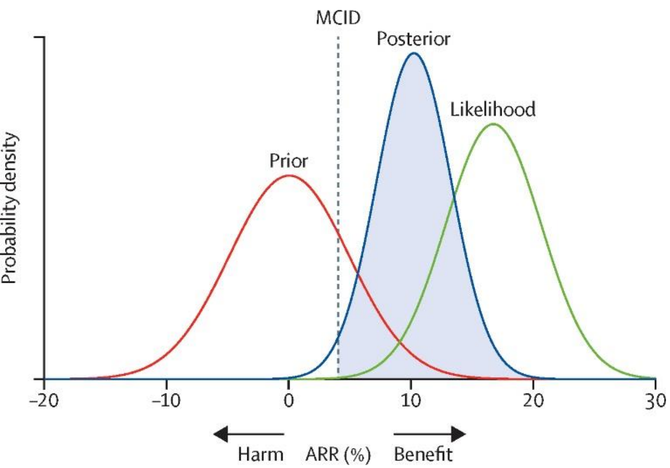
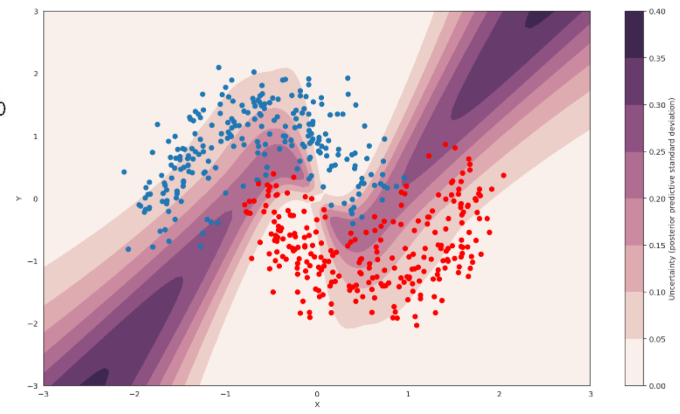
Bayesian methods

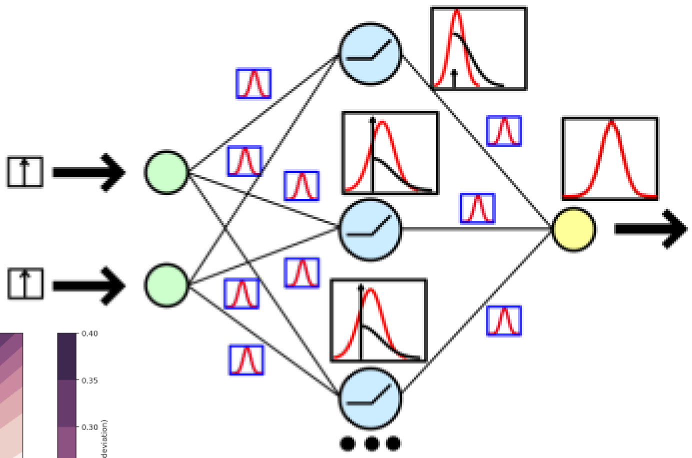
A way to incorporate our, *expectations and prior knowledge*into our decision making

or

How can we answer a question from very limited and uncertain data?







Gusztáv Schay
schay.gusztav@med.semmelweis-univ.hu

The dichotomic decision tree

Question (from real life)

↓

Transform into Y/N form (maybe dissect)

H0 setup

We use a well known situation as our assumption. (some reference)

↓

We set up the threshold, i.e. how much we "stick to H0" -> this is called the **significance level.**

α or α_{max}

↓

Calculate the conditional probability P(at least such a deviation | H0)

↓

Decision

P < sign.

P >= sign.

H0 is rejected

H0 is kept

Is not rejected

the decision is probabilistic

There is always the possibility of making a mistake, but we don't have a better way. Uncertainity is built into Nature. But, *on the long run* we can be optimal.

the Decision

		H0 is kept	**H0 is rejected**
The Truth (maybe at some time we get to know it)	H0 is true	correct decision	Type I. error α
	H0 is false	Type II. error β	correct decision

„The best MD is the pathologist, knowing everything correctly. It is a pity that it is too late…"

We can set the tolerable level of α with the significance level.

Power of a test: If we have a known alternative to H0 which is the true setting, then with what probability will be the (wrong) H0 get rejected.

All our *statements and probabilities*are conditional!
„hypo thesis" = „the one set at the bottom"

$P(A|H_0,C)$: not just H0, but also the (experimental) conditions have to be considered.

C is not easy to determine!
-> we need to have *inter-subjective consensus.*
(everyday has to accept certain starting points and axioms to be able to have science, etc.)

But: in **C** we also have the subjectum, individuum included!

P(rain| **C**) orP(the operation was succesful| **C**)

We use intuitive, in *frequentist way not definable* „probability" terms.

„I think today will be *probably* very nice weather."
„Don't worry, the operation *must have been* all right"

-> These considerations should be able to get included into our decision making!

Take care, here we actuallty have
P(A|C) and k_A |C
but we usually omit the C in the formulas..

Frequentist definition: $P_A = \lim_{N \to \infty} k_A$

subjective probability:

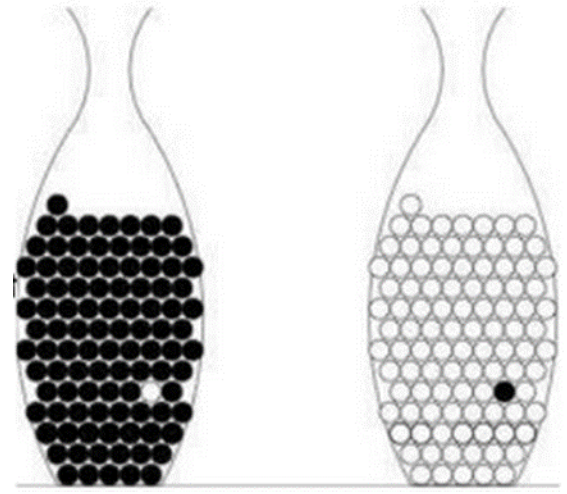
P_A = „***Degree of belief*** *in the happening of A*"

-> the intersubjective agreement is mandatory!
It is only possible to define a reasonable P-value IF
P(A|C)will be assessed *by everybody to the same level who knows C*.
(P will be set to the same number by everybody being in the exactlt same situation C)

-> all collegeues -reading the same lab report, having the same professional knowledge and training -would set the probabilities on the subsets of Ω event space (e.g. the possible diagnoses!) to the same values.

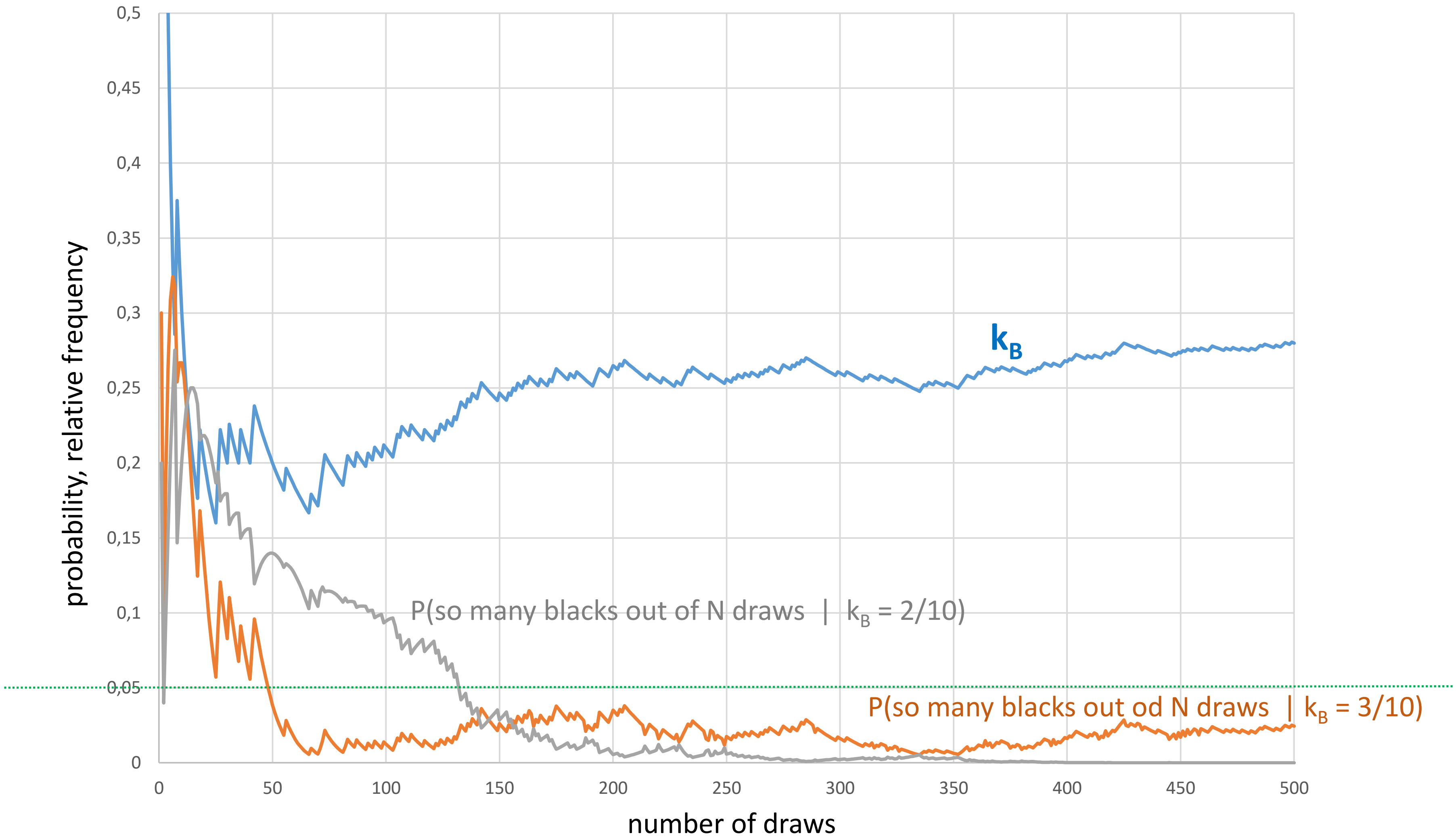
how can we compare our „probabilities”? Is then everything relative?

No, we use a common base, the **urn modell**.
every probability can be expressed by an urn.



P(A|C) = P(Black drawn| we play with the given Urn)

The dichotomic decision is sometimes problematic. (rejecting all possibilities, or accepting all…)



The Truth: we play with a 3/10 Urn.

The subjective probability can be defined exactly with mathematical rigor and axioms. Following the laws of logic and common sense it is possible to construct an **algebra for statements**.

$$\sim\sim a = a, \quad (1)$$

$$a \cdot b = b \cdot a, \quad (2) \qquad a \vee b = b \vee a, \quad (2')$$

$$a \cdot a = a, \quad (3) \qquad a \vee a = a, \quad (3')$$

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c = a \cdot b \cdot c, \quad (4)$$

$$a \vee (b \vee c) = (a \vee b) \vee c = a \vee b \vee c, \quad (4')$$

$$\sim(a \cdot b) = \sim a \vee \sim b, \quad (5)$$

$$\sim(a \vee b) = \sim a \cdot \sim b, \quad (5')$$

$$a \cdot (a \vee b) = a, \quad (6) \qquad a \vee (a \cdot b) = a. \quad (6')$$

~ = not
. = or
v = and

(1) and 5 more form the above laws are independent, the rest can be deduced.

do NOT memorize them! it is enough to know they exist.

I. $(cb)|a = F(c|ba\ ,\ b|a)$
II. $\sim b|a = S(b|a)$

$F(x,y)$ and
$S(x)$ are functions to be determined

-> if we seek to find the simplest functions fulfilling the rules then we may use
$F(x,y) = F(x)*F(y)$, $F(x):=x$ and $S(x) := 1-x$.

-> this leads us back formally to the Kolmogorov axioms already known from frequentist calculus!

$b|a + \sim b|a = 1$
and
$a|a + \sim a|a = 1$
sure and impossible events or statements.

do NOT memorize them!
for us it is enough that there IS an exact definition

The Bayesian probability theory is broad:
it does NOT mandant the existence of infinite repeatability
BUT in that case if the frequentist probability exists, the Bayesian probability gives the same value. (due to intersubjective agreement)

There is no formal change in the calculus!

now we can calculate with the possibility of statements!

1) For every p ***statement*** *it must be true that*, $0 \leq P(p) \leq 1$
   if p is surely true then $P(p) = 1$
   if p is surely impossible then $P(p)=0$
2) if p and q are mutually exclusive statements then $P(p \text{ or } q) = P(p) + P(q)$

$P(\text{not-}p) = 1 - P(p)$

It is also possible to standardize the degree of belief

Everything is gamble

A gamble in *fair way* has the price equal to it's expected benefit.

A frequentist connection (and inter-personal agreement)

in a *fair game on the long run* both parties are in balance

we package out degree of belief into a behavioral game:

what *rational* gamble are we willing to take?

(rational: the balance is 0, we just play for the fun of the game)

-> it is similar to the urn model, but makes inter-personal comparison easier:

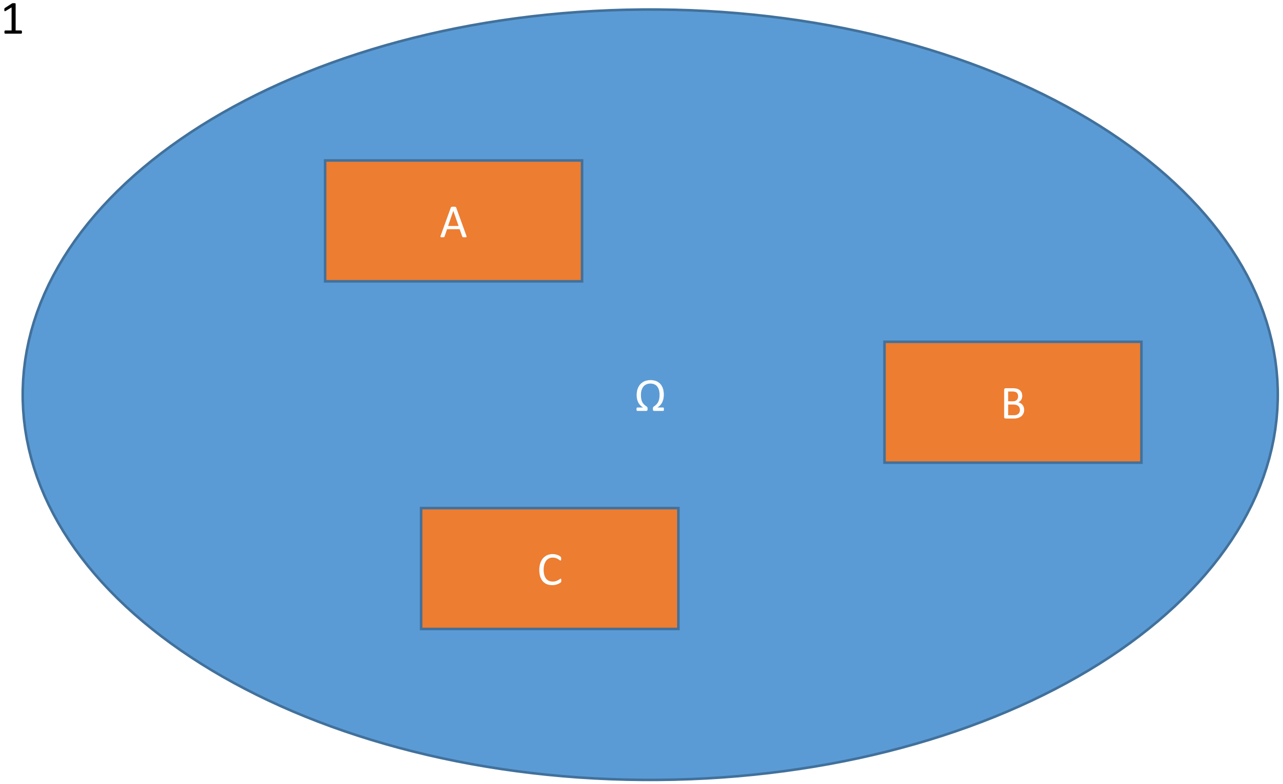
e.g. 1:100000 is the probability of a severe side effect of the drug.

1:x-means P=1/(1+x) is the probability of the gambled event/statement.

our estimated probabilities must be coherent:

$P(A)+P(B)+P(C) <= 1$

$P(\Omega)=1$



see the „Dutch book argument” -> only coherent probability estimations can yield a fair game. With incoherent probabilities somewhere will be sure gain and at other a sure loss.

Why do we need this?

There is now a way to assign probability to Hypotesis.

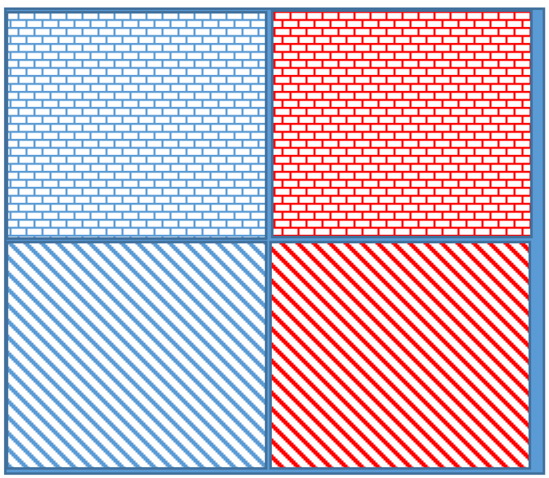
now we can assign probabilities to possible diagnoses!
and we can select, and arrange the diagnoses based on the probability distribution defined on the space of possible diagnoses.

we can even tell how many bits of information is gained by a certain diagnosis.
-> see information theory lecture

Conditional probability

P(A | B) = the probability of A ***given that*** condition B has occurred / is true.

e.g. : the patient has fever, *given that* she/he is COVID-19 infected.
I get the grade 5 in statistics *given that* I have attended every lecture.



We are interested only in a subset of Ω, and we need the relative frequency in that subset only.

P(blue | stripes) = the number of blues among the ones with stripes = 1 blue AND stripes / 2 stripes = ½

remarks:

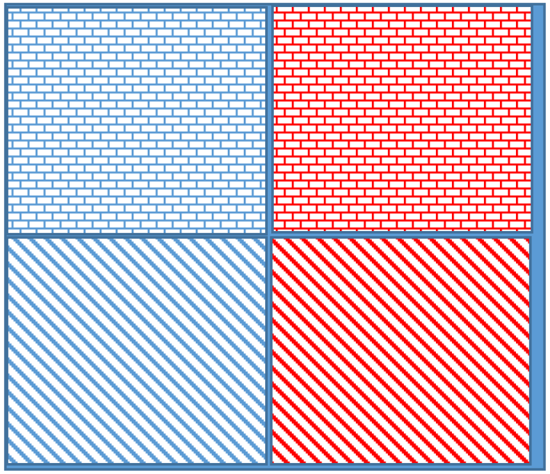
1. For independent A and B we have P(A|B=P(A)
2. For any two events P(AB) = P(A|B)*P(B) Bayes-theorem, or product law

Bayes theorem

conditional probability

$$P(A|H) = \frac{P(A \cdot H)}{P(H)}$$

we have two “events” A,H
H can now be a hypothesis!



P(AH) = P(blue AND stripes)
P(H) = P(stripes)

P(A|H) = P(blue, if we know it is with stripes)

now rearrange: P(H) = P(AH)/P(A|H)
-> probability of H can be calculated if we know the joint and the conditional probabilities.

Bayes theorem

the conditional probability $$P(A|H) = \frac{P(A \cdot H)}{P(H)}$$

if P(A) and P(H) exists, then the joint probability can be calculated in two ways:

$$\mathrm{P}(\mathrm{A} \cdot H) = P(A|H) * P(H) = P(H|A) * P(A)$$

we can rearrange: $$P(H|A) = \frac{P(A|H) * P(H)}{P(A)}$$

thus we have “flipped” the conditional probability (now A is the condition)

This enables us to calculate the probability of H **after** we observed the event A happening!

posterior probability

Take care: for P(A) we need to know EVERY way in which it can happen!

example: (you do NOT have to be able to repeat it!)

$$P(U|A) = \frac{P(A|U) * P(U)}{P(A)}$$

Let the probability of our patient having gastric ulcer be P(U)=15%, thus this is our degree of belief now. (could be just the prevalence if there is no better idea)

We demand a PCR test for helicobacter pylori, which is in 92% positive in gastric ulcer. We get a positive result.

We know the sensitivity=0.95, specificity=0.91 for this test, while the populational prevalence of h.pylori infection is 2%.

	+	-	
h.pylori infected	1.90%	0.10%	2%
not infected	8.82%	89.18%	98%

10.72% is the probability of a + test result independently of any other condition

P(A), i.e. the + result can happen in ***two ways*** *(mutually exclusive)*

- U is true
- U is false

$P(A)=P(AU)+P(A\bar{U})$

thus: $P(A) = P(U)*P(A|U) + P(\bar{U})*P(A|\bar{U})$

if the patient does not have gastric ulcer, we can use the general data of the PCR test

$$P(U|A) = \frac{0.92 \ * 0.15}{0.15 * 0.92 + (1 - 0.15) * 0.1072} = 0.602 = 60.2\%$$

posterior probability

thus **AFTER** the positive test the degree of belief regarding gastric ulcer is greatly increased

-> this is similar to the PPV, NPV calculations.

-> prevalence has a great influence

If we only have two mutual exclusive events then the calculation is simple. This is used extensively in epidemiology, as P(A), here the existence of the Disease, can happen only through two ways:
Diseased/Helathy and Test+/-

$$LR_+ = \frac{P(+|Diseased)}{P(+|Healthy)} \qquad O_{disease} = \frac{P(Diseased)}{P(not-Diseased)} = \frac{P(Diseased)}{P(Healthy)}$$

$$O_{diseased\,|+} = \frac{P(Diseased|+)}{P(Healthy|+)} = \frac{\frac{P(Diseased\ and\ +)}{P(+)}}{\frac{P(Healthy\ and+)}{P(+)}} = \frac{P(Diseased\ and+)}{P(Healthy\ and+)} =$$

$$\frac{P(+|Diseased) * P(Diseased)}{P(+|Diseased) * P(Healthy)} = LR_+ * O_{disease}$$

Same way for the calculation for LR_

$P(Diseased|+) * P(+) = P(Diseased\ and\ +) = P(+|Diseased) * P(Diseased)$

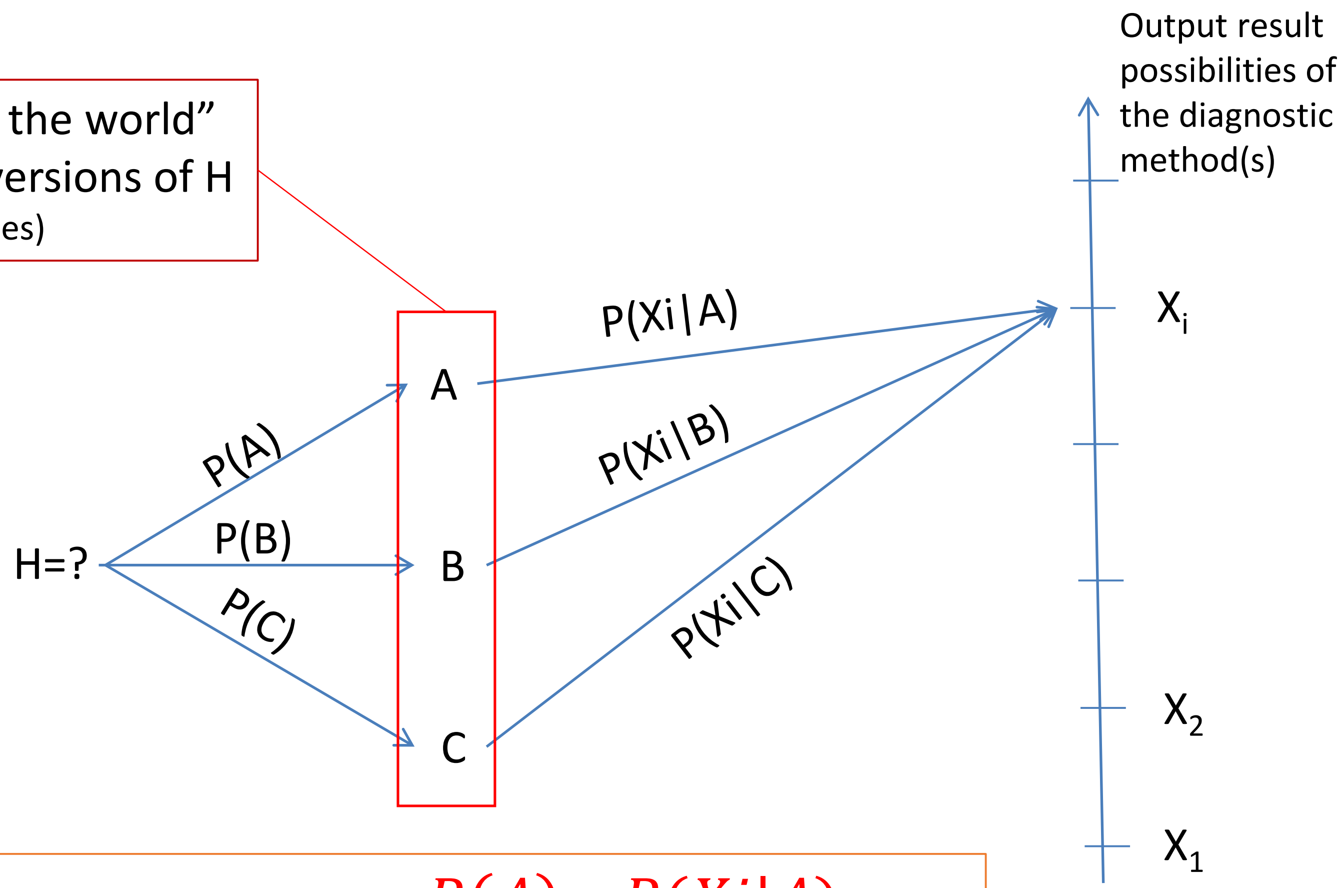
we have a lot of conditional probabilities, where one can be a hypothesis as well.

prevalence = D/ALL			incidence = NEW cases over t time / number at risk		
se	**+|D**	**prevalence indep**	incidence RATE = incidence / t time		
sp	**-|H**		RR_D	(D|R+)/(D|R-)	PPV/(1-NPV)
false neg rate 1-se	**-|D**		RR_H	(H|R+)/(H|R-)	(1-PPV)/NPV
false pos rate 1-sp	**+|H**		LR+	+|D / +|H	se/(1-sp)
PPV	**D|+**	**prevalence DEP**	LR-	-|D / -|H	(1-se)/sp
NPV	**H|-**		OR_D	(O_D|R+) / (O_D|R-)	RR_D / RR_H
false alarm rate 1-PPV	**H|+**		OR_H	O_H|R+ / O_H|R-	RR_H / RR_D
false reassurance rate 1-NPV	**D|-**		O_D = D/H	O_D_post = O*LR	
D: diseased			R+ risk factor present		O = p/1-p
H: healthy			R- risk factor NOT present		p = O/1+O
+/- Test result					

diagnostic test cheat sheet (c) G.Schay

With the Bayes theorem we can do reverse inference

„states of the world"
possible versions of H
(like diagnoses)



$$P(A|X = Xi) = \frac{P(A) * P(Xi|A)}{\sum_{k=A,B,C} P(k) * P(Xi|k)}$$

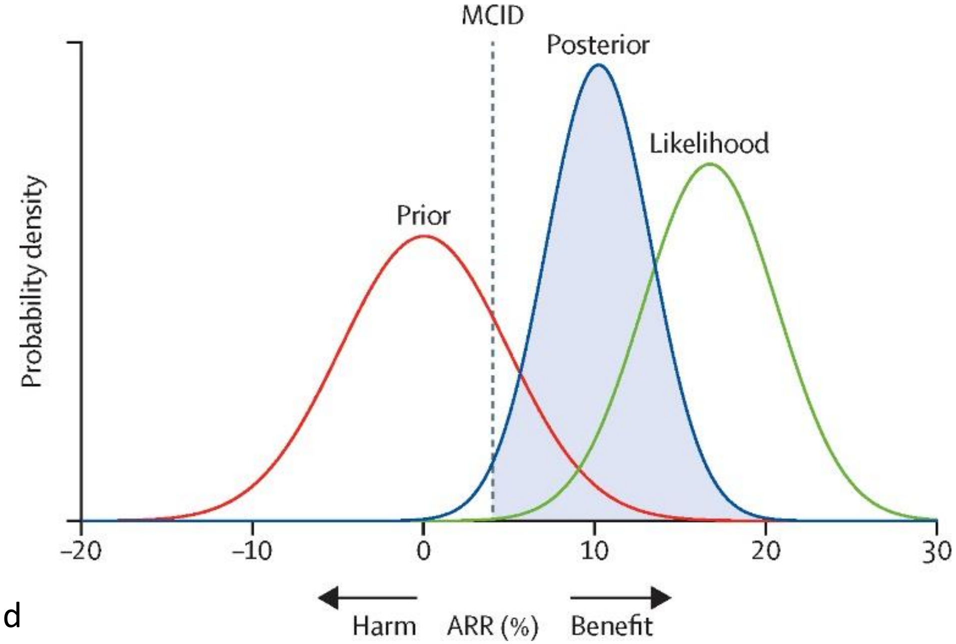
it is also possible to use a continuous H-space

$$\boldsymbol{P(A|X=Xi)} = \frac{\boldsymbol{P(A)} * P(Xi|A)}{\sum_{k=A,B,C} \boldsymbol{P(k)} * P(Xi|k)}$$

$$\boldsymbol{f(h|x)} = \frac{\boldsymbol{f(h)} * P(x|h)}{\int_h \boldsymbol{f(h)} P(x|h) dh}$$

We can update the prior distribution using the output of experiments, diagnostics to get the **posterior distribution**.

we must know the "forward" conditional probabilities (Likelihood).

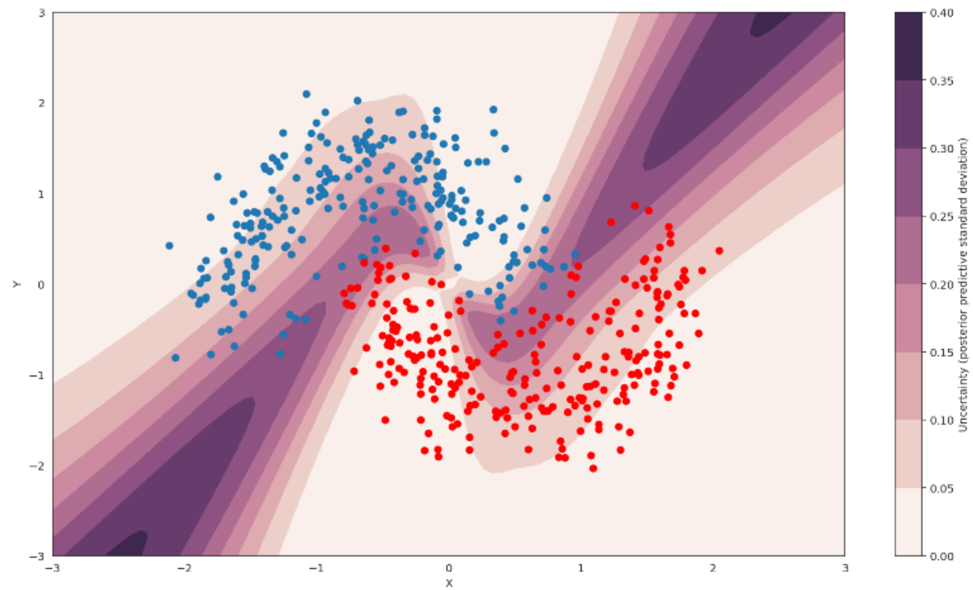


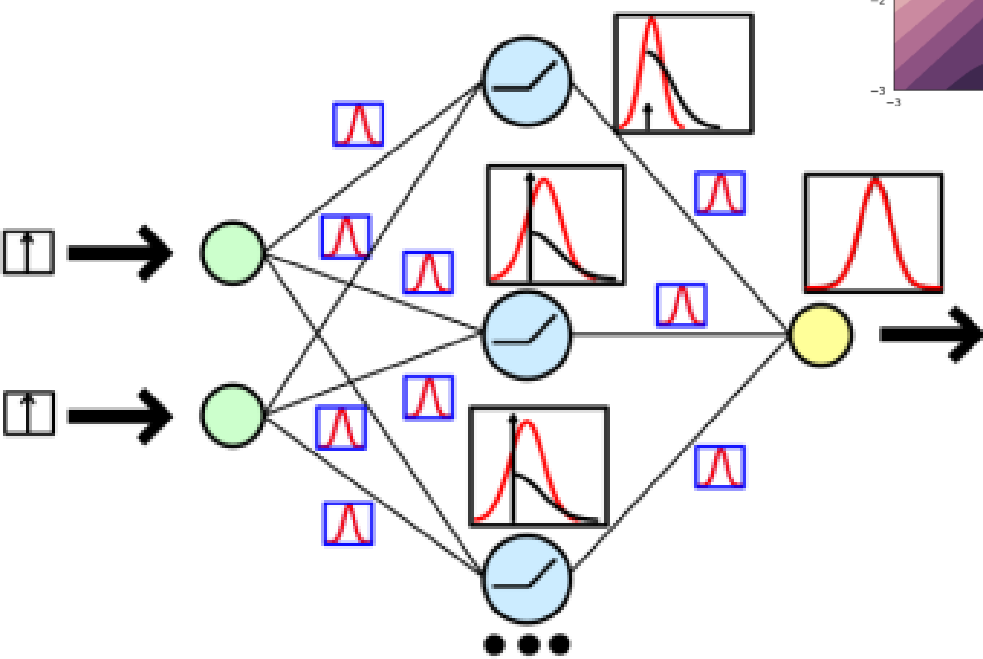
Remark: there are two important conditions:

- The prior is existent (with the Bayesian definition it is likely to be OK)
- we can calculate or at least estimate the Likelihoods, and cover the full Ω.

Medical applications

Decision supporting systems
Categorization
Research
Artificial Intelligence
(neural networks)





Decision supporting systems – decision theory

The Bayesian method helps to get insight by updating the prior after new experimental results arrive.

how to decide?

benefit

Decision -> Benefit?

We make the decision which maximizes the expected benefit.

for **possibilities** we have preferences for which we take two axioms as base:

completeness: $1 \preccurlyeq 2 \ \ or\ 2 \preccurlyeq 1$

transitivity: $1 \preccurlyeq 2 \ \ and\ 2 \preccurlyeq 3 \ \Longrightarrow 1 \preccurlyeq 3$

here the $\preccurlyeq$ is a relation in which we a re willing to "pay" to get the greater benefit.

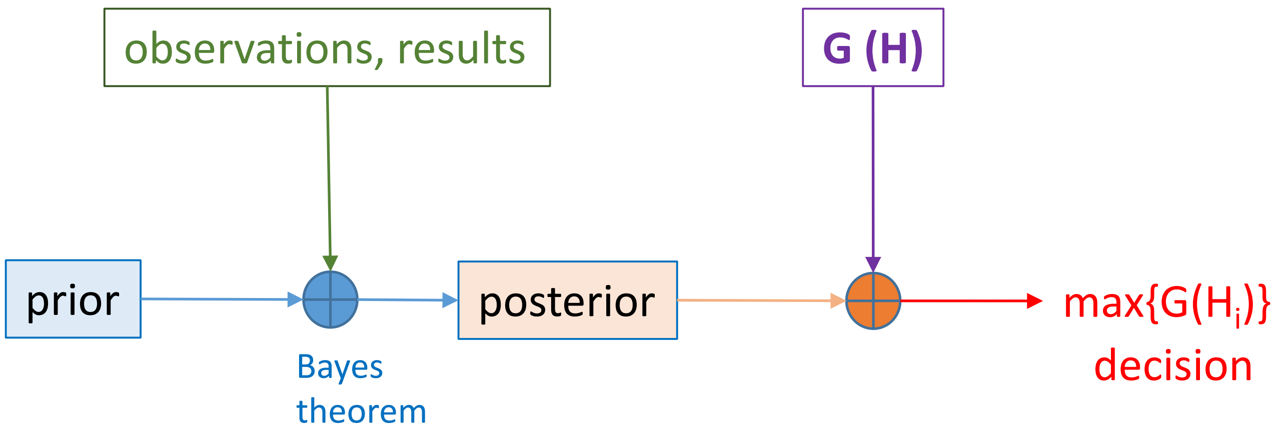
We may not know every possibility, but they can be *racionally extended* if new possibilities (therapies) become available.

the preferences can be mathematically expressed by a **utility function (G)** which renders numbers to possibilities (it can simply be the expected benefit, but can be more complex.)

vNM (von Neumann-Morgernstern) expected utility theory: G can be the "price" of a "lottery"

-> more in the advanced course

Bayesian decision making graph



-> more in the advanced course