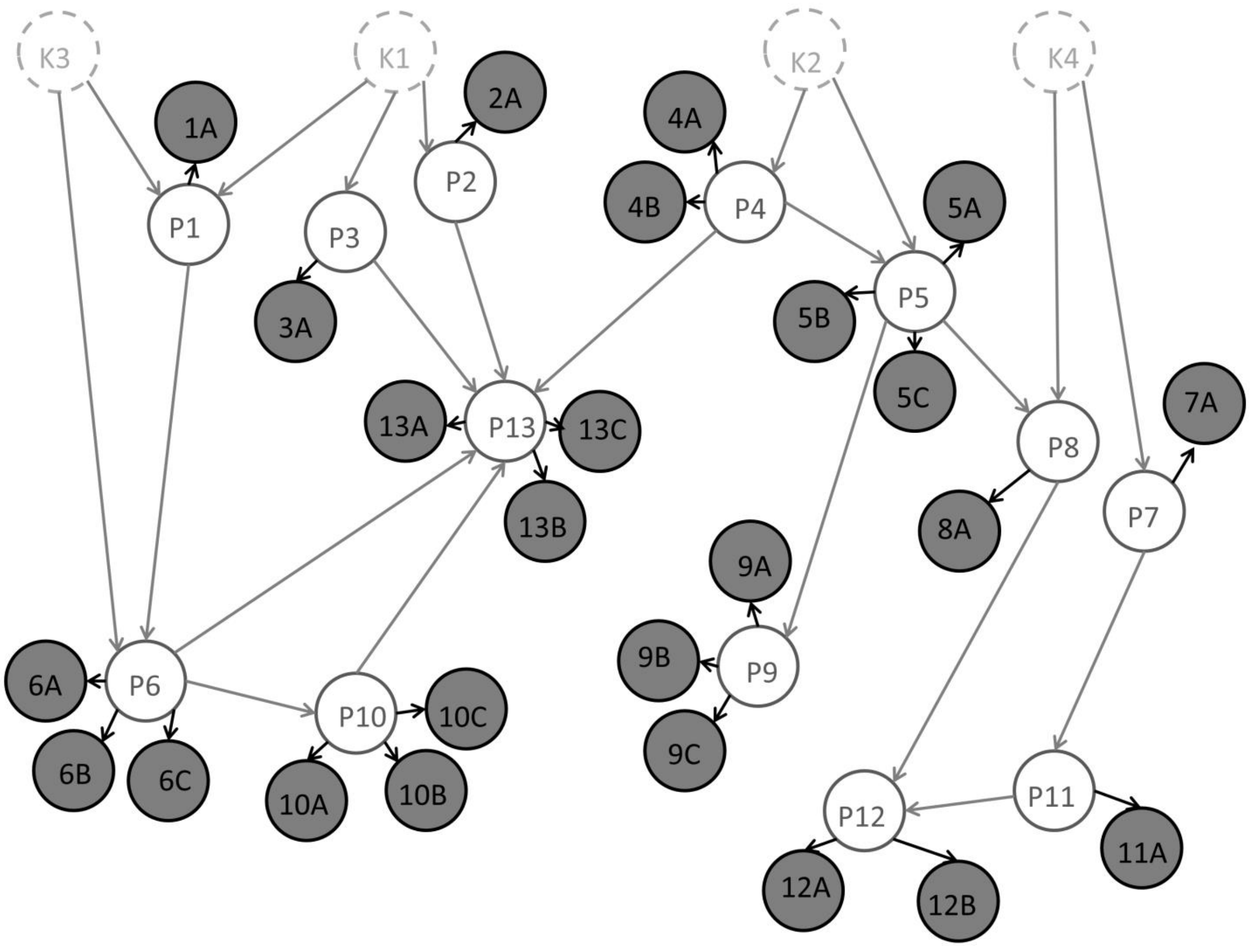
Motivating examples:

Integration of (spatial) modeling and decision analysis.

Collect data to resolve uncertainties and make informed decisions.

Motivation
(a petroleum exploration example)

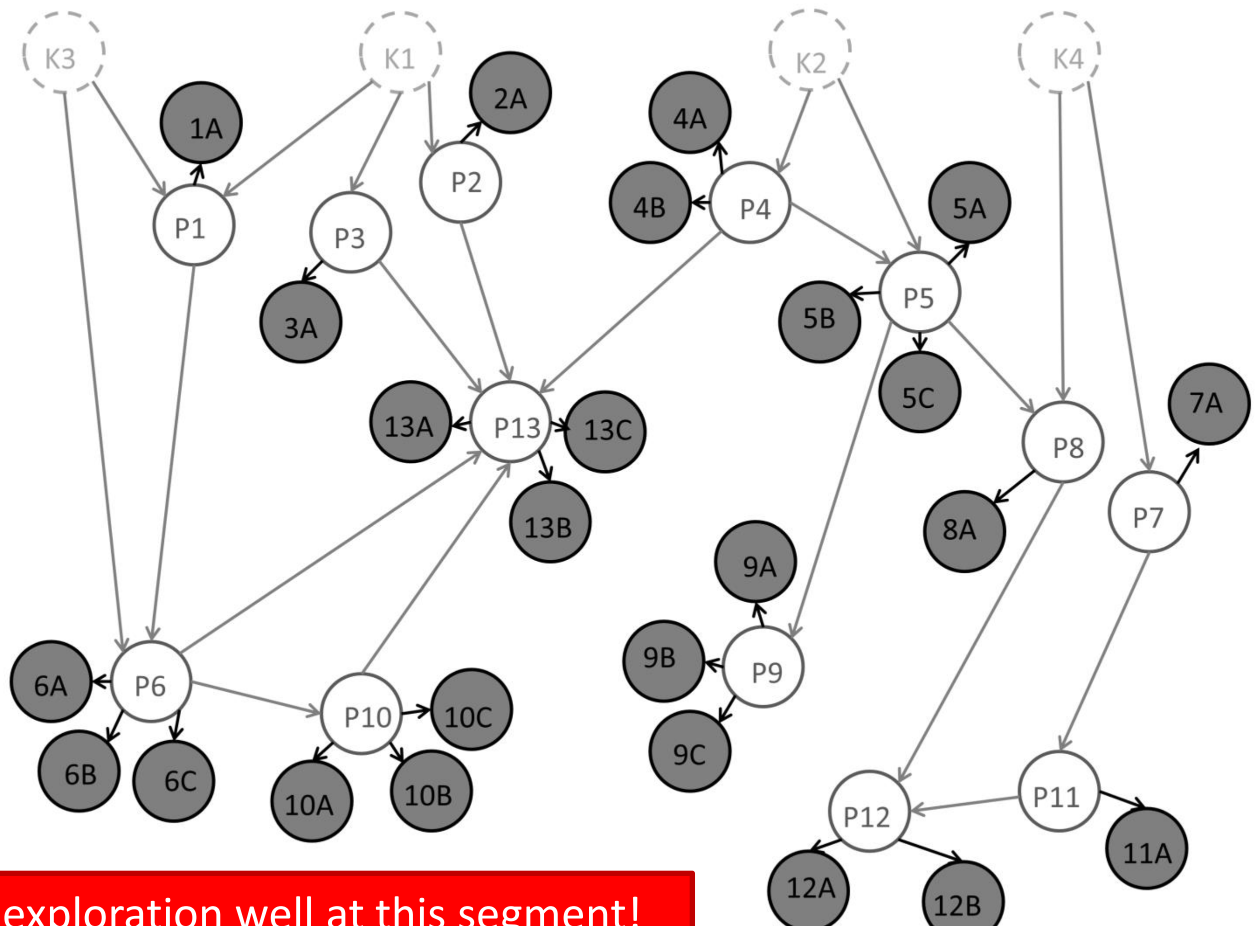
Gray nodes are petroleum reservoir segments where the company aims to develop profitable amounts of oil and gas.



Martinelli, G., Eidsvik, J., Hauge, R., and Førland, M.D., 2011, Bayesian networks for prospect analysis in the North Sea, *AAPG Bulletin*, 95, 1423-1442.

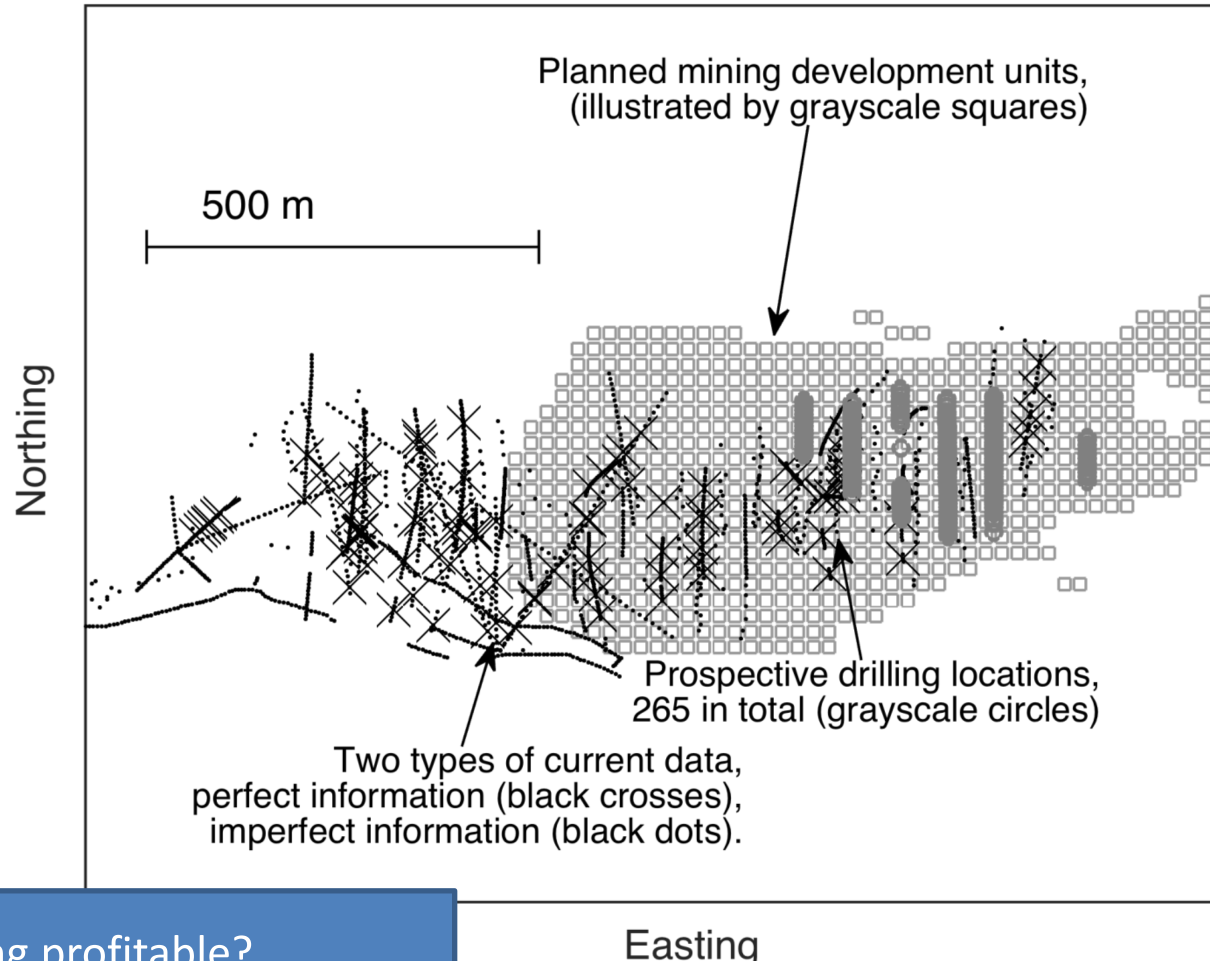
Motivation
(a petroleum exploration example)

Gray nodes are petroleum reservoir segments where the company aims to develop profitable amounts of oil and gas.



Drill the exploration well at this segment!
The value of information is largest.

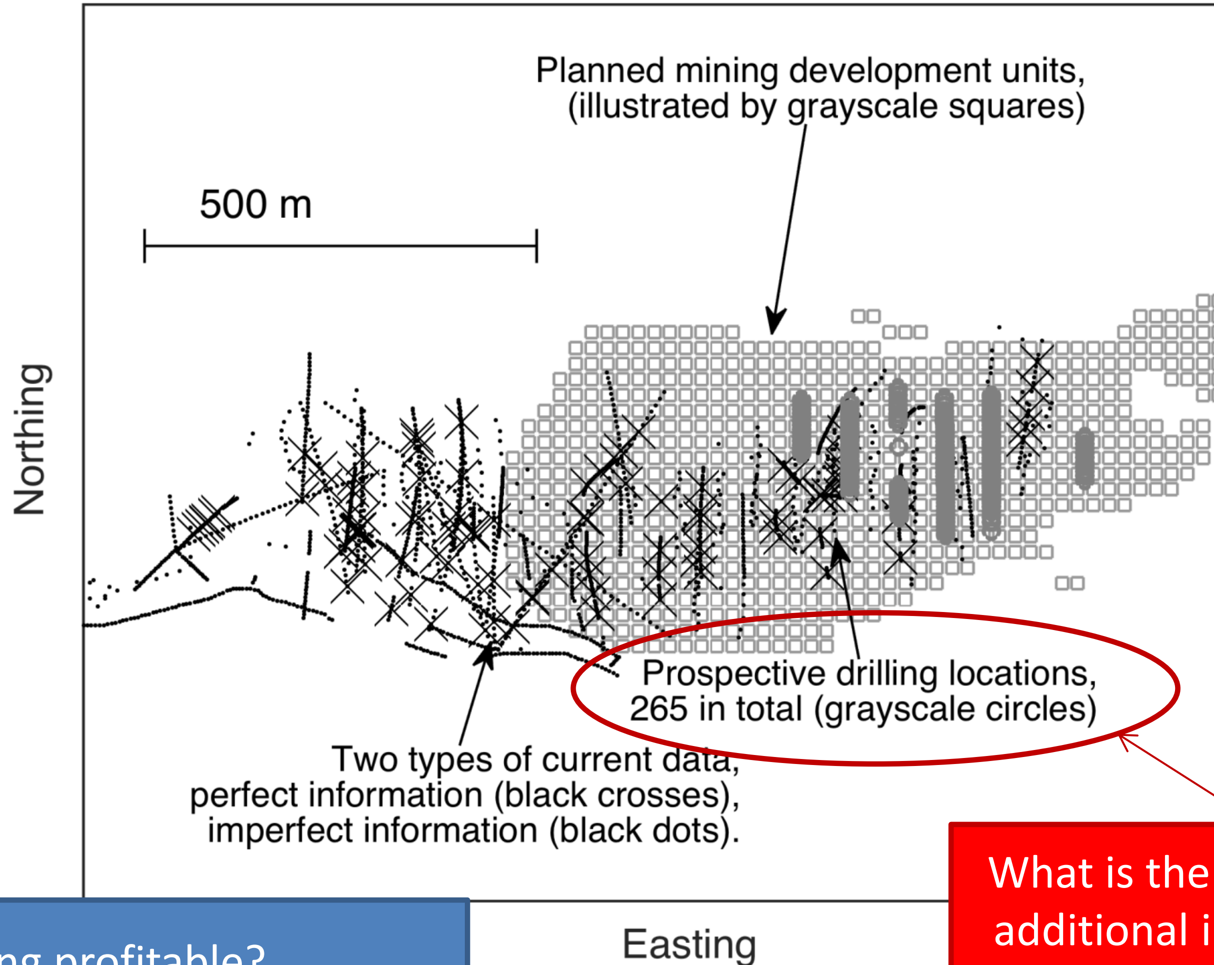
Motivation (an oxide mining example)



Is mining profitable?

Eidsvik, J. and Ellefmo, S.L., 2013, The value of information in mineral exploration within a multi-Gaussian framework, *Mathematical Geosciences*, 45, 777-798.

Motivation (an oxide mining example)

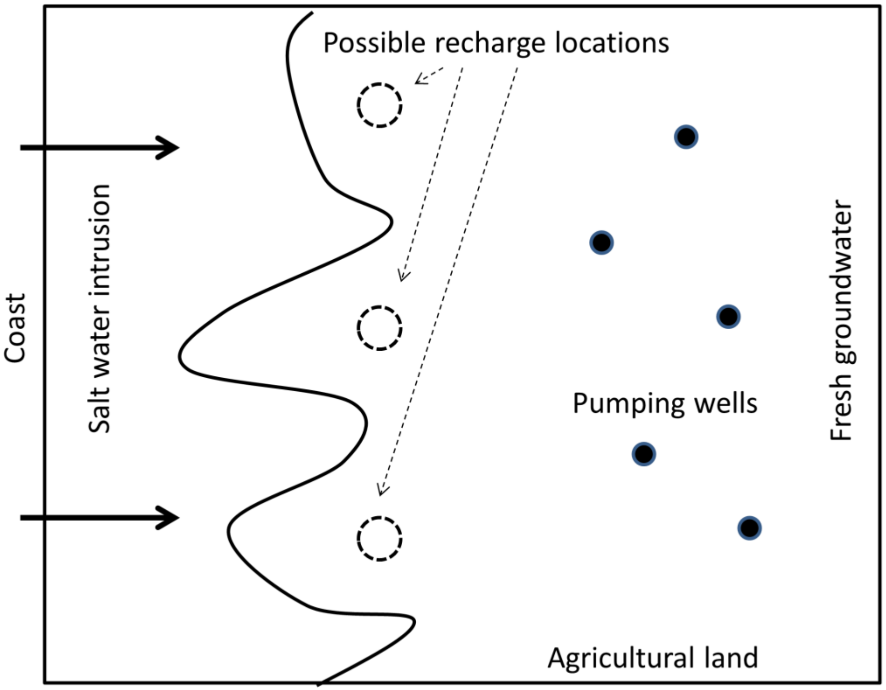


What is the value of this additional information?

Is mining profitable?

Motivation
(a groundwater example)

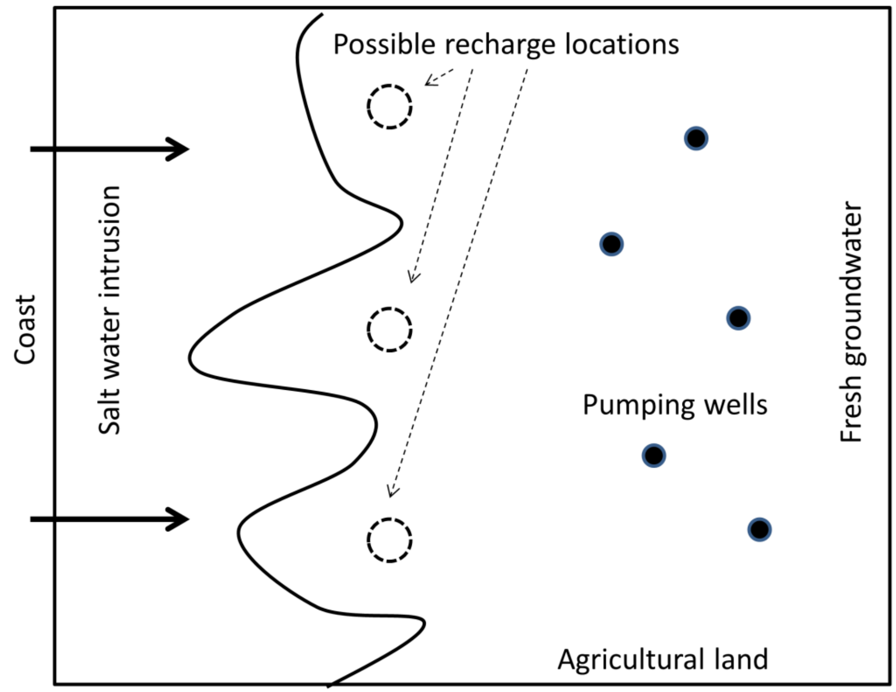
Which recharge location is better to prevent salt water intrusion?



Trainor-Guitton, W.J., Caers, J. and Mukerji, T., 2011, A methodology for establishing a data reliability measure for value of spatial information problems, *Mathematical Geosciences*, 43, 929-949.

Motivation
(a groundwater example)

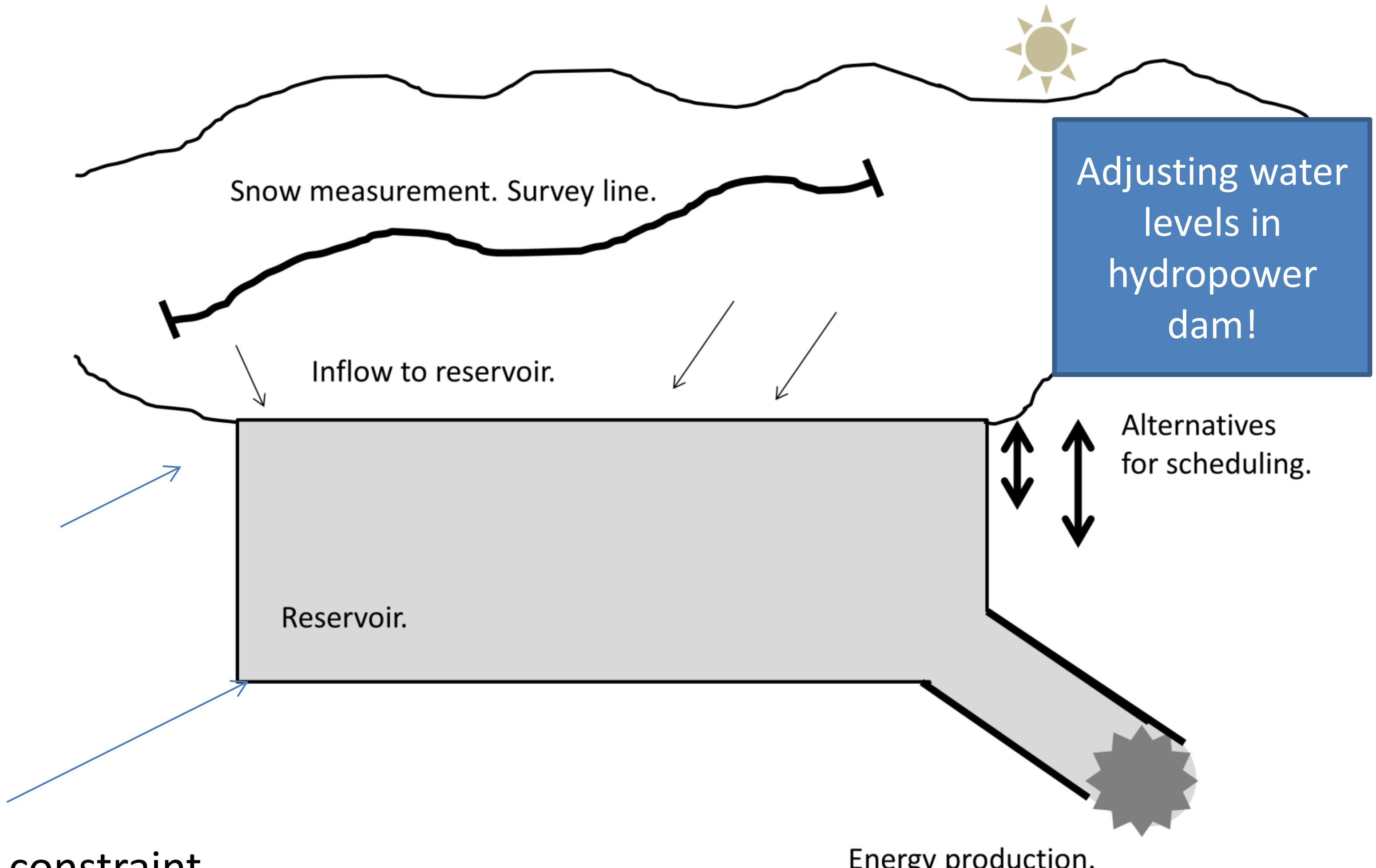
Which recharge location is better to prevent salt water intrusion?



Is it worthwhile to acquire electromagnetic data before making the decision about recharge?

Motivation
(a hydropower example)

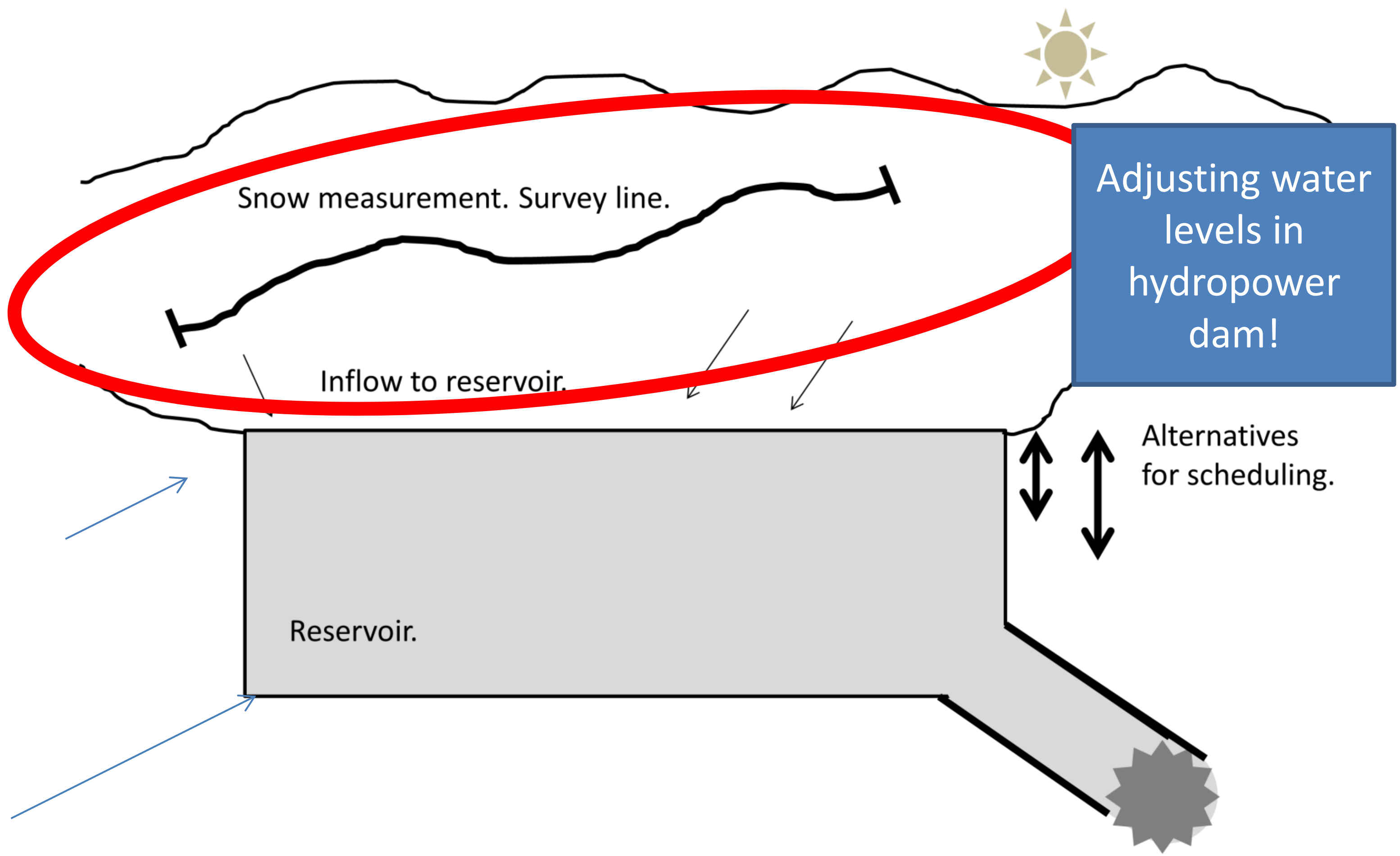
No money for water going over the dam.



Lower limit constraint.

Ødegård, H.L., Eidsvik, J. and Fleten, S.E., 2019, Value of information analysis of snow measurements for the scheduling of hydropower production, *Energy Systems*, 10, 1-19.

Motivation
(a hydropower example)



No money for water going over the dam.

Lower limit constraint.

Acquire snow measurements?

Other applications

- Farming and forestry – how to set up surveys for improved harvesting decisions.
- Biodiversity – where to monitor different biological variables for sustainability.
- Environmental – how monitor where pollutants are, to minimize risk or damage.

- Robotics - where should drone (UAV) or submarine (AUV) go to collect valuable data?
- Industry reliability – how to allocate sensors to ‘best’ monitor state of system?
- Internet of things – which sensors should be active now?

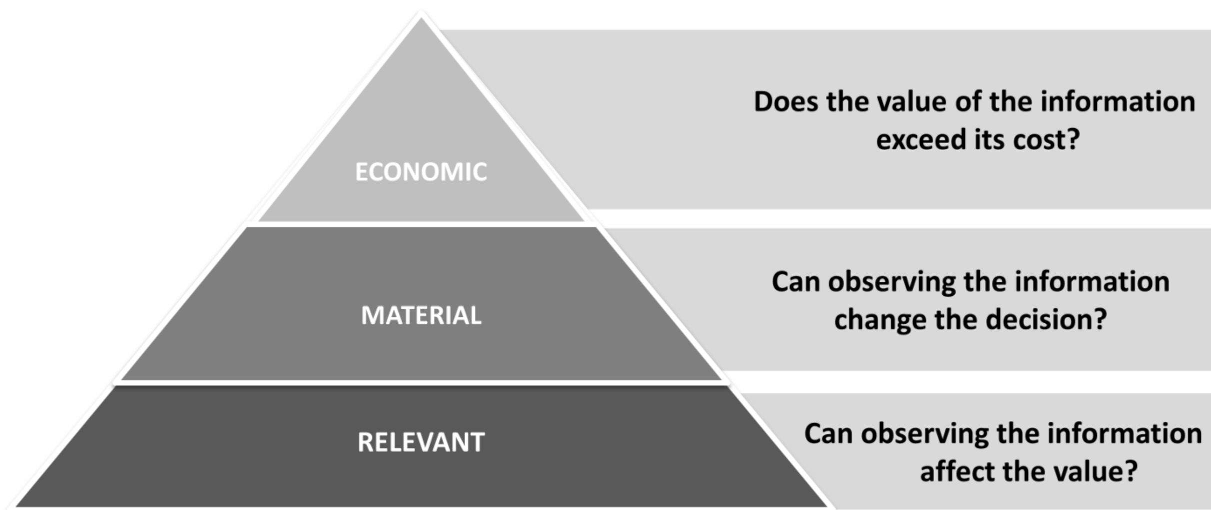
Which data are valuable?

Five Vs of big data:
- Volume
- Variety
- Velocity
- Veracity
- **Value**

We must acquire and process data that has value!
There is often a clear question that one aims to answer, and data should help us.

Value of information (VOI)

In many Earth science applications we consider purchasing more data before making difficult decisions under uncertainty. The value of information (VOI) is useful for quantifying the value of the data, before it is acquired and processed.



This pyramid of conditions - VOI is different from other information criteria (entropy, variance, prediction error, etc.)

Information gathering

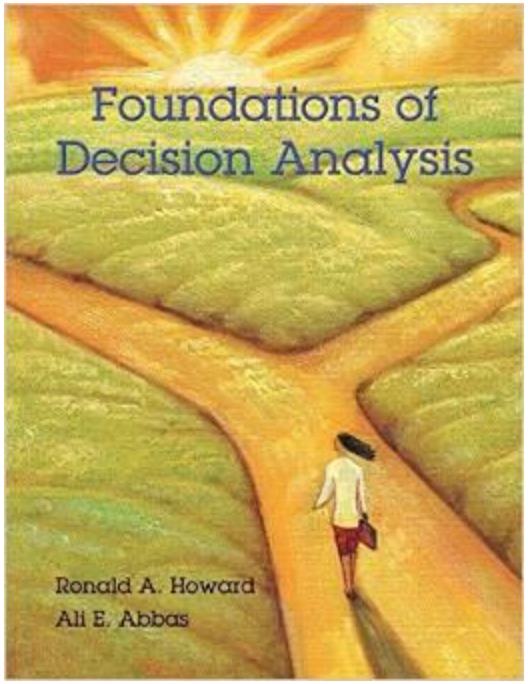
Why do we gather data?

To make better decisions!
To answer some kind of questions!
Reject or strengthen hypotheses!

We will use a **decision theoretic** perspective, but the methods are easily adapted to other criteria or value functions (2nd lecture).

Decision analysis (DA)

Decision analysis attempts to guide a decision maker to clarity of action in dealing with a situation where one or more decisions are to be made, typically in the face of uncertainty.



Howard, R.A. and Abbas, A., 2015, *Foundations of Decision Analysis*, Prentice Hall.

Framing a decision situation

Rules of actional thought. (Howard and Abbas, 2015)

- Frame your decision situation to address the decision makers true concerns.
- Base decisions on maximum expected utility.

'…systematic and repeated violations of these principles will result in inferior long-term consequences of actions and a diminishes quality of life…'

(Edwards et al., 2007, Advances in decision analysis: From foundations to applications, Cambridge University Press.)

Pirate example
(For motivating decision analysis and VOI)

Pirate example

- **Pirate example**: A pirate must decide whether to dig for a treasure, or not. The treasure is absent or present (uncertainty).

?

Pirate makes decision based on preferences and maximum **utility** or **value**!

- Digging cost.
- Revenues if he finds the treasure .

Pirate example

$a \in \{0,1\}$

- **Pirate example**: A pirate must decide whether to dig for a treasure, or not. The treasure is absent or present (uncertainty).

$x \in \{0,1\}$

?

Pirate makes decision based on preferences and maximum **utility** or **value**!

- Digging cost.
- Revenues if he finds the treasure .

$$\max_{a \in \{0,1\}} \{E(v(x,a))\}$$

Mathematics of decision situation:

- **Alternatives**

$$a \in \{0,1\} = A$$

- **Uncertainties (probability distribution)**

$$x \in \{0,1\} = \Omega \qquad p(x=1) = 0.01$$

- **Values**

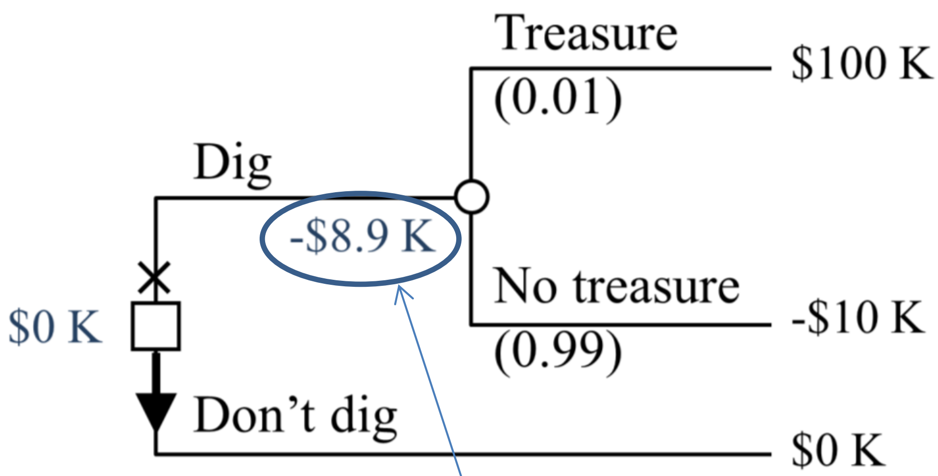
$$v = v(x,a)$$

$$v(x=0, a=1) = -10000 \qquad v(x=1, a=1) = 100000 \qquad v(x, a=0) = 0$$

- **Maximize expected value**

$$a^* = \arg\max_{a \in A} \left\{ E\left(v(x,a)\right) \right\}$$

Pirate's decision situation



Risk neutral!

$$E\left(u\left(v_{dig}\right)\right) = E\left(v_{dig}\right) = 0.01(100000) + 0.99(-10000) = -8900$$

Pirate example

- **Pirate example**: A pirate must decide whether to dig for a treasure, or not. The treasure is absent or present (uncertainty).

- Pirate can collect **data** before making the decision, if the experiment is worth its price!

- Perfect information. Clairvoyant!

- Imperfect information. Detector!

Value of information (VOI)

- VOI analysis is used to compare the additional value of making informed decisions with the price of the information.
- If the VOI exceeds the price, the decision maker should purchase the data.

VOI=Posterior value – Prior value

VOI – Pirate considers clairvoyant

$$PV = 0 = \$0K$$

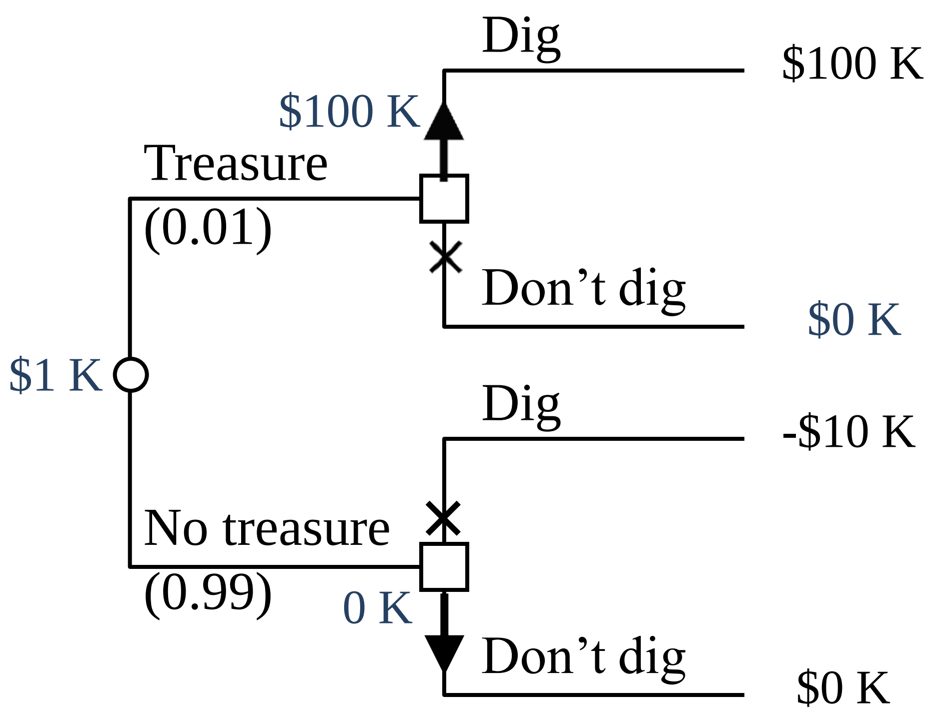
$$PoV(x) = \sum_x \max_{a \in A} \{v(x,a)\} p(x)$$

$$= \left(0.01 \cdot \max\{0,100\}\right) + \left(0.99 \cdot \max\{0,-10\}\right) = \$1K$$

$$VoI(x) = PoV(x) - PV = 1 - 0 = \$1K$$

Conclusion: Consult clairvoyant if (s)he charges less than $1000.

PoV – decision tree, perfect information



Pirate example - detector

- **Pirate example**: A pirate must decide whether to dig for a treasure, or not. The treasure is absent or present (uncertainty).

- Pirate can collect **data** before making the decision, if the experiment is worth its price!

Pirate makes decision based on preferences and maximum expected **value**!
- Digging cost.
- Revenues if he finds the treasure .

Pirate example - detector

$a \in \{0,1\}$

- **Pirate example**: A pirate must decide whether to dig for a treasure, or not. The treasure is absent or present (uncertainty).

- Pirate can collect **data** with a detector before making the decision, if this experiment is worth its price!

$x \in \{0,1\}$

$y \in \{0,1\}$

Pirate makes decision based on preferences and maximum expected **value**!
- Digging cost.
- Revenues if he finds the treasure .

$$\max_{a\in\{0,1\}} \{E(v(x,a) \mid y)\}$$

Detector experiment

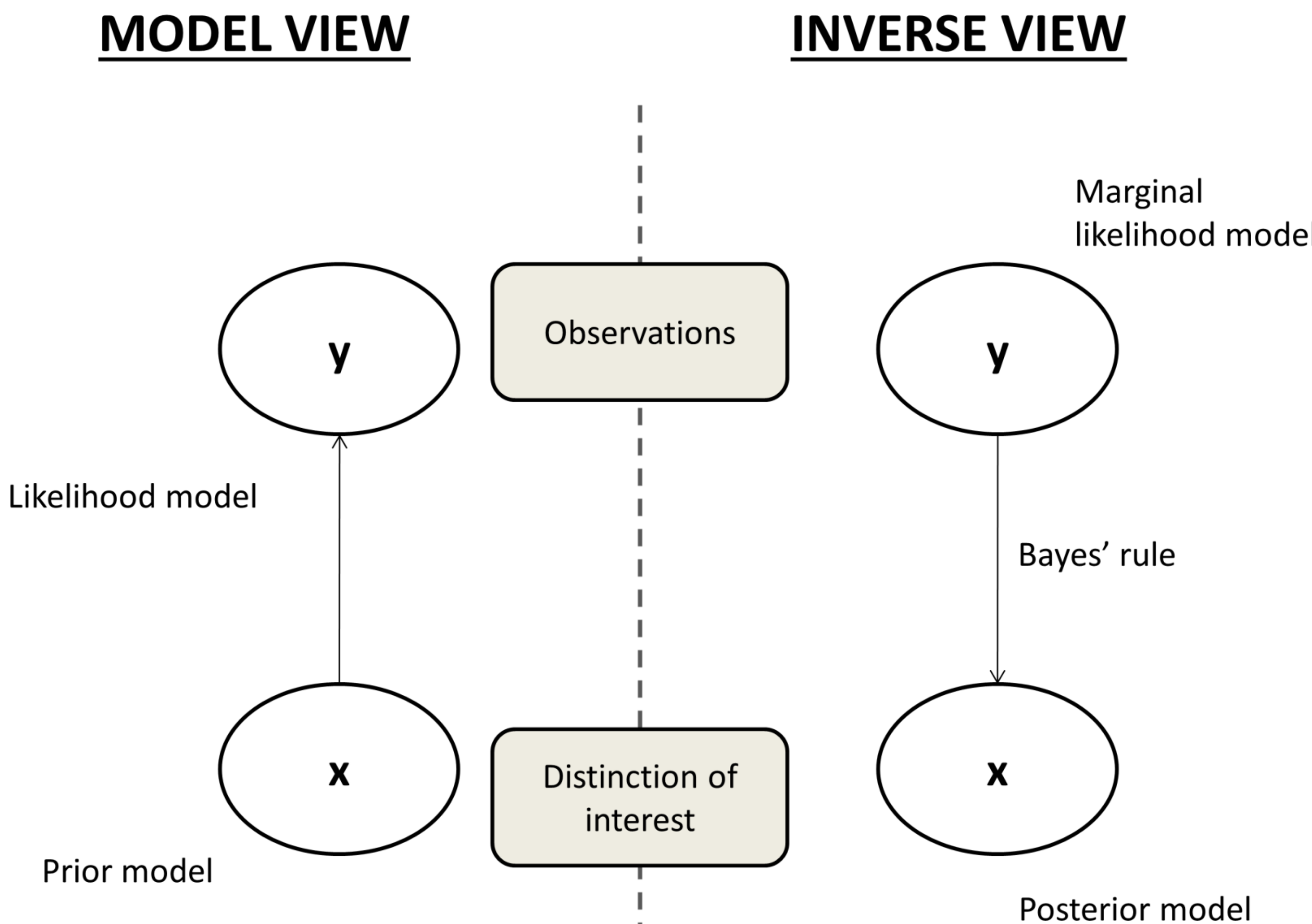
Accuracy of test:

$$p(y=0\,|\,x=0) = p(y=1\,|\,x=1) = 0.95$$

Should the pirate pay to do a detector experiment?

Does the VOI of this experiment exceed the price of the test?

Bayes rule - Detector experiment



Bayes rule - Detector experiment

Likelihood:

$$p(y=0\,|\,x=0)=p(y=1\,|\,x=1)=0.95$$

Marginal likelihood:

$$p\left(y=1\right)=p\left(y=1\,|\,x=0\right)p\left(x=0\right)+p\left(y=1\,|\,x=1\right)p\left(x=1\right)$$
$$=0.05\cdot 0.99+0.95\cdot 0.01=0.06$$

Posterior:

$$p(x=1\,|\,y=1)=\frac{p(y=1\,|\,x=1)\,p(x=1)}{p(y=1)}=\frac{0.95\cdot 0.01}{0.06}\approx 0.16=16/100.$$

$$p(x=1\,|\,y=0)=\frac{p(y=0\,|\,x=1)\,p(x=1)}{p(y=0)}=\frac{0.05\cdot 0.01}{0.94}\approx 0.0005=5/10000.$$

VOI – Pirate considers detector test

$$PoV(y) = \sum_{y} \max_{a \in A} \left\{ E\left(v\left(x, a\right) \mid y \right) \right\} p(y)$$

$$= \left(0.06 \cdot \max\left\{ 0, \left(100 \cdot 0.16\right) + \left(-10 \cdot 0.84\right) \right\} \right)$$

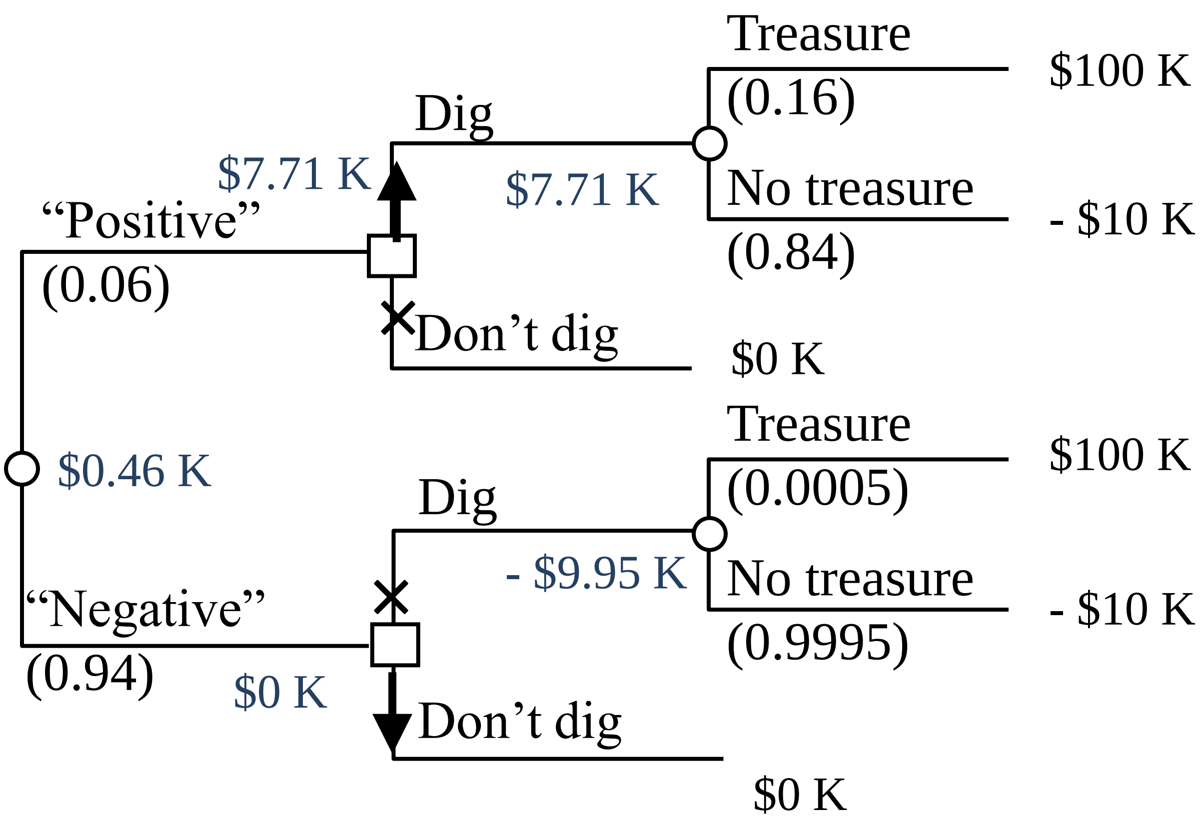
$$+ \left(0.94 \cdot \max\left\{ 0, \left(100 \cdot 0.0005\right) + \left(-10 \cdot 0.9995\right) \right\} \right)$$

$$= \left(0.06 \cdot \max\left\{ 0, 7.71 \right\} \right) + \left(0.94 \cdot \max\left\{ 0, -9.95 \right\} \right) = \$0.46K.$$

$$VoI(y) = PoV(y) - PV = 0.46 - 0 = \$0.46K$$

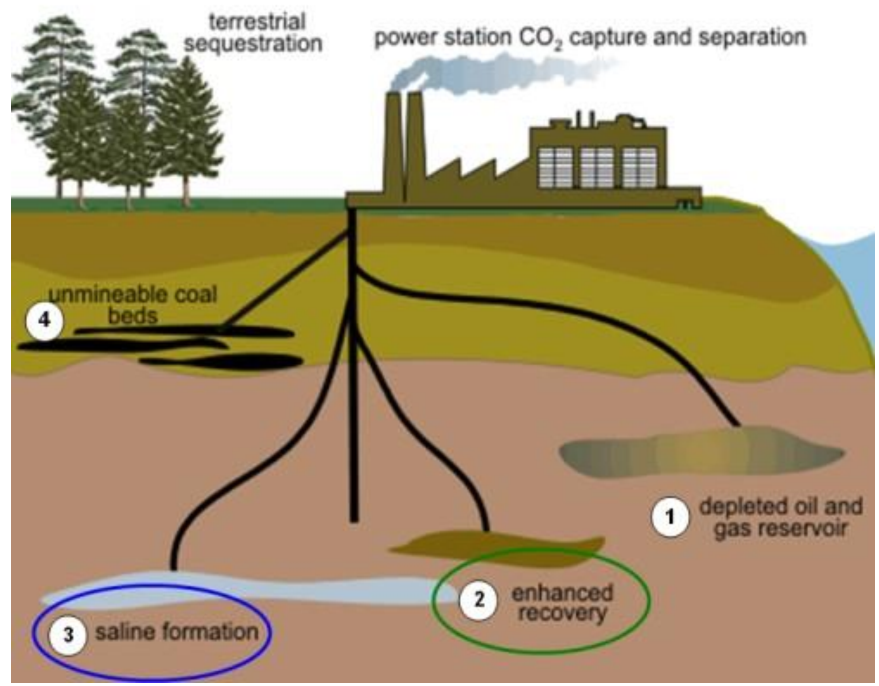
Conclusion: Purchase detector testing if its price is less than $460.

PoV - imperfect information



Exercise: CO2 sequestration

CO2 is sequestered to reduce carbon emission in the athmosphere and defer global warming.
Geological sequestration involves pumping CO2 in subsurface layers, where it will remain, unless it leaks to the surface.



VOI for CO2 sequestration

The decision maker can proceed with CO2 injection or suspend sequestration. The latter incurs a tax of 80 monetary units. The former only has a cost of injection equal to 30 monetary units, but the injected CO2 may leak (x=1). If leakage occurs, there will be a fine of 60 monetary units (i.e. a cost of 90 in total). Decision maker is risk neutral.

$$p(x=1)=0.3 \qquad p(x=0)=0.7$$

Data: Geophysical experiment, with binary outcome, indicating whether the formation is leaking or not.

$$p(y=0\,|\,x=0)=0.95 \qquad p(y=1\,|\,x=1)=0.9$$

Exercise:

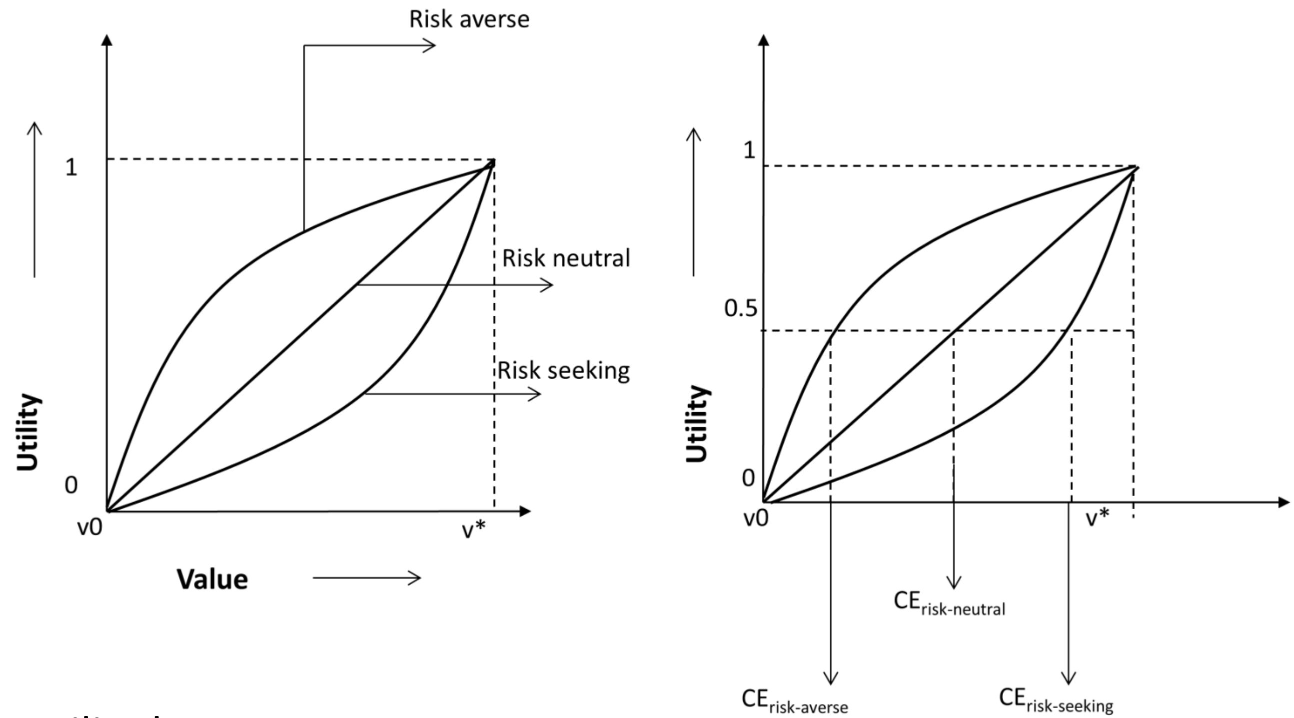
1. **Draw the decision tree without information.**
2. **Draw the decision tree with perfect information (clairvoyance).**
3. **Compute the VOI of perfect information.**
4. **Draw the decision tree with the geophysical experiment.**
5. **Compute conditional probabilities, expected values and the VOI of geophysical data.**

Value of information (VOI)
- More general formulation

- VOI analysis is used to compare the additional value of making informed decisions with the price of the information.
- If the VOI exceeds the price, the decision maker should purchase the data.

VOI=Posterior value – Prior value

Risk and utility functions

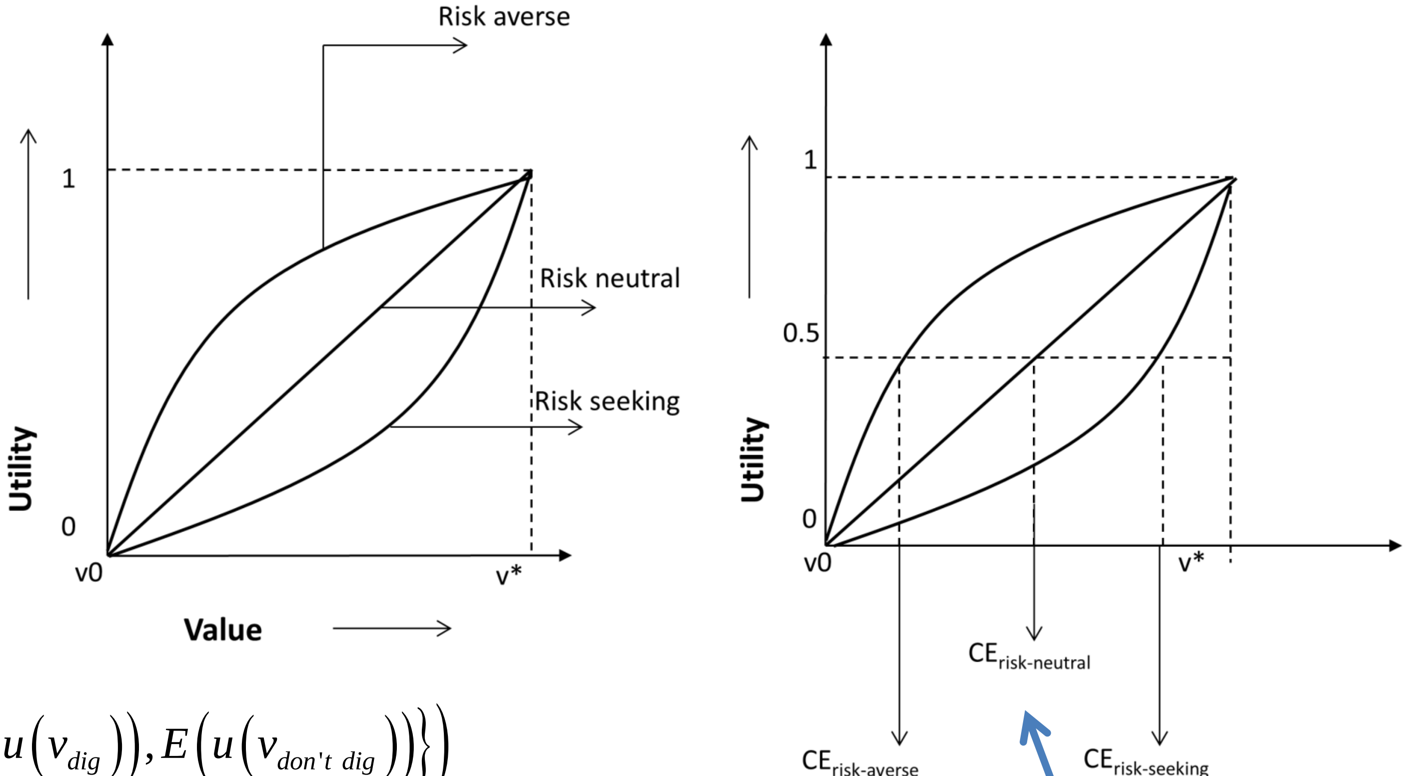


Exponential and linear utility have constant risk aversion coefficient:

$$\gamma = -{}^{u''(v)}\Big/_{u'(v)}$$

Certain equivalents (CE)

Utilities are mathematical. The certain equivalent is a measure of how much a situation is worth to the decision maker. (It is measured in value).



$$CE = u^{-1}\left(\max\left\{E\left(u\left(v_{dig}\right)\right), E\left(u\left(v_{don't\ dig}\right)\right)\right\}\right)$$

What is the value of indifference? How much would the owner of a lottery be willing to sell it for?

VOI - Clairvoyance

Price *P* of experiment makes the equality.

$$\sum_x \max_{a\in A}\left\{v\left(x,a\right)-P\right\} p\left(x\right)=\max_{a\in A}\left\{E\left(v\left(x,a\right)\right)\right\}$$

$$\rightarrow P=VOI=\sum_x \max_{a\in A}\left\{v\left(x,a\right)\right\} p\left(x\right)-\max_{a\in A}\left\{E\left(v\left(x,a\right)\right)\right\}$$

VOI=Posterior value – Prior value

Assuming risk neutral decision maker!

Value of information- Imperfect

Price of indifference.

$$\sum_y \max_{a\in A}\left\{E\left(v\left(x,a\right)-P \mid y\right)\right\} p\left(y\right)=\max_{a\in A}\left\{E\left(v\left(x,a\right)\right)\right\}$$

$$\sum_y \max_{a\in A}\left\{E\left(v\left(x,a\right)-P \mid y\right)\right\} p\left(y\right)=\max_{a\in A}\left\{E\left(v\left(x,a\right)\right)\right\}$$

$$\rightarrow P = VOI = \sum_y \max_{a\in A}\left\{E\left(v\left(x,a\right) \mid y\right)\right\} p\left(y\right)-\max_{a\in A}\left\{E\left(v\left(x,a\right)\right)\right\}$$

VOI=Posterior value – Prior value

Assuming risk neutral decision maker!

Properties of VOI

a) VOI is always positive

- Data allow better, informed decisions.

$$\max\left\{0, \sum_i v_i\right\} \le \sum_i \max\{0, v_i\}$$

b) If value is in monetary units ,VOI is in monetary units.

c) Data should be purchased if VOI > Price of experiment P.

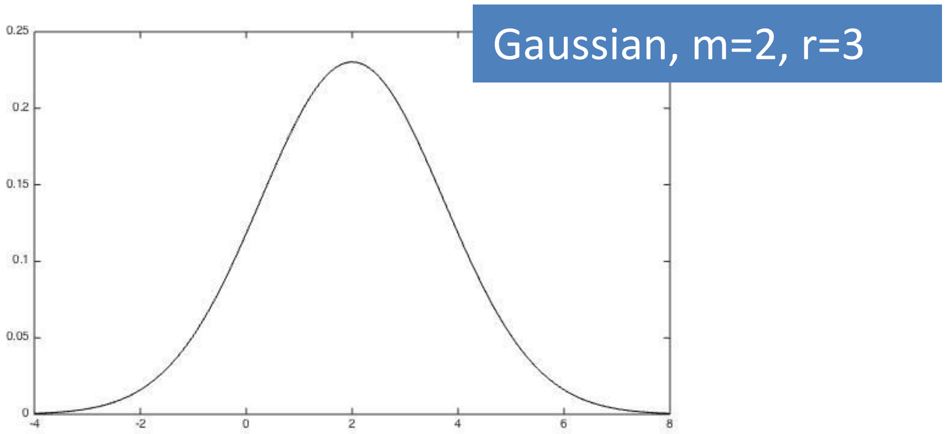
d) VOI of clairvoyance is an upper bound for any imperfect information gathering scheme.

e) When we compare different experiments, we purchase the one with largest VOI compared with the price:

$$\arg\max\{VOI_1 - P_1, VOI_2 - P_2\}$$

Gaussian model for profits

$$p(x) = \frac{1}{\sqrt{2\pi r^2}} \exp\left(-\frac{(x-m)^2}{2r^2} \right)$$



Uncertain profit of a project is Gaussian distributed.

VOI for Gaussian

Uncertain project profit is Gaussian distributed.
Invest or not?
The decision maker asks a clairvoyant for perfect information, if the VOI is larger than her price.

$$VOI(x) = \mathrm{PosteriorValue}(x) - \mathrm{PriorValue}$$

$$PV = \max\{0, E(x)\}, \qquad E(x) = m$$

$$PoV(x) = E(\max\{0, x\}) = \int \max\{0, x\}\, p(x)\, dx$$

VOI for Gaussian

Result:

$$E\left(\max\{0,x\}\right)=\int \max\{0,x\}\,p(x)\,dx=\int_0^\infty xp(x)\,dx=\int_{-m/r}^{\infty}(m+rz)\phi(z)\,dz$$

$$=m\int_{-m/r}^{\infty}\phi(z)\,dz+r\int_{-m/r}^{\infty}z\phi(z)\,dz=m\left(1-\Phi\left(-m/r\right)\right)+r\phi\left(-m/r\right)$$

$$=m\Phi\left(m/r\right)+r\phi\left(m/r\right),$$

VOI for Gaussian

Gaussian cdf

Gaussian pdf

Result:

$$VOI(x) = m\Phi\left(\frac{m}{r}\right) + r\phi\left(\frac{m}{r}\right) - \max\{0, m\}$$

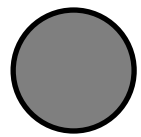
The analytical form facilitates computing, and eases the study of VOI properties as a function of the parameters.

$$m = 0,$$

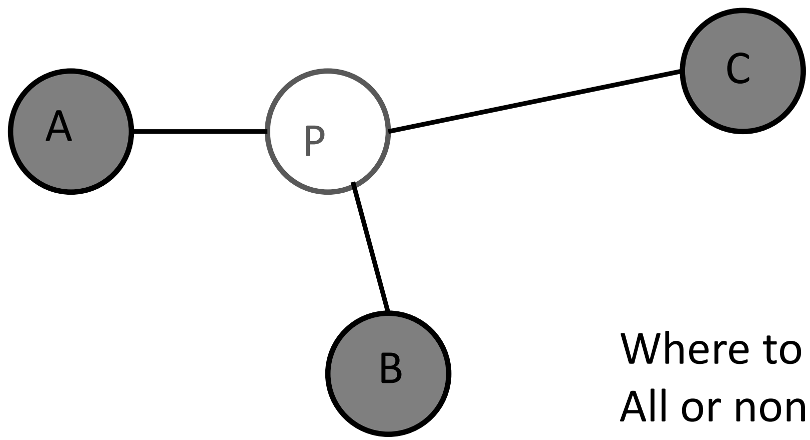
$$VOI(x) = r\phi(0) = \frac{r}{\sqrt{2\pi}}$$

More uncertain, more valuable information.

What if several projects or treasures?



What if several projects or treasures?



Where to invest?
All or none? Free to choose as many as profitable? One at a time, then choose again?

Where should one collect data? All or none? One only? Or two? One first, then maybe another?

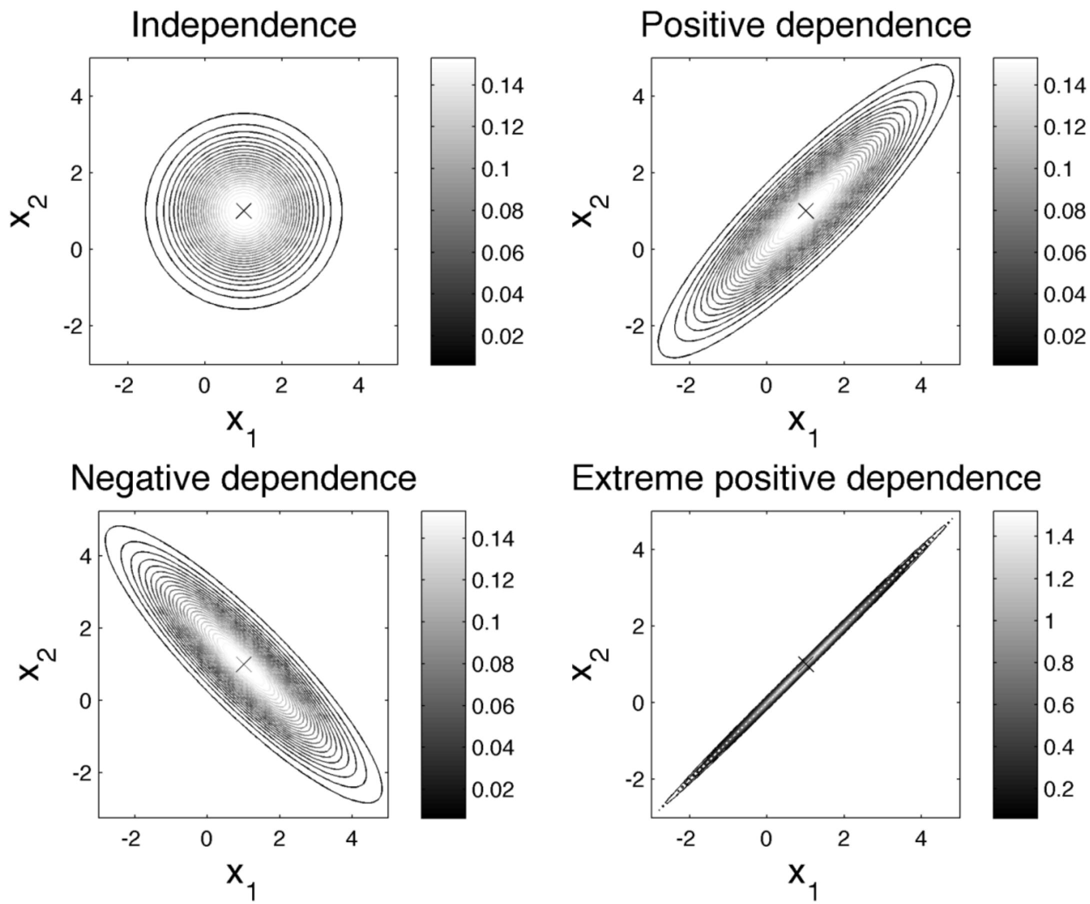
VOI and Earth sciences

- **Alternatives are spatial**, often with high flexibiliy in selection of sites, control rates, intervention, excavation opportunities, harvesting, etc.
- **Uncertainties are spatial**, with multi-variable interactions . Often both discrete and continuous.
- **Value function is spatial**, typically involving coupled features, say through differential equations. It can be defined by «physics» as well as economic attributes.
- **Data are spatial**. There are plenty opportunities for partial, total testing and a variety of tests (surveys, monitoring sensors, electromagnetic data, , etc.)

Two-project example

Two correlated projects with uncertain profits.

Decision maker considers investing in project(s).



Gaussian projects example

- **Alternatives**
  - Do not invest in project 1 (a1=0) - Invest in project 1 (a1=1)
  - Do not invest in project 2 (a2=0) - Invest in project 1 (a2=1)
  - Decision maker is free to select both, if profitable: Four sets of alternatives.
- **Uncertainty** (random variable)
  - Profits are *bivariate Gaussian*.
    Assume mean 0, variance 1 and fixed correlation.
- **Value** decouples to sum of profits, if positive.
- **Information gathering**
  - Report can be written about one project (assume perfect).
  - Report can be written about both projects (assume imperfect).

Gaussian projects example

$$\boldsymbol{x} = \left(x_1, x_2\right)$$

Prior model for profits: $p\left(\boldsymbol{x}\right) = N\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right),$ $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$

$$PV = \sum_{i=1}^{2} \max\left\{0, E\left(x_i\right)\right\} = 0 + 0 = 0$$

$$PoV\left(\boldsymbol{y}\right) = \sum_{i=1}^{2} \int \max\left\{0, E\left(x_i \mid \boldsymbol{y}\right)\right\} p\left(\boldsymbol{y}\right) d\boldsymbol{y}$$

$$\mathrm{VOI}\left(\boldsymbol{y}\right) = PoV\left(\boldsymbol{y}\right) - PV$$

Gaussian projects example

$$PV = \sum_{i=1}^{2} \max\left\{0, E\left(x_i\right)\right\} = 0 + 0 = 0$$

Must solve the integral expression!

$$PoV\left(\boldsymbol{y}\right) = \sum_{i=1}^{2} \int \max\left\{0, E\left(x_i \mid \boldsymbol{y}\right)\right\} p\left(\boldsymbol{y}\right) d\boldsymbol{y}$$

$$\mathrm{VOI}\left(\boldsymbol{y}\right) = PoV\left(\boldsymbol{y}\right) - PV$$

Need marginal for data!

Need conditonal expectation!

Perfect information about 1 project

$$y = x_1$$
$$p\left(x_1\right) = N\left(0,1\right)$$
$$E\left(x_1 \mid x_1\right) = x_1$$
$$E\left(x_2 \mid x_1\right) = \rho x_1$$

Get information about second project because of correlation!

$$Var\left(x_1 \mid x_1\right) = 0, Var\left(x_2 \mid x_1\right) = 1 - \rho^2$$

$$PoV\left(x_1\right) = \int_0^\infty x_1 p\left(x_1\right) dx_1 + \int_0^\infty \left|\rho\right| x_1 p\left(x_1\right) dx_1$$

$$= \left(1 + \left|\rho\right|\right) \Big/ \sqrt{2\pi}$$

Imperfect information, both projects

$$\boldsymbol{y} = \boldsymbol{x} + N\left(\boldsymbol{0}, \tau^2 \boldsymbol{I}\right)$$

$$p\left(\boldsymbol{y}\right) = N\left(\boldsymbol{0}, \tau^2 \boldsymbol{I} + \boldsymbol{\Sigma}\right) = N\left(\boldsymbol{0}, \boldsymbol{C}\right)$$

$$E\left(\boldsymbol{x} \mid \boldsymbol{y}\right) = \boldsymbol{\Sigma}\boldsymbol{C}^{-1}\boldsymbol{y}$$

Reduction in variances large, VOI is large.

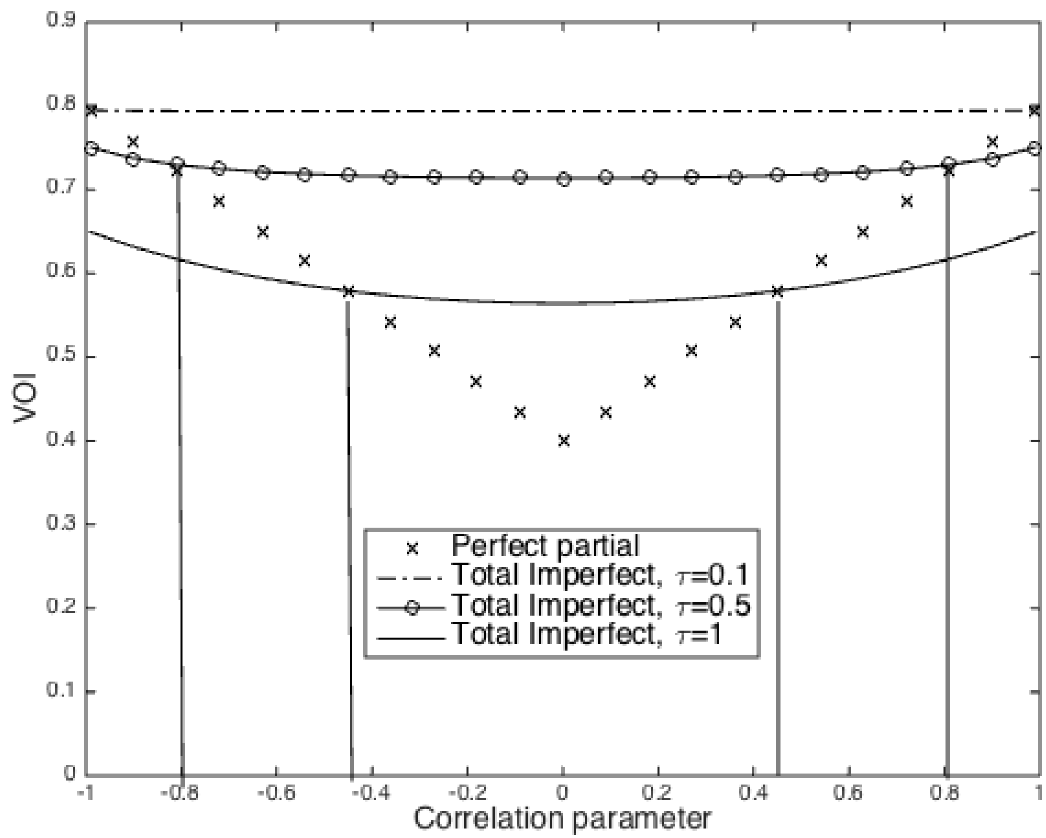
$$Var\left(\boldsymbol{x} \mid \boldsymbol{y}\right) = \boldsymbol{\Sigma} - \boldsymbol{R}, \qquad \boldsymbol{R} = \boldsymbol{\Sigma}\boldsymbol{C}^{-1}\boldsymbol{\Sigma}$$

$$PoV\left(\boldsymbol{y}\right) = \sum_{i=1}^{2} \int \max\left\{0, E\left(x_i \mid \boldsymbol{y}\right)\right\} p\left(\boldsymbol{y}\right) d\boldsymbol{y} = \left(\sqrt{R_{1,1}} + \sqrt{R_{2,2}}\right) \Big/ \sqrt{2\pi}$$

Gaussian projects results

$$PoV(\boldsymbol{y}) = \left(\sqrt{R_{1,1}} + \sqrt{R_{2,2}}\right) \Big/ \sqrt{2\pi}, \qquad \boldsymbol{R} = \boldsymbol{\Sigma C}^{-1}\boldsymbol{\Sigma}$$

$$PoV(x_1) = \left(1 + |\rho|\right) \Big/ \sqrt{2\pi}$$

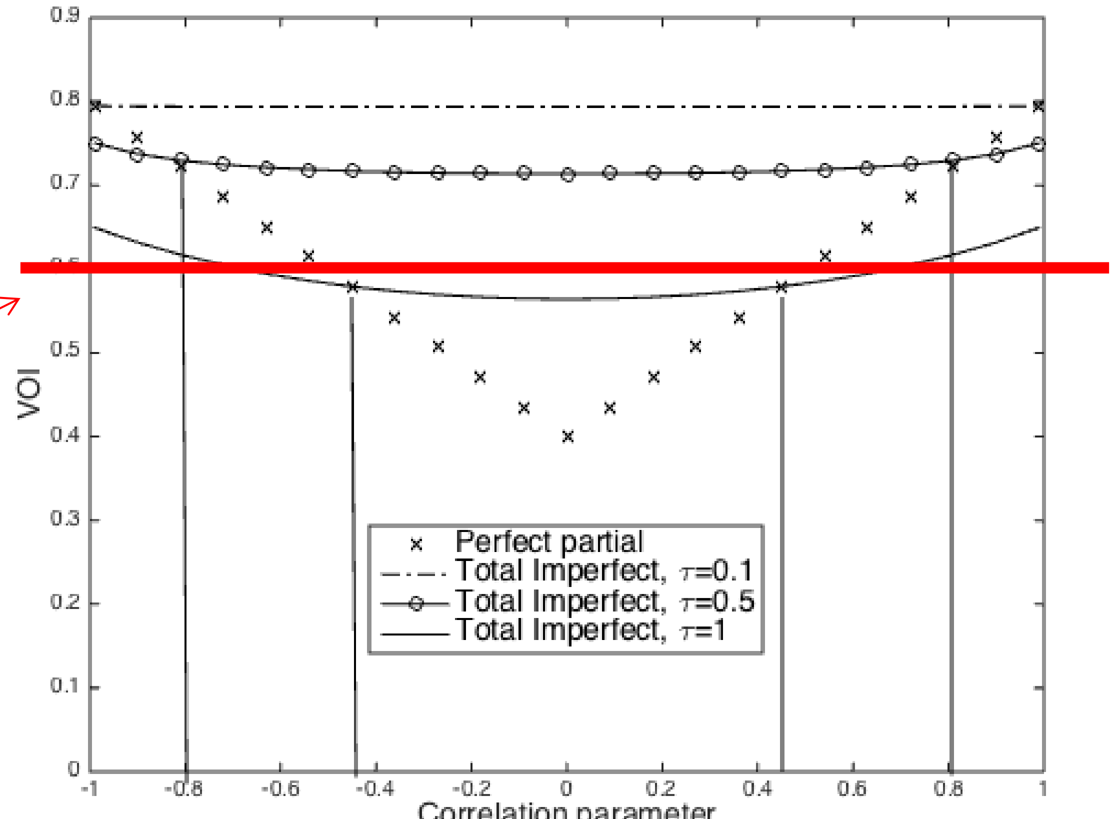


Gaussian projects results

$$PoV(\boldsymbol{y}) = \left(\sqrt{R_{1,1}} + \sqrt{R_{2,2}}\right) \Big/ \sqrt{2\pi}, \qquad \boldsymbol{R} = \boldsymbol{\Sigma} \boldsymbol{C}^{-1} \boldsymbol{\Sigma}$$

$$PoV(x_1) = \left(1 + |\rho|\right) \Big/ \sqrt{2\pi}$$

Price of test(s).



Insight from Gaussian projects

Dependence matters – the more correlation, the larger VOI.
The relative increase is very clear for partial information. It is also larger when there is more measurement noise. (With perfect total information, dependence does not matter.)

Decision maker must compare the VOI with the price of information, and purchase the data if the VOI exceeds the price.

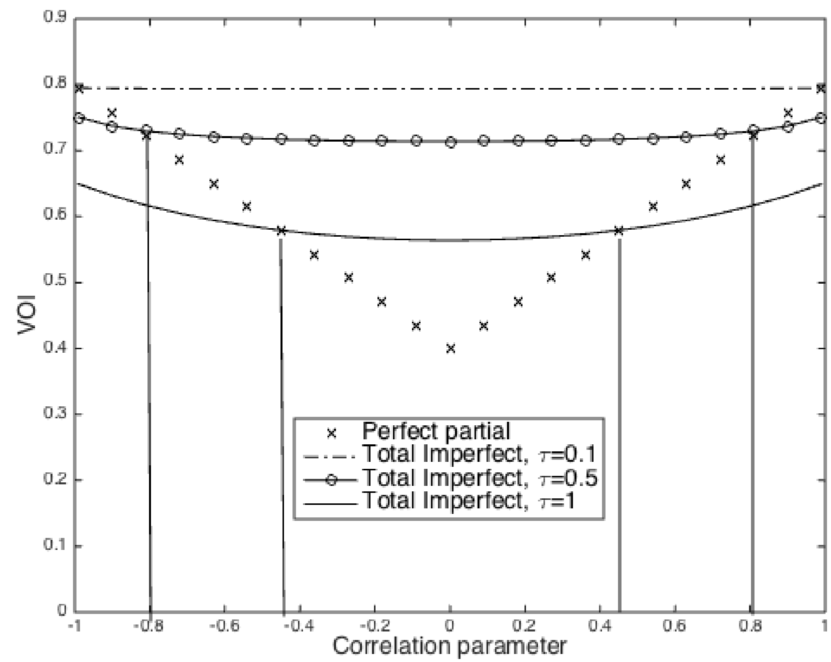
Exercise : Two Gaussian projects

Consider the bivariate Gaussian projects example, with prior mean 0 and variance 1, correlation 0.7 (and 0.1) and measurement noise st dev 1.

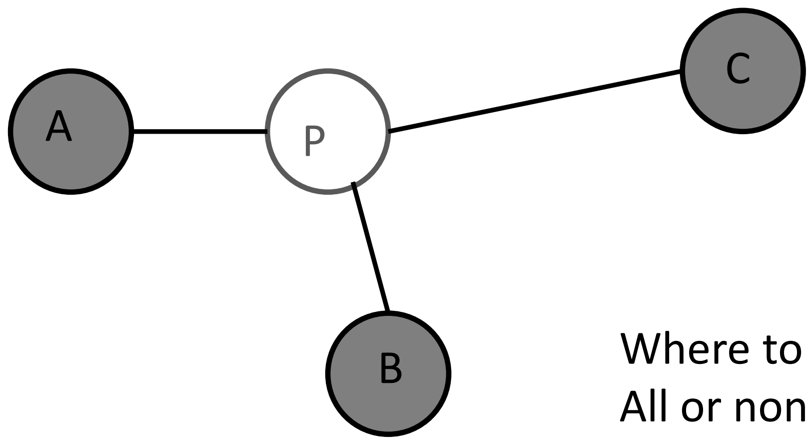
- Study the *decision regions* for no testing, partial perfect or total imperfect testing:

$$\arg\max\left\{VOI_{1,2}-\left(P_{1,2,imp}\right),VOI_{1}-P_{1,perf},0\right\}$$

Decision regions are visual plots,
with the price of one perfect on the x-axis,
and the price of two imperfect on the y-axis.



What if several projects or treasures?



Where to invest?
All or none? Free to choose as many as profitable? One at a time, then choose again?

Where should one collect data? All or none? One only? Or two? One first, then maybe another?

VOI and Earth sciences

- **Alternatives are spatial**, often with high flexibiliy in selection of sites, control rates, intervention, excavation opportunities, harvesting, etc.

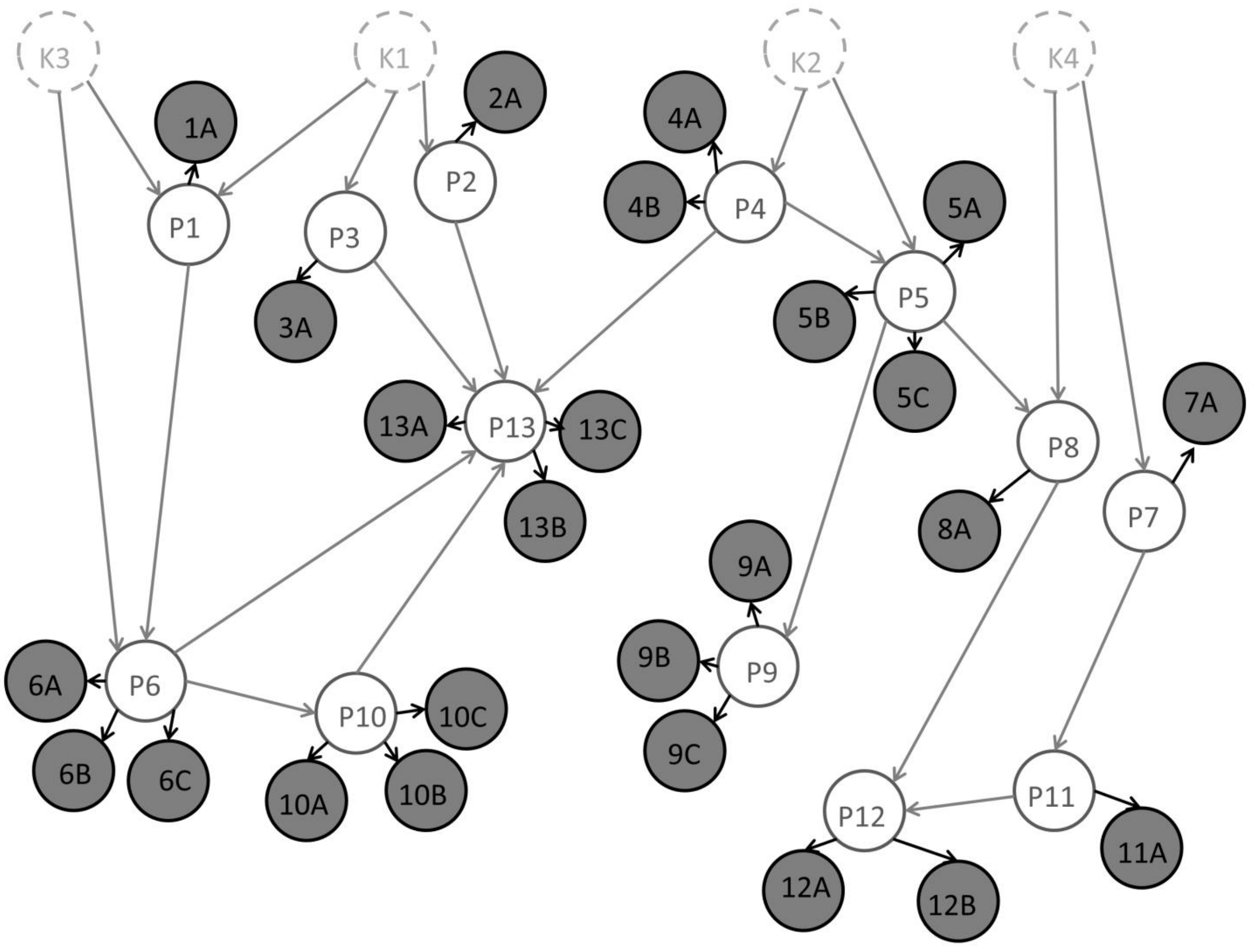
- **Uncertainties are spatial**, with multi-variable interactions . Often both discrete and continuous.

- **Value function is spatial**, typically involving coupled features, say through differential equations. It can be defined by «physics» as well as economic attributes.

- **Data are spatial**. There are plenty opportunities for partial, total testing and a variety of tests (surveys, monitoring sensors, electromagnetic data, , etc.)

Dependence? Does it matter?

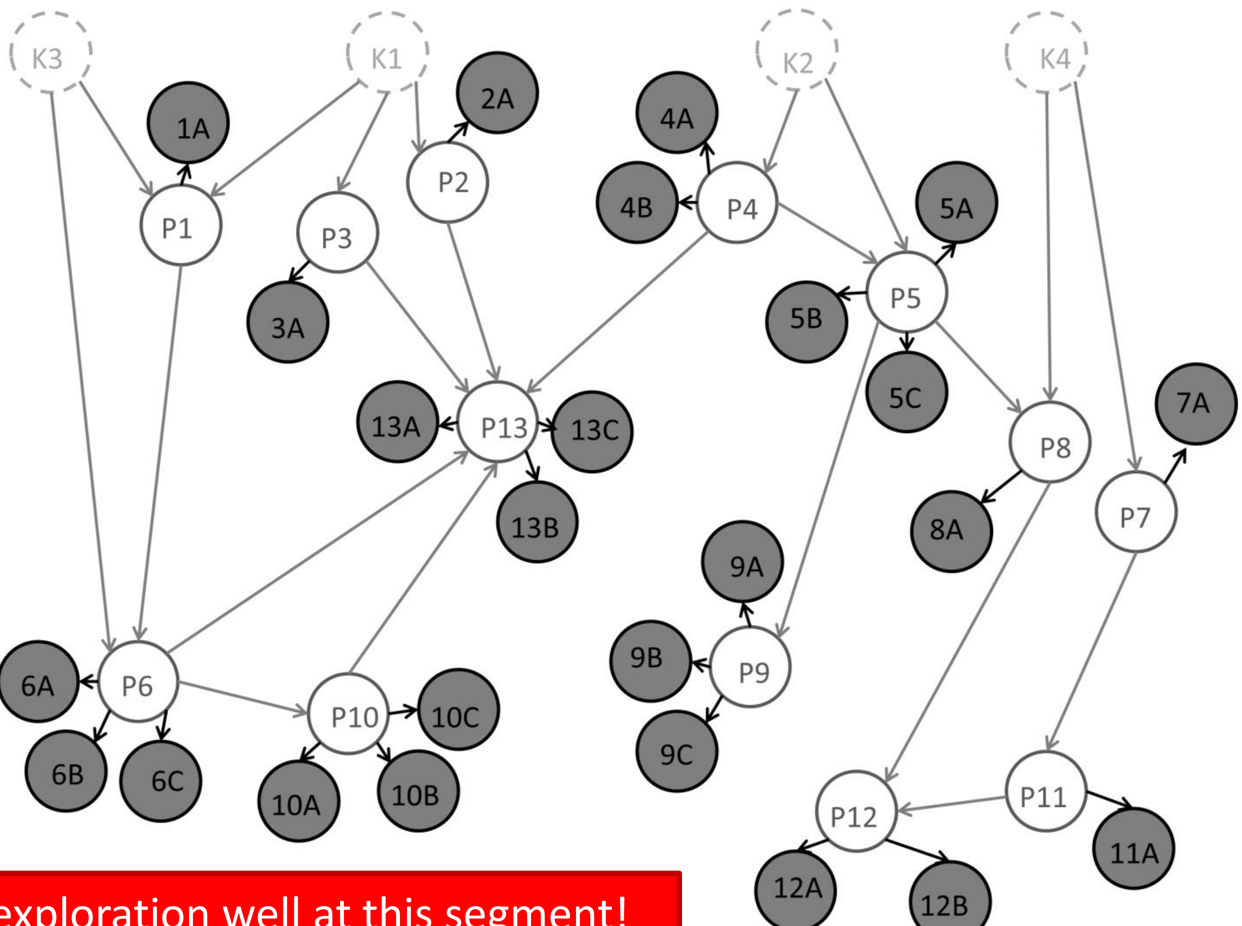
Gray nodes are petroleum reservoir segments where the company aims to develop profitable amounts of oil and gas.



Martinelli, G., Eidsvik, J., Hauge, R., and Førland, M.D., 2011, Bayesian networks for prospect analysis in the North Sea, *AAPG Bulletin*, 95, 1423-1442.

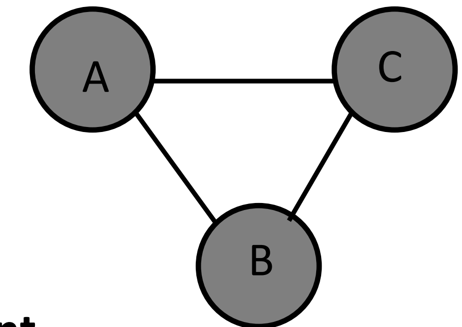
Dependence? Does it matter?

Gray nodes are petroleum reservoir segments where the company aims to develop profitable amounts of oil and gas.



Drill the exploration well at this segment!
The value of information is largest.

Joint modeling of multiple variables



Spatial variables are often not independent!

To study if dependence matter, we need to model the **joint** properties of uncertainties.

- What is the probability that variable A is 1 and, at the same time, variable B is 1 ?
- What is the probability that variable C is 0, and both A and B are 1 ?

Joint pdf

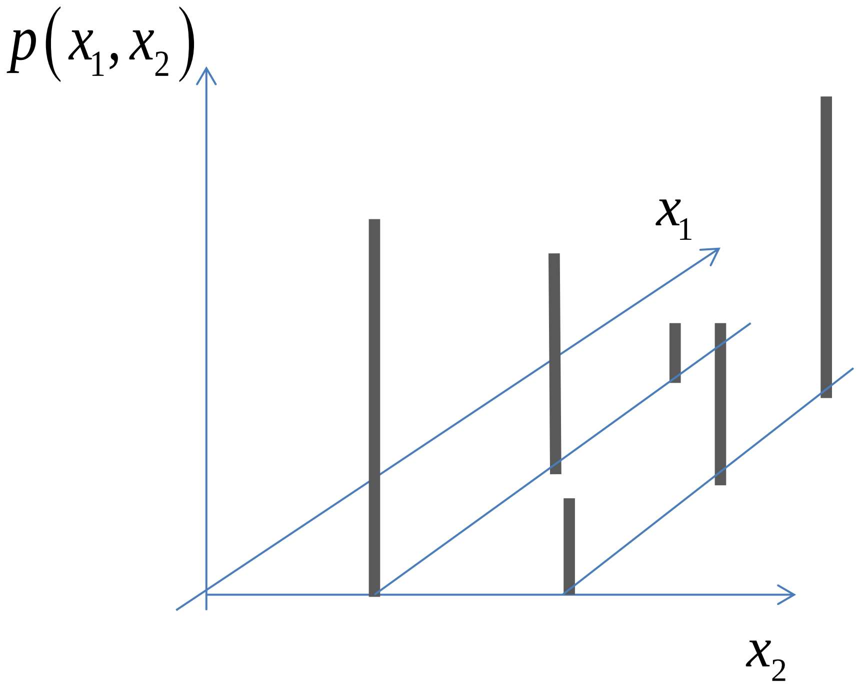
$$p(\mathbf{x}) = p(x_1, \ldots, x_n)$$

Discrete sample space:

$$p(\mathbf{x}) \geq 0, \qquad \mathbf{x} \in \Omega,$$

$$\sum_{x_1 \in \Omega_1} \cdots \sum_{x_n \in \Omega_n} p(\mathbf{x}) = 1.$$

Probability mass function (pdf)

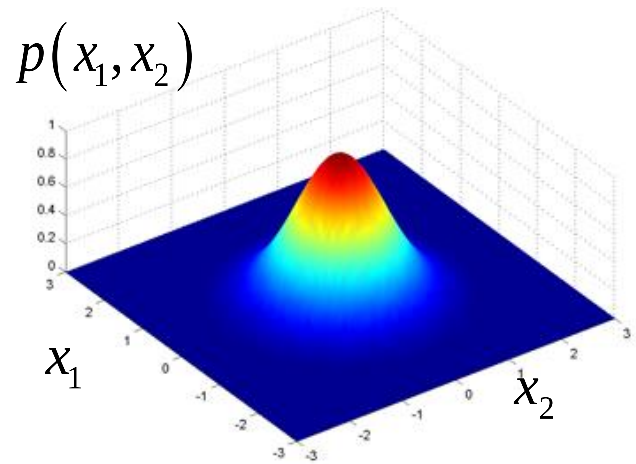


Continuous sample space:

$$p(\mathbf{x}) \geq 0, \qquad \mathbf{x} \in \Omega,$$

$$\int_{x_1 \in \Omega_1} \cdots \int_{x_n \in \Omega_n} p(\mathbf{x}) dx_1 \ldots dx_n = 1.$$

Probability density function (pdf)



Multivariate statistical models

The joint probability mass or density function (**pdf**) defines all probabilistic aspects of the distribution!

$$p(\boldsymbol{x}) = N(\mathbf{0}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\downarrow$$

$$E(\boldsymbol{x}) = \boldsymbol{\mu} = \int \boldsymbol{x} p(\boldsymbol{x}) d\boldsymbol{x},$$

$$Var(\boldsymbol{x}) = \boldsymbol{\Sigma} = \int (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^t \, p(\boldsymbol{x}) d\boldsymbol{x},$$

$$E(f(\boldsymbol{x})) = \int f(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}.$$

Marginal and conditional probability

$$\boldsymbol{x} = (\boldsymbol{x}_{\mathbb{K}}, \boldsymbol{x}_{\mathbb{L}})$$

$$p(\boldsymbol{x}_{\mathbb{K}}) = \int p(\boldsymbol{x})\, d\boldsymbol{x}_{\mathbb{L}}$$

Marginalization in joint pdf.

$$p(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}}) = \frac{p(\boldsymbol{x})}{p(\boldsymbol{x}_{\mathbb{L}})} = \frac{p(\boldsymbol{x})}{\int p(\boldsymbol{x})\, d\boldsymbol{x}_{\mathbb{K}}}$$

Conditioning in joint pdf.

Conditional mean and variance

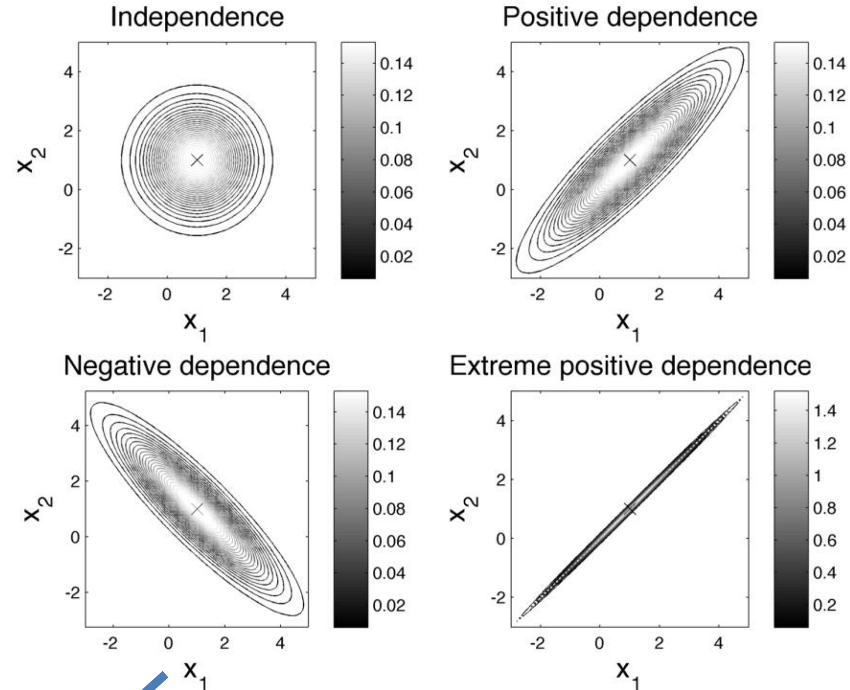
$$E(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}}) = \int \boldsymbol{x}_{\mathbb{K}}\, p(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}})\, d\boldsymbol{x}_{\mathbb{K}},$$

$$Var(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}}) = \int \left(\boldsymbol{x}_{\mathbb{K}} - E(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}})\right)\left(\boldsymbol{x}_{\mathbb{K}} - E(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}})\right)^{t} p(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}})\, d\boldsymbol{x}_{\mathbb{K}}.$$

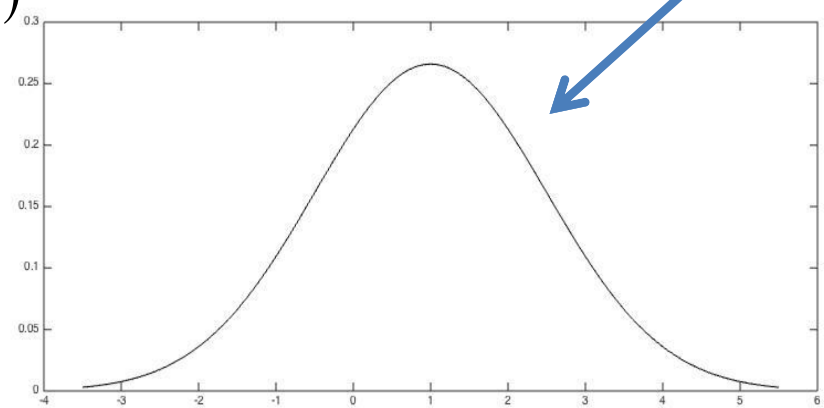
Marginalization

$$\boldsymbol{x} = (\boldsymbol{x}_{\mathbb{K}}, \boldsymbol{x}_{\mathbb{L}})$$

$$p(\boldsymbol{x}_{\mathbb{K}}) = \int p(\boldsymbol{x})\, d\boldsymbol{x}_{\mathbb{L}}$$



$p(x_1)$



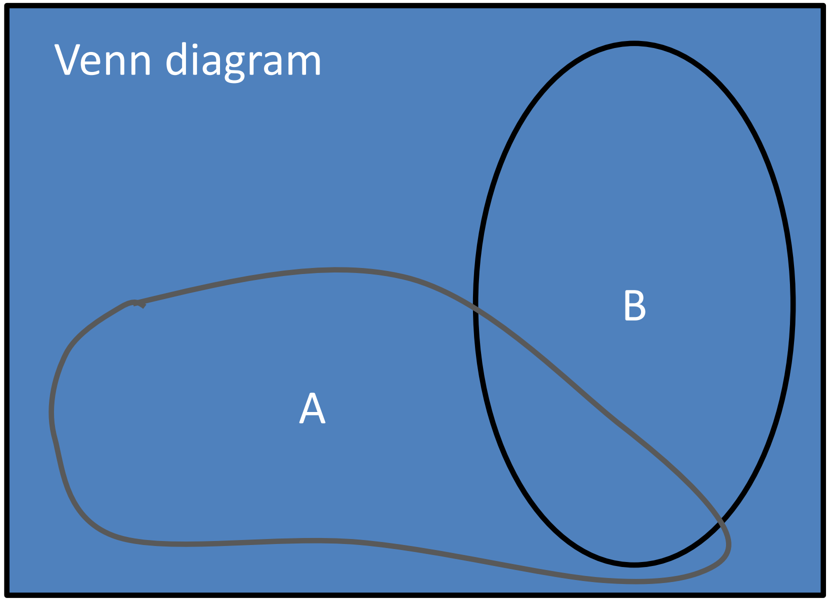
x_1

Conditional probability

$$p(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}}) = \frac{p(\boldsymbol{x})}{p(\boldsymbol{x}_{\mathbb{L}})} = \frac{p(\boldsymbol{x})}{\int p(\boldsymbol{x})\, d\boldsymbol{x}_{\mathbb{K}}}$$

$$p(A \mid B) = \frac{Area(A \cap B)}{Area(B)}$$

$$B = (A \cap B) \cup (A^C \cap B)$$



Conditional probability

$$p\left(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}}\right) = \frac{p\left(\boldsymbol{x}\right)}{p\left(\boldsymbol{x}_{\mathbb{L}}\right)} = \frac{p\left(\boldsymbol{x}\right)}{\int p\left(\boldsymbol{x}\right) d\boldsymbol{x}_{\mathbb{K}}}$$

$$p\left(\boldsymbol{x}_{\mathbb{K}}\right) \neq p\left(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}}\right)$$

$$p\left(\boldsymbol{x}_{\mathbb{K}}\right) = p\left(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}}\right)$$

Independence!

Must hold for all outcomes and for all subsets!
Unrealistic in most applications!

Modeling by conditional probability

The **joint** pdf can be difficult to model directly.
Instead we can build the joint pdf from **conditional** distributions.

$$p(\boldsymbol{x}) = p(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}})\, p(\boldsymbol{x}_{\mathbb{L}})$$

$$p(\boldsymbol{x}) = p(x_1)\, p(x_2 \mid x_1) \ldots p(x_n \mid x_{n-1}, \ldots, x_1)$$

Holds for any ordering of variables.

Modeling by conditional probability

Modeling by conditionals is done by conditional statements, not joint assessment:

- What is likely to happen for variable K when variable L is 1?
- What is the probability of variable C being 1 when variables A and B are both 0?

Such statements might be easier to specify,
and can more easily be derived from physical principles.

$$p(\boldsymbol{x}) = p(\boldsymbol{x}_{\mathbb{K}} \mid \boldsymbol{x}_{\mathbb{L}})\, p(\boldsymbol{x}_{\mathbb{L}})$$

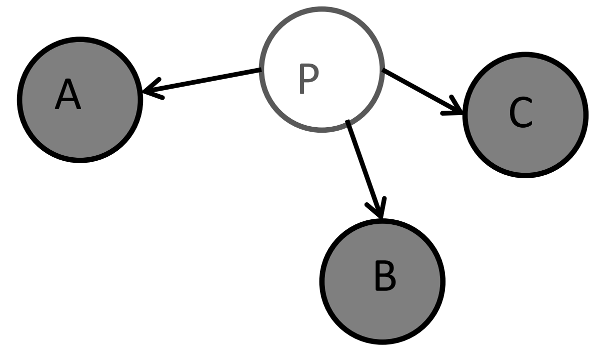
Modeling by conditional probability

$$p(\boldsymbol{x}) = p(x_1)\,p(x_2 \mid x_1)\ldots p(x_n \mid x_{n-1},\ldots,x_1)$$

Holds for any ordering of variables. Some conditioning variables can often be skipped.
Conditional independence in modeling.
This simplifies modeling and interpretation! And computing!

Modeling by conditional probability

Conditional independence:



$$p\left(x_A, x_B, x_C \mid x_P\right) = \prod_{i \in \{A,B,C\}} p\left(x_i \mid x_P\right)$$

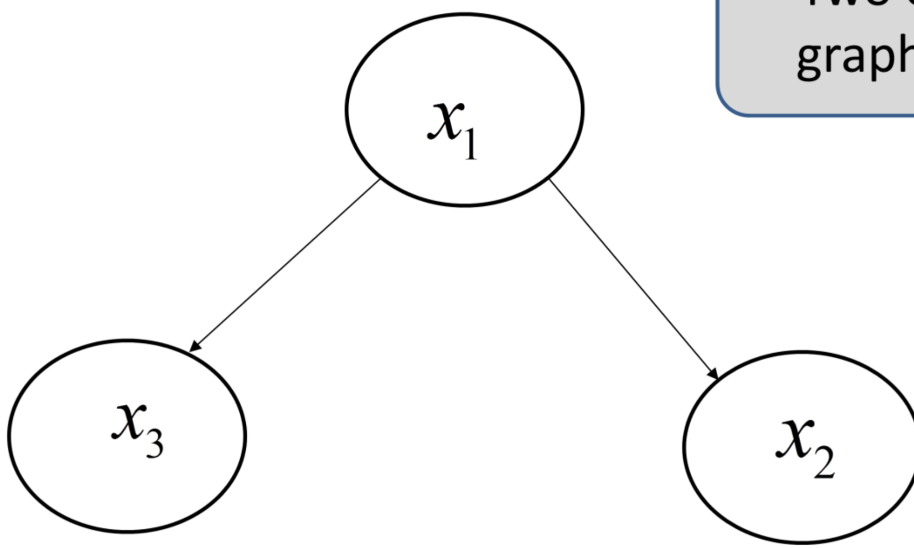
- What is the chance of success at B, when there is success at parent P?
- What is the chance of success at B, when there is failure at parent P?

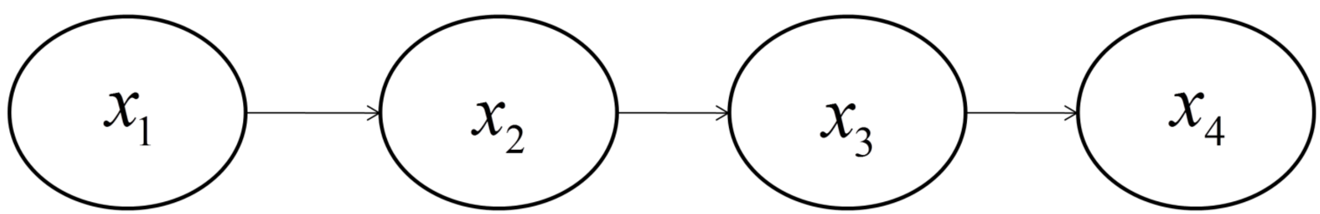
$$p\left(x_B = 1 \mid x_P = 1\right) = 0.9$$
$$p\left(x_B = 1 \mid x_P = 0\right) = 0$$

Must set up models for all nodes, using marginals for root nodes, and conditionals for all nodes with edges.

Bayesian networks and Markov chains

Two examples of graphical models

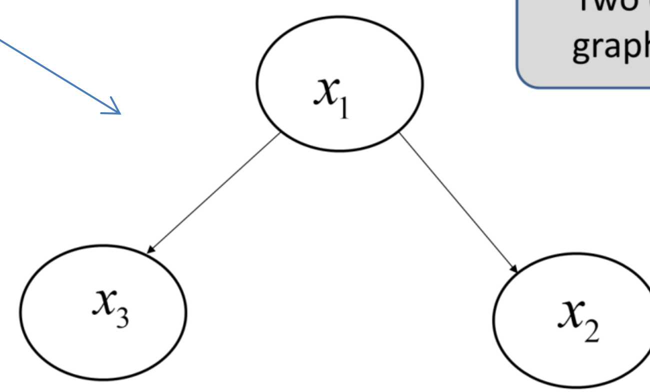


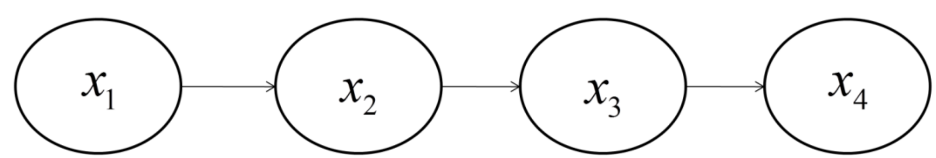


Bayesian networks and Markov chains

$$p(\boldsymbol{x}) = p(x_1, x_2, x_3) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_1)$$

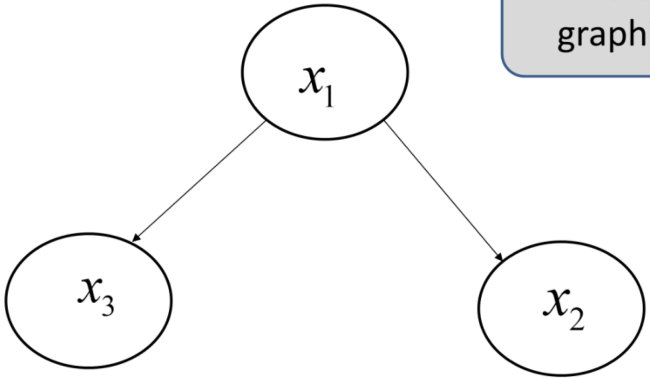
Two examples of graphical models

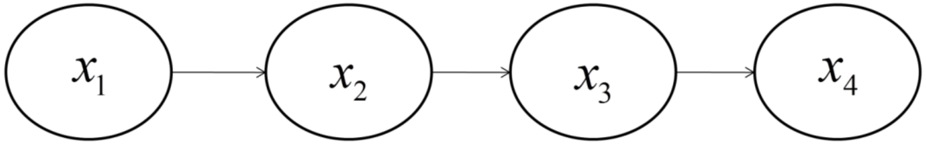




Bayesian networks and Markov chains

Two examples of graphical models

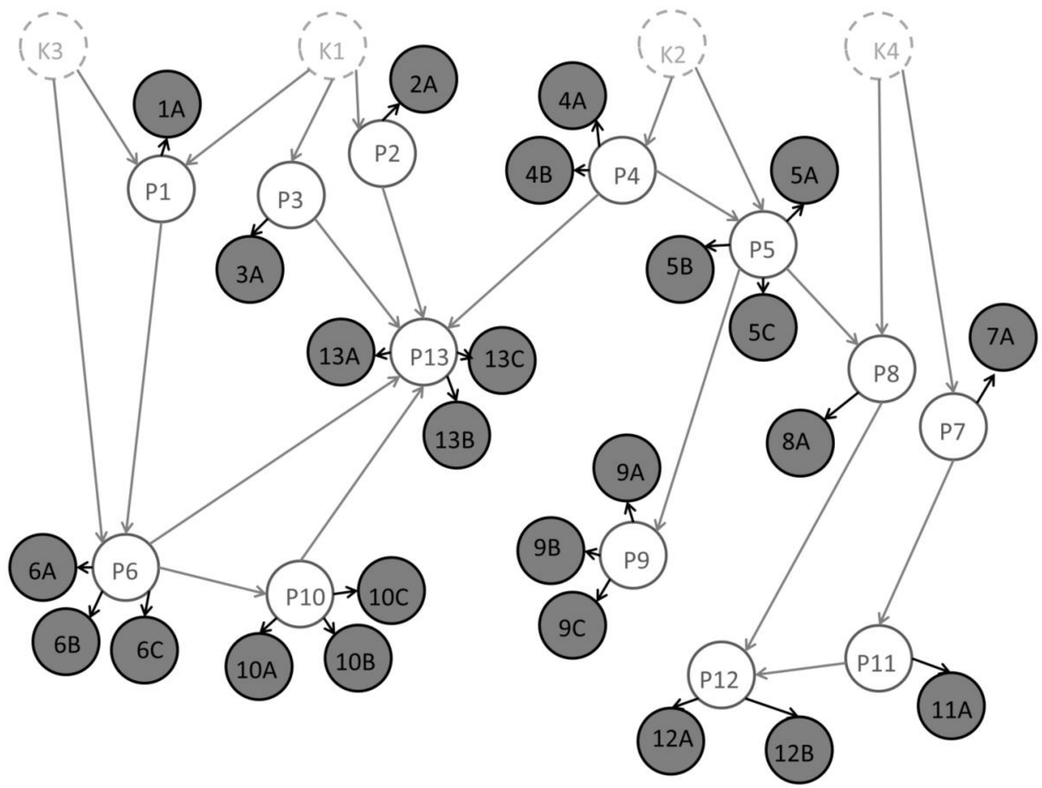




$$p(\boldsymbol{x}) = p(x_1, x_2, x_3, x_4) = p(x_1)\,p(x_2 \mid x_1)\,p(x_3 \mid x_2)\,p(x_4 \mid x_3)$$

Larger networks - computation

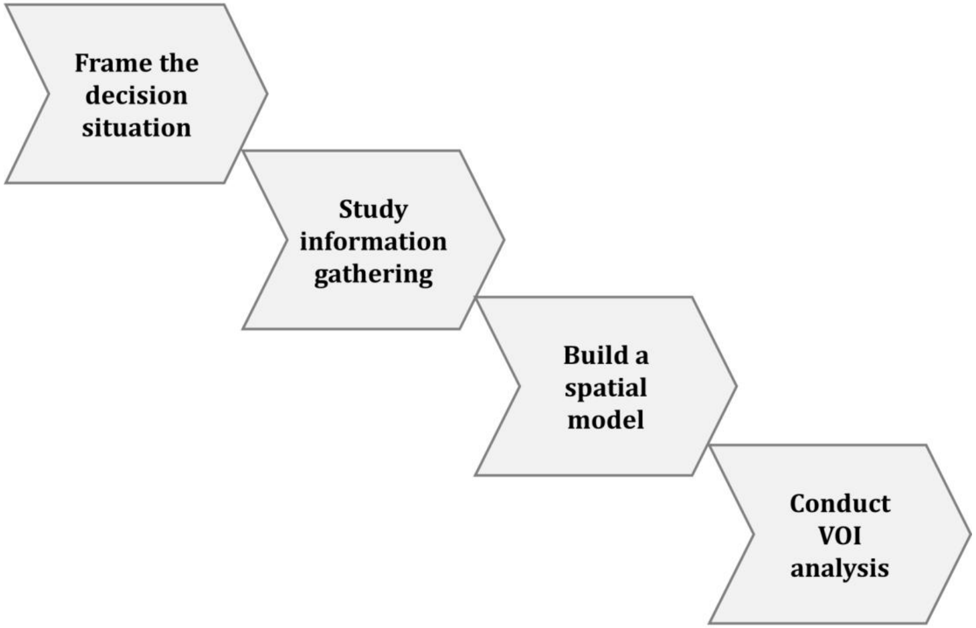
Algorithms have been developed for efficient marginalization, conditioning for Bayesian network models (Junction tree algorithm).



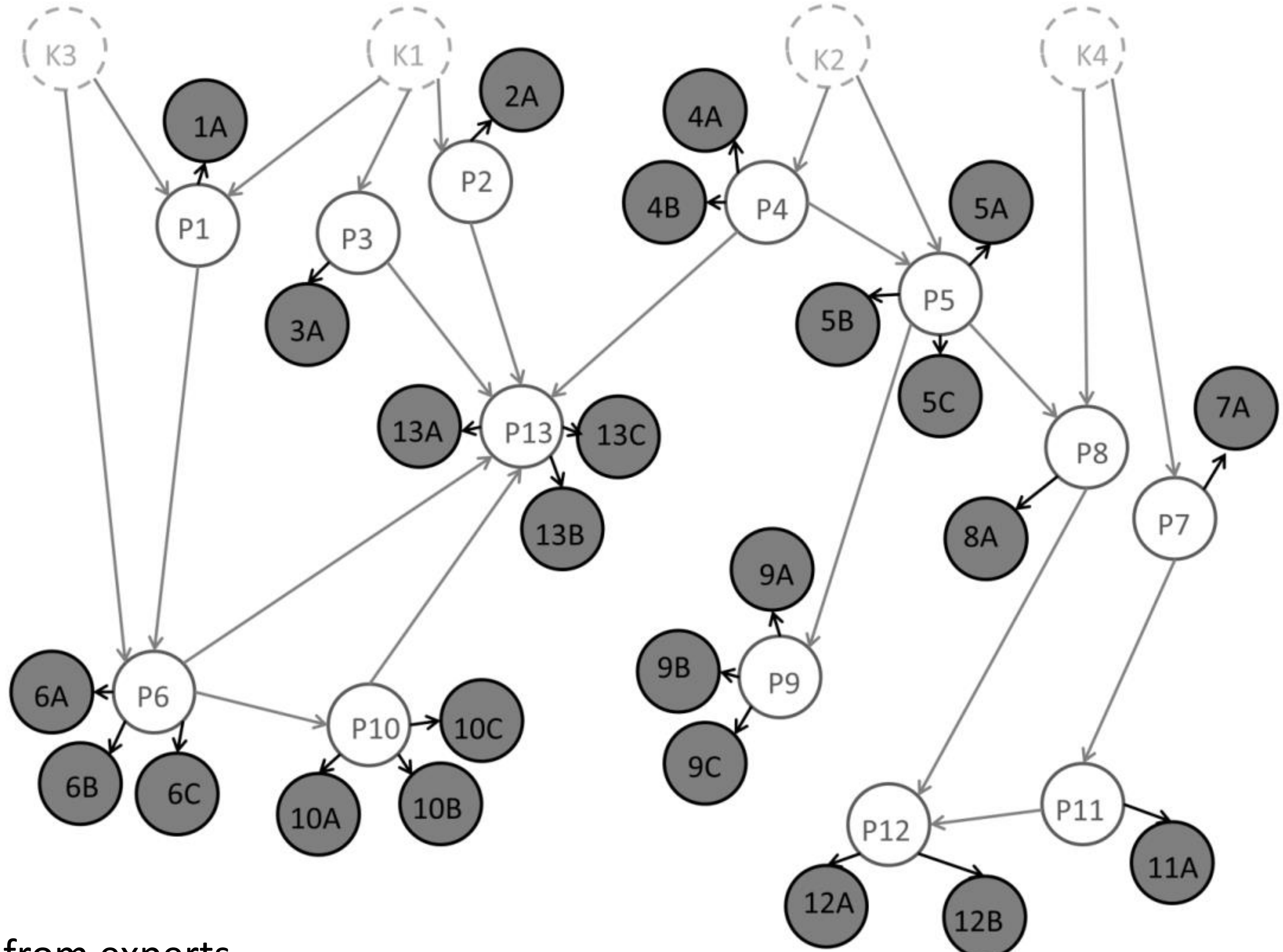
Martinelli, G., Eidsvik, J., Hauge, R., and Førland, M.D., 2011, Bayesian networks for prospect analysis in the North Sea, *AAPG Bulletin*, 95, 1423-1442.

VOI workflow

- Develop prospects separately. Shared costs for segments within one prospect.
- Gather information by exploration drilling. One well.
- Model is a Bayesian network model elicited from expert geologists in this area.
- VOI analysis done by exact computations for Bayesian networks. Conducted for all possible single exploration wells.



Bayesian network , Kitchens

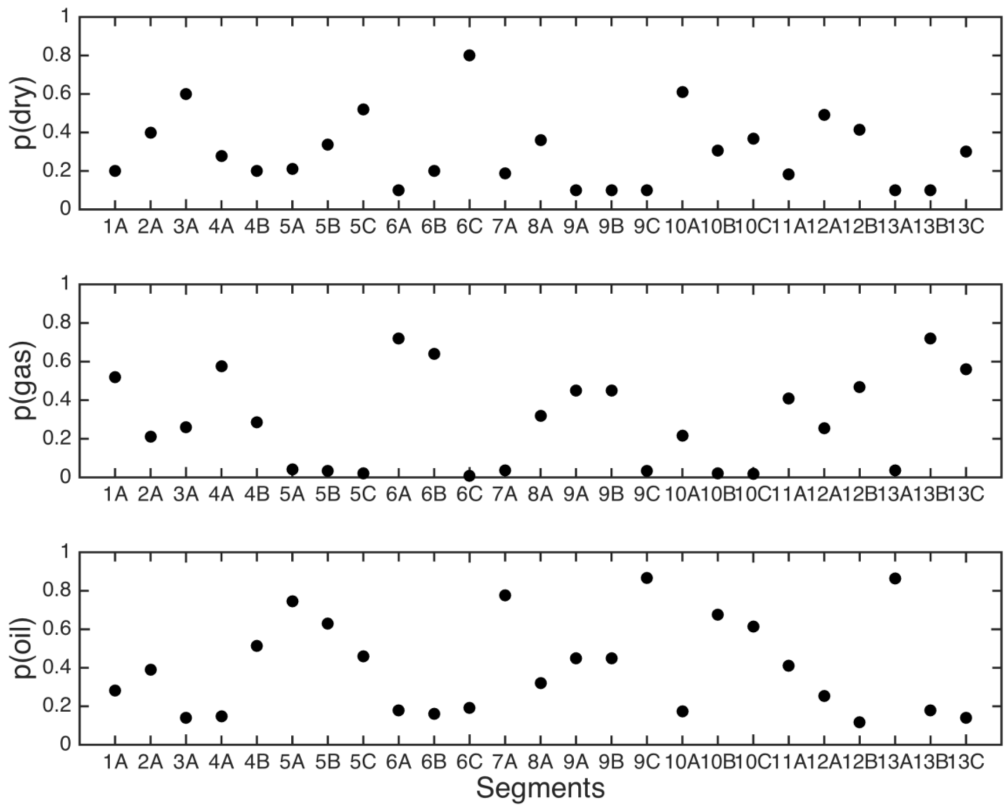


Model elicited from experts.
Migration from kitchens.
Local failure probability of migration.

Prior marginal probabilities

Three possible classes at all nodes:

- Dry
- Gas
- Oil



Prior values

Development fixed cost.
Infrastructure at prospect r.

$$PV = \sum_{r=1}^{13} \max \left\{ 0, \sum_{i \in \mathrm{Pr}} IV(x_i) - DFC \right\}$$

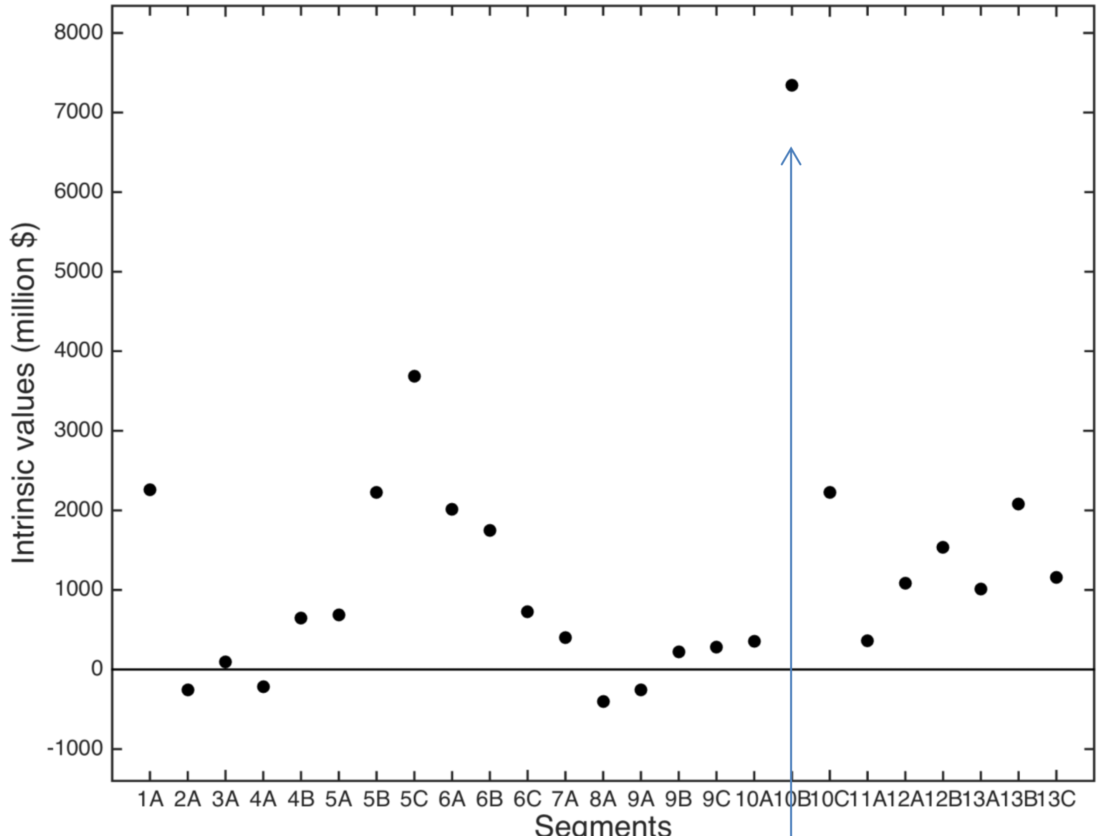
$$IV(x_i) = \sum_{k=1}^{3} \left(\mathrm{Rev}_{i,k}\, p(x_i = k) - \mathrm{Cost}_{i,k}\, p(x_i = k) \right) - \mathrm{Cost}_{i,0}$$

Revenues of oil/gas,
0 otherwise.

Cost if dry,
0 otherwise.

Cost of drilling
segment i.

Values



Most lucrative. But might not be most informative.

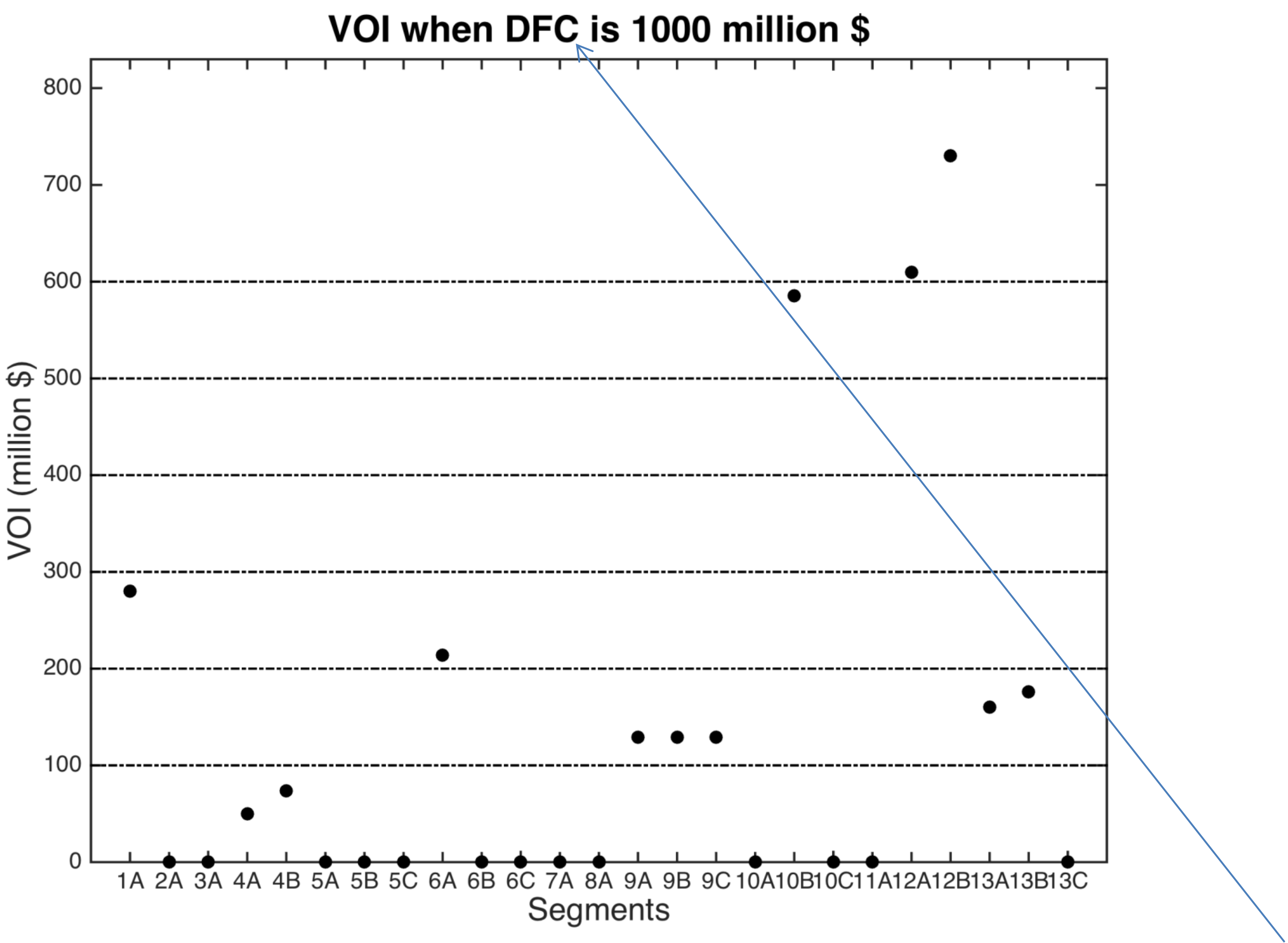
Posterior values and VOI

$$PoV\left(x_{\mathbb{K}}\right)=\sum_{l=1}^{3}\sum_{r=1}^{13}\max\left\{0,\sum_{i\in \Pr} IV\left(x_i \mid x_{\mathbb{K}}=l\right)-DFC\right\}p\left(x_{\mathbb{K}}=l\right)$$

$$VOI\left(x_{\mathbb{K}}\right)=PoV\left(x_{\mathbb{K}}\right)-PV$$

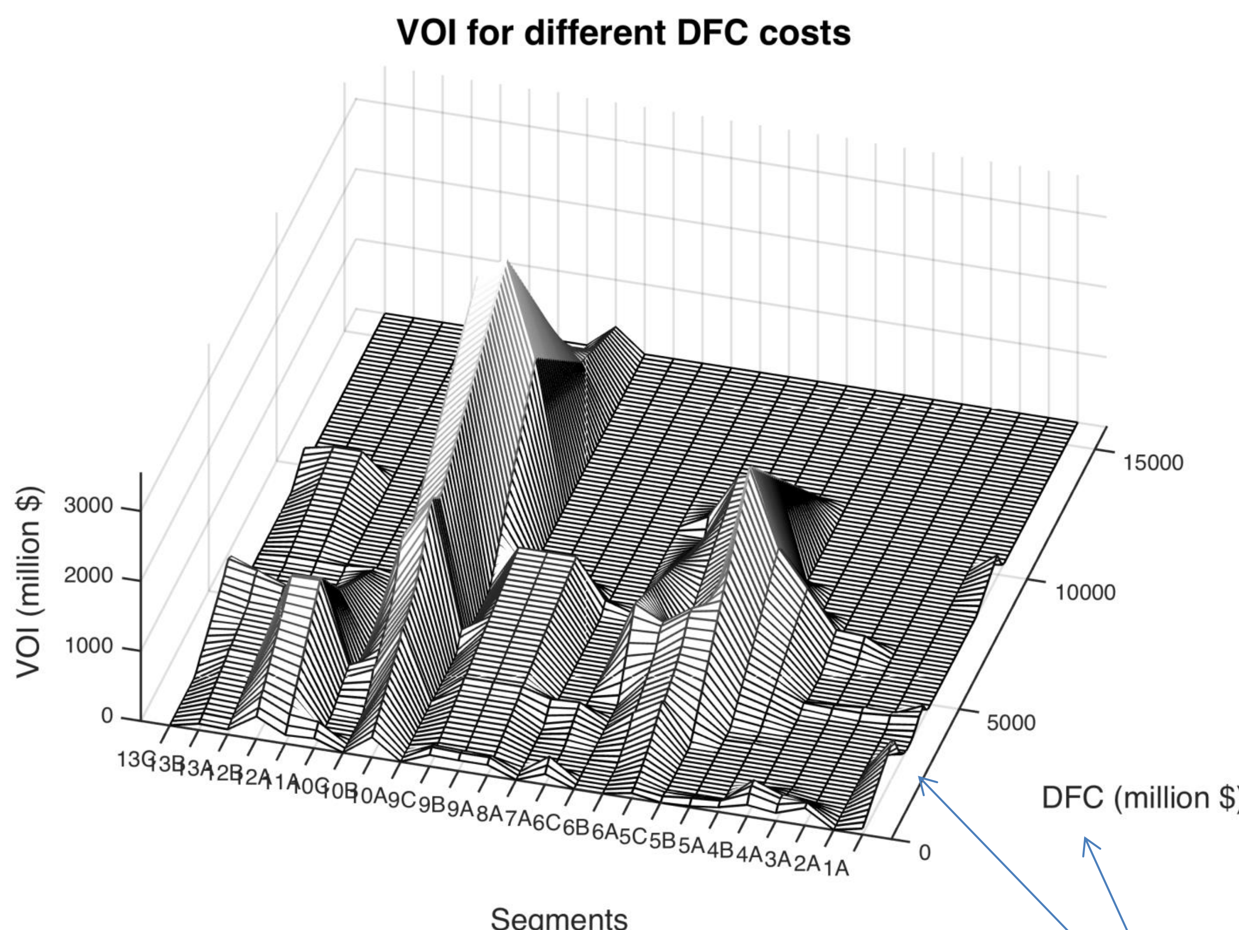
Data acquired at single well.

VOI single wells



Development fixed cost.

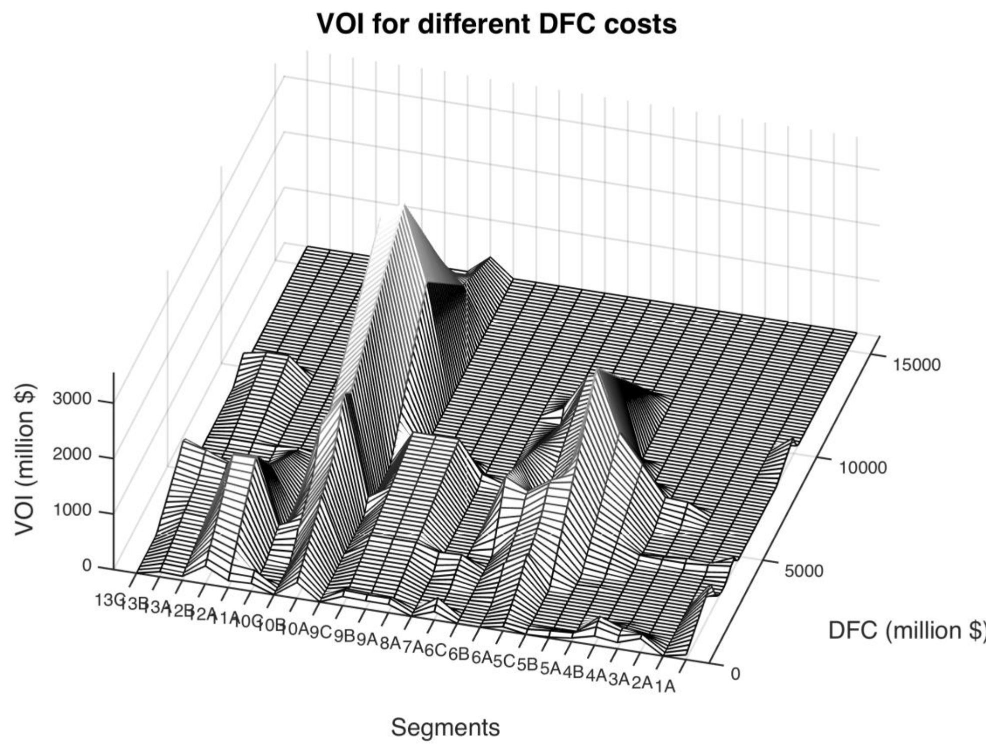
VOI for different costs



Development fixed cost.

VOI for different costs

- For each segment VOI starts at 0 (for small costs), grows to larger values, and decreases to 0 (for large costs).
- VOI is smooth for segments belonging to the same prospect. Correlation and shared costs.
- VOI can be multimodal as a function of cost, because the information influences neighboring segments, at which we are indifferent of at other costs.



Take home from this exercise:

- VOI is not largest at the most lucrative prospects.
- VOI is largest where more data are likely to help us make better decisions.
- VOI also depends on whether the data gathering can influence neighboring segments – data propagate in the Bayesian network model.

- Compare with price? Or compare different data gathering opportunities, and provide a basis for discussion.

Markov chains

Markov chains are special graphs, defined by initial probabilities and transition matrices.

$$p(\boldsymbol{x}) = p(x_1, x_2, \ldots, x_n) = p(x_1)\, p(x_2 \mid x_1) \ldots p(x_n \mid x_{n-1})$$

$$p(x_1 = k), \qquad k = 1, \ldots, d$$

$$p(x_{i+1} = l \mid x_i = k) = P(k, l), \qquad k, l = 1, \ldots, d$$

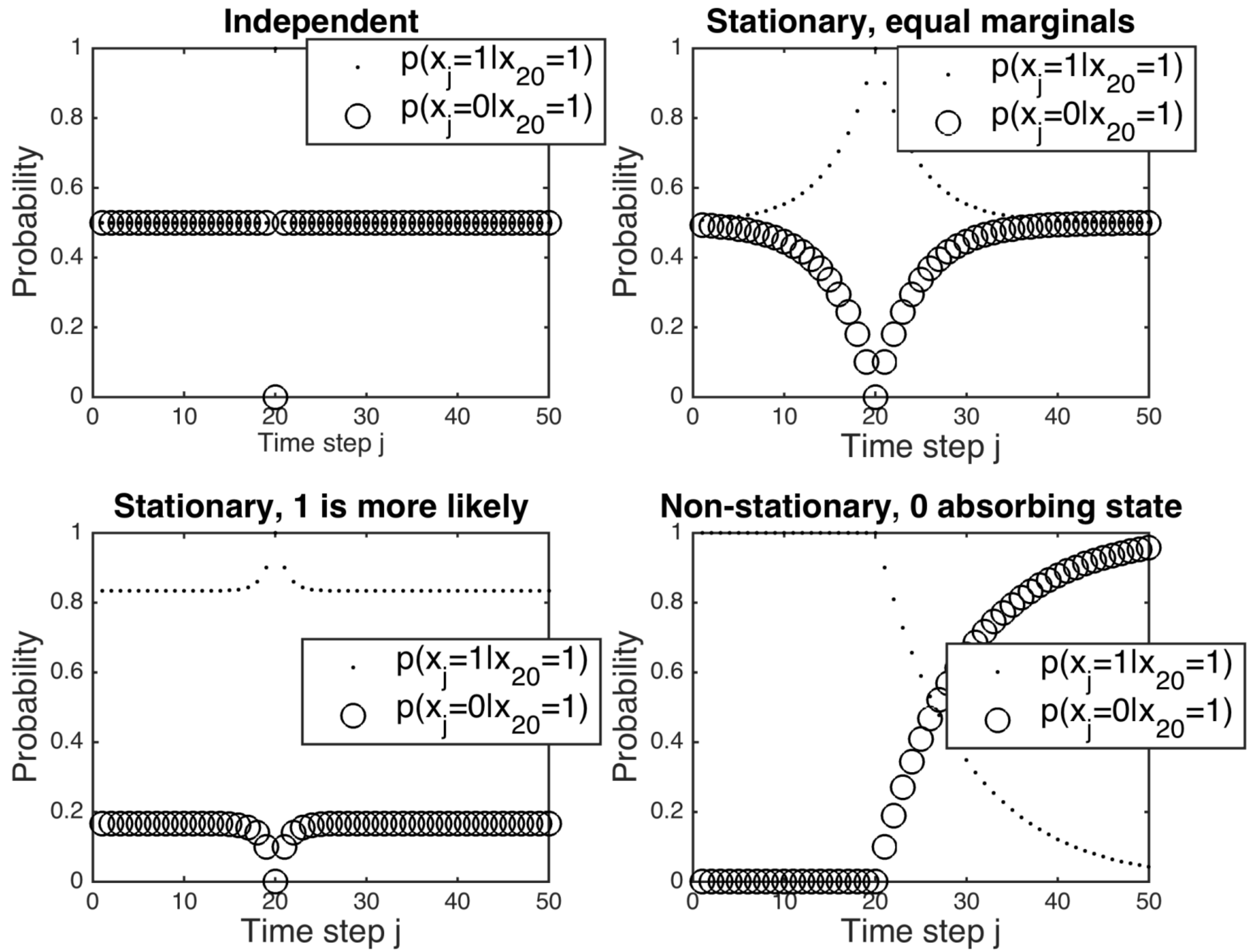
$$d = 2$$

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \qquad P = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \qquad P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \qquad P = \begin{bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{bmatrix}$$

Independence

Absorbing

Markov chains (given perfect information)



Avalanche decisions and sensors

Suppose that parts along a road are at risk of **avalanche**.
- One can remove risk by clearing roads, at a cost.
- Otherwise, the repair cost depends on the unknown risk class (low or high).

Data, sensor at a particular location, can help classify the risk class and hence improve the decisions made regarding cleaning / wait and see.

Avalanche decisions - risk analysis

n=50 identified locations along railroad track, at increasing altitude and risk of **avalanche**.
One can remove risk entirely by cost 100 000.
If it is not removed, the repair cost, at each location, depends on the unknown risk class:

$$C_j, \qquad j \in \{1,2\},$$

$$\mathrm{C}_1 = 0, \mathrm{C}_2 = 5000,$$

Decision maker must choose whether to

i) **clean tracks** up front, with fixed price.

ii) **wait and see**, with the uncertain price.

The decision is based on the minimization of expected costs.

Prior value:

$$PV = \max\left\{-100000, -5000\sum_{i=1}^{50} p\left(x_i = 2\right)\right\}$$

Clean up front

Expected value when they wait and see.

Markovian model for risk of avalanche

Risk tends to start in lower class (1), and then move to higher class (2).
If risk class 2 is reached, it will stay there until location 50 (absorbing state).

$$x_i \in \{1, 2\}, \qquad i = 1, \dots, 50,$$

$$p(x_1 = 1) = 0.99,$$

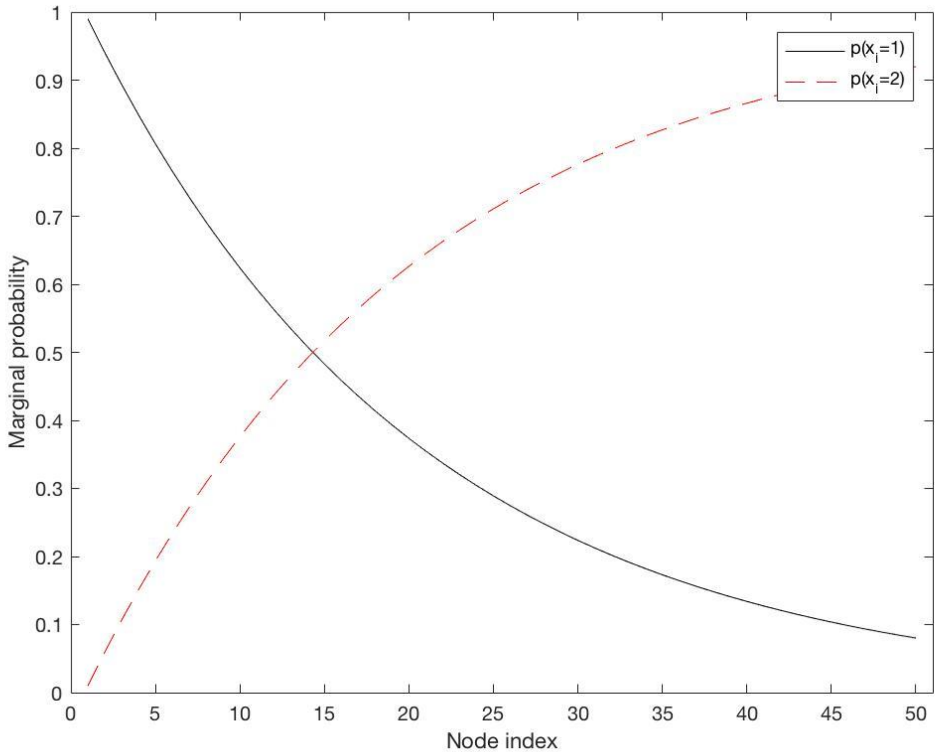
$$p(x_1 = 2) = 0.01,$$

$$P = \begin{bmatrix} p & (1-p) \\ 0 & 1 \end{bmatrix}$$

Results – marginals

$$p\left(x_i = l\right) = \sum_{k=1}^{2} p\left(x_{i-1} = k\right) p(x_i = l \mid x_{i-1} = k), \qquad l = 1, 2, \qquad i = 1, \ldots, n$$

$$p\left(x_i = 1\right) = p \cdot 0.99^{i-1} \qquad i = 1, \ldots, n$$



Markovian model for risk of avalanche

- Install a sensor at one location, getting perfect information at that node.
- Compute conditional probabilities.

$$p\left(x_i = k \mid x_j = l\right), \qquad i = 1, \ldots, 50$$

Results – conditionals (forward)

$$p\left(x_i = k \mid x_j = l\right) = \sum_{q=1}^{2} p\left(x_i = k, x_{i-1} = q \mid x_j = l\right) = \sum_{q=1}^{2} P\left(q,k\right) p\left(x_{i-1} = q \mid x_j = l\right)$$

$$p\left(x_i = 1 \mid x_j = 1\right) = 0.99^{i-j} \qquad i \geq j,$$

$$p\left(x_i = 2 \mid x_j = 2\right) = 1 \qquad i \geq j$$

Results – conditionals (backward)

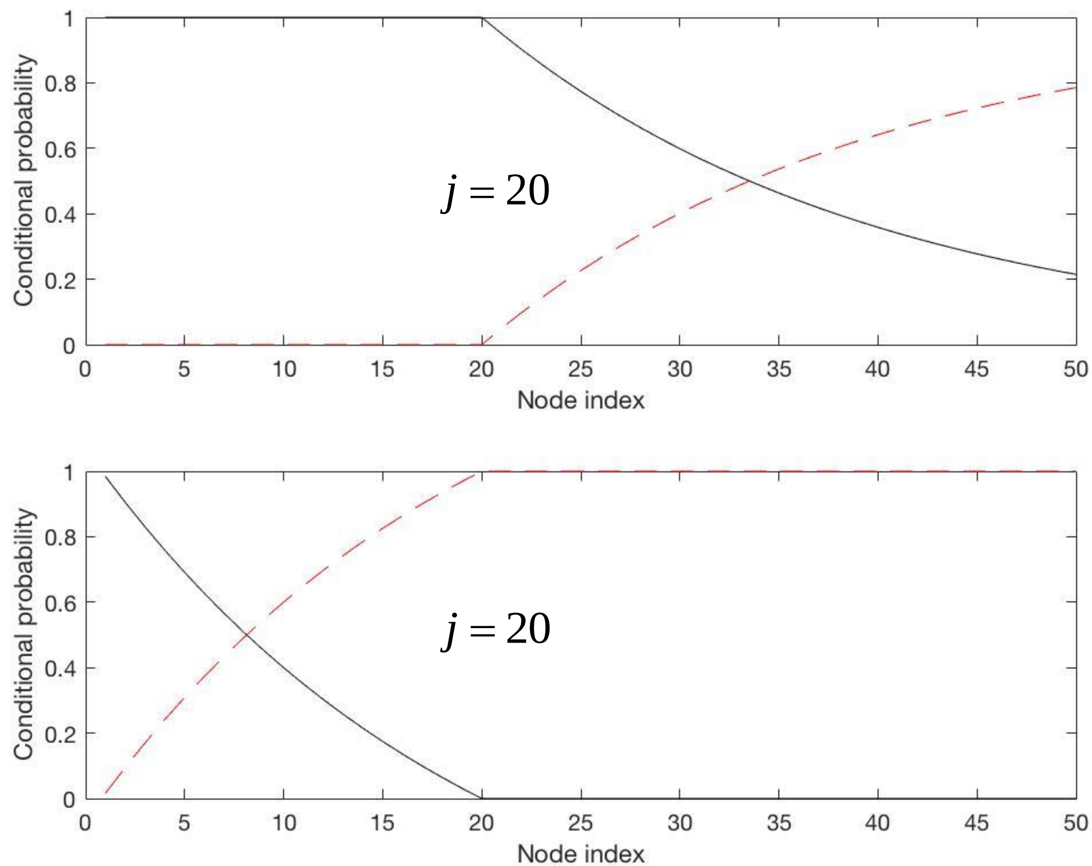
$$p\left(x_i = k \mid x_{i+1}, \ldots, x_n\right) = p\left(x_i = k \mid x_{i+1} = l\right) = \frac{p\left(x_i = k, x_{i+1} = l\right)}{p\left(x_{i+1} = l\right)} = \frac{P\left(k,l\right) p\left(x_i = k\right)}{p\left(x_{i+1} = l\right)}$$

$$p\left(x_i = k \mid x_j = l\right) = \sum_{q=1}^{2} p\left(x_i = k \mid x_{i+1} = q\right) p\left(x_{i+1} = q \mid x_j = l\right)$$

$$p\left(x_i = 1 \mid x_j = 1\right) = 1 \qquad i < j,$$

$$p\left(x_i = 2 \mid x_j = 2\right) = \frac{p\left(x_i = 2\right)}{p\left(x_j = 2\right)} \qquad i < j$$

Results – conditional probabilities



Markovian model for risk of avalanche

- Plan to install a sensor at one location, getting perfect information at that location.

- Compute the posterior value, with sensor location at one location.
  Compute the VOI.
- What is the optimal sensor location, if the goal is to improve risk decisions?

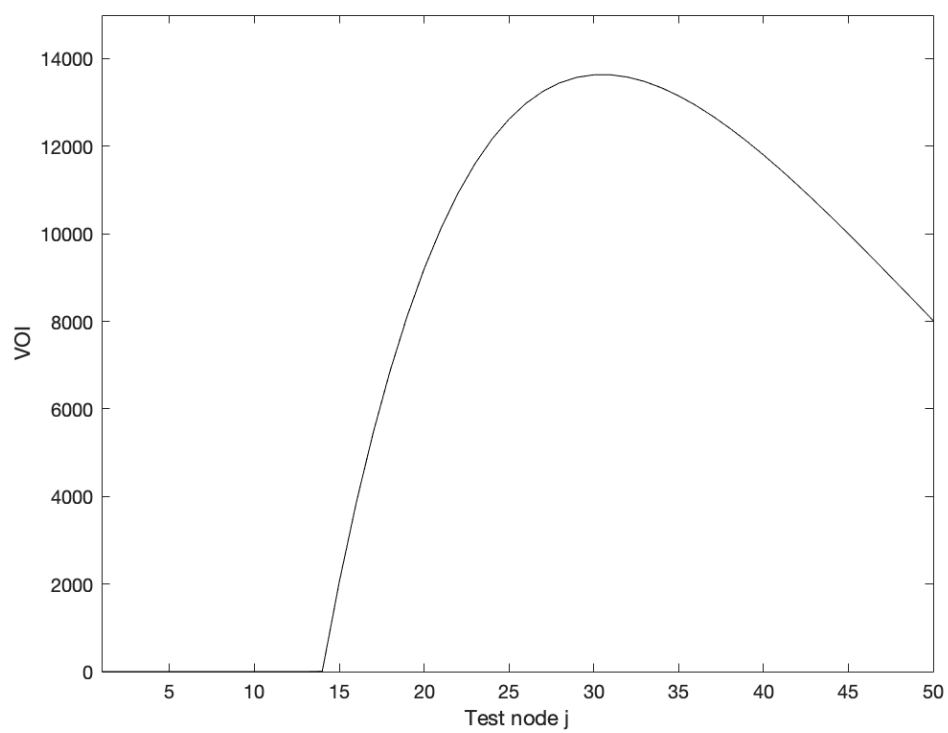
Results – VOI

$$PV = \max\left\{-100000, -5000\sum_{i=1}^{50} p\left(x_i = 2\right)\right\}$$

$$PoV\left(x_j\right) = \sum_{k=1}^{2} \max\left\{-100000, -5000\sum_{i=1}^{50} p\left(x_i = 2 \mid x_j = k\right)\right\} p\left(x_j = k\right)$$

$$VOI\left(x_j\right) = PoV\left(x_j\right) - PV$$

Best location near *j=30*.
The VOI is about 13000



Plan for course

Time	Topic
Lecture 1	Introduction and motivating examples
	Elementary decision analysis and the value of information
	Multivariate statistical modeling, dependence, graphs
	Value of information analysis for dependent models
Lecture 2	Spatial statistics, spatial design of experiments
	Value of information analysis in spatial decision situations
	Examples of value of information analysis in Earth sciences
Lecture 3	Computational aspects
	Sequential decisions and sequential information gathering
	Examples from mining and oceanography

Every day: Small exercise half-way, and computer project at the end.

Project - Avalanche risk

n=50 identified locations along railroad track, at risk of **avalanche**.

Decision maker must choose whether to

i) **clean tracks** up front, with fixed price.

ii) **wait and see**, with the uncertain price.

$$PV = \max\left\{-100000, -5000\sum_{i=1}^{50} p\left(x_i = 2\right)\right\}$$

$$PoV\left(x_j\right) = \sum_{k=1}^{2} \max\left\{-100000, -5000\sum_{i=1}^{50} p\left(x_i = 2 \mid x_j = k\right)\right\} p\left(x_j = k\right)$$

Is the VOI sensitive to *pp=0.95 (0.5, 0.9, 0.99)* ?

Is the optimal sensor location sensitive to *p=0.95 (0.5, 0.9, 0.99)* ?

Implement in Python, R or Matlab.

$$x_i \in \{1,2\}, \qquad i = 1,\ldots,50,$$

$$p\left(x_1 = 1\right) = 0.99, \qquad P = \begin{bmatrix} p & 1-p \\ 0 & 1 \end{bmatrix}$$

$$p\left(x_1 = 2\right) = 0.01,$$