Časopis pro pěstování matematiky a fysiky

Jaroslav Janko

Rozvoj theorie neparametrických testů ve statistické indukci

Časopis pro pěstování matematiky a fysiky, Vol. 74 (1949), No. 2, 62--74

Persistent URL: http://dml.cz/dmlcz/123049

Résumé. — Výtah.

Revue des travaux publiés en 1935-1948 sur les chaînes de Markoff et problèmes voisins.

BOHUSLAV HOSTINSKÝ, Brno.

Le problème primitif relatif à une chaîne simple de Markoff consiste à étudier les probabilités de passage $P_{ik}^{(n)}$ en fonction des indices i, k et de l'indice d'itération n. Introduisons des variables continues au lieu des indices. Ce changement nous amène à considérer d'abord l'équation de SMOLUCHOWSKI (9a) ou l'équation plus générale de CHAPMAN (9), où Φ est la densité de probabilité de passage. Admettons plus généralement que la probabilité de passage (sur un segment de droite) soit représentée par $\Phi(x, Y, s, t)$, où x est la position initiale (à l'époque s) et où Y désigne une partie du segment considéré; le point se trouve à l'intérieur de Y à l'époque t. Φ satisfait a l'équation (11). POSPÍŠIL a donné en 1935 une première solution de (11) sous la forme d'une série infinie. Il reste à interpréter, au point de vue du Calcul des probabilités, la signification des termes de cette série. La Théorie des chaînes et surtout la recherche des solutions de l'équation (11) doit servir à perfectionner l'expression analytique des lois statistiques qui régissent le développement des systèmes physiques.

ROZVOJ THEORIE NEPARAMETRICKÝCH TESTŮ VE STATISTICKÉ INDUKCI.

JAROSLAV JANKO, Praha.

Při řešení problémů statistické indukce se většinou předpokládalo, že kumulativní distribuční funkce základního souboru závisí známým způsobem na jistých parametrech, jejichž velikost se odhaduje. Tato theorie případu parametrického byla vybudována pracemi R. A. FISHERA, J. NEYMANA, E. PEARSONA, A. WALDA, S. WILKSE a řady dalších badatelů. Byla aplikována na většinu známých frekvenčních funkcí, které se vyskytují v praxi statistické, a její užívání bylo usnadněno množstvím tabulek.

Theorie statistických testů se věnovala v poslední době zvýšenou měrou problémům, kde není možno předpokládati určitou funkční formu rozdělení četností základního souboru, a snažila se podati řešení, které platí pro všechny základní soubory se spojitými kumulativními distribučními funkcemi. Problémy tohoto druhu se nazývají neparametrickými. O jejich řešení chci podati krátký přehled. Jednotné ucelené theorie tu ještě nemáme. Mohu se dotknouti tedy hlavních problémů a jejich řešení obsažených v příslušných pojednáních, na něž odkazuji. Vyčerpávající přehled této theorie do roku 1943 podal SCHEFFÉ (25) s podrobným uvedením literatury.

V každém problému statistické indukce se předpokládá, že kumulativní distribuční funkce $F(x)$ patří do nějaké dané třídy Ω distribučních funkcí $F \in \Omega$. Mají-li náhodné proměnné $X_1, X_2, \ldots, X_N$ kumulativní distribuční funkce $F(X_1), F(X_2), \ldots, F(X_N)$, pak říkáme, že určité hodnoty $x_1, x_2, \ldots, x_N$ tvoří náhodný výběr rozsahu N ze základního souboru s kumulativní distribuční funkcí $F(x)$. V případě procesu náhodného výběru se nazývá N-rozměrný Euklidovský prostor R_N prostorem výběrovým W. Je-li Ω k-parametrický svazek funkcí, nazývá se problém parametrickým; jinak neparametrickým. V theorii neparametrických problémů se vhodně užívá roztřídění kumulativních distribučních funkcí do čtyř tříd:

Ω_0 bude třída všech jednoproměnných kumulativních distribučních funkcí, t. j. třída všech monotonních neklesajících funkcí $F(x)$, pro něž $F(-\infty) = 0$, $F(+\infty) = 1$, $F(x) = F(x + 0)$. Ke každé funkci $F \in \Omega_0$ si můžeme představit příslušnou náhodnou proměnnou X takovou, že $P(X \leq x) = F(x)$. K některým účelům potřebujeme odlišit třídu $\Omega^{(0)}$ degenerovaných kumulativních distribučních funkcí daných rovnicemi $F(x) = 0$ pro $x < x_0$ a $F(x) = 1$ pro $x \geq x_0$, kde x_0 je reálné číslo. Potom

Ω_1 bude třída nedegenerovaných kumulativních distribučních funkcí $\Omega_1 = \Omega_0 - \Omega^{(0)}$,

Ω_2 bude třída spojitých $F(x)$,

Ω_3 bude třída všech absolutně spojitých $F(x)$, t. j. všech $F(x)$, pro která existuje funkce hustoty pravděpodobnosti $f(x)$ taková, že

$$F(x) = \int_{-\infty}^{x} f(t)\, \mathrm{d}t \tag{1}$$

a konečně

Ω_4 bude třída všech $F(x)$, které lze vyjádřit ve tvaru (1) s $f(x)$ spojitou.

Pokroky v řešení neparametrických problémů se uskutečnily za omezení, že kumulativní distribuční funkce patří do jedné z uvedených tříd Ω_ν ($\nu = 0, 1, 2, 3, 4$).

Mějme tedy kumulativní distribuční funkci N proměnných, která je členem nějaké dané třídy Ω v N-rozměrném prostoru čili $F_N \in \Omega_\nu$.

Je-li ω daná podtřída v Ω, pak statistická hypotéza je tvrzení, že $F_N \in \omega$. Test hypotézy spočívá v tom, že se zvolí t. zv. kritický obor testu w ve výběrovém prostoru W a padne-li výběrový bod $E(x_1, x_2, \ldots, x_N)$ do oboru w, hypotéza se zamítne. Volba kritického oboru w se obyčejně provádí tak, že se zvolí určitá kladná konstanta α, a existují-li obory w, pro něž pravděpodobnost, že výběrový bod E padne do w, vypočtená z kumulativní distribuční funkce F_N, je rovna α pro všechny $F_N \in \omega$, pak se volba omezí na tuto třídu. V parametrickém případě zavedl pro tyto obory J. Neyman název „obory podobné“, t. j. podobné výběrovému prostoru vzhledem ke všem $F_N \in \omega$. Mezi všemi existujícími „podobnými obory“ pro zvolené α se pak ve speciálním problému zvolí takový obor w, který má nejvhodnější silofunkci. V neparametrickém případě je obyčejně provedena volba určitého w ze svazku podobných oborů pomocí statistického koeficientu zvoleného intuitivně. K získání podobných oborů navrhl obecnou methodu R. A. FISHER a nazývá se methodou znáhodnění (randomisační). Její podstatu lze podle SCHEFFÉHO vyložiti takto:

Uvažujme množinu S takových permutací souřadnic $x_1, x_2, \ldots, x_N$, bodu E, které ponechají beze změny všechny kumulativní distribuční funkce F_N v ω. Jejich počet nechť je s; celkový počet všech permutací těchto souřadnic je $N!$. Pro kterýkoliv bod E ve výběrovém prostoru W definujeme nyní množinu $\{E'\}$, která obsahuje s bodů, které jsme dostali, když jsme na souřadnice bodu E provedli permutace množiny S. Hodnota kumulativní distribuční funkce F_N je tedy táž ve všech s bodech E, vzniklých z E pro všechny $E \in W$ a všechny $F_N \in \omega$. Těch s bodů z množiny $\{E'\}$ bude od sebe odlišných až na případy, kdy bod E leží v určitém oboru, který označíme W_0 a který závisí na množině S permutací určených třídou ω, a budou vždy obsaženy ve svazku všech diagonálních nadrovin $x_i = x_j$, kde $i \neq j$. K sestrojení kritického oboru w při randomisační methodě se zvolí celé kladné číslo $r < s$ a pro každé E, které není ve W_0, se vezme r bodů z příslušné množiny $\{E'\}$ do oboru w a zbývajících $s - r$ bodů zůstane vně w. Obor w je pak obor podobný s konstantou $\alpha = \dfrac{r}{s}$, jak ukázal SCHEFFÉ (26).

O tom, kterých r bodů z podsouboru s bodů se má vzíti dovnitř kritického oboru w se rozhoduje obyčejně pomocí intuitivně zvoleného statistického koeficientu T. Je to nějaká funkce výběru $T(E)$, která nezávisí na kumulativní distribuční funkci F_N. Vezme se tedy r bodů z podsouboru $\{E'\}$, které dávají hodnoty $T(E')$ v určitém oboru (obyčejně r největších nebo nejmenších hodnot), do kritického oboru a tyto hodnoty se pak nazývají „významnými“.

Specielním případem obecných method randomisačních je methoda pořadí (25), která využívá zvláštního druhu symetrie F_N. Nechť jsou $F_N \in \omega$ úplně symetrické v každé z jistých podmnožin souřadnic, tedy na př. v t množinách o $n_1, n_2, \ldots, n_t$ souřadnicích, kde $\sum_{i=1}^{t} n_i = N$.

Vezměme, jak to činí SCHEFFÉ, třeba souřadnice počítané tak, že F_N je úplně symetrická v množině $x_{p_i+1}, x_{p_i+2}, \ldots, x_{p_i+n_i}$

$$\left(p_i = \sum_{j=1}^{i-1} n_j;\ i = 2, 3, \ldots, t;\ p_1 = 0\right) \text{ pro všechny } F_N \in \omega.$$

Množina permutací S je tedy zobecněna utvořením všech $n_i!$ permutací na n_i souřadnicích $x_{p_i+1}, \ldots, x_{p_i+n_i}$ $(i = 1, 2, \ldots, t)$, takže celkový počet permutací v S je $s = n_1!\, n_2! \ldots n_t!$. V souhlase s i-tou množinou souřadnic, ve které F_N je symetrická, rozdělíme výběrový prostor W v $n_i!$ oborů definovaných vztahy $x_{p_i+1} < x_{p_i+2} < \ldots < x_{p_i+n_i}$ a ostatními nerovnostmi, které dostaneme permutací indexů v této nerovnosti. Označme tyto obory $w_{i,k}$ $(k = 1, 2, \ldots, n_i!)$. Dále nechť je

$$w_{k_1, k_2, \ldots k_t} = w_{1k_1} \cdot w_{2k_2} \cdot w_{3k_3} \cdot \ldots w_{tk_t}$$

tedy $w_{k_1,k_2,\ldots,k_t}$ je částí W společnou oborům uvedeným na pravé straně. Tímto postupem se rozděluje výběrový prostor W na s disjunktních oborů $w_{k_1k_2\ldots k_t}$, které označíme jednoduše w_α $(\alpha = 1, 2, \ldots, s)$. Množina těchto oborů pokrývá celý výběrový prostor W až na obor W_0, v němž některé souřadnice jsou sobě rovné. Řekneme, že výběrový bod E má pořadí R_α, jestliže spadá do w_α. Můžeme pak mluviti o „pořadí" jako o náhodné proměnné, která nabývá s možných hodnot R_α. Kritický obor w se sestrojí methodou pořadí, kde w je svazek r oborů w_α. Pořadí R_α odpovídající r oborům w_α, které tvoří kritický obor w, se nazývají významnými. Statistický koeficient $T(E)$, jehož se užije za kriterium pro rozhodnutí, která jsou významná pořadí, bude jen funkcí pořadí.

Vidíme, že methoda pořadí zjednodušuje problém testování statistických hypotéz tím, že nekonečný N-rozměrný výběrový prostor W je nahražen konečným prostorem o s bodech, když $F_N \in \Omega_2$.

Jedním ze základních problémů v neparametrické statistické indukci je problém odhadu kumulativní distribuční funkce $F(x)$, o níž je známo jen, že je spojitá. Řeší se pomocí výběrové kumulativní distribuční funkce $F_N(x)$ tak, že se sestrojí obory přijatelnosti, v nichž $F(x)$ zcela leží s určitou pravděpodobností α. Lze postupovati tak (46), že se definují funkce

$$L_1(x) = F(x) + \delta_1[F(x)]$$
$$L_2(x) = F(x) - \delta_2[F(x)],$$

kde $\delta_1(x)$ a $\delta_2(x)$ jsou spojité funkce nezáporné, definované na uzavřeném intervalu $\langle 0, 1\rangle$. Naším úkolem jest určit pravděpodobnost, že $F_N(x)$ bude ležet v tomto pásu; že tedy

$$L_2(x) \leqq F_N(x) \leqq L_1(x) \tag{2}$$

pro všechna x, a pak dosáhnouti pomocí $\delta_1(x)$ a $\delta_2(x)$ toho, aby tato pravděpodobnost byla α. Potom chceme, aby obor uvnitř pásu (2) tvořil obor přijatelnosti, takže všechna $F(x)$, pro která $F_N(x)$ je v oboru přijatelnosti, budou tvořit obor spolehlivosti $R(F_N)$. Tento obor spolehlivosti je možno skutečně prakticky sestrojiti, neboť lze ukázati (46), že obor spolehlivosti je také určen pásem schodovitých funkcí $F_N^{(1)}(x)$ a $F_N^{(2)}(x)$ takových, že $F_N^{(2)}(x) \leqq F_N(x) \leqq F_N^{(1)}(x)$.

Je tu ještě důležitá otázka, jak volit $\delta_1(x)$ a $\delta_2(x)$, na kterou dosud nebyla dána odpověď. Je však užitečné volit obě konstantní; jestliže budou obě rovna téže konstantě Δ, potom $F_N^{(1)}(x) = F_N(x) + \Delta$, $F_N^{(2)}(x) = F_N(x) - \Delta$

až na nutnou úpravu, aby hranice nešly nad jedničku nebo pod nulu. Máme-li pak tabulku hodnot Δ jakožto funkce α, můžeme snadno sestrojit obory spolehlivosti. Takovou tabulku skutečně podal Kolmogorov (12) a rozšířil ji Smirnov (29). Kolmogorov ukázal, že limita pravděpodobnosti

$$\lim_{N\to\infty} P\{\sup[F(x) - F_N(x)]\sqrt{N} \leqq \lambda\} = \Phi(\lambda),$$

kde funkce

$$\Phi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k \cdot e^{-2k^2\lambda^2}$$

je známá rychle konvergující řada Dirichletova.

Přibližné pásy spolehlivosti pro $F(x)$ k danému α budou tedy vymezeny funkcemi $F_N(x) + \frac{\lambda_\alpha}{\sqrt{N}}$ a $F_N(x) - \frac{\lambda_\alpha}{\sqrt{N}}$, při čemž λ_α je určeno tak, že $\Phi(\lambda) = \alpha$.

Také Wald a Wolfowitz (38) ukázali, jak stanovit exaktní pásy spolehlivosti při malém rozsahu výběru ($N = 6$), a udali methodu, jak najíti pravděpodobnost, že $F_N(x)$ bude v oboru přijatelnosti, které lze užíti na obecná $\delta_1(x)$ a $\delta_2(x)$ a výběry konečného rozsahu.

Postup, jehož bylo užito k odhadu $F(x)$, může sloužiti také k testování hypotéz o distribuční funkci základního souboru. Máme-li testovat na př. hypotézu $F_0(x)$, můžeme při velkém N užíti Kolmogorova výsledku a stanovíme, zda $\sup|F_0(x) - F_N(x)|\sqrt{N}$ přesahuje λ_α, které odpovídá zvolenému koeficientu spolehlivosti α.

Máme-li testovat, zda dva výběry rozsahu N_1 a N_2 jsou z téhož základního souboru, říkáme, že se jedná o problém dvou výběrů. Test můžeme založit na $\sup[F_{N_1}(x) - F_{N_2}(x)]$, kde $F_{N_1}(x)$ a $F_{N_2}(x)$ jsou distribuční funkce těch dvou výběrů. Jsou-li oba výběry velké, můžeme užíti výsledku Smirnovova, který ukázal, že limita pravděpodobnosti

$$\lim_{N\to\infty} P\{\sup[F_{N_1}(x) - F_{N_2}(x)]\sqrt{M} \leqq \lambda\} = \Phi(\lambda),$$

(kde $M = \frac{N_1 N_2}{N_1 + N_2}$ a $\Phi(\lambda)$ je funkce zavedená Kolmogorovem), když N_1 a N_2 se blíží nekonečnu tak, že jejich poměr zůstává mezi dvěma pevnými kladnými čísly.

Při řešení problému dvou výběrů se velmi dobře uplatnil rozvoj theorie rozdělení iterací. Máme-li posloupnost dvou druhů prvků S a M, v níž je n_1 písmen S a n_2 písmen M (na příklad $SMMSMMMSSMSSSM$), pak rozumíme iterací posloupnost prvků téhož druhu omezenou prvky jiného druhu (kromě prvního prvku, který není omezen zleva žádným prvkem, a posledního, který není omezen zprava). Tak je v našem případě jedna iterace S délky jedna, za ní je iterace M délky dvě, pak opět jedna iterace S délky jedna je následována iterací M délky tři atd. Rozdělení iterací dvou druhů prvků odvodil Mood (16) a většina problémů distribučních sem spadajících byla rozřešena. Theorie rozdělení iterací má také velký význam při řešení otázky, zda výběrové hodnoty $x_1, x_2, \ldots$ $\ldots, x_N$ v pořadí, jak byly získány, jsou nahodilé. V tomto případě uvažujeme třídu Ω_2 všech N-rozměrných spojitých kumulativních distribučních funkcí $F(x_1, x_2, \ldots, x_N)$ a nulová hypotéza tvrdí, že

$$F(x_1, x_2, \ldots, x_N) = F(x_1)F(x_2) \ldots F(x_N).$$

V praxi je důležité, aby otázka nezávislosti byla zkoumána před aplikováním theorie náhodného výběru. Důležité jsou testy nezávislosti, spočívající na statistických číslech seřazených, jež dostaneme, jestliže pozorované hodnoty výběrové $x_1, x_2, \ldots, x_N$ seřadíme v rostoucím pořadí velikosti od nejmenší hodnoty do největší, které pak označíme $x_{(1)} < < x_{(2)} < \ldots < x_N$, a na iteracích.

Jsou to především testy užívající iterací nad a pod mediánem. Budiž $x_1, x_2, \ldots, x_{2n+1}$ nějaký výběr hodnot, které byly ze základního souboru vyňaty v uvedeném pořadí, takže to nejsou statistická čísla seřazená. V této posloupnosti označme třeba a každé číslo x menší než medián $x_{(n+1)}$ a b každé číslo x větší než medián. Vynecháme-li medián z uvedené posloupnosti, zbývá $2n$ hodnot x, které jsou nyní nahraženy nějakým seskupením n písmen a a n písmen b. Bude zde r_{1i} iterací a délky i a r_{2i} iterací b délky i. Vyjde-li se z předpokladu nezávislosti, redukuje se theorie pravděpodobnosti těch r_{1i} a r_{2i} na uvažování $(2n)!$ permutací n prvků a a n prvků b. Problém řešil Mood a jeho základní theorie rozdělení iterací užil Mosteller (17). Vzal v úvahu délku L nejdelší iterace prvků a jako kriteria pro testování nahodilosti. Významné jsou zde velké hodnoty L. Pro prvky b existuje podobné kriterium. Kromě toho uvažoval kriterium L' definované jako délka nejdelší iterace prvků a nebo b a tabeloval některé kritické hodnoty. Mezi testy tohoto typu lze zařaditi také Wald-Wolfowitzův U-test (37). Statistický koeficient U je celkový počet iterací v nějaké posloupnosti V o N prvcích, která se sestrojí tak, že se uspořádá všech $n_1 + n_2 = N$ prvků vzestupně podle velikosti. Posloupnost $V = v_1, v_2, \ldots, v_{n_1+n_2}$ se definuje nyní tak, že $v_i = 0$, patří-li prvek do množiny $x_1, x_2, \ldots, x_{n_1}$, a $v_i = 1$, patří-li do množiny $y_1, y_2, \ldots, y_{n_2}$. Při problému dvou výběrů mají za nulové hypotézy, t. j. hypotézy, že oba výběry jsou z téhož základního souboru, všechna V touž pravděpodobnost. Jsou-li dva základní soubory různé, pak některé malé intervaly mají větší pravděpodobnost při jednom rozdělení četností než při druhém, z čehož vyplývá zmenšení průměrného celkového počtu iterací. Jsou tedy malé hodnoty tohoto statistického koeficientu kritické. Rozdělení jeho je známé (37) a pro $N \to \infty$ při pevném $\frac{n_1}{n_2} = k$ je rozdělení U asymptoticky normální s průměrem $\frac{2n_1}{1+k}$ a rozptylem $\frac{4kn_1}{(1+k)^3}$.

Nejsou-li n_1 a n_2 velká, lze kritické body poměrně snadno stanoviti. Máme skutečně pro praxi vhodné tabulky, které podal SWED a EISENHART (30) a OLMSTEAD (20).

Mohou býti dána také jiná vhodná kriteria, ale ode všech takových testů se žádá, aby byly konsistentní. Tato vlastnost ovšem musí býti definována tak, aby jí bylo možno použíti také v problémech neparametrických. WALD a WOLFOWITZ definovali konsistenci takto:

Nějaký test je konsistentní, blíží-li se jednotce pravděpodobnost zamítnutí nulové hypotézy, když je nesprávná (čili doplněk pravděpodobnosti chyby typu II), roste-li počet výběrů do nekonečna. Konsistentní charakteristikou se, jak známo, nazývá nějaká funkce pozorování, která konverguje stochasticky k parametru základního souboru, když počet výběrů roste do nekonečna. Testuje-li se hypotéza o parametru základního souboru pomocí nějakého konsistentního odhadu, bude test konsistentním také podle definice Wald-Wolfowitzovy.

K posouzení statistického koeficientu U uvažme ještě, že se může vyskytnout nějaká mimořádně dlouhá iterace v jedné nebo v obou proměnných. Mohla by zajisté býti rovněž indikátorem pro to, že hypotéza by měla býti zamítnuta. Ale tato iterace může býti doprovázena velkým počtem iterací délky 1, takže pak hodnota U není kriticky nízká. Skutečně bylo také navrženo, aby se dostatečně dlouhá iterace považovala za indikaci pro zamítnutí nulové hypotézy. Mohlo by se však vyskytnout více iterací poměrně dlouhých, ale žádná by nebyla kriticky dlouhá. Za tohoto kriteria by nulová hypotéza nebyla zamítnuta, ale hodnota U by byla malá třeba kriticky.

Aby byla odstraněna libovůle ve volbě kriteria, pokusil se WOLFOWITZ rozšířit methodu poměru věrohodností na statistické koeficienty spočívající na pořadí (44) a naznačil, že nemusí zůstat omezena na pořadová statistická čísla, nýbrž že ji lze patrně rozšířit na statistické koeficienty obecného Fisherova znáhodňovacího typu. Je ovšem třeba odvoditi v souvislosti s touto Fisherovou methodou obecné a konstruktivní methody, jimiž bychom dostali kritické obory a maximální věrohodnost.

Při řešení problému nahodilosti bylo také užito iterací z binomického nebo multinomického základního souboru. Studoval je hlavně BORTKIEWICZ, MISES, WISHART a HIRSCHFELD, COCHRAN a MOOD. Iterace z binomického základního souboru se liší od iterací dvou druhů prvků v tom, že n_1 a n_2, jichž jsme nahoře užili, jsou náhodné proměnné. Násobíme-li tedy obecně distribuční formuli platnou pro nějaké pevné n_1 a n_2 pravděpodobností této specielní množiny n_1 a n_2, tedy výrazem $\binom{n_1+n_2}{n_1} \cdot p_1^{n_1} \cdot p_2^{n_2}$

a sečteme přes n_1 a n_2, dostaneme příslušnou distribuční formuli iterací z binomického základního souboru. Iterací tohoto druhu bylo na př. užito ke zkoumání t. zv. „setrvačnosti počasí", t. j. otázky, zda suché měsíce následují po suchých a vlhké po vlhkých (GOLD, COCHRAN). Musím se také zmíniti o „iteracích nahoru a dolů", které mají důležité praktické použití zejména v kontrole jakosti výroby. Máme-li výběrové hodnoty $x_1, x_2, \ldots, x_N$ v tom pořadí, jak jsme je dostali, stanovíme z této posloupnosti hodnot nyní posloupnost $N - 1$ znamének kladných a záporných, definovanou takto: Je-li $x_i < x_{i+1}$, napíšeme $+$; je-li $x_i > x_{i+1}$, napíšeme $-$; ($i = 1, 2, \ldots, N - 1$). Uvažujeme nyní iterace kladných a záporných znamének, které se nazývají iterace nahoru, resp. dolů. Problémy rozdělení četností pro tyto iterace jsou obtížnější než pro iterace dvou druhů a výsledky tu nejsou zdaleka úplné.

Byly navrženy určité testy zakládající se na počtu délek iterací nahoru a dolů v posloupnosti. Navrhl to již FISHER, pak KERMACK a MC KENDRICK (10), WALLIS a MOORE. Testy v praxi nejobvyklejší jsou založeny na srovnání iterací různých délek s jejich očekávaným počtem nebo na vyskytnutí se mimořádně dlouhých iterací. Kriterium L'', kterého se zde užívá, je délka nejdelší iterace kladných a záporných znamének. Velké hodnoty L'' jsou významné. Theorii pravděpodobnosti L'' vypracovali LEVENE a WOLFOWITZ (14). Jiná kriteria pro testování nezávislosti v pořadové posloupnosti navrhli YOUNG (47), WALD a WOLFOWITZ (39), ANDERSON (1), PITMAN (24), HOTELLING a PABST (8).

Vrátíme-li se ještě několika slovy k problému dvou výběrů, můžeme říci, že první neparametrické řešení podal K. PEARSON a použil statistického koeficientu χ^2 a jeho rozdělení. Řešení založené na methodě randomizační navrhl PITMAN, který použil numerické hodnoty rozdílu výběrových průměrů $T = |\bar{x}_1 - \bar{x}_2|$. Uvažoval rozdělení četností $\bar{x}_1 - \bar{x}_2$ pro všechna možná rozdělení $n_1 + n_2$ pozorování do dvou skupin, které mají n_1 resp. n_2 prvků. Velké hodnoty T jsou významné. Pak byly navrženy různé testy užívající pořadí pro problém dvou výběrů. O dvou bylo dokázáno, že nejsou konsistentní; a to o testu Thompsonově podali tento důkaz WALD a WOLFOWITZ (37) a o MATHIESENOVĚ (15) BOWKER (3).

Praktické problémy nevyžadují vždycky, aby se předpokládalo jen, že kumulativní distribuční funkce je spojitá. Jsou případy, kdy o ní můžeme říci více; na příklad že je symetrická nebo jednovrcholová a podobně. Bylo by proto třeba vypracovati theorii pro takové případy. Tato theorie by jistě měla značný praktický význam.

Obecnou formulaci problému statistické indukce, která by byla dostatečně široká, aby obsáhla případ parametrický i neparametrický a zahrnovala odhad i testy, podal WALD v roce 1939 (34). Postupoval tím směrem, že položil základy obecné theorie statistických funkcí rozhodovacích, a to nejprve pro t. zv. případ klasický — nesekvenční — kde počet pozorování, na jejichž základě se má rozhodnutí udělat, je předem určen, t. j. rozsah náhodného výběru je pevně dán. V další práci rozšířil tuto theorii na případ t. zv. sekvenční, kde počet pozorování požadovaných pro rozhodnutí není předem určen, ale závisí na výsledku pozorování. Obecný problém statistických rozhodování formuloval takto:

Budiž $X = \{X^i\}$ $(i = 1, 2, \ldots)$ nekonečná posloupnost náhodných proměnných. Každé speciální pozorování proměnné X je dáno nekonečnou posloupností $x = \{x^i\}$ $(i = 1, 2, \ldots)$ reálných čísel, kde x^i značí pozorovanou hodnotu proměnné X^i. Prostor W všech možných nekonečných posloupností x se nazývá výběrovým prostorem a prvky x z W jsou výběrové body. Předpokládejme, že rozdělení pravděpodobnosti X, t. j. funkce míry pravděpodobnosti μ v prostoru W není známa, ale je známo, že patří do dané třídy Ω. Dále je dán prostor D, jehož prvky d představují možná rozhodnutí, která lze učinit v uvažovaném problému. Obyčejně každý prvek d z prostoru D bude sdružen s určitou podmnožinou ω z Ω a rozhodnutí d lze interpretovati jako přijetí hypotézy, že pravá funkce míry pravděpodobnosti μ je prvkem z ω. Fundamentálním problémem statistickým je nyní dáti pravidlo, podle něhož se učiní nějaké rozhodnutí, t. j. pravidlo, podle něhož se vybere specielní prvek d z prostoru D na základě pozorovaného výběrového bodu x. Znamená to tedy, že je dána třída Ω funkcí míry pravděpodobnosti a prostor D všech možných rozhodnutí d. Problém je sestrojit nějakou funkci $d(x)$, nazvanou statistická rozhodovací funkce, která přiřazuje každému výběrovému bodu x nějaký prvek $d(x)$ z prostoru D, takže se učiní rozhodnutí $d(x)$, když byl pozorován výběrový bod x. Je tedy statistická rozhodovací funkce definována pro všechny body x výběrového prostoru W a pro každý výběrový bod x je hodnota té funkce prvkem prostoru D.

Řešení tohoto problému podal v roce 1947 Wald za jistých omezení. Ale poslední nová práce Waldova odstraňuje některé dřívější zbytečně omezující předpoklady. Tak v dosavadní theorii předpokládal, že prostor Ω přípustných distribučních funkcí F je kompaktní, což obyčejně není v parametrickém případě splněno. Tato podmínka nyní v rozpojitém případě úplně odpadá a ve spojitém jest nahražena podmínkou separability Ω. Úvahy Waldovy jsou zcela abstraktní, ale začíná se pracovat na aplikaci této theorie v praktických problémech sestrojením hospodářsky nejvýhodnější kontroly studiem konkretních risikových funkcí (Satterthwaite, Špaček).

Snad je z tohoto krátkého nástinu patrno, že vyhlídky theorie, která nečiní předpokladů o tvaru rozdělení četností základního souboru budí zájem mezi theoretiky i praktiky. Když sledujeme vývoj pokusů o obecné theorie nejlepších testů a odhadů, vidíme, že byly inspirovány značně abstraktní práce matematické, do nichž vstupují problémy z theorie míry, úvahy ve funkcionálech, funkčních prostorech a v metrice prostorů. Svědčí o tom na příklad práce o theorii některých neparametrických hypotéz, kterou uveřejnili letos E. L. Lehmann a C. Stein (13). Odvodili optimální testy pro dva typy neparametrických hypotéz vůči určitým třídám alternativ. Tyto hypotézy autoři ilustrují tímto příkladem:

1. Simultánní rozdělení proměnných $x_1, x_2, \ldots, x_{n_1}$, $y_1, y_2, \ldots, y_{n_2}$ je invariantní za všech permutací proměnných.

2. Proměnné jsou rozděleny nezávisle a identicky. Ukázali, že theorie optimálních testů pro hypotézy prvního typu je táž jako theorie optimálních podobných testů pro hypotézy druhého typu. Definice, kterých užívají k řešení řady dalších ještě problémů jsou dosti obecné, takže lze theorie užíti na problémy spojité i rozpojité a pozorování, která nejsou nezávislá, tu nečiní potíží.

Mohl jsem se jen dotknouti hlavních míst na cestě prudkého vývoje theorie neparametrických testů v posledních letech. Podrobnosti a dalš odkazy lze najíti v citované literatuře. Chci tím upozorniti na vědecké pole, které v mnoha místech čeká na zpracování a vzbuditi zájem o hlubší studium prací, které tu znamenají rozvoj, jaký nemá časté obdoby v dějinách věd.

LITERATURA*)

1. R. L. ANDERSON: Distribution of the serail correlation coefficient, 1942.
2. G. BATEMAN: On the power function of the longest run as a test for randomness in a sequence of alternatives, Biometrika XXXV: I., II., 1948.
3. A. BOWKER: Note on consistency of a proposed test for the problem of two samples, 1944.
4. H. CRAMÉR: Mathematical methods of statistics, Princeton University Press, 1946.
5. R. A. FISHER: Statistical methods for research workers, Oliver & Boys, Edinburg and London, 1925.
6. M. FRIEDMANN: The use of ranks to avoid the assumption of normality implicit in the analysis of variance, Journal of the Amer. Statistical Assoc., 1937.
7. N. HOEFFDING: A class of statistics with assymptotically normal distribution, 1948.
8. H. HOTELLING — M. R. PABST: Rank correlation and tests of significance involving no assumptions of normality, 1936.
9. M. KENDALL: The advanced theory of statistics I., II., London 1945, 1948.
10. W. O. KERMACK — A. G. MC KENDRICK: Tests for randomness in a series of observations, Proc. Roy. Soc. Edinburg, 1937.
11. A. KOLMOGOROV: Confidence limits for an unknown distribution function, 1941.
12. A. KOLMOGOROV: Sulla determinatione empirica di una legge di distributione, Giornale Ist. Ital. Attuari, 1933.
13. E. L. LEHMANN — C. STEIN: On the theory of some nonparametric hypotheses, 1949.
14. H. LEVENE — J. WOLFOWITZ: The covariance matrix of runs up and down, 1944.
15. H. MATHIESEN: A method of testing the hypothesis that two samples are from the same population, 1943.
16. A. M. MOOD: The distribution theory of runs, 1940.
17. F. MOSTELLER: Note on an application of runs to quality control charts, 1941.
18. R. B. MURPHY: Nonparametric tolerance limits, 1948.

*) Práce, u nichž je uveden pouze letopočet, jsou uveřejněny v dotyčném ročníku časopisu „The Annals of Mathematical Statistics“.

19. J. NYEMAN: Basic ideas and some recent results of the theory of testing statistical hypotheses, Journal of Royal Stat. Soc., 1942.
20. P. S. OLMSTEAD: Distribution of sample arrangements for runs up and down, 1946.
21. E. PAULUS: A note on tolerance limits, 1943.
22. E. S. PEARSON: Some aspect of the problem of randomisation, Biometrika, 1937—1938.
23. K. PEARSON: On the probability that two independent distributions of frequency are really samples from the same population, Biometrika, 1911.
24. E. J. G. PITTMANN: Significance tests which may be applied to samples from any populations, Suppl. of J. Roy. Stat. Soc., 1937, a Biometrika, 1938.
25. A. SCHEFFÉ: Statistical inference in the non-parametric case, 1943.
26. A. SCHEFFÉ: On a measure problem arising in the theory of non-parametrics tests, 1943.
27. A. SCHEFFÉ: Non-parametric estimation I. Validation of order statistics, 1945.
28. W. A. SHEWHART: Statistical method from the viewpoint of quality control, 1939.
29. N. SMIRNOV: Ocenka raschozděnija meždu empiričeskimi krivymi raspredělenija v dvuch nezavisimych vyborkach, Bjulleten Moskovskovo Gosudarstvennovo universiteta — Matematika, II, **2**, 1939.
30. F. S. SWED — C. EISENHART: Tables for testing randomness of grouping in a sequence of alternatives, 1943.
31. W. R. THOMPSON: On confidence ranges for the median and other expectation distributions for populations of unknown distribution form, 1936.
32. W. R. THOMPSON: Biological applications of normal range and associated significance test in ignorance of original distribution form, 1938.
33. J. TUKEY: Non-parametric estimation I., II., III., 1947, 1948.
34. A. WALD: Contributions to the theory of statistical estimation and testing hypotheses, 1939.
35. A. WALD: Foundations of a general theory of sequential decision functions, Econometrica, 1947.
36. A. WALD: Statistical decision functions, 1949.
37. A. WALD — J. WOLFOWITZ: On a test whether two samples are from the same population, 1940.
38. A. WALD — J. WOLFOWITZ: Confidence limits for continuous distribution functions, 1939.
39. A. WALD — J. WOLFOWITZ: An exact test for randomness in the nonparametric case, based on serial correlation, 1943.
40. A. WALD — J. WOLFOWITZ: Statistical tests based on permutations of the observations, 1944.
41. S. WILKS: On the determination of sample sizes for setting tolerance limits, 1941.
42. S. WILKS: Statistical prediction with special reference to the problem of tolerance limit, 1942.
43. S. WILKS: Order statistics, Bulletin of the Am. Mathem. Soc., 1948.
44. J. WOLFOWITZ: Additive partition functions and a class of statistical hypotheses, 1942.
45. J. WOLFOWITZ: Asymptotic distribution of runs up and down, 1944.
46. J. WOLFOWITZ: Non-parametric statistical inference, Proceedings of the Berkeley symposium on mathematical statistics and probability, 1949.
47. L. C. YOUNG: On randomness in ordered sequences, 1941.

*

Summary. — Výtah.

Advances in the theory of non-parametric tests in statistical inference.

JAROSLAV JANKO, Praha.

One of the fundamental problems in the non-parametric statistical inference is the problem of estimation of cumulative distribution function $F(x)$ about which we only know that it is continous. It is solved by means of a sample cumulative distribution function $F_N(x)$ so that there are constructed the acceptance regions in which $F(x)$ lies with the certain probability α. In fact it is possible to construct practically the respective confidence region, for it is also determined, as has been shown by WOLFOWITZ through the belt of such two step-functions $F_N^{(1)}(x)$ and $F_N^{(2)}(x)$, that $F_N^{(2)}(x) \leqq F_N(x) \leqq F_N^{(1)}(x)$. It is useful to choose a belt of a constant width and equal on both sides, so that then $F_N^{(1)}(x) = F_N(x) + \Delta$ and $F_N^{(2)}(x) = F_N(x) - \Delta$ with the exception of a necessary correction consisting in it that the boundaries do not exced the figure one and do not fall below nought. If we then have a table of values Δ as function of α, we can easily construct the confidence regions. Such a table has been constructed by KOLMOGOROV and was extended by SMIRNOV. How to determine the exact confidence belts with a small sample size has been shown by WALD and WOLFOWITZ who determined a method about finding a probability that $F_N(x)$ would be within an acceptance region which is usable for generally given widths of the belt and for finite sample sizes.

The method which has been used in estimating $F(x)$, can also be used in testing the hypotheses of distribution function of the population. If we test whether the two samples with the sizes n_1 and n_2 have come from the same population, we say that it is a problem of two samples. In its solution the development of the distribution theory of runs (Mood) was made valid, with success. This theory is of great improtance with the solution of the question whether the sample values $x_1, x_2, \ldots, x_N$ in order sequence as drawn, are random. Different criteria for testing randomness have been suggested by YOUNG, ANDERSON, HOTELLING and PABST, PITMAN. About the two tests i. e. Thompson's and Mathiesen's it has been proved that they are not consistent. WALD and WOLFOWITZ have defined the consistency so as to be able to be also used in the non-parametric problems. According to them a test is consistent, if the probability of rejecting the null hypothesis, when it is false, approaches one while the sample number increases indefinitely.

Fourthermore we should quote a Wald-Wolfowitz's U-test which was used with the solution of the „two samples problem". Its frequency distribution has been determined. If the sample sizes n_1 resp. n_2 are not big there is no difficulty in obtaining a table of critical points (SWED and

EISENHART) If the total of the sizes of these two samples $N = n_1 + n_2$ increases indefinitely with constant $\frac{n_1}{n_2} = k$ the U-distribution approches the normal distribution asymptotically. It is necessary to attach great importance to the WALD'S general formulation of the problem of statistical inference which starts with a general theory of statistical decision functions.

QUELQUES PROBLÈMES ACTUELS CONCERNANT LES FONDEMENTS DES MATHÉMATIQUES.

STANISŁAW JAŚKOWSKI, Toruń.

Dans le vaste domaine appelé fondements des mathématiques ou logique des mathématiques, j'ai choisi deux groupes de problèmes qui paraissent avoir une importance particulière à cause de leurs conséquences pratiques.

I. Problèmes de décision.

1. Dans la logique des mathématiques, on appelle „méthode de décision" (decision method, Entscheidungsverfahren) une méthode générale de résoudre tous les problèmes d'un type donné. La notion de méthode générale est connue aux mathématiciens depuis l'antiquité; l'algorithme d'Euclide en fournit un exemple. La notation logique symbolique, la forme de système déductif formalisé, donnée aux théories mathématiques et enfin la distinction entre le système et les considérations méthodologiques appartenant au „méta-système" — voici les facteurs qui d'une part ont contribué au développement de quelques méthodes de décision et qui, de l'autre, ont permis de démontrer l'impossibilité d'en construire quelques autres.

Toutes les méthodes de décision obtenus par les logiciens sont du même type. On construit une théorie formalisée T et on définit une classe de formules (classe de problèmes) P par l'énumération des termes constants, des variables et des quantificateurs, qui tous peuvent apparaître dans les formules de la classe P. On décrit d'une manière univoque le procédé qui donne la réponse, pour chaque formule F de la classe P, si F est un théorème de la théorie T ou non. Le résultat négatif classique est dû à GÖDEL qui a montré en 1930 qu'il n'était pas possible de trouver une méthode de décision applicable à la classe des problèmes formulés à l'aide de la notion de nombre naturel et des opérations d'addition et de multiplication.

La méthode bien connue de vérifier les formules du calcul des propositions à l'aide de substitution des valeurs 1 et 0 pour les variables est le prototype des méthodes de décision modernes. Dans les dernières