Bayesian inference

Chapter 2

Disposition

- The product rule
- The sum rule
- Qualitative properties
- Numerical values
- Notations and Comments
- Summary

The product rule (1/2)

- Robot uses: $(AB \mid C) = F[(A \mid C),(B \mid AC)]$
- Following the desideratum of structural consistency we get equation (2.13)

$$F[F[x,y],z] = F[x,F[y,z]]$$

- Through derivation of F(x,y) we get a non-trivial solution for (2.13) by setting

$$\omega(F[x,y]) = \omega(x)\omega(y)$$

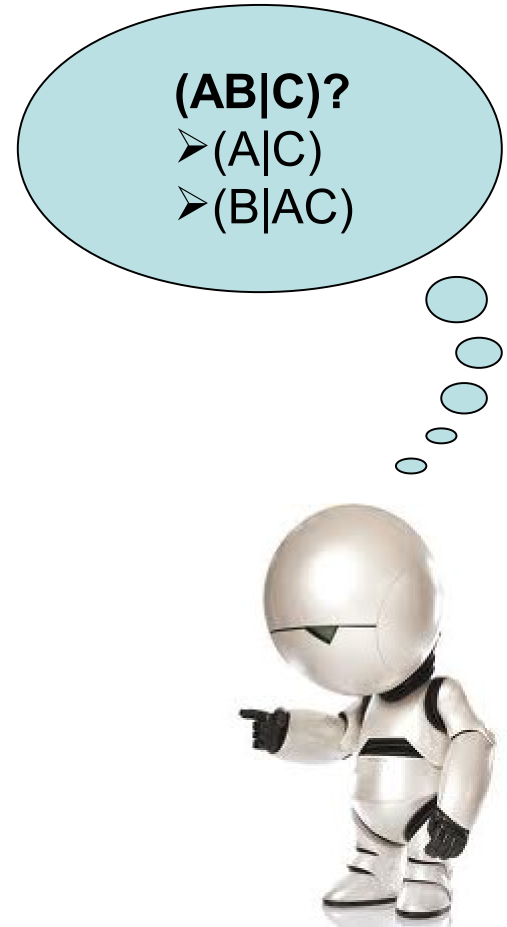
where

$$\omega(x) = \exp\left(\int^{x} \frac{dx}{H(x)}\right)$$

($H(x)$- arbitrary function that does not change sign)

- This leads to the **product rule**:

$$\omega(AB \mid C) = \omega(A \mid BC)\omega(B \mid C) = \omega(B \mid AC)\omega(A \mid C)$$



The product rule (2/2)

- Suppose A is true given C. Then

$$AB \mid C = B \mid C$$
$$A \mid BC = A \mid C$$
$$\omega\left(B \mid C\right) = \omega\left(A \mid C\right)\omega\left(B \mid C\right)$$

  which leads to ω(A|C)=1.
- Suppose A is impossible given C. Then

$$AB \mid C = A \mid C$$
$$A \mid BC = A \mid C$$
$$\omega\left(A \mid C\right) = \omega\left(A \mid C\right)\omega\left(B \mid C\right)$$

  which leads to ω(A|C)=0 (or ∞).
- ω(x) is continuous monotonic increasing (or decreasing) function on the interval [0,1] (or [1, ∞[).

  (because of desideratum II)

The sum rule (1/2)

$$\omega^m(A|C) + \omega^m(\overline{A}|C) = 1$$

- Let

  $$u = \omega(A|C), v = \omega(\overline{A}|C) = S(u)$$

  we then have

  $$S(0) = 1, S(1) = 0$$

- Applying the product rule we get the (2.40)

  $$\omega(A|C)S\left[\frac{\omega(A\overline{B}|C)}{\omega(A|C)}\right] = \omega(B|C)S\left[\frac{\omega(B\overline{A}|C)}{\omega(B|C)}\right]$$

- Jaynes now define a variable q(x,y) and a function J(q) as in (2.48-49). Through a rather vast derivation including geometric and Taylor expansion he gets a differential equation (2.57) with the solution (2.58)

  $$S = (1 - x^m)^{1/m}$$

- Using (2.58) we get (2.60):

$$\omega^m(A|C) + \omega^m(\overline{A}|C) = \omega^m(A|C) + [S(\omega(A|C))]^m = \omega^m(A|C) + \left[(1 - \omega^m(A|C))^{1/m}\right]^m = 1$$

The sum rule (2/2)

- We now define

  $$p(x) = \omega^m(x)$$

- Raising both sides of the product rule to the power of m (2.63) we get:

  $$p(AB \mid C) = p(AB \mid C)\,p(B \mid AC)$$ The product rule

  $$p(A \mid B) = \omega^m(A \mid B) = 0^m = 0$$ when A is impossible given B

  $$p(A \mid B) = \omega^m(A \mid B) = 1^m = 1$$ when A is certain given B

- Through some straight forward calculations (2.66) we get the **sum rule:**

  $$p(A + B \mid C) = 1 - p(\overline{A}\,\overline{B} \mid C) = \ldots = p(A \mid C) + p(B \mid C) - p(AB \mid C)$$

  (Note that there is an typo in the first step of (2.66))

- p(x) is a monotonic function.

$$p(A \mid B) = \omega^m(A \mid B)$$

Obeys the product rule and sum rule for any choice of a positive constant m!

Qualitative properties

- Lets assume that we have the prior information C=(A→B)

$$p(AB \mid C) = p(A \mid C)$$
$$p(A\overline{B} \mid C) = 0$$

- *If A is true then B is true* correspond to the product rule

$$p(B \mid AC) = p(AB \mid C) / p(A \mid C) = 1$$

  and similarly for *if B is false then A is false*

- *If B is true then A becomes more plausible* correspond to the product rule on the form

$$p(A \mid BC) = p(A \mid C)\, p(B \mid AC) / p(B \mid C) = p(A \mid C) / p(B \mid C) \geq p(A \mid C)$$

Logic is only a special case of the rules for plausible reasoning!

Numerical values (1/2)

- Two propositions, A_1 and A_2, are *mutually exclusive* given B if

$$p\left(A_1 A_2 \mid B\right) = 0$$

- The propositions $\{A_1, \ldots A_n\}$ are *exhaustive* given B if one and only one of them must be true:

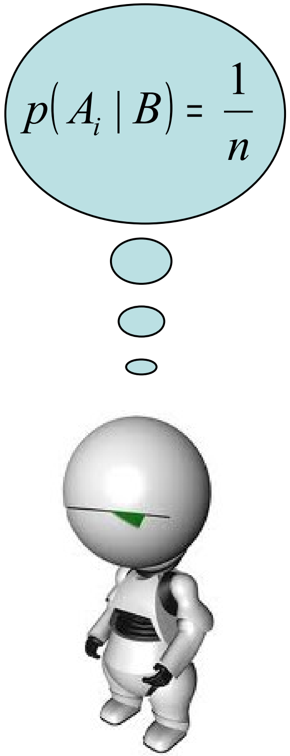
$$\sum_{i=1}^{n} p\left(A_i \mid B\right) = 1$$

- *Principle of indifference*: if $\{A_1, \ldots A_n\}$ are exhaustive and that the information of B is indifferent between A_i and A_j for all (i,j) then through desideratum IIIc we get (2.95):

$$p\left(A_1 \mid B\right) = p\left(A_2 \mid B\right) = \ldots = p\left(A_n \mid B\right) = \frac{1}{n}$$

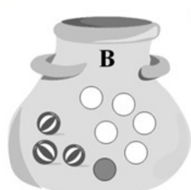
this result can be reached intuitively but Jaynes urge us to follow his reasoning which rearranging the propositions.

  - This holds independent of the choice of the function p(x).



Numerical values (2/2)

Probability of getting a white ball is $\frac{5}{9}$



- We shall now look at the plausibility x=A|B to be dependent of the fixed value of p and focus on the quantity of p which will henceforth be called the **probability**.
- The acctual value of the plausibility A|B is not necessary.
- *Bernoulli urn problem*: What is the probability that a ball drawn randomly from the urn marked B is white?
- Let the propositions $\{A_1, \ldots A_n\}$ are mutually exclusive and exhaustive given D and the proposition C can be defined to be true for m first of them (if not we can always rearrange the propositions A_i). Then, applying the sum rule we get:

$$p(C \mid D) = p(A_1 + A_2 + \ldots + A_m \mid D) = \sum_{i=1}^{m} p(A_i \mid D) = \frac{m}{n}$$

Notations and Comments

- The theorems established in this chapter hold for finite set and we shall let
  - P(A|B) denote the probability when the arguments are propositions.
  - f(r|np) denote the probability when the arguments are numbers.
- The theory is subjective because it depends on the analysts' knowledge and objective because it is independent on her personality.
- Gödel's theorem states that no mathematical system can prove its own consistency, but:
  - If the robot is set to calculate the probabilities P(B|E) and E={E_1, … E_n} and these include some contradiction the robot's program will crash!

Summary

- $P(A|C)$ denotes the *probability* of A given the knowledge of C.
- $P(A|C)=1$ if A is *certain* given C.
- $P(A|C)=0$ if A is *impossible* given C.
- $P(A|C)+P(\bar{A}|C)=1$
- $P(AB|C)=P(A|C)P(B|AC)$ – the *product rule*.
- $P(A+B|C)=P(A|C)+P(B|C)-P(AB|C)$ – the *sum rule*.
- If $C=(A\rightarrow B)$ the
  - $P(B|AC)=1$ is the same as the logic deduction A is true hence B is true.
  - $P(A|BC)\geq P(A|C)$ means that is B is true A is more plausible.
- $P(A_1,A_2|B)=0$ then A_1 and A_2 are *mutually exclusive*.
- $P(A_i|B)=1/n$ for $i=1,\ldots,n$ when $A_1,\ldots,A_n$ are *exhaustive* and B is *indifferent* between $A_1,\ldots,A_n$.

DON'T PANIC