60615A : Decision Analysis
Session 2 - Probabilistic Modeling I

Prof. Carolina Osorio

Slides of Prof. Erick Delage
Department of Decision Sciences
HEC Montréal

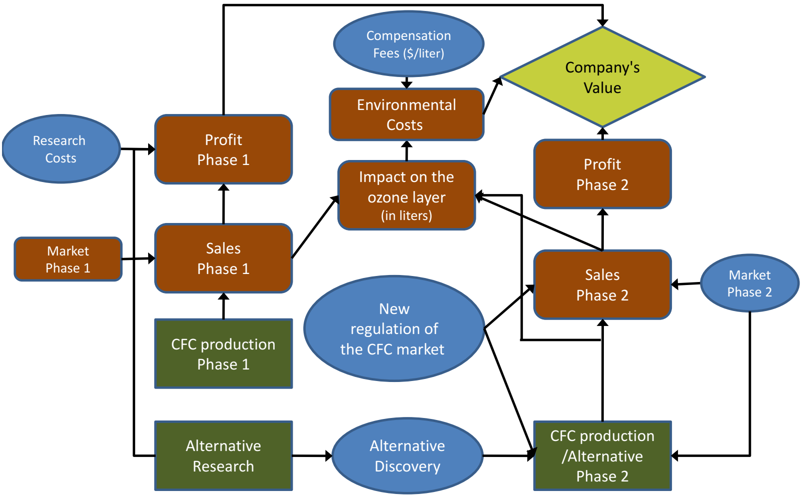
Outline

Outline

The presence of uncertainty

- It often happens that certain parameters of a problem are unknown at the moment of making a decision.
  - E.g. : weather, demand for a product, oil prices, amount of time between two volcanic eruptions, etc.
- We say that a parameter is uncertain when it can take different values and that the likelihood of each is computed using probability theory.
  - The uncertainty gives place to the notion of risk : i.e., each action has the potential to lead to positive consequences as well as negative consequences (with more or less chance).

Example of uncertainty at Dupont



Probability Definitions - Frequentist

- Definition : proportion of times an event will happen if we repeat the experience an infinite number of times.
    - Physical probability : we can study the experience's physical aspect and derive precisely the probability of each event.
        - Example : roll dice without bias, draw a ball from an urn, ...
    - Statistics : we consider the frequencies on a big sample of similar experiences.
        - Example : mortality rates, customer ratings, elections, ...

Probability Definitions - Subjective (Bayesian)

- Definition : the probabilities represent the level of conviction (or belief) of the individual and are updated when new information is available.
- Can be applied to any unconfirmed declaration, whether it relates to a random event or not, repeated or not.
- Example : What is the respective probability of each of the following events ?
  - The professor has more than 50$ in their pockets.
  - The Canadians of Montreal will qualify for the finals this year.

Probability Definitions - The difference

- Consider an urn containing 10 balls. Out of 10 balls drawn randomly from the urn (with replacement), 8 were red. What is the probability of randomly drawing a red ball from the urn ?
- Frequentist's answer :
  - Necessarily, the proportion is not 0.
  - Other than that, I can't claim anything.
  - I reject more easily the hypothesis that the probabilitiy is 0.1 rather than 0.8 (i.e., the proportion is 10% rather than 80%)
- Subjective's answer :
  - Originally, I believed that all proportions were as likely (equal probability on each proportion).
  - After the experience, it is likely to be close to 0.8
  - I perceive that the probability that there is 8 red balls is approximately 0.33
  - However, if the urn was red this probability would be higher.

Fundamental properties of probability

1. All probabilities are non-negative and lower or equal to 1.
2. The sum of the probabilities of mutually exclusive and collectively exhaustive events is 1.
3. The set of conditional probabilities must be coherent with the set of marginal probabilities.

$$\mathbb{P}(A\&B) = \mathbb{P}(A|B)\mathbb{P}(B) \quad \Rightarrow \quad \mathbb{P}(A|B) = \mathbb{P}(A\&B)/\mathbb{P}(B)$$

4. Decomposition of the probability of an event :

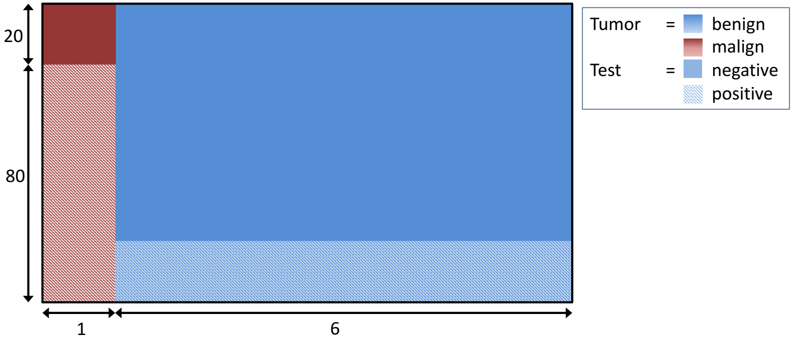
$$\mathbb{P}(A) = \mathbb{P}(A\&B_1) + \mathbb{P}(A\&B_2) + \ldots + \mathbb{P}(A\&B_n) \ ,$$

where $B_1, B_2, \ldots, B_n$ are <u>mutually exclusive</u> and <u>collectively exhaustive</u> events.

Not so easy to avoid violations

- «A certain biopsy test has the reputation of diagnosing properly the benign or malignant nature of a tumor 80% of the time : i.e., a malignant tumor will be accurately identified 80% of the time and a benign tumor will also be, 80% of the time. In general, tumors are 6 times more often benign than malignant.»
- The test just diagnosed a tumor as malignant. What is the probability that the tumor is really malignant ?
- Answer : 0.4 not 0.8. Why ? ? ?

Need to account for basic proportions



$$\text{Prop. tumor=malignant if test = positive} = \frac{\text{Area of «tumor=malignant \& test=positive»}}{\text{Area of «test=positive»}}$$

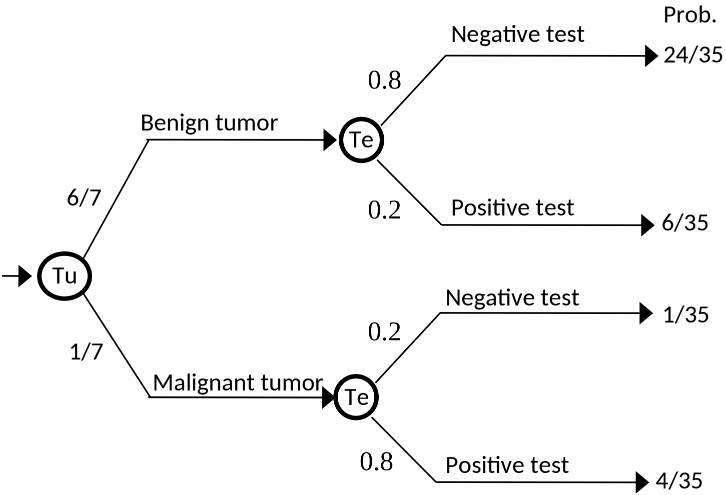
$$= \frac{1 \times 80}{1 \times 80 + 6 \times 20} = 0.4$$

Bayes' theorem

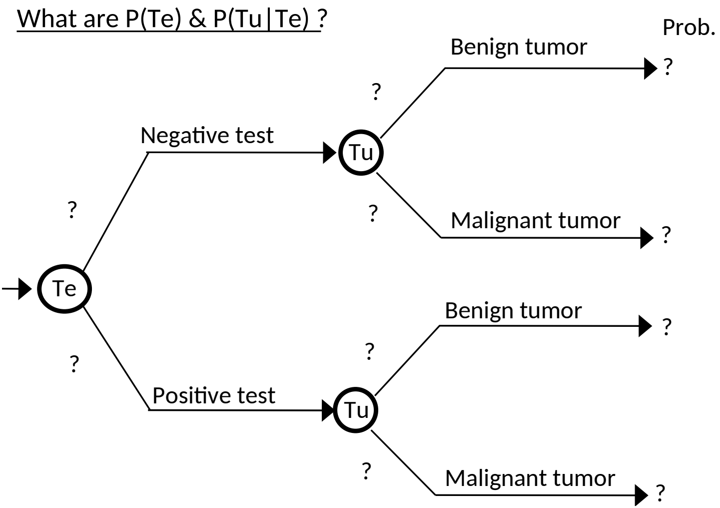
- Bayes' theorem tells us how to account for new information about random variables for which we already had information.
- In a case where we formulated $P(A)$ for $A = A_1, A_2, \cdots, A_n$ and receive the new information B
  - Characterize $P(B|A_1)$, $P(B|A_2), \cdots, P(B|A_n)$
  - Apply Bayes' theorem

$$
\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\
&= \frac{P(B|A)P(A)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots P(B|A_n)P(A_n)}
\end{aligned}
$$

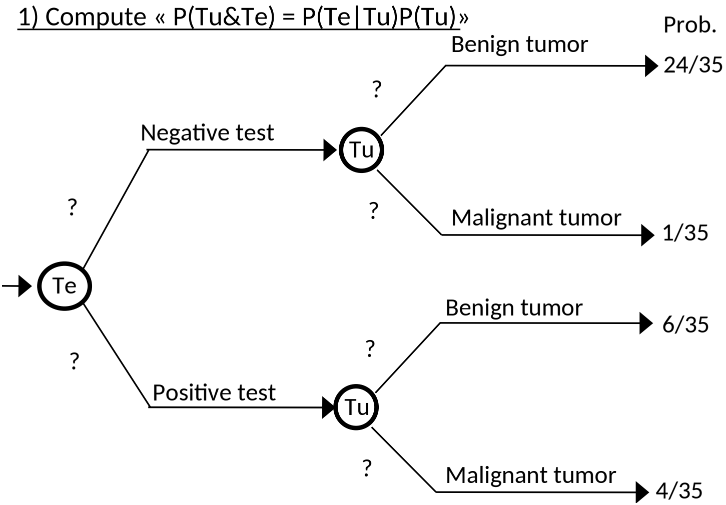
Analysis using a scenario tree



Analysis using a scenario tree

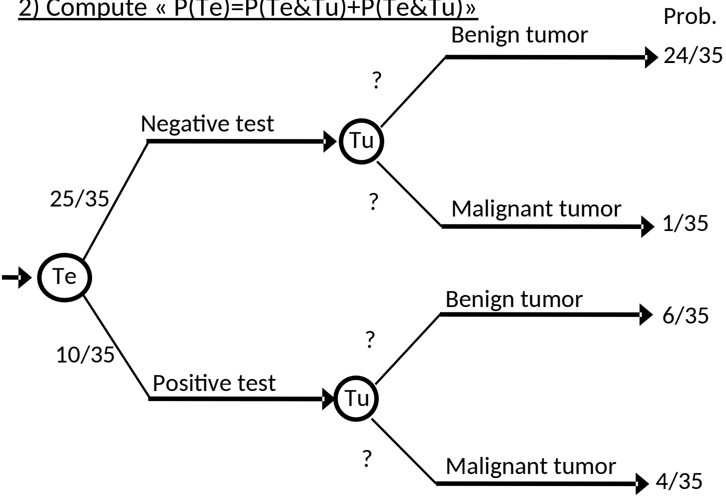


Analysis using a scenario tree

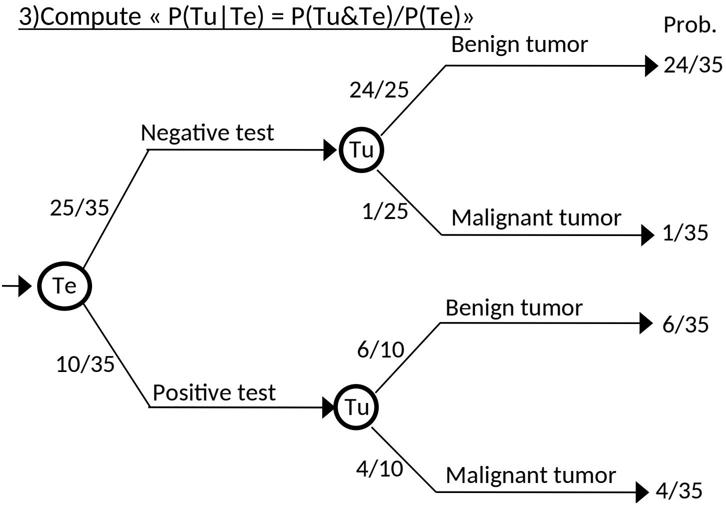


Analysis using a scenario tree

2) Compute « $P(Te)=P(Te\&Tu)+P(Te\&\overline{Tu})$ »



Analysis using a scenario tree

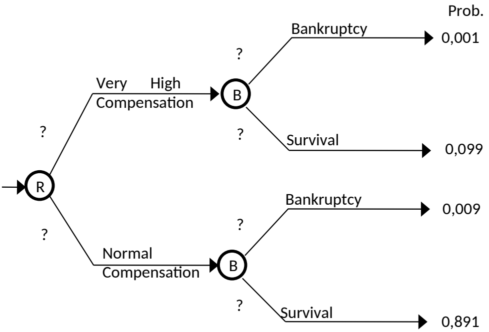


Independence between two events : $A \perp B$

The event B is independent of event A if knowing the status of A doesn't have an effect on the probability that B happens.

- «A independent of B» implies that «B independent of A»
- E.g. : Is the upcoming bankruptcy of a business independent of the compensation of its CEO ?

CEO's compensation	State of the company one year later	Number
Very High	Bankruptcy	1
Very High	Survival	99
Normal	Bankruptcy	9
Normal	Survival	891

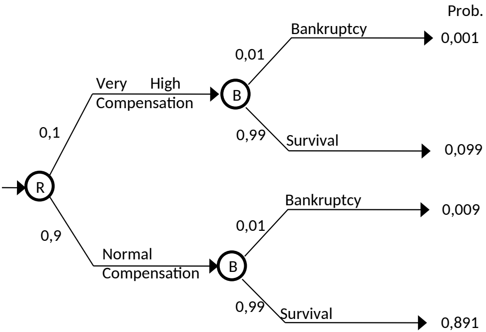


Independence between two events : $A \perp B$

The event B is independent of event A if knowing the status of A doesn't have an effect on the probability that B happens.

- «A independent of B» implies that «B independent of A»
- E.g. : Is the upcoming bankruptcy of a business independent of the compensation of a CEO ?

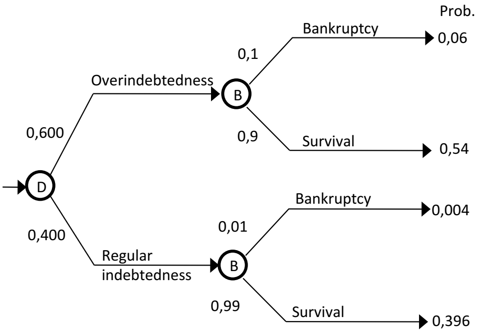
CEO's compensation	State of the company one year later	Number
Very High	Bankruptcy	1
Very High	Survival	99
Normal	Bankruptcy	9
Normal	Survival	891



Dependence between two events : $A \not\perp B$

The event A is dependent on the event B if knowing the status of A influences the probability that B happens.

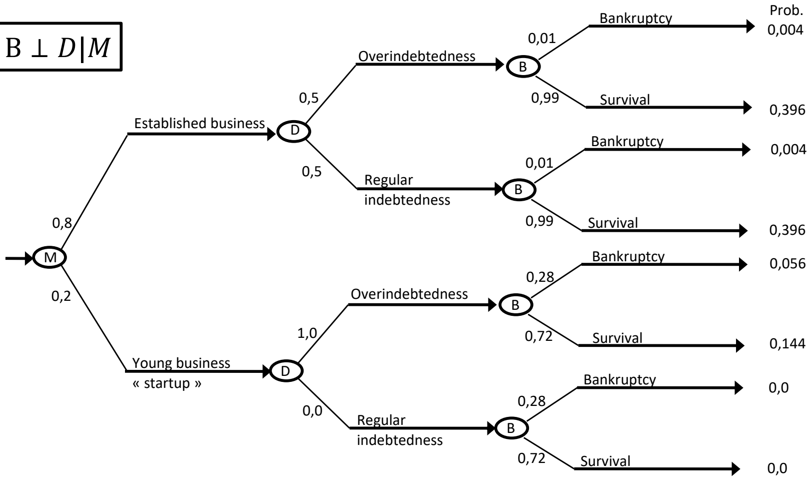
- «A is dependent on B» implies that «B is dependent on A»
- E.g. : Is the bankruptcy of a business dependent on the over-indebtedness of a company ?



Conditional independence : $A \perp B | C$

The relationship between over-indebtedness and bankruptcy can possibly be explained by the maturity of company :

$\boxed{B \perp D|M}$



Conditional independence : $A \perp C|B \nRightarrow A \perp C$

Conditional independence doesn't imply non-conditional independence. Take for example the following case :

- Consider throwing three unbaised six-faced dice, let A be the value of die #1, B the sum of dice #1 and #2, C the sum of the three dice.
- $C \perp A|B$ since if we know the sum of the first two dice, the value of die #1 doesn't provide us any additional information about the sum of the three dice.

$$
\begin{aligned}
P(C = k|A = i, B = j) &= P(C = k|B = j) \\
&= \begin{cases} 1/6 & \text{if } C - B \in \{1, \ldots, 6\} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

- $C \not\perp A$ since knowing the value of die #1 informs us of the result of the sum of the three : if $A = 6$ then $C > 8$.

Outline

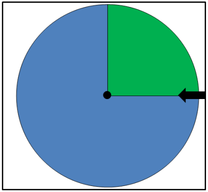
Determining a discrete probability - Verbal method

- You wish to learn the opinion of an expert on the probability that event A happens
- It is greatly discouraged to use the following method :
  - Ask «How probable is event A to happen ?»
  - Assign the probability according to the following table :

Answer	Value
Small chances	0,1
Not likely	0,15
I am Doubtful	0,25
There is a chance	0,5
It's possible	0,5
It's likely	0,65
There is a good chance	0,75
Most likely	0,85
Almost certain	0,95

Determining a discrete probability - Wheel of fortune

- You wish to learn the opinion of an expert on the probability that event A happens
- Consider two prizes of different values.
  - Prize 1 = Trip to Europe
  - Prize 2 = Nothing
- Present two bets :
  - First Bet :
    - He wins Prize #1 if A happens
    - He wins Prize #2 if A doesn't happen
  - Second Bet (wheel of fortune) :
    - He wins Prize #1 if the arrow stops on the green
    - He wins Prize #2 if the arrow stops on the blue
- Adjust the size of the green region so that the two bets are equivalent



Determining a continuous distribution

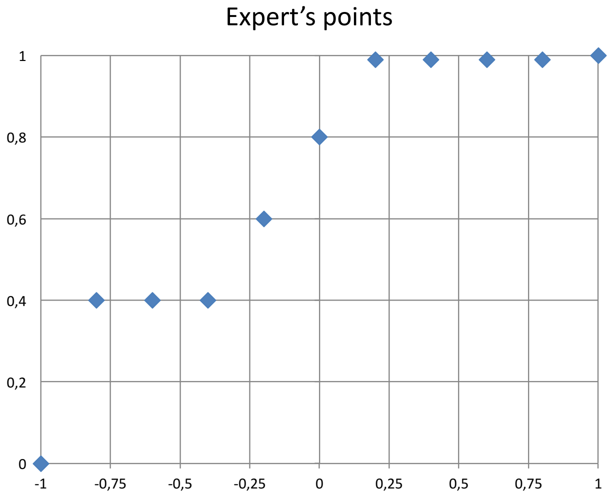
- You wish to know the opinion of an expert on the probabilities related to the realization of a random continuous variable Z.
  1. Determine the minimal & maximal value of the variable.
  2. Discretize the interval [min. value, max. value]
  3. For each value $z_1, z_2, \ldots$, use the wheel of probabilities to determine the probabilities $P(Z \leq z_1)$, $P(Z \leq z_2)$, ...
  4. Trace a continuous and et non-decreasing curve passing through the points found.
  5. Confirm the result with the expert.

Tracing the curve of a continuous distribution

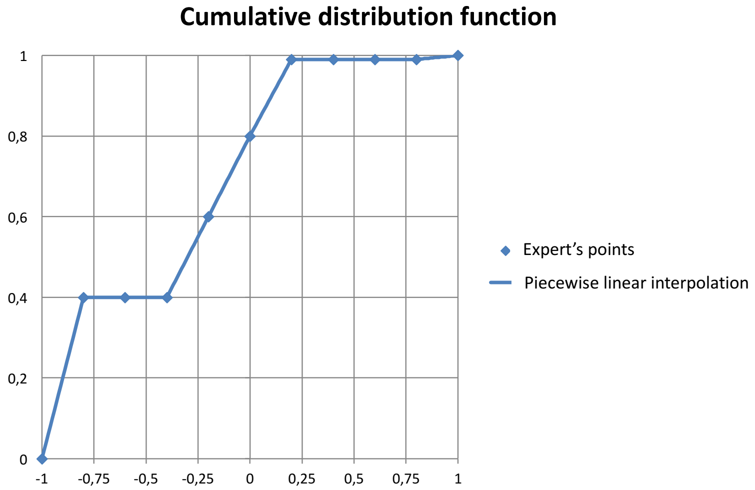
- Do a linear interpolation between known points
  - Advantage : this distribution respects the principle of simplicity (i.e «the most simple hypothesis is the most likely») since it is the one that maximizes the «disorganization» of Z while respecting the known percentiles.
- Fit a parameterized functional form
  - Advantage : the expert can correct the points to be coherent with the distribution they propose.
  - Advantage : allows for the calibration of abstract distribution parameters through the use of interpretable probabilities.

Tracing the curve of a continuous distribution

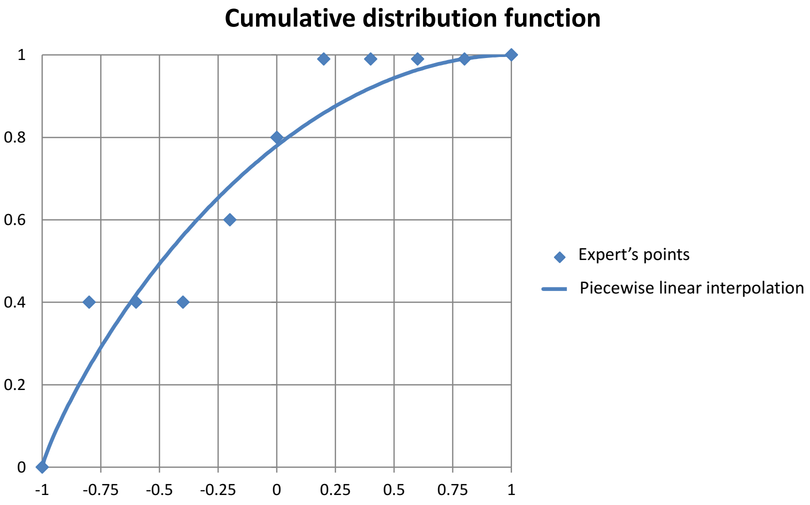
Expert's points	
a	P(X<=a)
-1	0
-0,8	0,4
-0,6	0,4
-0,4	0,4
-0,2	0,6
0	0,8
0,2	0,99
0,4	0,99
0,6	0,99
0,8	0,99
1	1



Tracing the curve of a continuous distribution



Tracing the curve of a continuous distribution



Mystery bag exercise (see Excel file)

- We present to an «expert» a bag containing a quantity of coins (1¢, 5¢, 25¢, 1$)
- Objective : Quantify the knowledge of this expert relative to the monetary value of the bag by following this procedure :
  - Establish the min & max of this value
  - Discretize the interval to obtain the values $\{z_1, z_2, \ldots, z_n\}$
  - Determine with the help of the wheel of fortune the amounts for which :
    $p_1 = P(\text{value of bag} \leq z_1)$, $p_2 = P(\text{value of bag} \leq z_2)$, $\ldots$, $p_n = P(\text{value of bag} \leq z_n)$.
  - Trace a curve passing through the points $(z_i,\ p_i)$.
- What is the expected value of the bag ?

Survey from session 1

Are you good «experts» ???

(see Survey Results)

Quality of a subjective distribution

Two criteria for assessing the quality :

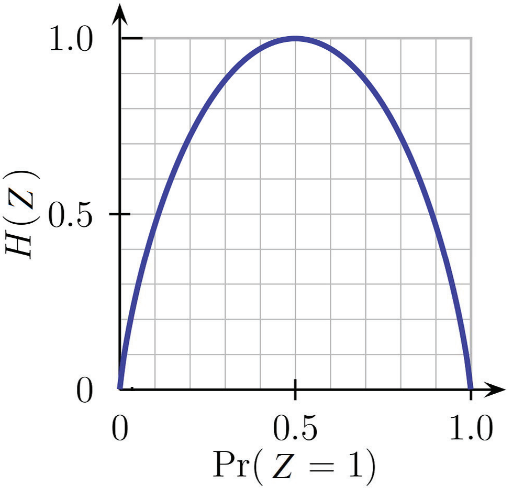
- The distribution should be «informative»
  - The prob. of an event should differ from 0.5, otherwise the expert doesn't know whether the event is more likely to happen or not.
  - The distribution should be concentrated and non uniform.
  - The amount of information is often considered inversely proportional to entropy :

$$H(Z) = \sum_i P(Z = z_i) \log_2(1/P(Z = z_i))$$

- The distribution should represent uncertainty <u>authentically</u>
  - In the long term, the events that the expert claimed to have a probability p should occur a proportion of the time that is close to $(100 * p)\%$.
  - This property is also called «good calibration»

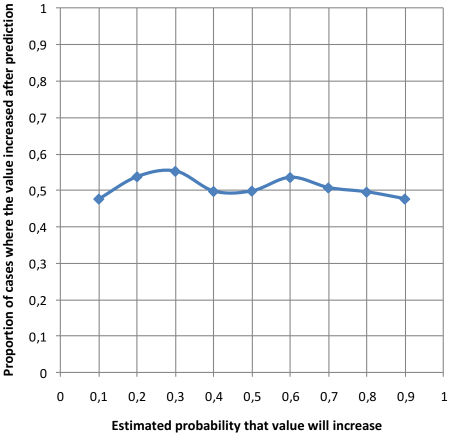
Example of Informativeness

Entropy of a $0-1$ (Bernoulli) random variable :



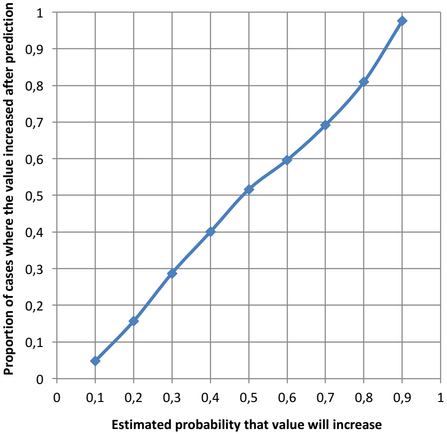
Example of Calibration I

Subjective probability that Microsoft's stock rises next day :



Example of Calibration II

Subjective probability that Microsoft's stock rises next day :



Determining a distribution over many variables

- You wish to know the opinion of an expert regarding the probabilities associated to the realization of many random variables W, X, Y, Z
  - Approach 1 : decompose using conditional probabilities

$$P(W, X, Y, Z) = P(W|X, Y, Z) \cdot P(X|Y, Z) \cdot P(Y|Z) \cdot P(Z)$$

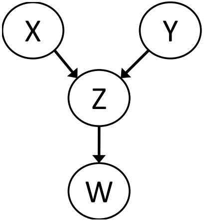
  - Difficulty : A distribution must be determined for each potential realization of the conditioned variables.

X	Y	Z	P(W=1|X,Y,Z)
0	0	0	?
0	0	1	?
0	1	0	?
0	1	1	?

X	Y	Z	P(W=1|X,Y,Z)
1	0	0	?
1	0	1	?
1	1	0	?
1	1	1	?

Determining a distribution over many variables

- Approach 2 : exploit the «local Markov property» of an influence diagram :
  - If we know neither more nor less than the value of all the variables that influences a variable X directly, then X is independent of all sets of variables that aren't downstream from it.
- Advantage : we can limit the amount of conditioned distribution that need to be estimated.
- Example :



$$P(W|X,Y,Z) = P(W|Z)\ , \quad P(Y|X) = P(Y)$$

$$\begin{aligned} P(W,X,Y,Z) &= P(W|X,Y,Z)P(Z|X,Y)P(Y|X)P(X) \\ &= P(W|Z)P(Z|X,Y)P(Y)P(X) \end{aligned}$$

Reduction of the estimation task

Consider that $W, X, Y, Z \in \{0, 1\}$, then the number of probabilities to estimate reduces from 15 to 8 when we exploit the independences described in the precedent diagram.

- With independence :
  - For $P(X)$, $P(Y)$: estimate $P(X = 1)$ & $P(Y = 1)$
  - For $P(Z|X, Y)$: estimate $P(Z = 1|X = 0, Y = 0)$, $P(Z = 1|X = 0, Y = 1)$, $P(Z = 1|X = 1, Y = 0)$ & $P(Z = 1|X = 1, Y = 1)$
  - For $P(W|Z)$: estimate $P(W = 1|Z = 0)$ & $P(W = 1|Z = 1)$
- Without independence :
  - For $P(X)$: estimate a probability
  - For $P(Y|X)$: estimate 2 probabilities
  - For $P(Z|X, Y)$: estimate 4 probabilities
  - For $P(W|X, Y, Z)$: estimate 8 probabilities