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.

DEFINITION OF THE BAYESIAN TASK

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 strategy $Q\colon X \to D$ which minimizes the Bayesian risk

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

The solution to the Bayesian task is the Bayesian strategy Q minimizing the risk.

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

Penalty function 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.

PROBLEMS DUE TO PENALTY FUNCTION

'Minimisation 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.

A PRIORI PROBABILITY OF SITUATIONS

It can be difficult to find probabilities $p_K(k)$, $k \in K$, which are needed for Bayesian solution.

Reasons:

1. Hidden state is random but $p_K(k)$, $k \in K$, are unknown. An object has not been analysed 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 analysing 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 OF ALLIED AIRPLANE?

Observation x describes the observed airplane.

$$\text{Hidden state} \begin{cases} k = 1 & \text{allied airplane} \\ k = 2 & \text{enemy airplane} \end{cases}$$

- 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, $0 \leq \alpha \leq 1$, that α is the probability of 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. 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

X - a set of pictures of written characters.

k - letter name (label), $k \in K$.

z - $z \in Z = \{1, 2, 3\}$ identifies the writer (this info is not known = unobservable intervention).

Task: Which letter is written in the picture x?

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

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

EXAMPLE—recognizing characters written by 3 persons (2)

- We can speak only about conditional probabilities $p_{X|K,Z}(x \mid k, z)$, 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

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

- However, we do not know how often it will be necessary to recognise 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—strategy divides the space of probabilities into convex cones.

NEYMAN–PEARSON TASK (1)

- Observation $x \in X$, two states: $k = 1$ (normal), $k = 2$ (dangerous).
- 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), $X = X_1 \cup X_2$.

NEYMAN–PEARSON TASK (2)

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

The strategy is characterised by two numbers:

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

NEYMAN–PEARSON TASK (3)

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

1. The conditional probability of the false negative is not larger than a predefined value ε.

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

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

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

under the conditions

$$\sum_{x \in X_1} p_{X|K}(x \,|\, 2) \leq \varepsilon \,, \quad X_1 \cap X_2 = \emptyset, \quad X_1 \cup X_2 = X \,.$$

NEYMAN–PEARSON TASK (4)

The fundamental result of Neyman–Pearson (1928, 1933) says:

For optimally separated X_1 and X_2 holds that a threshold value θ exist that each observation $x \in X$ for which the likelihood ratio

$$\frac{p_{X|K}(x \,|\, 1)}{p_{X|K}(x \,|\, 2)} < \theta$$

belongs to the set X_2 and otherwise $x \in X_1$.

Let us put the case of equality aside for pragmatic reasons.

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.

Simultaneously, the strategy minimises 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 optimisation task solved later in a single constructive framework.

MINIMAX TASK

Observations X are decomposed into subsets $X(k)$, $k \in K$, such that they minimise the number $\max_{k \in K} \omega(k)$.

Consider the following situation. Customer demands in advance that the PR algorithm will be evaluated by two tests:

Preliminary test (performed by the customer himself) checks the probability of a wrong decision $\omega(k)$ was for all states k.

The customer selects the worst state $k^* = \operatorname{argmax}_{k \in K} \omega(k)$.

Final test checks only those objects are checked that 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 of the final test.

The problem has not been widely known for the more general case, i.e., for the arbitrary number of object states.

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 characterised by four numbers:

$\omega(1)$ is a conditional probability of a wrong decision about the state $k = 1$,

$$\omega(1) = \sum_{x \in X_2} p_{X|K}(x \mid 1) \; ;$$

$\omega(2)$ is a conditional probability of a wrong decision about the state $k = 2$,

$$\omega(2) = \sum_{x \in X_1} p_{X|K}(x \mid 2) \; ;$$

WALD TASK (4)

$\chi(1)$ 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 (5)

- For such strategies, the requirements
  $\omega(1) \le \varepsilon$ and $\omega(2) \le \varepsilon$
  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 $\omega(1) \le \varepsilon$ and $\omega(2) \le \varepsilon$ is characterised 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 (6)

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 generalziation for > 2 states is given.

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.
  $p_{X|K,Z}(x \mid k, z)$ 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

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

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

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