ROZHODOVACÍ FUNKCE A PRINCIP MINIMAXU

VÁCLAV FABIAN, Praha.

(Došlo dne 2. března 1955.) DT : 519.24

Úkolem tohoto článku je informovat čtenáře o theorii rozhodovacích funkcí. Je patrna obtížnost definice optimálnosti rozhodovací funkce. Jednu z možných definic podává princip minimaxu, jehož význam je v článku blíže osvětlen. Minimax je také srovnán s některými jinými principy.

1. Úvod. V článku si budeme všímat v podstatě pouze jedné stránky theorie rozhodovacích funkcí, totiž toho, jak hodnotí statistické metody (= rozhodovací funkce) a nebudeme mluvit o problémech konstrukce rozhodovacích funkcí. To nám umožní pracovat s větší obecností a vyhnout se některým složitým matematickým pojmům a úvahám.

Proto se v našem článku na příklad neobjeví pojem striktní určenosti, jenž je důležitý pro otázky konstrukce rozhodovacích funkcí (a má též samostatný logický význam v theorii her).

Konečným úkolem je vzbudit zájem o studium theorie rozhodovacích funkcí, na niž se přenáší zcela neoprávněně nedůvěra, již mají někteří statistikové k minimaxovému řešení.

Tuto nedůvěru k minimaxovému řešení způsobují asi tyto okolnosti:

1. Povaha věci; nesnadnost definice optimálnosti rozhodovací funkce.

2. Nepochopení významu některých matematických pojmů (apriorní pravděpodobnosti, Bayesova řešení), jichž se užívá s výhodou k řešení některých otázek theorie rozhodovacích funkcí.

3. Odůvodňování minimaxového řešení analogií s theorií her, jež je podle mého názoru neoprávněné.

4. Neporozumění některým klasickým metodám, jež zdánlivě jednoznačně řeší některé statistické problémy bez obtíží a bez jistého prvku libovůle, již s sebou nese minimaxové řešení.

Obtížnost definice nejlepší rozhodovací funkce bude patrna v odstavci 5; omyly, o nichž jsme se právě zmínili v bodech 2 až 4, se budeme zabývat v odstavci 10.

2. Statistický problém. Nechť

$$X = X_1, X_2, X_3, \dots$$

je konečná nebo nekonečná posloupnost náhodných proměnných. Nechť $\mathfrak{M}$ je nějaká množina distribučních funkcí náhodné proměnné X (*pravá* distribuční funkce náhodné proměnné X je prvkem množiny $\mathfrak{M}$).

Nechť D je množina konečných rozhodnutí.[1])

Dále nechť je W nezáporná funkce definovaná na kartézském součinu množin $\mathfrak{M} \times D$, při čemž $W(F, d)$ značí ztrátu, která nastane, je-li F *pravá* distribuční funkce a d zvolené konečné rozhodnutí.

Budiž dále $c[(i_{11}, i_{12}, \dots, i_{1n_1}), (i_{21}, i_{22}, \dots, i_{2n_2}), \dots, (i_{k1}, i_{k2}, \dots, i_{kn_k})]$ cena pozorování k skupin náhodných proměnných $(X_{i_{11}}, X_{i_{12}}, \dots, X_{i_{1n_1}}), (X_{i_{21}}, X_{i_{22}}, \dots X_{i_{2n_2}}), \dots$ $\dots, (X_{i_{k1}}, X_{i_{k2}}, \dots, X_{i_{kn_k}})$. Budiž konečně $\mathbf{D}$ nějaká množina rozhodovacích funkcí[2]) (prvkům z $\mathbf{D}$ budeme říkat použitelné rozhodovací funkce).

Potom nazveme pětici

$$(\mathfrak{M}, D, W, c, \mathbf{D})$$

statistickým problémem.

Řešením daného konkretního statistického problému je určení jednoho prvku $d \in D$, to jest určení konečného rozhodnutí. Je-li F *pravá* distribuční funkce a platí-li

$$W(F, d) \leq W(F, d')$$

pro všechna $d' \in D$, bude nejlepším rozhodnutím prvek d.

Tato úvaha není bohužel přímo použitelná, neboť neznáme *pravou* distribuční funkci. O té pouze víme, že je prvkem množiny $\mathfrak{M}$. Je proto důležité, aby při formulaci statistického problému byla množina $\mathfrak{M}$ definována tak, aby neobsahovala žádnou distribuční funkci, o níž předem víme, že není *pravou* distribuční funkcí náhodné proměnné X. Čím menší je množina $\mathfrak{M}$, tím lépe nás informuje o *pravé* distribuční funkci. Nevhodnou (neodpovídající skutečnosti) volbou množiny $\mathfrak{M}$ může být na příklad silně ovlivněno minimaxové řešení statistického problému.

Protože nejvýhodnější volba konečného rozhodnutí závisí na neznámé *pravé* distribuční funkci, snažíme se ji odhadnout pozorováním některých členů posloupnosti X. Obecně můžeme říci, že zvětšením počtu pozorovaných náhodných proměnných zlepšíme svou informaci o F. Na druhé straně však takové zvětšení počtu pozorování není vždy možné, nebo není účelné vzhledem k nákladům, jež je třeba na taková pozorování vynaložit. Cenu těchto pozorování

---

[1]) Říkáme konečné rozhodnutí, abychom odlišili prvek $d \in D$, jehož vybrání znamená řešení konkretního statistického problému, od jiných rozhodnutí, jež činíme během rozhodovacího procesu (na příklad, které náhodné proměnné mají být pozorovány).

[2]) V dalších dvou odstavcích řekneme, co je to rozhodovací funkce.

nám udává funkce c, při čemž zdánlivá složitost, s níž je tato funkce definována, je někdy vhodná, neboť bývá na příklad nákladnější pozorovat n náhodných proměnných jednu po druhé, než všech n současně.

Zhruba můžeme říci, že pravidlo, které stanoví pozorování, jež mají být provedena, a způsob, jak podle těchto pozorování má být určeno konečné rozhodnutí $d \in D$, nazýváme rozhodovací funkcí. Pro statistiku jakožto vědu není pak úkolem určit vhodné $d \in D$, ale vhodnou, respektive nejlepší rozhodovací funkci $\delta \in \mathbf{D}$. Proto musí být věnována velká péče otázkám hodnocení rozhodovacích funkcí a stanovení pojmu *nejlepší* rozhodovací funkce.

Theorie rozhodovacích funkcí studuje vlastnosti rozhodovacích funkcí a tím pomáhá při jejich hodnocení (má ovšem i jiné výsledky, týkající se konstrukce rozhodovacích funkcí; o nich však v tomto článku nechceme mluvit). Bohužel bez dalších principů není sama ve většině případů schopna určit pojem *nejlepší* rozhodovací funkce pro daný statistický problém.

Protože někdy můžeme o některých rozhodovacích funkcích předem říci, že jsou nevhodné, není třeba uvažovat všechny rozhodovací funkce a stačí se omezit jen na nějakou jejich množinu $\mathbf{D}$.

Poznamenáme ještě, že zřejmě předpoklad nezápornosti funkce W není nijak omezující, zrovna tak jako název ztráta. Je ovšem možné, že by se theorie rozhodovacích funkcí zdála méně pesimistickou, kdybychom používali váhové funkce s jiným významem.

Uvedeme konkretní příklad statistického problému. Předpokládejme, že chemická továrna chce postavit zděnou věž. Materiál musí vyhovovat určitým podmínkám vzhledem k zamýšlenému použití věže. Je stanovena nejvyšší přípustná nassáklivost a nejvyšší přípustná četnost p_0 nevyhovujících cihel ve zdivu. Továrna se má rozhodnout, zda použije ke stavbě cihel, jež jí byly dodány (označíme soubor těchto cihel $\mathfrak{A}$), či zda dodávku vrátí a vyžádá si jinou. Ztráta (jak ji továrna odhaduje) bude 0, jestliže věž bude vystavěna z vyhovujícího materiálu, bude 1 000 000, bude-li věž vystavěna z nevyhovujícího materiálu a bude 10 000, jestliže odřeknutím dodávky se stavba zdrží. Nassáklivost cihly se zjišťuje experimentem, při němž se cihla zničí. Celkový náklad (včetně ceny zničených cihel) za zjištění nassáklivosti u n cihel závisí jen na počtu těchto cihel a je roven n.

Rozhodnutí v tomto problému může zřejmě být dosaženo statistickou metodou. Modelem tohoto problému je pak statistický problém $(\mathfrak{M}, D, W, c, \mathbf{D})$, kde:

$$X = X_1, X_2, X_3, \ldots ,$$

a $X_i = 0$ je-li i-tá cihla ve výběru bez vracení z $\mathfrak{A}$ vyhovující, $X_i = 1$ v opačném případě.

Označíme-li F_p distribuční funkci náhodné proměnné X za předpokladu, že p je četnost nevyhovujících cihel v $\mathfrak{A}$, je $\mathfrak{M}$ množina všech distribučních funkcí F_p pro $p \in \langle 0, 1 \rangle$.

D obsahuje dva prvky d_1 (vystavět věž) a d_2 (odřeknout dodávku).

$$\begin{aligned} W(F_p, d_1) &= 0\,, && \text{je-li} \quad p \leq p_0\,, \\ W(F_p, d_1) &= 1\,000\,000\,, && \text{je-li} \quad p > p_0\,, \\ W(F_p, d_2) &= \quad 10\,000 && \text{pro každé } p\,. \end{aligned}$$

D definujeme jako množinu všech rozhodovacích funkcí, které nevyžadují více pozorování, než je rozdíl počtu cihel v $\mathfrak{A}$ a počtu cihel potřebných ke stavbě.

3. Prosté rozhodovací funkce. Všimli jsme si již v předchozím odstavci pojmu rozhodovací funkce.

Je-li tedy ($\mathfrak{M}$, D, W, c, **D**) daný statistický problém a δ rozhodovací funkce, pak na počátku rozhodovacího procesu musí δ určit buď skupinu náhodných proměnných s indexy $(i_{11}, i_{12}, \ldots, i_{1n_1})$, jež má být pozorována, nebo přímo již nějaké rozhodnutí $d \in D$. Toto prvé (pomocné) rozhodnutí označme $\delta(\emptyset)$. Je tedy buď $\delta(\emptyset) = (i_{11}, i_{12}, \ldots, i_{1n_1})$ a rozhodovací proces bude pokračovat pozorováním náhodných proměnných $X_{i_{11}}, X_{i_{12}}, \ldots, X_{i_{1n_1}}$, nebo $\delta(\emptyset)$ je konečné rozhodnutí $\delta(\emptyset) \in D$ a rozhodovací proces je ukončen právě tímto rozhodnutím $\delta(\emptyset)$.

Označme, je-li $e = (i_1, i_2, \ldots, i_n)$ množina indexů, symbolem X_e vektorovou náhodnou proměnnou $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$.

Bylo-li $\delta(\emptyset) = e_1 = (i_{11}, i_{12}, \ldots, i_{1n_1})$, rozhodovací proces pokračoval a rozhodovací funkce δ musí učinit další rozhodnutí $\delta(e_1, X_{e_1})$, závisící na hodnotě napozorované náhodné proměnné X_{e_1}. $\delta(e_1, X_{e_1})$ je buď konečné rozhodnutí (a proces končí rozhodnutím $\delta(e_1, X_{e_1})$), nebo je $\delta(e_1, X_{e_1})$ skupina indexů $e_2 = (i_{21}, i_{22}, \ldots, i_{n_2})$ a proces pokračuje pozorováním náhodné proměnné X_{e_2}.

Jestliže po k-tém kroku máme napozorovány náhodné proměnné X_{e_1}, $X_{e_2}, \ldots, X_{e_k}$, musí nám rozhodovací funkce δ určit další rozhodnutí

$$\delta(e_1, e_2, \ldots, e_k, X_{\bigcup_{i=1}^{k} e_i})\,,$$

což je opět buď skupina indexů

$$e_{k+1} = (i_{k+1,1}, i_{k+1,2}, \ldots, i_{k+1,n_{k+1}})$$

(a rozhodovací proces pokračuje pozorováním náhodné proměnné $X_{e_{k+1}}$) anebo konečné rozhodnutí z množiny D (a proces je ukončen).

Obyčejně se požaduje, aby rozhodovací funkce s pravděpodobností 1 (pro každé $F \in \mathfrak{M}$) vybrala konečné rozhodnutí po konečném počtu pozorování.

Konečně se také předpokládá, že nepozorujeme žádnou náhodnou proměnnou dvakrát (můžeme ovšem pozorovat několik náhodných proměnných se stejnou distribuční funkcí).

Rozhodovací funkce, jež jsme nyní podrobněji popsali, budeme nazývat prostými rozhodovacími funkcemi. Formálně lze definovat prostou rozhodovací funkci na příklad takto:

Označme symbolem $\mathbf{E}$ systém všech konečných množin přirozených čísel. Nechť pro $e \in \mathbf{E}$ značí x_e funkční hodnotu náhodné proměnné X_e. Symboly $e_1, e_2, \ldots$ označme prvky z $\mathbf{E}$. Je-li δ funkce definovaná na množině $\mathbf{M}$ zobrazující tuto množinu do $\mathbf{E} \cup D$ a platí-li: $\emptyset \in \mathbf{M}$

a

$$\{[e_1, e_2, \ldots, e_k, x_{\bigcup_{i=1}^{k} e_i}] = \mathbf{m} \in \mathbf{M}\ ,\quad \delta(\mathbf{m}) = e_{k+1} \in \mathbf{E}\} \Rightarrow$$

$$\Rightarrow \{e_{k+1} \cap \bigcup_{i=1}^{k} e_i = \emptyset\ ,\quad [e_1, e_2, \ldots, e_k, e_{k+1}, x_{\bigcup_{i=1}^{k+1} e_i}] \in \mathbf{M}\}\ ,$$

nazveme δ prostou[3]) rozhodovací funkcí.

Pro danou prostou rozhodovací funkci δ a danou hodnotu x náhodné proměnné X můžeme zřejmě určit konečné rozhodnutí $d = d(\delta, x)$ a k skupin $e_1, e_2, \ldots, e_k$ indexů napozorovaných náhodných proměnných.

Je pak

$$Z(F, \delta, x) = W(F, d) + c(e_1, e_2, \ldots, e_k)^{4})$$

celková ztráta za předpokladu, že *pravá* distribuční funkce je F, náhodná proměnná X nabyla hodnoty x a statistik vybral pro řešení problému rozhodovací funkci δ. Definujeme-li

$$Z(F, \delta) = \int Z(F, \delta, x)\, \mathrm{d}F(x)^{5})$$

(kde neoznačený integrační obor znamená, že se integruje přes celý obor integrační proměnné), označuje nám $Z(F, \delta)$ očekávanou hodnotu ztráty za předpokladu, že *pravá* distribuční funkce je F a zvolená rozhodovací funkce je δ.

Je patrné, že pro hodnocení je každá rozhodovací funkce dostatečně popsána funkcí $Z(F, \delta, x)$, místo níž můžeme také uvažovat náhodnou proměnnou $Z(F, \delta, X)$.

V theorii rozhodovacích funkcí je však zavedeno hodnocení rozhodovacích funkcí nikoliv podle náhodné proměnné $Z(F, \delta, X)$, ale podle její očekávané hodnoty $Z(F, \delta)$. Toto zjednodušení se zdá být dosud vhodným a nebylo, pokud vím, napadáno. (Věc totiž zůstává i tak dosti složitá.)

---

[3]) non-randomized

[4]) Jestliže pro dané x rozhodovací funkce δ nevybere po konečném počtu pozorování žádné $d \in D$, definujeme $Z(F, \delta, x) = +\infty$.

[5]) Předpokládejme, že množina $\mathbf{D}$ byla tak zvolena, že pro každé $\delta \in \mathbf{D}$ má tento integrál smysl. Explicitní charakterisace funkcí δ, pro něž $Z(F, \delta, x)$ je měřitelná (F) pro každé $F \in \mathfrak{M}$, není jednoduchá [16].

4. Znáhodněné rozhodovací funkce. Dříve než použijeme důležitého pojmu ztrátové funkce k hodnocení rozhodovacích funkcí, rozšíříme pojem rozhodovací funkce zavedením znáhodněných rozhodovacích funkcí. V odstavci devátém ukážeme na příkladě smysl a důležitost tohoto rozšíření.

Označme (pro daný statistický problém) symbolem R množinu všech rozhodovacích funkcí, pro něž $Z(F, \delta)$ existuje pro každé $F \in \mathfrak{M}$. Budiž $\mathbf{R}$ σ-algebra podmnožin mnoziny R, obsahující všechny vzory borelovských množin při zobrazení $Z(F,*)$ pro každé $F \in \mathfrak{M}$. Potom, je-li δ^* pravděpodobnostní míra na $\mathbf{R}$, nazveme δ^* znáhodněnou rozhodovací funkcí.

δ^* pak určuje tento postup při řešení statistického problému: Pomocí náhodného mechanismu vybereme jedno $\delta \in R$ tak, že pravděpodobnost, že $\delta \in A \in \mathbf{R}$, je rovna $\delta^*(A)$. Další postup pak určuje vybraná prostá rozhodovací funkce δ.

Nyní je přirozené rozšířit definici ztrátové funkce i na znáhodněné rozhodovací funkce takto:

$$Z(F, \delta^*) = \int Z(F, \delta) \, d\delta^*(\delta) .$$

Je patrné, že můžeme každou prostou rozhodovací funkci δ ztotožnit se znáhodněnou funkcí δ^*, jež připisuje pravděpodobnost 1 množině sestávající z jediného prvku δ. Prosté rozhodovací funkce budou tedy tvořit podsystém systému všech znáhodněných funkcí, které budeme nazývat nyní jednoduše rozhodovacími funkcemi a pro něž budeme používat též označení δ.

Poznamenáme ještě, že znáhodnění rozhodovacích funkcí je možno provésti ještě jiným způsobem, avšak je možno ukázat, že oba způsoby jsou za jistých předpokladů ekvivalentní [17].

V dalších odstavcích se budeme rozhodovacími funkcemi zabývat v podstatě již jen prostřednictvím jejich ztrátových funkcí.

5. Částečné uspořádání rozhodovacích funkcí. Přípustnost rozhodovacích funkcí. Úplnost a úplnost v podstatě systémů rozhodovacích funkcí. Budiž opět $(\mathfrak{M}, D, W, c, \mathbf{D})$ statistický problém. Ze zásady hodnocení rozhodovacích funkcí podle jejich ztrátových funkcí plynou tyto definice:

Definice 1. *Dvě rozhodovací funkce δ_1, δ_2 považujeme za ekvivalentní a píšeme $\delta_1 \sim \delta_2$, jestliže $Z(F, \delta_1) = Z(F, \delta_2)$ pro každé $F \in \mathfrak{M}$.*

Definice 2. *Říkáme, že δ_1 je lepší rozhodovací funkce než δ_2 a píšeme $\delta_1 \precsim \delta_2$, jestliže $Z(F, \delta_1) \leq Z(F, \delta_2)$ pro každé $F \in \mathfrak{M}$ a není $\delta_1 \sim \delta_2$.*

Definice 3. *Rozhodovací funkce $\delta \in \mathbf{D}$ je přípustná $(\mathbf{D})$, jestliže žádná $\delta' \in \mathbf{D}$ není lepší než δ.*

Význam těchto tří definic je jistě zcela jasný. Připomeneme pouze, že přípustnost $(\mathbf{D})$ je podstatně závislá na $\mathbf{D}$. V dalších definicích si budeme všímat systémů rozhodovacích funkcí. Snahou jest, najíti pokud možno nejmenší systém rozhodovacích funkcí s vhodnými vlastnostmi a tím usnadnit definici nejlepší rozhodovací funkce.

Definice 4. *Systém* $\mathbf{D}_0 \subset \mathbf{D}$ *je úplný* ($\mathbf{D}$), *jestliže ke každé* $\delta \in \mathbf{D} - \mathbf{D}_0$ *existuje* $\delta_1 \in \mathbf{D}_0$ *tak, že* $\delta_1 \prec \delta_0$.

Definice 5. *Systém* $\mathbf{D}_0 \subset \mathbf{D}$ *je v podstatě úplný* ($\mathbf{D}$), *jestliže ke každé* $\delta \in \mathbf{D} - \mathbf{D}_0$ *existuje* $\delta_1 \in \mathbf{D}_0$ *tak, že je buď* $\delta_1 \prec \delta$ *nebo* $\delta_1 \sim \delta$.

Definice 6. *Systém* $\mathbf{D}_0$ *je minimální (v podstatě) úplný* ($\mathbf{D}$), *jestliže je (v podstatě) úplný* ($\mathbf{D}$) *a žádný jeho vlastní podsystém nemá tuto vlastnost.*

Při hledání nejlepší rozhodovací funkce se můžeme zřejmě omezit na nějaký v podstatě úplný[6]) systém rozhodovacích funkcí. Je-li tento systém minimální v podstatě úplný ($\mathbf{D}$), dosáhli jsme současně maximálního zjednodušení otázky nejlepší rozhodovací funkce na podkladě částečného uspořádání, zavedeného definicemi 1 a 2.

Jestliže pro daný statistický problém existuje v podstatě úplný ($\mathbf{D}$) *systém, obsahující právě jeden prvek* δ, *je zřejmě možno nazvat tuto* δ *nejlepší rozhodovací funkcí.*

Taková situace nastane však bohužel zřídka. Bude potom třeba nějakého dalšího principu pro definici nejlepší rozhodovací funkce. Od každé obecné definice budeme pak přirozeně požadovat, aby v případě existence minimálního úplného systému byla nejlepší rozhodovací funkce jeho prvkem.

Podat takovou obecnou definici jistě nebude snadnou věcí. Jakmile nějaký minimální v podstatě úplný ($\mathbf{D}$) systém bude obsahovat alespoň dva prvky, potom, ať zvolíme za nejlepší rozhodovací funkci kteroukoliv $\delta \in \mathbf{D}$, bude k ní existovat $\delta_1 \in \mathbf{D}$ tak, že $Z(F, \delta_1) < Z(F, \delta)$ alespoň pro jedno $F \in \mathfrak{M}$. Protože *pravou* distribuční funkci neznáme, může hoření vztah platit právě pro ni.

Nějaký (v podstatě) úplný ($\mathbf{D}$) systém vždy existuje, nemusí však existovat minimální (v podstatě) úplný systém. Za jistých předpokladů však minimální úplný ($\mathbf{D}$) systém existuje a je totožný se systémem všech přípustných ($\mathbf{D}$) rozhodovacích funkcí [18]. Z jeho existence pak snadno plyne existence minimálních v podstatě úplných ($\mathbf{D}$) systémů. Mezi dvěma takovými systémy lze zřejmě definovat prosté zobrazení, kterým jsou si přiřazeny vždy dvě ekvivalentní rozhodovací funkce.

6. Apriorní pravděpodobnost. Bayesovo řešení. Jestliže ξ je pravděpodobnostní míra na σ-algebře $\mathbf{S}$ podmnožin z $\mathfrak{M}$ a $\mathbf{S}$ obsahuje všechny vzory borelovských množin při zobrazení $Z(*, \delta)$ pro každé $\delta \in \mathbf{D}$, nazveme ξ apriorní pravděpodobností. Vzhledem k této pravděpodobnosti můžeme definovat Bayesovu očekávanou hodnotu ztráty

$$Z_\xi(\delta) = \int Z(F, \delta) \, d\xi(F) .$$

Existuje-li rozhodovací funkce, jež minimalisuje Bayesovu očekávanou hod-

---

[6]) Někdy, nebude-li nebezpečí nedorozumění, budeme psáti prostě *úplný* místo *úplný* ($\mathbf{D}$) a podobně.

notu ztráty, budeme ji nazývat Bayesovým řešením vzhledem k apriorní pravděpodobnosti ξ a budeme ji označovat δ_ξ. Splňuje tedy δ_ξ vztah

$$Z_\xi(\delta_\xi) = \min_{\delta \in \mathbf{D}} Z_\xi(\delta) .$$

Pojem apriorní pravděpodobnosti a Bayesova řešení byl velmi často napadán i hájen v diskusích plných nedorozumění (viz na příklad [14], str. 75—76).

Právem bývá napadán tento postup: Zvolí se apriorní pravděpodobnost ξ (na př. připisující každému $F \in \mathfrak{M}$ stejnou pravděpodobnost) a prohlásí se δ_ξ za nejlepší řešení statistického problému. Skutečně, tento postup, pokud nemůže odůvodnit výběr speciální pravděpodobnosti ξ, je těžko pochopitelný.

Neprávem se však napadá samotné zavedení pojmů apriorní pravděpodobnost a Bayesovo řešení, neboť v matematice jistě nedospějeme k nesprávným výsledkům pouhým zavedením pojmů.

V některých případech je však apriorní pravděpodobnost ξ matematickým modelem jisté reálné vlastnosti statistického problému. V takovém případě se zdá být vhodnou definice označující δ_ξ za nejlepší rozhodovací funkci.

Tak na příklad byl proveden tento experiment: Byla hozena mince; padl-li líc, byla pozorována náhodná proměnná X s distribuční funkcí F_1. Padl-li rub, byla pozorována náhodná proměnná X s distribuční funkcí F_2. Na podkladě pozorování náhodné proměnné X má statistik určit vybranou distribuční funkci F_1. Zřejmě, zvolíme-li $\xi(\{F_i\}) = \frac{1}{2}$, bude δ_ξ vhodnou rozhodovací funkcí. Existují též netriviální příklady [5].

Abychom se vyhnuli zmatku, který často vzniká zaměňováním dvou různých pojmů, totiž apriorní pravděpodobnosti (tak jak jsme ji my definovali) a speciální apriorní pravděpodobnosti, jež je modelem nějaké konkretní informace o statistickém problému, speciálně o množině $\mathfrak{M}$, budeme poslední pojem označovat slovy konkretní apriorní pravděpodobnost.

Bohužel ve většině případů nemůžeme buď vůbec předpokládat existenci konkretní apriorní pravděpodobnosti a i v těch případech, kdy tato konkretní apriorní pravděpodobnost existuje, ji často nebudeme znát a tedy nebudeme moci použít Bayesova řešení.

Ještě jednu poznámku k případu, kdy známe konkretní apriorní pravděpodobnost ξ. Může se pak stát, že Bayesovo řešení vzhledem ke ξ neexistuje, neboť pro každé $\delta \in \mathbf{D}$

$$Z_\xi(\delta) > \inf_{\delta' \in \mathbf{D}} Z_\xi(\delta') .$$

V takovém případě se můžeme minimální ztrátě přiblížit libovolně blízko, neboť ke každému kladnému ε existuje taková rozhodovací funkce δ, že

$$Z_\xi(\delta) < \inf_{\delta' \in \mathbf{D}} Z_\xi(\delta') + \varepsilon .$$

Nevyžaduje tedy tento případ (jehož obdobu najdeme v příštím odstavci) zvláštního řešení.

7. Minimax. Jestliže nemáme žádné předběžné informace o pravé distribuční funkci (až na množinu $\mathfrak{M}$), navrhuje se toto řešení. Označme

$$r(\delta) = \sup_{F \in \mathfrak{M}} Z(F, \delta)$$

a nazývejme toto číslo risiko rozhodovací funkce δ.[7])

Je-li tedy r risiko rozhodovací funkce δ, můžeme říci, že při použití rozhodovací funkce δ je očekávaná hodnota ztráty menší nebo rovna r pro všechna $F \in \mathfrak{M}$. (A r je nejmenší číslo, pro něž toto tvrzení platí.)

Minimaxovou rozhodovací funkcí nazvu takovou rozhodovací funkci $\delta \in \mathbf{D}$, pro niž platí

$$r(\delta) \leq r(\delta') \quad \text{pro všechna} \quad \delta' \in \mathbf{D}\,.$$

Za nejlepší rozhodovací funkci pak pokládám každou minimaxovou přípustnou (**D**) *rozhodovací funkci.* Jestliže taková funkce neexistuje, je možno uvažovat způsobem podobným jako na konci předchozího odstavce.

Snadno se zjistí, že existuje-li jediná minimaxová rozhodovací funkce, je přípustná (**D**).

8. Subminimax. Často pro minimaxovou rozhodovací funkci δ^* je $Z(F, \delta^*) = r(\delta^*)$ pro každé $F \in \mathfrak{M}$ a tedy ztrátová funkce nabývá stálé dosti velké hodnoty. Proto se navrhuje hledat tak zvaná subminimaxová řešení, t. j. rozhodovací funkce $\delta \in D$, pro něž $r(\delta) \leqq r(\delta^*) + \varepsilon$, kde ε je malé kladné, a která značně zmenšují ztrátovou funkci pro *mnohá* $F \in \mathfrak{M}$ ([8]). Zajímavý příklad jedné subminimaxové rozhodovací funkce uvádí Robbins [15].

Hodges a Lehmann [11] uvažují statistický problém, v němž je znám odhad ξ konkretní apriorní pravděpodobnosti. Zvolím kladné ε takové, aby zvýšení risika o ε nepřineslo podstatnou škodu a hledám funkci $\delta^{**} \in \mathbf{D}$, jež minimalisuje Bayesovu ztrátu vzhledem ke ξ za podmínky $r(\delta^{**}) \leq r(\delta^*) + \varepsilon$.

Autoři propočítávají příklad takového řešení a ukazují, že se jim v daném případě na úkor malého zvětšení risika podařilo oproti minimaxové rozhodovací funkci podstatně snížit Bayesovu ztrátu vzhledem ke ξ.

Ačkoliv tyto úvahy jistě zatím nemohou být použity k definici optimálnosti rozhodovací funkce (problém nejvýhodnějšího ε), znamenají další objasnění problematiky.

9. Příklad. Budiž $X = X_1$ jednorozměrná náhodná proměnná, množina $\mathfrak{M}$ nechť obsahuje dva prvky F_1 a F_2, množina D dva prvky d_1 a d_2, cena $c = 0$ identicky,

$$W(F_i, d_j) = \begin{cases} 0, & \text{je-li } i = j\,, \\ 1, & \text{je-li } i \neq j\,, \end{cases}$$

[7]) Pozměnil jsem obvyklé názvy zavedených pojmů. Obvykle bývá *risikovou funkcí* nazývána naše *ztrátová funkce* a pro naše *risiko* nebývá zaváděno žádné slovní označení.

a X nechť nabývá hodnot $-1, 0$ podle F_1 a hodnot $0, 1$ podle F_2, každé hodnoty s pravděpodobností $\frac{1}{2}$. **D** budiž systém všech rozhodovacích funkcí.

Definujme dvě prosté rozhodovací funkce δ_1, δ_2:

$$\delta_i(\emptyset) = \{1\},$$
$$\delta_i(\{1\}, -1) = d_1, \quad \delta_i(\{1\}, 1) = d_2,$$
$$\delta_1(\{1\}, \ 0) = d_1, \quad \delta_2(\{1\}, 0) = d_2$$

a δ_α (pro $\alpha \in \langle 0, 1 \rangle$) jako znáhodněnou rozhodovací funkci, jež s pravděpodobností α rozhoduje jako δ_1 a s pravděpodobností $1 - \alpha$ rozhoduje jako δ_2.

Je zřejmé, že $\{\delta_\alpha\}_{\alpha \in \langle 0,1 \rangle}$ je minimální v podstatě úplný (**D**) systém rozhodovacích funkcí. Současně je také systémem všech přípustných funkcí. Mimo to je každá δ_α Bayesovým řešením vzhledem k apriorní pravděpodobnosti, jež každému F_i připisuje pravděpodobnost $\frac{1}{2}$.

Je

$$Z(F_i, \delta_i) = \begin{cases} 0, & \text{je-li } i = j, \\ \frac{1}{2}, & \text{je-li } i \neq j, \end{cases}$$
$$Z(F_1, \delta_\alpha) = \frac{1-\alpha}{2}, \quad Z(F_2, \delta_\alpha) = \frac{\alpha}{2},$$
$$Z_\xi(\delta_\alpha) = \tfrac{1}{2}[\xi_1(1-\alpha) + (1-\xi_1)\,\alpha], \text{[8]}$$
$$Z_{(\frac{1}{2}, \frac{1}{2})}(\delta_\alpha) = \tfrac{1}{4} \text{ pro všechna } \alpha.$$

Bayesovým řešením pro $\xi = (\xi_1, \xi_2)$, $\xi_1 > \xi_2$ je δ_1, pro $\xi = (\xi_1, \xi_2)$, $\xi_2 > \xi_1$ je δ_2. Pro $\xi = (1, 0)$ je Bayesovým řešením též rozhodovací funkce δ, jež bez ohledu na pozorování vybírá rozhodnutí d_1, t. j. $\delta(\emptyset) = d_1$. Tato rozhodovací funkce již není přípustná (**D**).

Minimaxová přípustná rozhodovací funkce existuje jediná: $\delta_{\frac{1}{2}}$, s risikem $\frac{1}{4}$ a konstantní ztrátovou funkcí $Z(F_i, \delta_{\frac{1}{2}}) = \frac{1}{4}$ pro $i = 1, 2$.

Je patrné, že při omezení se na prosté rozhodovací funkce by minimální risiko bylo $\frac{1}{2}$. Na druhé straně jsou však známy případy, kdy systém všech prostých rozhodovacích funkcí je úplný [6].

10. Některá nedorozumění týkající se minimaxového řešení a některých klasických metod. V celém odstavci nechť δ^* značí minimaxovou přípustnou rozhodovací funkci.

Nejtriviálnější námitka je ta, že minimaxová rozhodovací funkce minimalisuje risiko, ačkoliv by měla minimalisovat ztrátu $Z(F, \delta)$, kde F je *pravá* distribuční funkce. O této námitce jsme již mluvili v odstavci 5. Víme také, že když existuje pro daný statistický problém taková rozhodovací funkce $\delta' \in$ **D**, pro kterou $Z(F, \delta') \leq Z(F, \delta)$ pro každé $\delta \in$ **D** a $F \in \mathfrak{M}$, potom $\delta' \sim \delta^*$ a δ^* minimalisuje tedy i $Z(F, \delta)$, kde F je *pravá* distribuční funkce.

[8]) Apriorní pravděpodobnost $\xi = (\xi_1, \xi_2)$ je dána čísly $\xi_1 = \xi(\{F_1\})$, $\xi_2 = \xi(\{F_2\}) = 1 - \xi_1$.

Za jistých dosti obecných podmínek je δ^* Bayesovým řešením vzhledem k tak zvané nejméně příznivé apriorní pravděpodobnosti ξ, jež je definována vztahem $Z_\xi(\delta_\xi) \geq Z_\eta(\delta_\eta)$ pro všechny apriorní pravděpodobnosti η, při čemž δ_ξ a δ_η jsou Bayesova řešení vzhledem ke ξ a k η.

Tato vlastnost má dvojí význam:

1. Vede k odůvodnění minimaxového řešení v theorii her, kdy minimaxové řešení je Bayesovým řešením za předpokladu, že protivník hraje co nejlépe, t. j. pro nás nejméně příznivě.

2. Je důležitá pro hledání minimaxové rozhodovací funkce v daném statistickém problému.

Tato vlastnost však též působí, jak se domnívám, dvojí nedorozumění:

1. Svádí k odůvodňování minimaxového řešení analogií s theorií her. ([16], str. 25: „...we can hardly say, that Nature wishes to maximize $Z(F, \delta)$. Nevertheless, since Nature's choice is unknown to the experimenter, it is perhaps not unreasonable for the experimenter to behave as if Nature wanted to maximize the risk".) Taková analogie podle mého názoru neexistuje a takové pokusy o odůvodnění jsou právem napadány.

2. Namítá se, že minimaxové řešení je nevhodné, neboť je Bayesovým řešením k nejméně příznivé apriorní pravděpodobnosti ξ. Nemůžeme však obecně předpokládat (a nám jde o obecnou definici optimálnosti), že ξ je konkretní pravděpodobnost, neboť jednak konkretní pravděpodobnost vůbec nemusí existovat, nebo existuje, ale neznáme ji, nebo konečně ji známe, ale může být různá od ξ. Tato námitka, myslím, není oprávněná. Známe-li konkretní apriorní pravděpodobnost η, použijeme Bayesova řešení δ_η. Minimaxového řešení použijeme jen tehdy, jestliže nám konkretní pravděpodobnost není známa, nebo neexistuje-li. Jestliže neznáme konkretní apriorní pravděpodobnost η a nepoužijeme Bayesova řešení vzhledem k η, nemůže to být vytýkáno ze stejných důvodů, jako nebývá vytýkáno, že nevybereme právě to $d \in D$, pro něž $W(F, d) = \inf_{\delta' \in D} W(F, d')$, je-li F *pravá* distribuční funkce. Přitom je patrno z odstavce 7, že k zavedení minimaxové rozhodovací funkce nás vedly jiné důvody než její vlastnost Bayesova řešení vzhledem k nejméně příznivé apriorní pravděpodobnosti. Ve skutečnosti při definici minimaxové rozhodovací funkce jsme pojmu apriorní pravděpodobnosti a tím méně pojmu konkretní apriorní pravděpodobnosti vůbec nepoužili.

Další námitky bývají tyto: 1. Princip minimaxu vede k rozhodovacím funkcím, které z jistých důvodů, daných konkretní situací, jsou nevhodné (složitý sekvenční postup a pod.). 2. $Z(F, \delta)$ bývá velká právě pro ta $F \in \mathfrak{M}$, která nepřicházejí prakticky v úvahu. V prvém případě byla tedy zřejmě špatně definována množina $\mathbf{D}$, v druhém případě byla buď špatně definována množina $\mathfrak{M}$, nebo jsme měli použít Bayesova řešení. Ve druhém případě je ovšem věc poněkud komplikovaná, neboť existuje ještě třetí možnost: Konkretní apriorní pravděpodobnost sice neznáme, přesto však se o některých prvcích $F \in \mathfrak{M}$ domníváme, že jsou pravděpodobnější, než jiné. Jestliže tyto informace o množině $\mathfrak{M}$ jsou takového rázu, že je můžeme vyjádřit odhadem ξ konkretní apriorní pravděpodobnosti η, může snad k vhodným výsledkům vést cesta naznačená v odstavci 8. Uvědomíme si, že na tuto metodu možno pohlížeti jako na zobecnění minimaxového i Bayesova řešení statistického problému. Zřejmě ε rovnému nule odpovídá minimaxové řešení a $\varepsilon = +\infty$ Bayesovo.

Konečně se někteří statistikové domnívají, že již před theorií rozhodovacích funkcí statistika studovala mnohé problémy, že došla poměrně jednoduchými metodami k elegantním řešením a že je možno tímto způsobem postupovat dále.

V těchto případech šlo však vždy spíše o to, najít vůbec nějakou rozhodovací funkci a nezkoumala se otázka optimálnosti. Jestliže se tato otázka zkoumala, byla často optimálnost ne příliš vhodně definována.

Připomeneme při této příležitosti, že mnohé metody (jako na příklad testy shody, analysa rozptylu) byly a jsou soustavně aplikovány na jiné problémy, než pro které byly odvozeny, neboť tak zvanou nulovou hypothesu bývá obyčejně možno předem prohlásit za neplatnou a její zamítnutí je jen věcí rozsahu výběru. Když potom konstruujeme rozhodovací funkci pro přesně formulovaný statistický problém, dostáváme podstatně jiné výsledky.

Všimněme si nyní některých klasických (a nesekvenčních) řešení, jak bývají uváděna v moderních učebnicích matematické statistiky.

Testování hypothes. V případě dvou jednoduchých hypothes podává řešení věta Neymann-Pearsonova, jež pro danou chybu prvního druhu α udává nejlepší rozhodovací funkci δ_α (minimalisující chybu druhého druhu β). $\{\delta_\alpha\}$ je minimální v podstatě úplný ($\mathbf{D}$) systém rozhodovacích funkcí (je-li $\mathbf{D}$ systém všech nesekvenčních rozhodovacích funkcí a je-li $W(F_i, d_j) = \left\langle \begin{matrix} 0, & i = j \\ 1, & i \neq j \end{matrix} \right)$. Věta Neymann-Pearsonova dává tedy stejné výsledky jako theorie rozhodovacích funkcí. Minimaxové řešení bude taková δ_α, pro niž bude $\alpha = \beta$.

Pro složené hypothesy jedna z navrhovaných metod je princip největší věrohodnosti. Jsou však známy případy, v nichž tento princip vede ke zcela nesmyslným rozhodovacím funkcím [12].

V případě testování jednoduché hypothesy $\Theta = \Theta_0$ proti složené alternativní hypothese $\Theta \neq \Theta_0$ navrhovali někteří autoři použíti lokálně nejlepších testů. Tento princip se zdá být nepřijatelný, neboť ve většině statistických problémů bude závažnost chybného rozhodnutí růst s rostoucí vzdáleností pravého parametru od Θ_0. Birnbaum [3] konstruoval takový lokálně nejlepší test, jenž pro každé Θ splňující podmínku $|\Theta - \Theta_0| > c$ (kde Θ je pravá hodnota parametru a c je konstanta závisející jen na velikosti testu α) je horší než všechny ostatní testy.

Odhady. V theorii bodového odhadu neznámého parametru je největší důraz kladen na nestranné odhady. Mezi těmito odhady často existuje odhad s minimálním rozptylem; takový odhad se pokládá za nejlepší.

Theorie rozhodovacích funkcí nedává tak jednoduché a jednoznačné řešení (úplný systém rozhodovacích funkcí obsahuje víc než jeden prvek). Snadno však zjistíme proč. Za obecných podmínek totiž platí, že každá přípustná rozhodovací funkce je Bayesovým řešením vzhledem k nějaké apriorní pravděpodobnosti ([16], Theorem 3.20, str. 101). GIRSHICK a SAVAGE [9] dokázali dále, že žádné Bayesovo řešení není nestranným odhadem, je-li $W(\Theta, g) = \lambda(\Theta)(g - \Theta)^2$, kde g je odhad parametru Θ a $\lambda(\Theta)$ je kladná funkce. To však znamená, že za těchže podmínek žádný nestranný odhad není přípustný.

Všimněme si ještě principu invariance ([5], [12]). Zdálo by se, že tento intuitivně velmi přijatelný princip by mohl být vhodným požadavkem na rozhodovací funkce a že by mohl vhodně zmenšit nějaký úplný (**D**) systém rozhodovacích funkcí a tím usnadnit definici optimálnosti. Jestliže bychom nahradili původní systém **D** systémem $\bar{\mathbf{D}}$ všech invariantních $\delta \in \mathbf{D}$, pak by možná minimaxové řešení tohoto pozměněného statistického problému mělo lepší vlastnosti. Dá se však bohužel ukázat, že tento postup není obecně možný, neboť 1. za jistých předpokladů bychom dostali minimaxové řešení původního problému, 2. za jistých předpokladů by však toto řešení vedlo k nepřípustné (**D**) rozhodovací funkci [12], [4]. Tím není ovšem vyloučena použitelnost principu invariance ve speciálních případech a to zejména ke konstrukci minimaxových řešení [13].

Naposledy si všimneme tak zvané metody velkých výběrů. Rozsáhlých výběrů použije vhodná rozhodovací funkce tehdy, bude-li cena pozorování malá ve srovnání s hodnotami funkce W. Jinak ovšem posuzování a aplikace rozhodovací funkce δ_N (vyžadující pevný počet pozorování N) podle limitních vlastností posloupnosti $\{\delta_N\}$, může být nanejvýš heuristickou pomůckou; to je právě podstata metody velkých výběrů.

11. Konstrukce minimaxových rozhodovacích funkcí. Jenom krátce se zmíníme o tom, že konstrukce minimaxových rozhodovacích funkcí pro daný statistický problém bývá značně obtížná. Zjednodušení se dosáhne uvažováním nějakého v podstatě úplného systému. Přitom velký význam má použití sufficientní a transitivní posloupnosti statistik [1].

Upozorníme ještě na dosti značnou nesnáz spočívající v tom, že minimaxová rozhodovací funkce podstatně závisí na definici daného statistického problému, zejména na množině $\mathfrak{M}$, která se bude v praktických aplikacích často měnit. Proto bude obtížné podat řešení dosti obecná, která by se mohla často aplikovat. Tak při přejímací kontrole určitého výrobku nebude třeba předem odhadovati četnost vadných výrobků celým intervalem $\langle 0, 1 \rangle$ a pro odhad této četnosti nebude tedy minimaxovou rozhodovací funkcí odhad Hodges-Lehmanův [10].

Takovéto konstrukční nesnáze však nemají mít podle mého názoru vliv na definici optimálnosti rozhodovací funkce. Možná ostatně, že se během doby některé obtíže konstrukce odstraní a v jiných případech že se najdou vhodné aproximativní metody hledání minimaxové rozhodovací funkce.

Zatím ovšem, když má statistik řešit daný statistický problém a nezná nejvhodnější rozhodovací funkci, je oprávněn navrhnouti jinou rozhodovací funkci, při čemž by měl znát alespoň nějaký odhad její ztrátové funkce. Na odmítnutí řešit takový problém se totiž můžeme dívat jako na speciální rozhodovací funkci s jistými důsledky a s určitou ztrátovou funkcí. Odmítnutí se pak může ukázat horší, než některá z rozhodovacích funkcí, jež statistik zná a může použít.

14. Závěr. Ukázali jsme, jak theorie rozhodovacích funkcí hodnotí jednotlivé rozhodovací funkce podle jejich ztrátových funkcí a jak z tohoto hodnocení vyplývá částečné uspořádání rozhodovacích funkcí.

Protože toto částečné uspořádání většinou neurčuje nejlepší prvek, je k jeho stanovení třeba dalšího principu. Takovým principem je princip minimaxového řešení.

Minimax není ovšem jediným možným řešením. Ukázali jsme význam a použití Bayesova řešení, je-li známa konkretní apriorní pravděpodobnost, a zmínili jsme se o kompromisních, t. zv. subminimaxových řešeních.

Zabývali jsme se též některými jinými požadavky, na př. nestranností a invariancí, a ukázali jsme, že nemohou být přijaty za obecné kriterium vhodnosti rozhodovacích funkcí.

LITERATURA

[1] *R. R. Bahadur*: Sufficiency and Statistical Decision Functions. Annals of Mathematical Statistics *25*, str. 423—462.

[2] *Robert E. Bechhofer*: A Single Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. Annals of Mathematical Statistics *25*, str. 16—39.

[3] *Allan Birnbaum*: Admissible Tests for the Mean of a Rectangular Distribution. Annals of Mathematical Statistics *25*, str. 157—161.

[4] *David Blackwell*: On the Translation Parameter Problem of Discrete Variables. Annals of Mathematical Statistics, *22*, str. 393—399.

[5] *V. Fabian, A. Špaček*: Experience in Statistical Decision Problems. Čech. mat. žurn. *6* (81), 1956.

[6] *Dvoretzky, Wolfowitz, Wald*: Elimination of Randomization in Certain Statistical Decision Procedures and Zero-sum Two-person Games. Annals of Mathematical Statistics *22*, str. 1—21.

[7] *D. Blackwell, M. A. Girshick*: Theory of Games and Statistical Decisions. John Wiley and Sons, 1954.

[8] *P. Frank* and *J. Kiefer*: Almost Subminimax and Biased Minimax Procedures. Annals of Mathematical Statistics *22*, str. 465—468.

[9] *M. A. Girshick* and *L. J. Savage*: Bayes and Minimax Estimates for Quadratic Loss Functions. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (1951), str. 149—158.

[10] *J. L. Hodges* and *E. L. Lehmann*: Some Problems in Minimax Estimation. Annals of Mathematical Statistics *21*, str. 182—197.

[11] *J. L. Hodges* and *E. L. Lehmann*: Using Previous Experiance in Reaching Statistical Decisions. Annals of Mathematical Statistics *23*, str. 396—407.

[12] *E. L. Lehmann*: Some Principles of the Theory of Testing Hypotheses. Annals of Mathematical Statistics *21*, str. 1—28.

[13] *E. L. Lehmann* and *C. M. Stein*: The Admissibility of Certain Invariant Statistical Tests Involving a Translation Parameter. Annals of Mathematical Statistics *24*, str. 473—479.

[14] *D. V. Lindley*: Statistical Inference. Journal of the Royal Statistical Society, Series B (Methodolical) *XV*, str. 21—65.

[15] *Herbert Robbins*: Asymptotically Subminimax Solutions of Compound Statistical Decisions Problems. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (1951), str. 131—148.

[16] *Abraham Wald*: Statistical Decision Functions, New York 1950.

[17] *A. Wald* and *J. Wolfowitz*: Two Methods of Randomization in Statistics and the Theory of Games. Annals of Mathematics *53*, str. 581—586.

[18] *A. Wald* and *J. Wolfowitz*: Characterization of the Minimal Complete Class of Decision Functions when the Number of Distributions and Decisions is Finite. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (1951), str. 149—158.

[19] *J. Wolfowitz*: On ε-Classes of Decision Functions. Annals of Mathematical Statistics *22*, str. 461—464.