Non-Bayesian decision making

Václav Hlaváč

Czech Technical University, Faculty of Electrical Engineering
Department of Cybernetics, Center for Machine Perception
121 35 Praha 2, Karlovo nám. 13, Czech Republic

hlavac@fel.cvut.cz, http://cmp.felk.cvut.cz

Lecture plan

- Penalties and probabilities which do not suffice for Bayesian task.
- Task formulation of prototype non-Bayesian tasks.
- Unified formalism leading to a solution—the pair of dual tasks of linear programming.
- Solution to non-Bayesian tasks.

Courtesy: Vojtěch Franc.

BAYESIAN TASK (rehearsal)

Bayesian task of statistical decision making seeks for

- sets X, K and D,
  function $p_{XK} \colon X \times K \to \mathbb{R}$ and
  function $W \colon K \times D \to \mathbb{R}$
- a Bayesian strategy $Q \colon X \to D$ which minimizes the Bayesian risk

$$R(Q) = \sum_{x \in X} \sum_{k \in K} p_{XK}(x,k)\, W(k, Q(x)) \,.$$

Typical use: minimization of the probability of a wrong classification.

Bayesian approach, limitations

Despite the generality of Bayesian approach, there are many tasks which cannot be expressed within Bayesian framework. Why?

- It is difficult to establish a penalty function. E.g., it does not assume values from the totally ordered set.
- A priori probabilities $p_K(k)$, $k \in K$, are not known or cannot be known because k is not a random event.
- Conditional probabilities $p(x|k)$ are difficult to express.

Non-Bayesian formulations can be natural for practical tasks

- Even is the case in which events are random and all involved probabilities are known, it is sometimes of advantage to approach the problem as a non-Bayesian one.
- In a practical tasks, it can be more intuitive for a customer to express desired classifier properties as allowed rate of false positives (false alarms) and false negatives (overlooked danger).

Example:

In quality check of electronic components (e.g., tantalum capacitors), it is common for the customers to express allowed number of rejects (false negatives, overlooked danger) in pieces per million (ppm).

Way out (only in several special cases): Non-Bayesian formulations

Good news

- There are several practically useful non-Bayesian tasks for which a solution similar to Bayesian tasks exist.
- These non-Bayesian tasks can be expressed in a general framework of linear programming in which the solution is easy and intuitive.

Bad news

- The class of non-Bayesian tasks covers only a subset of possible tasks.
- Nothing can be said about the task and its solution if it does not belong non-Bayesian tasks.

Penalty function in Bayesian decision making

- Decision is rated by a real number which corresponds to a penalty function value.
- The quality of the decision has to be expressed in ‘compatible units’.
- Values of a penalty function have to constitute an ordered set. The addition and multiplication has exist for this set.

Problems due to penalty function

'Minimization of the mathematical expectation of the penalty' requires that the penalty assumes the value in the totally ordered set (by relation $<$ or $\geq$) and multiplication by a real number and addition are defined.

An example—Russian fairy tales hero

When he turns to the left, he loses his horse, when he turns to the right, he loses his sword, and if he turns back, he loses his beloved girl.

Is the sum of p_1 horses and p_2 swords is less or more than p_3 beloved girls?

- Often various losses cannot be measured by the same unit even in one application.
- Penalty for false positive (false alarm) and false negative (overlooked danger) might be incomparable.

Example: decisions while curing a patient

$x \in X$ parameters (features, observations) measured on a patient

$k \in K = \{$healthy, seriously_sick$\}$

$d \in D = \{$do not cure, apply_a_drug$\}$

Penalty function $W : K \times D \to \mathbb{R}$

K \ D	do not cure	apply_a_drug
healthy	correct decision	small health damage
seriously_sick	death possible	correct decision

How to assign real number to a penalty?

Note: Health insurances do not have this problem . . .

A priori probability of situations

It can be difficult to find probabilities $p_K(k)$, $k \in K$,
which are needed for Bayesian formulation. Recall $p(x, k) = p(x|k)\, p(k)$.

Reasons:

1. Hidden state is random but $p_K(k)$, $k \in K$, are unknown. An object has not been analyzed sufficiently. Two options:

   (a) Formulate the task not in the Bayesian framework but in another one that does not require statistical properties of the object which are unknown.

   (b) She or he will start analyzing the object thoroughly and gets *a priori* probabilities which are inevitable for the Bayesian solution.

2. Hidden state is not random and that is why the *a priori* probabilities $p_K(k)$, $k \in K$, do not exist and thus it is impossible to discover them by an arbitrary detailed exploration of the object. Non-Bayesian methods must be used.

An example—enemy or allied airplane?

- Observation x describes the observed airplane.
- Two hidden states $\begin{cases} k = 1 & \text{allied airplane,} \\ k = 2 & \text{enemy airplane.} \end{cases}$
- The conditional probability $p_{X|K}(x|k)$ can depend on the observation x in a complicated manner but it exists and describes dependence of the observation x on the situation k correctly.
- *A priori* probabilities $p_K(k)$ are not known and even cannot be known in principle because it is impossible to say about any number α, $0 \le \alpha \le 1$, that α is the probability of the occurrence of an enemy plane.
- Consequently $p_K(k)$ do not exist since the frequency of experiment result does not converge to any number which we are allowed to call probability. The hidden state k is not a random event.

Beware of a pseudosolution

Refers to the airplane example.

- If *a priori* probabilities are unknown the situation is avoided by supposing that *a priori* probabilities are the same for all possible situations, e.g., the occurrence of an enemy plane has the same probability as the occurrence of an allied one.
- It is clear that it does not correspond to the reality even if we assume that an occurrence of a plane is a random event.
- Missing logical arguments are quickly substituted by a pseudo-argument by referencing, e.g., to C. Shannon thanks to the generally known property that an uniform probability distribution has the highest entropy.
- It happens even if this result does not concern the studied problem in any way.

Conditional probabilities of observations

Motivating example: Recognizing characters written by 3 persons

- Given:
  - X is a set of pictures of written characters x.
  - k is a name of a character (label), $k \in K$.
  - $z \in Z = \{1, 2, 3\}$ identifies the writer (this info is not known $\Rightarrow$ it is an unobservable intervention).
- Task: Recognize, which character is written in the picture x?

We can talk about the penalty function $W(k, d)$ and *a priori* probabilities $p_K(k)$ of individual characters.

We cannot talk about conditional probabilities $p_{X|K}(x \mid k)$ because the appearance x of a character depends not only on the character label but also on a non-random intervention (i.e., who wrote it).

Example: Recognizing characters written by three persons (2)

- We can speak only about conditional probabilities $p_{X|K,Z}(x \,|\, k, z)$, i.e., how a character looks like if it was written by a certain person.
- If the intervention z would be random and $p_Z(z)$ would be known for each z then it would be possible to speak also about probabilities

$$p_{X|K}(x \,|\, k) = \sum_{z=1}^{3} p_Z(z) p_{X|K,Z}(x \,|\, k, z) \; .$$

- However, we do not know how often it will be necessary to recognize pictures written by this or that person.
- Under such uncertain statistical conditions an algorithm ought to be created that will secure the required recognition quality of pictures independently on the fact who wrote the letter. The concept of *a priori* probabilities $p_Z(z)$ of the variable z cannot be used because z is not random and a probability is not defined for it.

Formulations of Non-Bayesian tasks
Introduction

- Let us introduce several known non-Bayesian tasks (and several new modifications to them).
- The whole class of non-Bayesian tasks has common features:
  - There is one formalism for expressing tasks and their solution (dual tasks of linear programming).
  - Similarly as for Bayesian tasks: The strategy divides the space of probabilities into convex cones.

Girl detection example (1)

- A person is characterized by the:
  - Observed height and weight,

    $$x \in X = \underbrace{\{\text{small}, \text{mid}, \text{tall}\}}_{\text{height}} \times \underbrace{\{\text{skinny}, \text{light}, \text{mid}, \text{heavy}\}}_{\text{weight}}$$

  - Class label = sex, $k \in K = \{\text{boy}, \text{girl}\}$.
- Apriori probabilities are unknown.
- Conditional probabilities are known:

$p_{X|K}(x|\text{girl})$

	skinny	light	mid	heavy
small	0.197	0.145	0.094	0.017
mid	0.077	0.299	0.145	0.017
tall	0.000	0.009	0.000	0.000

$p_{X|K}(x|\text{boy})$

	skinny	light	mid	heavy
small	0.011	0.005	0.011	0.011
mid	0.005	0.071	0.408	0.038
tall	0.000	0.016	0.255	0.169

Girl detection example (2)

Goal: Find a decision strategy $Q\colon X \to K$ allowing to detect girls among persons based on the observed height and weight.

The decision strategy Q splits the set X into two subsets:

$$\left.\begin{aligned} X_{\text{boy}} &= \{x \mid Q(x) = \text{boy}\} \\ X_{\text{girl}} &= \{x \mid Q(x) = \text{girl}\} \end{aligned}\right\} \text{ such that } \left\{\begin{aligned} & X_{\text{boy}} \cup X_{\text{girl}} = X\,, \\ & X_{\text{boy}} \cap X_{\text{girl}} = \emptyset\,. \end{aligned}\right.$$

Example of a strategy $Q(x) \to k$

small	girl	girl	girl	boy
mid	girl	girl	boy	boy
tall	boy	boy	boy	boy
	skinny	leight	mid	heavy

Neyman–Pearson task, two classes only (1)

- Observation $x \in X$, two states: $\begin{cases} k = 1 & \text{normal,} \\ k = 2 & \text{dangerous.} \end{cases}$
- The probability distribution of the observation x depends on the state k to which the object belongs. $p_{X|K}(x \mid k)$, $x \in X$, $k \in K$ are known.
- Given observation x, the task is to decide if the object is in the normal or dangerous state.
- The set X is to be divided into two such subsets X_1 (normal states) and X_2 (dangerous states), $X = X_1 \cup X_2$, $X = X_1 \cap X_2 = \emptyset$.

Neyman–Pearson task (2)

The observation x can belong to both states $\Rightarrow$ there is no faultless strategy.

The strategy is characterized by two numbers:

- Probability of the false positive (false alarm)
$\omega(1) = \sum\limits_{x \in X_2} p_{X|K}(x \,|\, 1)$.
- Probability of the false negative (overlooked danger)
$\omega(2) = \sum\limits_{x \in X_1} p_{X|K}(x \,|\, 2)$.

Girl detection example (3)

The decision strategy Q is characterized by the probability of:

$$\omega(\text{girl}) = \sum_{x \in X_{girl}} p_{X|K}(x|\text{boy}) \ldots \text{false alarm (boy recognized as a girl).}$$

$$\omega(\text{boy}) = \sum_{x \in X_{boy}} p_{X|K}(x|\text{girl}) \ldots \text{overlooked girl (girl recognized as a boy).}$$

ω(girl)

	skinny	light	mid	heavy
small	0.197	0.145	0.094	0.017
mid	0.077	0.299	0.145	0.017
tall	0.000	0.009	0.000	0.000

ω(boy)

	skinny	light	mid	heavy
small	0.011	0.005	0.011	0.011
mid	0.005	0.071	0.408	0.038
tall	0.000	0.016	0.255	0.169

Neyman–Pearson task (3)

A strategy is sought in the Neyman–Pearson task, i.e.,
a decomposition of X into $X_1 \subset X$, $X_1 \cup X_2 = X$, $X_2 \subset X$, $X_1 \cap X_2 = \emptyset$,
such that:

1. The conditional probability of the false negative is not larger than $\varepsilon \in (0, 1)$, which is a prescribed limit on the probability of overlooked danger.

$$\sum_{x \in X_1} p_{X|K}(x \,|\, 2) \leq \varepsilon \, .$$

2. A strategy has to have minimal conditional probability of the false positive

$$\sum_{x \in X_2} p_{X|K}(x \,|\, 1)$$

subject to the condition

$$\sum_{x \in X_1} p_{X|K}(x \,|\, 2) \leq \varepsilon \, .$$

Trivial example: men/women by weight only (1)

PŘÍKLAD (DETEKTOR ŽEN PODLE VÁHY)

$x \in X = \{M, L, S, T\}$ ... váha osoby

M ~ muší váha
L ~ lehká váha
S ~ střední váha
T ~ těžká váha

$y \in Y = \{$MUŽ, ŽENA$\}$ ... skrytý stav

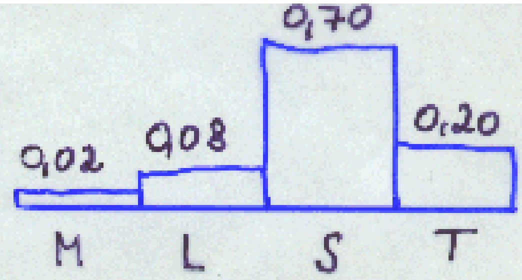
$\varepsilon = 0{,}20$ ... pravd. přehlédnutí ženy je max 20%

$Q: X \to Y$ ... Nayman-Pearsonova strategie

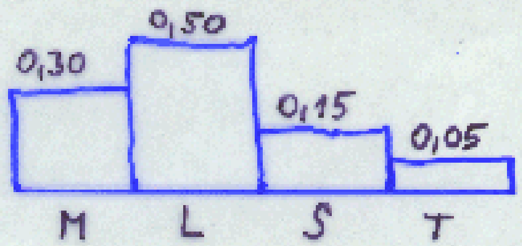
(max počet všech strategií $2^{|X|} = 16$)

Trivial example: men/women by weight only (2)

$p(x \mid y = \text{MUŽ})$



$p(x \mid y = \text{ŽENA})$



$$\min_{x \in X_{\text{ŽENA}}} \sum p(x \mid y = \text{MUŽ})$$

za podmínky

$$\sum_{x \in X_{\text{MUŽ}}} p(x \mid y = \text{ŽENA}) \le 0{,}20$$

PŘÍPUSTNÉ STRATEGIE: 1.–4.

OPTIMÁLNÍ STRATEGIE: 4.

					PŘEHLEDN. ŽENA	FALEŠNÝ POPLACH
1.	ŽENA	ŽENA	ŽENA	ŽENA	0,00	1,00
2.	ŽENA	ŽENA	ŽENA	MUŽ	0,05	0,80
3.	ŽENA	ŽENA	MUŽ	ŽENA	0,15	0,30
4.	ŽENA	ŽENA	MUŽ	MUŽ	0,20	0,10
5.	ŽENA	MUŽ	ŽENA	ŽENA	0,50	0,92
16.	MUŽ	MUŽ	MUŽ	MUŽ	1,00	0,00

Neyman–Pearson task (4)

Solution: Neyman–Pearson (1928, 1933)

The optimal strategy separates observation sets X_1 and X_2 according to a likelihood ratio by a threshold value θ

$$Q^* = \begin{cases} k = 1 & \text{if} \quad \frac{p_{X|K}(x\,|\,1)}{p_{X|K}(x\,|\,2)} > \theta\,, \\ k = 2 & \text{othewise.} \end{cases}$$

The rule is a special case of division of a space of probabilities into convex cones, i.e., it corresponds to Bayesian strategy.

Solving Non-Bayesian tasks using linear programming

Decision strategies $Q\colon X \to K$ can be equivalently represented by a function $\alpha\colon X \times K \to \{1, 0\}$ which satisfies

$$\sum_{k \in K} \alpha(x,k) = 1\,, \quad \forall x \in X\,,\ \alpha(x,k) \in \{0,1\}\,,\ \forall x \in X\,,\ \forall k \in K\,.$$

Example for the 'Girl detection' problem:

$\alpha(x, \text{boy})$

	skinny	leight	mid	heavy
small	0	0	0	1
mid	0	0	1	1
tall	1	1	1	1

$\alpha(x, \text{girl})$

	skinny	leight	mid	heavy
small	1	1	1	0
mid	1	1	0	0
tall	0	0	0	0

Neyman-Pearson's task solved by LP

To avoid integer programming, the relaxed stochastic strategy $\alpha\colon X \times K \to \langle 0, 1\rangle$ is introduced, which satisfies $\alpha(x,1) + \alpha(x,2) = 1$, $\forall x \in X$, $\alpha(x,k) \geq 0$, $\forall x \in X$, $\forall k \in \{1,2\}$.

<table>
<tr><th>Linear Programming relaxation</th><th>Original formulation</th></tr>
<tr><td>

$$\alpha^* = \operatorname*{argmin}_{\alpha} \sum_{x \in X} \alpha(x,2) p_{X|K}(x|1) \,,$$

subject to

$$\sum_{x \in X} \alpha(x,1) p_{X|K}(x,2) \quad \leq \quad \varepsilon \,,$$

$$\begin{aligned} \alpha(x,1) + \alpha(x,2) \quad &= \quad 1 \,, \quad \forall x \in X \,, \\ \alpha(x,1) \quad &\geq \quad 0 \,, \quad \forall x \in X \,, \\ \alpha(x,2) \quad &\geq \quad 0 \,, \quad \forall x \in X \,. \end{aligned}$$

</td><td>

$$Q^* = \operatorname*{argmin}_{X_1, X_2} \sum_{x \in X_2} p_{X|K}(x|1) \,,$$

subject to

$$\sum_{x \in X_1} p_{X|K}(x|2) \leq \varepsilon \,.$$

</td></tr>
</table>

Neyman-Pearson's task as LP (2)

The dual formulation of the Neyman-Pearson's task

$$(t(x)^*, \tau^*) = \operatorname*{argmax}_{t(x),\tau} \left(\sum_{x \in X} t(x) - \varepsilon\, \tau \right)$$

subject to

$$\begin{aligned} t(x) - \tau\, p_{X|K}(x|2) &\leq 0\,, \quad x \in X\,, \\ t(x) - p_{X|K}(x|1) &\leq 0\,, \quad x \in X\,, \\ \tau &\geq 0\,. \end{aligned}$$

Příklad J. Matase, pravděpodobnost ženy (1)

PŘÍKLAD: DETEKCE ŽEN (NEYMAN-PEARSON)

Rozhodovací strategie $Q : X \to C$ se charakterizuje pravděpodobnostmi

$$\varepsilon_1 = \sum_{x \in X_{\check{z}ena}} p(x \mid mu\check{z})$$ FALEŠNÝ POPLACH

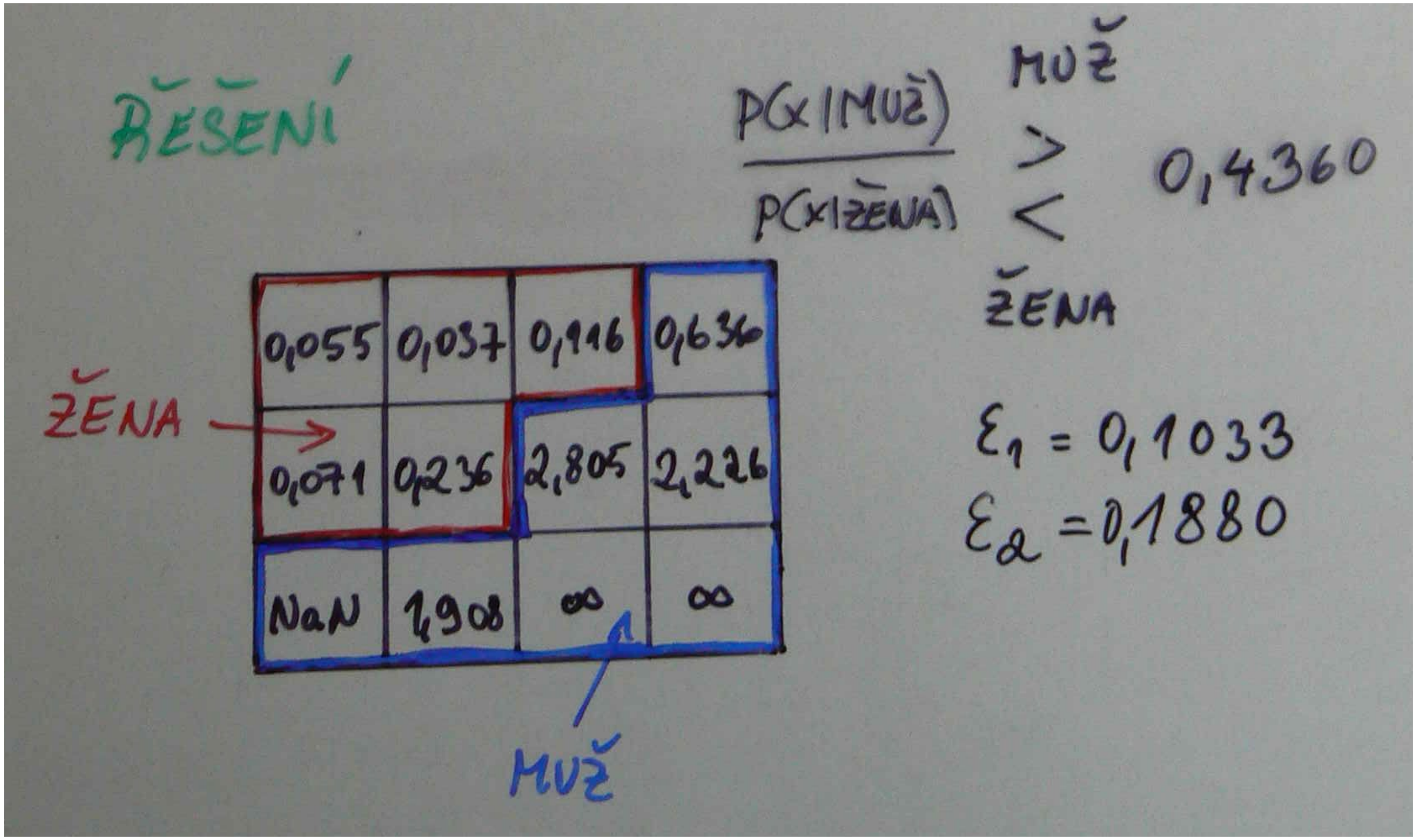
$$\varepsilon_2 = \sum_{x \in X_{mu\check{z}}} p(x \mid \check{z}ena)$$ PŘEHLÉDNUTÁ ŽENA

Hledáme $Q = \{X_{mu\check{z}}, X_{\check{z}ena}\}$ takovou, že

$$\min_{X_{mu\check{z}}, X_{\check{z}ena}} \sum_{x \in X_{\check{z}ena}} p(x \mid mu\check{z})$$

za podmínky $$\sum_{x \in X_{mu\check{z}}} p(x \mid \check{z}ena) \le 0{,}2$$

Příklad J. Matase, pravděpodobnost ženy (2)



Generalised Neyman-Pearson task for two dangerous states

$k = 1$ corresponds to the set X_1;

$k = 2$ or $k = 3$ correspond to the set X_{23}.

Seeking a strategy with the conditional probability of the false positives (overlooked dangerous states) both $k = 2$ and $k = 3$ is not larger than the beforehand given value ε.

Simultaneously, the strategy minimizes the false negatives (false alarms), $\sum\limits_{x \in X_{23}} p_{X|K}(x \,|\, 1)$ under conditions

$$\sum_{x \in X_1} p_{X|K}(x \,|\, 2) \le \varepsilon \,, \; \sum_{x \in X_1} p_{X|K}(x \,|\, 3) \le \varepsilon \,, \; X_1 \cap X_{23} = \emptyset \,, \; X_1 \cup X_{23} = X \,.$$

The formulated optimization task solved later in a single constructive framework.

Minimax task, introduction

- Selects the strategy according to the worst case scenario.
- Observations X are decomposed into subsets $X(k)$, $k \in K$, such that they minimize the number $\max_{k \in K} \omega(k)$.
- Consider a customer who demands that the PR algorithm will be evaluated by two tests in advance:

  Preliminary test (performed by the customer himself) checks the probability of a wrong decision $\omega(k)$ for all states k. The customer selects the worst state $k^* = \operatorname{argmax}_{k \in K} \omega(k)$.

  Final test checks only those objects which are in the worst state. The result of the final test will be written in the protocol and the final evaluation depends on the protocol content. The algorithm designer aims to achieve the best result in the final test.
- The problem has not been widely known for the more general case, i.e., for the arbitrary number of object states.

Minimax task (2)

- $x \in X$ are observable parameters.
- $k \in K$ are hidden states.
- $Q\colon X \to K$ is the sought strategy given by the decomposition $X = X_1 \cup X_2 \cup \ldots \cup X_{|K|}$
- Each strategy is characterized by $|K|$ numbers

$$\omega(k) = \sum_{k \notin X(k)} p(x|k) \ ,$$

i.e., conditional probabilities of a wrong decision under the condition that the correct hidden state is k.

Minimax task (3)

Minimax task formulation

A strategy Q^* is sought which minimizes

$$\max_{k \in K} \omega(k)$$

- The solution decomposes the space of probabilities into convex cones.
- Notice that the case $|K| = 2$ is the Neyman-Pearson task in which convex cones degenerate to 1D case – a likelihood ratio.
- Notice that the strategy belongs to the Bayesian family.

Wald task (motivation)

- A tiny part of Wald sequential analysis (1947).
- Neyman task lacks symmetry with respect to states of the recognized object. The conditional probability of the false negative (overlooked danger) must be small, which is the principal requirement.
- The conditional probability of the false positive (false alarm) is a subsidiary requirement. It can be only demanded to be as small as possible even if this minimum can be even big.
- It would be excellent if such a strategy were found for which both probabilities would not exceed a predefined value ε.
- These demands can be antagonistic and that is why the task could not be accomplished by using such a formulation.

Wald task (2)

Classification in three subsets X_0, X_1 and X_2 with the following meaning:

- if $x \in X_1$, then $k = 1$ is chosen;
- if $x \in X_2$, then $k = 2$ is chosen; and finally
- if $x \in X_0$ it is decided that the observation x does not provide enough information for a safe decision about the state k.

Wald task (3)

A strategy of this kind will be characterized by four numbers:

- $\omega(1)$ is a conditional probability of a wrong decision about the state $k = 1$,
  $\omega(1) = \sum_{x \in X_2} p_{X|K}(x \,|\, 1)$,
- $\omega(2)$ is a conditional probability of a wrong decision about the state $k = 2$,
  $\omega(2) = \sum_{x \in X_1} p_{X|K}(x \,|\, 2)$.
- $\chi(1)$ is a conditional probability of a indecisive situation under the condition that the object is in the state $k = 1$, $\chi(1) = \sum_{x \in X_0} p_{X|K}(x \,|\, 1)$.
- $\chi(2)$ is a conditional probability of the indecisive situation under the condition that the object is in the state $k = 2$, $\chi(2) = \sum_{x \in X_0} p_{X|K}(x \,|\, 2)$.

Wald task (4)

- For such strategies, the requirements
  $\omega(1) \leq \varepsilon$ and $\omega(2) \leq \varepsilon$
  are not contradictory for an arbitrary non-negative value ε because the strategy $X_0 = X$, $X_1 = \emptyset$, $X_2 = \emptyset$ belongs to the class of allowed strategies too.
- Each strategy fulfilling $\omega(1) \leq \varepsilon$ and $\omega(2) \leq \varepsilon$ is characterized by how often the strategy is reluctant to decide, i.e., by the number $\max\big(\chi(1), \chi(2)\big)$.
- Strategy which minimizes $\max\big(\chi(1), \chi(2)\big)$ is sought.

Wald task (5)

Solution (without proof) of this task for two states only is based on the calculation of the likelihood ratio

$$\gamma(x) = \frac{p_{X|K}(x\,|\,1)}{p_{X|K}(x\,|\,2)}\ .$$

Based on comparison to 2 thresholds θ_1, θ_2, $\theta_1 \leq \theta_2$ it is decided for class 1, class 2 or the solution is undecided.

In the SH10 book, there the generalization for > 2 states is given.

Wald task, example
Men, women, weight, height

PŘÍKLAD (DETEKTOR MUŽŮ A ŽEN PODLE VÁHY)

$x \in X = \{M, L, S, T\}$ ... váha osoby

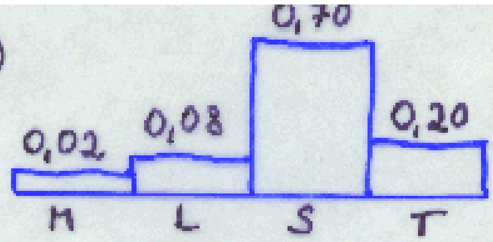
$y \in Y = \{\text{MUŽ}, \text{ŽENA}\}$ ... skrytý stav

$\varepsilon = 0{,}10$ ... pravd. chybného rozhodnutí je max 10% pro muže i ženy.

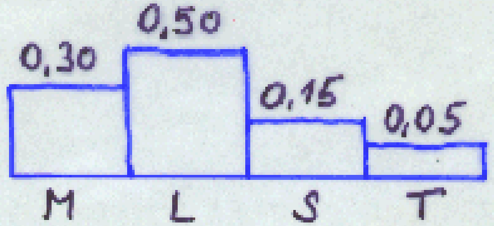
$Q: X \to Y \cup \{\text{NEVÍM}\}$ ... hledaná strategie (jedna z $3^{|X|}$)

Wald task, example (2)
Men, women, weight, height

$p(x \mid y = \text{MUŽ})$



$p(x \mid y = \text{ŽENA})$



$$\omega(\text{MUŽ}) = \sum_{x \in X_{\text{ŽENA}}} p(x \mid y = \text{MUŽ})$$

$$\omega(\text{ŽENA}) = \sum_{x \in X_{\text{MUŽ}}} p(x \mid y = \text{ŽENA})$$

$$\chi(\text{MUŽ}) = \sum_{x \in X_0} p(x \mid y = \text{MUŽ})$$

$$\chi(\text{ŽENA}) = \sum_{x \in X_0} p(x \mid y = \text{ŽENA})$$

						ω(MUŽ)	ω(ŽENA)	χ(MUŽ)	χ(ŽENA)
	1.	ŽENA	ŽENA	ŽENA	ŽENA	1,00	0,00	0,00	0,00
	2.	MUŽ	MUŽ	MUŽ	MUŽ	0,00	1,00	0,00	0,00
	3.	NEVÍM	NEVÍM	NEVÍM	NEVÍM	0,00	0,00	1,00	1,00
	4.	ŽENA	NEVÍM	NEVÍM	MUŽ	0,02	0,05	0,78	0,65
OPTIMÁLNÍ →	5.	ŽENA	ŽENA	NEVÍM	MUŽ	0,10	0,05	0,70	0,15
	6.	ŽENA	ŽENA	MUŽ	MUŽ	0,10	0,20	0,00	0,00
	7.								
	81.								

Linnik tasks = decisions with non-random interventions

- In previous non-Bayesian tasks, either the penalty function or *a priori* probabilities of the states don't make sense.
- In Linnik tasks, even the conditional probabilities $p_{X|K}(x \,|\, k)$ do not exist.
- Due to Russian mathematician J.V. Linnik from 1966.
- Random observation x depends on the object state and on an additional unobservable parameter z. The user is not interested in z and thus it need not be estimated. However, the parameter z must be taken into account because conditional probabilities $p_{X|K}(x \,|\, k)$ are not defined.
- Conditional probabilities $p_{X|K,Z}(x \,|\, k, z)$ do exist.

Linnik tasks (2)

- Other names used for Linnik tasks:
  - Statistical decisions with non-random interventions.
  - Evaluations of complex hypotheses.
- Let us mention two examples from many possibilities:
  - Testing of complex hypotheses with random state and with non-random intervention
  - Testing of complex hypotheses with non-random state and with non-random interventions.

Linnik task with random state and non-random interventions (1)

- X, K, Z are finite sets of possible observation x, state k and intervention z.
- $p_K(k)$ be the *a priori* probability of the state k. $p_{X|K,Z}(x \,|\, k, z)$ be the conditional probability of the observation x under the condition of the state k and intervention z.
- $X(k)$, $k \in K$ decomposes X according to some strategy determining states k.

  The probability of the incorrect decision (quality) depends on z

$$\omega(z) = \sum_{k \in K} p_K(k) \sum_{x \notin X(k)} p_{X|K,Z}(x \,|\, k, z) \,.$$

Linnik task with random state and non-random interventions (2)

- The quality ω^* of a strategy $(X(k),\, k \in K)$ is defined as the probability of the incorrect decision obtained in the case of the worst intervention z for this strategy, that is

$$\omega^* = \max_{z \in Z} \omega(z)\ .$$

- ω^* is minimised, i.e.,

$$\big(X^*(k), k \in K\big) = \operatorname*{argmin}_{(X(k),k\in K)} \max_{z \in Z} \sum_{k \in K} p_K(k) \sum_{x \notin X(k)} p_{X|K,Z}(x \,|\, k, z)\ .$$

Linnik task with non-random state and non-random interventions (1)

- Neither the state k nor intervention z can be considered as a random variable and consequently *a priori* probabilities $p_K(k)$ are not defined.
- Quality ω depends not only on the intervention z but also on the state k

$$\omega(k, z) = \sum_{x \notin X(k)} p_{X|K,Z}(x \,|\, k, z) \,.$$

Linnik task with non-random state and non-random interventions (2)

- The quality ω^*

$$\omega^* = \max_{k \in K} \max_{z \in Z} \omega(k, z) ,$$

- The task is formulated as a search for the best strategy in this sense, i.e., as a search for decomposition

$$\big(X^*(k), k \in K\big) = \operatorname*{argmin}_{(X(k), k \in K)} \max_{k \in K} \max_{z \in Z} \sum_{x \notin X(k)} p_{X|K,Z}(x \,|\, k, z) .$$