Non-Bayesian decision making

Václav Hlaváč

Czech Technical University in Prague
Czech Institute of Informatics, Robotics and Cybernetics
160 00 Prague 6, Jugoslávských partyzánů 1580/3, Czech Republic
http://people.ciirc.cvut.cz/hlavac, vaclav.hlavac@cvut.cz
also Center for Machine Perception, 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: Michail I. Schlesinger, Vojtěch Franc

Bayesian task (rehearsal)

Bayesian task of statistical decision making seeks for

- sets X, Y and D,
- statistical model, i.e joint probability function as a function of two variables $p_{XY} \colon X \times Y \to \mathbb{R}$
- penalty function $W \colon Y \times D \to \mathbb{R}$

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

$$R(Q) = \sum_{x \in X} \sum_{y \in Y} p_{XY}(x, y)\, W(y, Q(x))\,.$$

Typical use: minimizing 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 $p_Y(y)$, $y \in Y$, are not known or cannot be known because y is not a random event.
- Class-conditional probabilities $p(x|y)$ 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 practical tasks, it can be more intuitive for a customer to express desired classifier properties as the acceptable rate of false positives (false alarms) and false negatives (overlooked danger).

Example:

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

Way out (only in a few 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 small subset of possible tasks. *Analogy: a few known positive islands in a sea of unknown.*
- Nothing can be said about the task and its solution if it does not belong to non-Bayesian tasks.

Penalty function in Bayesian decision making

- The 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 the multiplication by a real number and the addition are defined.

An example from a Russian fairy tale: A hero comes to the place where he must make a decision.

When the hero turns to the left, he loses his horse, when he turns to the right, he loses his sword, and when 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 a single application.
- The penalty for the false positive (false alarm) and the false negative (overlooked danger) might be incomparable.

Example: Decisions when curing a patient

- $x \in X$, i.e. observable parameters (features, observations) measured on a patient
- $y \in Y$, i.e. hidden parameters $=$ {healthy, seriously sick}
- $d \in D$, i.e. decisions $=$ {do not cure, apply a drug}

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

Y \ D	do not cure	apply a drug
healthy	correct decision	small health damage
seriously sick	death possible	correct decision

This examples illustrates the difficulty with assigning a penalty (real number) to a decision.

Note: Health insurance companies overcome this problem by assessing everything by money.

Troubles with *a priori* probability of situations

It can be difficult to find *a priori* probabilities $p_Y(y)$, $y \in Y$, which are needed in Bayesian formulation. Recall $p(x, y) = p(x|y)\, p(y)$.

Reasons:

1. Hidden state is random but *a priori* probabilities $p_Y(y)$, $y \in Y$, are unknown. An object has not been analyzed sufficiently.

   The hint has two options:

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

   (b) Start analyzing the object thoroughly and get *a priori* probabilities, which are inevitable for the Bayesian solution.

2. Hidden state is not random. Consequently, the *a priori* probabilities $p_Y(y)$, $y \in Y$, do not exist. It is impossible to discover them by an arbitrary detailed exploration of the object. The hint is to check if one of non-Bayesian methods can be used.

An example: Enemy or allied airplane?

- Observation x describes the observed airplane, e.g. in a radar signal.
- Two hidden states $\begin{cases} y = 1 & \text{allied airplane,} \\ y = 2 & \text{enemy airplane.} \end{cases}$
- The class-conditional probability (also named likelihood) $p_{X|Y}(x|y)$ might depend on the observation x in a complicated manner. Nevertheless, it exists and describes the dependence of the observation x on the situation y correctly.
- *A priori* probabilities $p_Y(y)$ are not known. They 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 enemy plane occurrence.
- Consequently $p_Y(y)$ do not exist since the frequency of the experiment result does not converge to any number, which can be named the probability. The hidden state y is not a random event.

Beware of a pseudosolution

Refers to the airplane example.

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

Conditional probabilities of observations

Motivating example: Recognizing characters written by three persons

- Given:
  - X is a set of pictures of hand written characters x.
  - y is a name (label) of a character, $y \in Y$.
  - $z \in Z = \{1, 2, 3\}$ 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 $W(y, d)$ and *a priori* probabilities $p_Y(y)$ of individual characters.

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

Example: Recognizing characters written by three persons (2)

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

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

- However, we do not know how often it will be necessary to recognize pictures handwritten by this or that person.
- Under such uncertain statistical conditions, the algorithm ought to be created that will secure the required recognition quality of pictures independently of 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 original extensions of them by M. I. Schlesinger).
- The whole class of solvable non-Bayesian tasks has two common features:
  1. There is one formalism for expressing tasks and their solution (dual tasks of linear programming).
  2. Similarly to Bayesian tasks: The decision strategy divides the space of probabilities into convex cones.

Basic concepts; Female detection example (1)

- A person is characterized by the:
  - Observed parameters, the height and the weight in our case,

    $$x \in X = \underbrace{\{\text{short}, \text{average}, \text{tall}\}}_{\text{height}} \times \underbrace{\{\text{lean}, \text{light}, \text{average}, \text{heavy}\}}_{\text{weight}}$$

  - Class label = sex, i.e. $y \in Y = \{\text{male}, \text{female}\}$.
- *A priori* probabilities are unknown.
- Class-conditional probabilities are known, e.g.:

$p_{X\|Y}(x\,\|\,\text{female})$				
	lean	light	average	heavy
short	0.197	0.145	0.094	0.017
average	0.077	0.299	0.145	0.017
tall	0.001	0.008	0.000	0.000

$p_{X\|Y}(x\,\|\,\text{male})$				
	lean	light	average	heavy
short	0.011	0.005	0.011	0.011
average	0.005	0.071	0.408	0.038
tall	0.002	0.014	0.255	0.169

Basic concepts; Female detection example (2)

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

The decision strategy q splits the set X of observations into two subsets:

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

Example of one possible strategy $Q(x) \to y$

	lean	light	average	heavy
short	female	female	female	male
average	female	female	male	male
tall	male	male	male	male

Basic concepts; Female detection example (3)

The decision strategy $q \in Q\colon X \to Y$ is characterized by the probability of:

- $\omega(\text{FP}) = \sum_{x \in X_{\text{female}}} p_{X|Y}(x|\text{male}) \ldots$ false positive,
  (also false alarm or Type 2 error, i.e. a male recognized wrongly as a female).
- $\omega(\text{FN}) = \sum_{x \in X_{\text{male}}} p_{X|Y}(x|\text{female}) \ldots$ false negative
  (also overlooked danger or Type 1 error, i.e. a female recognized wrongly as a male).

Class-conditional probabilities are repeated from slide 15. The color label female (red) or male (blue) was added for one possible strategy the for the illustration.

(female)

	lean	light	average	heavy
short	0.197	0.145	0.094	0.017
average	0.077	0.299	0.145	0.017
tall	0.000	0.009	0.000	0.000

(male)

	lean	light	average	heavy
short	0.011	0.005	0.011	0.011
average	0.005	0.071	0.408	0.038
tall	0.000	0.016	0.255	0.169

Neyman–Pearson task, two classes only (1)

Proposed by Jerzy Neyman and Egon Pearson in the paper ‘On the Problem of the Most Efficient Tests of Statistical Hypotheses’ in 1933.

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

Neyman–Pearson task (2)

The observation x can belong either to normal state or to dangerous state.

$\Rightarrow$ There is no faultless strategy.

The strategy is characterized by two numbers $\omega(\mathsf{FP})$, $\omega(\mathsf{FN})$:

- Class-conditional probability of the false positive (false alarm)
  $\omega(\mathsf{FP}) = \sum_{x \in X_2} p_{X|Y}(x \,|\, 1)$.
- Class-conditional probability of the false negative (overlooked danger)
  $\omega(\mathsf{FN}) = \sum_{x \in X_1} p_{X|Y}(x \,|\, 2)$.

Neyman–Pearson task (3)

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

1. The class-conditional probability of the false negative is not larger than the prescribed limit (threshold) ε, $0 < \varepsilon < 1$, on the probability of the false negative (overlooked danger). This yields eligible strategies.

$$\omega(\mathsf{FN}) = \sum_{x \in X_1} p_{X|Y}(x \,|\, 2) \leq \varepsilon \, .$$

2. A strategy q^* is sought among eligible strategies, which has the class-conditional probability of the false positive is the smallest.

$$q^* = \mathop{\mathrm{argmin}}_{q: X \to Y} \omega(\mathsf{FP}) = \mathop{\mathrm{argmin}}_{q: X \to Y} \sum_{x \in X_2} p_{X|Y}(x \,|\, 1) \quad \text{subject to the condition}$$

$$\omega(\mathsf{FN}) = \sum_{x \in X_1} p_{X|Y}(x \,|\, 2) \leq \varepsilon \, .$$

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

We start the explanation even with a simpler example using only one observation: weight. The person height is not taken into account for a moment.

The ‘Females detection example by the weight only’ formulation:

- The observation x (weight of the person), $x \in X = \{\text{lean}, \text{leight}, \text{average}, \text{heavy}\}$.
- The hidden state $y \in Y = \{\text{male}, \text{female}\}$.
- $\varepsilon = 0.2$ meaning that the prescribed class-conditional probability of the false negative (overlooking a female) is 20% at most.
- The decision strategy q has to be the Neyman-Pearson strategy.
- The number of all possible strategies is $2^{|X|} = 2^4 = 16$.

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

$p(x|y = \text{male})$

lean	leight	average	heavy
0.02	0.08	0.7	0.2

$p(x|y = \text{female})$

lean	leight	average	heavy
0.3	0.5	0.15	0.05

$$\min \sum_{x \in X_{\text{female}}} p(x|y = \text{male}) \quad \textit{i.e., false alarm}$$

under the condition $$\sum_{x \in X_{\text{male}}} p(x|y = \text{female}) \leq 0.20 \quad \textit{i.e., overlooked female}$$

eligible solutions (rows 1–4)

optimal solution (row 4)

	lean	light	average	heavy	overlooked female	false alarm
1	female	female	female	female	0.00	1.00
2	female	female	female	male	0.05	0.80
3	female	female	male	female	0.15	0.30
4	female	female	male	male	0.20	0.10
5	female	male	female	female	0.50	0.92
	⋮				⋮	
16	male	male	male	male	1.00	0.00

Neyman–Pearson task; the solution (4)

- The solution is by J. Neyman and E. Pearson (1928, 1933).
- The optimal strategy q^* separates the observation sets X_1 and X_2 according to the threshold θ from the likelihood ratio $L(x)$ value .
- The optimal strategy q^* is obtained by selecting the minimal θ, for which

$$q^* = \begin{cases} y = 1 \quad \text{(normal state)} & \text{if} \quad L(x) = \dfrac{p_{X|Y}(x \mid 1)}{p_{X|Y}(x \mid 2)} > \theta\,, \\ y = 2 \quad \text{(dangerous state)} & \text{otherwise.} \end{cases}$$

The rule is the special case of the division of a space of probabilities into convex cones, i.e., it corresponds to Bayesian strategy. It was explained in the 'Bayesian' lecture.

Female detection example, Neyman-Pearson (4)

A

	$p_{X\|Y}(x \mid \text{female})$			
	lean	light	average	heavy
short	0.197	0.145	0.094	0.017
average	0.077	0.299	0.145	0.017
tall	0.001	0.008	0.000	0.000

B

	$p_{X\|Y}(x \mid \text{male})$			
	lean	light	average	heavy
short	0.011	0.005	0.011	0.011
average	0.005	0.071	0.408	0.038
tall	0.002	0.014	0.255	0.169

C

	$L(x) = p(x \mid \text{male}) / p(x \mid \text{female})$			
	lean	light	average	heavy
short	0.056	0.034	0.117	0.647
average	0.065	0.237	2.814	2.235
tall	2.000	1.750	∞	∞

D

	rank order of $p(x \mid \text{male}) / p(x \mid \text{female})$			
	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	12

E

Strategy #	θ	Round(θ)
12	empty	empty
11	2.906897	2.907
10	2.524544	2.525
9	2.117647	2.118
8	1.875	1.875
7	0.551471	0.551
6	0.442259	0.442
5	0.17724	0.177
4	0.090978	0.091
3	0.004549	0.005
2	0.04516	0.045
1	0.027919	0.028

- Taking into account known class-conditional probabilities (Table A and Table B), the likelihood ratio $L(x)$ is calculated. It yields Table C.
- We order likelihood ratios $L(x)$ from the smallest one to the biggest one. The appropriate rank as shown in Table D.
- There are 12 possible strategies. Let us set the threshold θ values for the strategies. The value θ for the particular strategy lies in between the two limit values $L(x)$ in Table C corresponding to neighboring ranks.
- Let us set θ as the average of limit values $L(x)$ corresponding to the pair of neighboring ranks. Let us label strategies by the higher rank number of the pair.

Female detection example, Neyman-Pearson (5)

- In this example, different likelihood ratios $L(x)$ produce 11 possible strategies out of 12 $L(x)$ entries.
- Strategy $\#12$ is not possible and is shown as 'empty' in the tables.
- We display these 11 possible strategies.
- Table entry colors (female$=$red; male$=$blue) denote a particular strategy. If $L(x)$ of the entry is above the threshold θ then the entry is labeled as female.
- The next job is to select the optimal strategy.

$L(x) = p(x \mid \text{male}) / p(x \mid \text{female})$

	lean	light	average	heavy
short	0.056	0.034	0.117	0.647
average	0.065	0.237	2.814	2.235
tall	2.000	1.750	∞	∞

rank order of $p(x \mid \text{male}) / p(x \mid \text{female})$

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	12

Strategy 1

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 2

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 3

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 4

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 5

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 6

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 7

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 8

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 9

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 10

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 11

	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Female detection example, Neyman-Pearson (6)

For each possible strategy, we calculated $\omega(FP)$ and $\omega(FN)$. We selected from eligible strategies $(\omega(FP) \leq 0.2)$ the optimal strategy $\#5$ because of the minimal $\omega(FN)$.

Strategy 3	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 4	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 5	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

Strategy 6	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	empty

$p_{X\|Y}(x\,\|$female$)$	lean	light	average	heavy
short	0.197	0.145	0.094	0.017
average	0.077	0.299	0.145	0.017
tall	0.001	0.008	0.000	0.000

$p_{X\|Y}(x\,\|$male$)$	lean	light	average	heavy
short	0.011	0.005	0.011	0.011
average	0.005	0.071	0.408	0.038
tall	0.002	0.014	0.255	0.169

$L(x)$ = $p(x$ | male$)$ / $p(x$| female$)$	lean	light	average	heavy
short	0.056	0.034	0.117	0.647
average	0.065	0.237	2.814	2.235
tall	2.000	1.750	∞	∞

rank order of $p(x$ |male$)$ / $p(x$| female$)$	lean	light	average	heavy
short	2	1	4	6
average	3	5	10	9
tall	8	7	11	12

Strategy #	θ	Round(θ)	ω(FP)	ω(FN)
12	empty	empty		
11	2.906897	2.907	0.831	0.000
10	2.524544	2.525	0.576	0.000
9	2.117647	2.118	0.168	0.145
8	1.875	1.875	0.130	0.162
7	0.551471	0.551	0.128	0.163
6	0.442259	0.442	0.114	0.171
5	0.17724	0.177	0.103	0.188
4	0.090978	0.091	0.032	0.487
3	0.004549	0.005	0.021	0.581
2	0.04516	0.045	0.016	0.658
1	0.027919	0.028	0.005	0.855

Strategy # = higher rank of the rank pair

orange color denotes the optimal strategy

Solving Non-Bayesian tasks using linear programming (LP)

Decision strategies $Q: X \to Y$ can be equivalently represented by a function $\alpha: X \times Y \to \{1, 0\}$ satisfying

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

Values of α for the 'Female detection' example, cf. slide 15:

(x, male)

	lean	light	average	heavy
short	0	0	0	1
average	0	0	1	1
tall	1	1	1	1

(x, female)

	lean	light	average	heavy
short	1	1	1	0
average	1	1	0	0
tall	0	0	0	0

Neyman-Pearson task solved by LP

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

Original formulation	**Linear programming relaxation**
$$Q^* = \underset{X_1,X_2}{\operatorname{argmin}} \sum_{x \in X_2} p_{X\|Y}(x\|1) \ ,$$ subject to $$\sum_{x \in X_1} p_{X\|Y}(x\|2) \leq \varepsilon \ .$$	$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} \sum_{x \in X} \alpha(x,2)\, p_{X\|Y}(x\|1) \ ,$$ subject to $$\begin{aligned} \sum_{x \in X} \alpha(x,1) p_{X\|Y}(x,2) &\leq \varepsilon \ , \\ \alpha(x,1) + \alpha(x,2) &= 1 \ , \quad \forall x \in X \ , \\ \alpha(x,1) &\geq 0 \ , \quad \forall x \in X \ , \\ \alpha(x,2) &\geq 0 \ , \quad \forall x \in X \ . \end{aligned}$$

Neyman-Pearson task as LP (2)

The dual formulation of the Neyman-Pearson task

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

subject to

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

Female detection example, Neyman-Pearson (1)

The decision strategy $q\colon x \to y$ is characterized by two class-conditional probabilities $\omega(\text{male})$, $\omega(\text{female})$ and the likelihood ratio threshold θ; $\theta = 0.2$ in this example.

$$\omega(\text{female}) = \sum_{x \in X_{\text{female}}} p(x|\text{male}) \qquad \ldots \text{ false positive}$$

$$\omega(\text{male}) = \sum_{x \in X_{\text{male}}} p(x|\text{female}) \qquad \ldots \text{ false negative (overlooked girl)}$$

We seek the decision strategy $q^*\colon x \to \{\text{male}, \text{female}\}$

$$q^* = \operatorname*{argmin}_{X_{\text{male}}, X_{\text{female}}} \sum_{x \in X_{\text{female}}} p(x|\text{male})$$

under the condition

$$\sum_{x \in X_{\text{male}}} p(x|\text{female}) \leq 0.2$$

Generalised Neyman-Pearson task for two dangerous states

- $y = 1$ corresponds to the set X_1;
- $y = 2$ or $y = 3$ correspond to the set X_{23}.
- We seek a strategy with the class-conditional probability of the false positives (overlooked dangerous states) both $y = 2$ and $y = 3$ is not larger than a value ε given in advance.
- Simultaneously, the strategy minimizes the false negatives (false alarms), $\sum\limits_{x \in X_{23}} p_{X|Y}(x \,|\, 1)$ under the conditions

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

The formulated optimization task will be 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 $X(y)$, $y \in Y$, such that they minimize the number $\max_{y \in Y} \omega(y)$.
- Consider a customer who demands that the pattern recognition algorithm will be evaluated by two tests in advance:

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

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

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

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

Minimax task (3)

Minimax task formulation

A strategy q^* is sought which minimizes $\max_{y \in Y} \omega(y)$

- The solution decomposes the space of probabilities into convex cones.
- Notice that the case $|Y| = 2$ 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-Pearson task lacks symmetry with respect to states of the recognized object. The class-conditional probability of the false negative (overlooked danger) must be small, which is the principal requirement.
- The class-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 class-conditional 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 $y = 1$ is chosen;
- if $x \in X_2$, then $y = 2$ 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 y.

Wald task (3)

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

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

Wald task (4)

- 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 characterized by how often the strategy is reluctant to decide, i.e., by the number $\max\big(\chi(1), \chi(2)\big)$.
- The strategy q^* 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

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

Based on comparison to two 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 explained.

Wald task, example, male/female according to weight (1)

Example – Discriminating women and men according to their weight

- $x \in X = \{\text{lean, light, average, heavy}\} \cup X_0$ – the weight of a person $\cup$ the weight corresponding to the hidden state `not known`

  If $x \in X_0$ it is decided that the observation x does not provide enough information for a safe decision about the state y, i.e. if the observed person is a male or a female.

- $y \in Y = \{\text{male, female}\}$ – the hidden state

- $\varepsilon = 0.10$ – the probability of the allowed erroneous decision is maximally 10%, both for males and females.

- $Q\colon X \to Y \cup \{\texttt{not known}\}$ – the sought decision strategy; one from $3^{|X|} = 3^4 = 81$ options.

Wald task, example, male/female according to weight (2)

$$p(x|y = \text{male})$$

lean	leight	average	heavy
0.02	0.08	0.7	0.2

$$p(x|y = \text{female})$$

lean	leight	average	heavy
0.3	0.5	0.15	0.05

$$w(\text{female}) = \sum_{x \in X_{\text{male}}} = p(x|y = \text{female})$$

$$w(\text{male}) = \sum_{x \in X_{\text{female}}} = p(x|y = \text{male})$$

$$\chi(\text{male}) = \sum_{x \in X_0} = p(x|y = \text{male})$$

$$\chi(\text{female}) = \sum_{x \in X_0} = p(x|y = \text{female})$$

Optimal solution

1	Female	Female	Female	Female
2	Male	Male	Male	Male
3	Not known	Not known	Not known	Not known
4	Female	Not known	Not known	Male
5	**Female**	**Female**	**Not known**	**Male**
6	Female	Female	Male	Male
	...			
81				

ω(Male)	ω(Female)	χ(Male)	χ(Female)
1.00	0.00	0.00	0.00
0.00	1.00	0.00	0.00
0.00	0.00	1.00	1.00
0.02	0.05	0.78	0.65
0.10	**0.05**	**0.70**	**0.15**
0.10	0.20	0.00	0.00
...			

Linnik tasks = decisions with non-random interventions

- In previous non-Bayesian tasks, either the penalty function or *a priori* probabilities of the states do not make sense.
- In Linnik tasks, even the class-conditional probabilities $p_{X|Y}(x \mid y)$ do not exist.
- Proposed by the Russian mathematician J.V. Linnik in 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 class-conditional probabilities $p_{X|Y}(x \mid y)$ are not defined.
- Class-onditional probabilities $p_{X|Y,Z}(x \mid y, z)$ 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-randominterventions (1)

- X, Y, Z are finite sets of possible observation x, state y and intervention z.
- $p_Y(y)$ be the *a priori* probability of the state y. Let $p_{X|Y,Z}(x \mid y, z)$ be the class-conditional probability of the observation x under the condition of the state y and intervention z.
- $X(y)$, $y \in y$ decomposes X according to some strategy determining states y.
  The probability of the incorrect decision (quality) depends on z

$$\omega(z) = \sum_{y \in Y} p_Y(y) \sum_{x \notin X(y)} p_{X|Y,Z}(x \mid y, z) \,.$$

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

- The quality ω^* of a strategy $(X(y),\, y \in Y)$ 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.,

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

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

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

$$\omega(y,z) = \sum_{x \notin X(y)} p_{X|Y,Z}(x \mid y, z) \,.$$

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

- The quality ω^*

$$\omega^* = \max_{y \in Y} \max_{z \in Z} \omega(y, z) \, ,$$

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

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