Silviu Guiaşu; Abe Shenitzer
Princip maxima entropie

Pokroky matematiky, fyziky a astronomie, Vol. 31 (1986), No. 4, 214--216,217--224

Persistent URL: http://dml.cz/dmlcz/138890

Princip maxima entropie

Silviu Guiasu a Abe Shenitzer

Matematické modelování a variační principy

Jistě není třeba zdůrazňovat důležitost variačních problémů v matematice a jejích aplikacích. Výčet variačních úloh různých stupňů obtížnosti je velmi dlouhý: od slavných starověkých úloh na hledání minima a maxima přes variační úlohy analytické mechaniky a teoretické fyziky až k variačním problémům moderního operačního výzkumu. Zatímco maximalizace nebo minimalizace funkce či funkcionálu je rutinní záležitostí, některé speciální variační problémy dávají řešení, jež buď sjednocují původně nijak spolu nesouvisející výsledky, anebo překvapivě dobře vystihují výsledky našich experimentů. Takové variační problémy se nazývají variační principy. Zda struktura našeho světa je v souladu s variačními principy či nikoliv, je filozofický problém. Avšak odhalovat a aplikovat variační principy lze považovat za spolehlivou strategii k lepšímu chápání části této struktury. V aplikované matematice vytváříme model reality vyzdvižením jistých souvislostí a (nevyhnutelně) zanedbáním jiných. Jednou z cest k vytvoření přesvědčivého a užitečného modelu je získat jej jako řešení nějakého variačního problému.

Záměrem tohoto článku je dodat jisté argumenty ve prospěch myšlenky, že by variační problém maximalizace entropie měl být povýšen na variační princip.

Entropie jako míra neurčitosti

Někdy se variační princip realizuje v maximalizaci nebo minimalizaci funkce či funkcionálu, jejichž význam není bezprostředně zřejmý. V takových případech je použití variačního principu oprávněno vlastnostmi jeho řešení. Důležitým příkladem je princip nejmenší akce v analytické mechanice. Zde tzv. „akce“ nemá žádnou okamžitou a přirozenou fyzikální interpretaci, avšak řešení (Hamiltonovy kanonické rovnice) dává přesně zákon pohybu. V případě principu maxima entropie však sama funkce, která je maximalizována, totiž entropie, má pozoruhodné vlastnosti opravňující ji k tomu, aby byla považována za dobrou míru množství neurčitosti obsažené v pravděpodobnostním rozdělení.

Nechť $\bar{p} = (p_1, \ldots, p_m)$ je konečné pravděpodobnostní rozdělení, tj. m-tice reálných čísel splňujících vztahy

$$(1) \qquad p_k \geqq 0\,, \quad (k = 1, \ldots, m)\,; \quad \sum_{k=1}^{m} p_k = 1\,.$$

SILVIU GUIASU, ABE SHENITZER: *The Principle of Maximum Entropy.* The Mathematical Intelligencer, Vol. 7, No. 1, pp. 42—48.

Číslo p_k může představovat pravděpodobnost k-tého výsledku pravděpodobnostního pokusu nebo pravděpodobnost nabývání k-té možné hodnoty pro konečnou diskrétní náhodnou veličinu.

Entropie přiřazená pravděpodobnostnímu rozdělení (1) je číslo

$$H_m(\bar{p}) = H_m(p_1, \ldots, p_m) = -\sum_{k=1}^{m} p_k \ln p_k \,, \tag{2}$$

kde klademe $0 \,.\, \ln 0 = 0$ kvůli zajištění spojitosti funkce $-x \ln x$ v počátku. Pro každé kladné celé číslo $m \geqq 2$ je H_m funkce definovaná na množině pravděpodobnostních rozdělení vyhovujících podmínce (1).

Entropie má několik vlastností se zajímavými interpretacemi. Zmíníme se o některých z nich.

1. $H_m(\bar{p}) \geqq 0$, je spojitá a invariantní vzhledem ke každé permutaci indexů.

2. Má-li $\bar{p}$ pouze jedinou složku různou od nuly (tj. rovnou jedné), je $H_m(\bar{p}) = 0$.

3. $H_m(p_1, \ldots, p_m) = H_{m+1}(p_1, \ldots, p_m, 0)$.

4. $H_m(p_1, \ldots, p_m) \leqq H_m(1/m, \ldots, 1/m)$; rovnost nastává, právě když $p_k = 1/m$ ($k = 1, \ldots, m$).

5. Je-li $\bar{\pi} = (\pi_{1,1}, \ldots, \pi_{m,n})$ sdružené pravděpodobnostní rozdělení, jehož marginální pravděpodobnostní rozdělení jsou $\bar{p} = (p_1, \ldots, p_m)$, resp. $\bar{q} = (q_1, \ldots, q_n)$, pak

$$H_{mn}(\pi_{1,1}, \ldots, \pi_{m,n}) = H_m(p_1, \ldots, p_m) + \sum_{k=1}^{m} p_k H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k) \,, \tag{3}$$

kde podmíněná entropie $H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k)$ se počítá pouze pro takové hodnoty k, pro něž $p_k \neq 0$.

6. Při výše uvedeném značení platí

$$\sum_{k=1}^{m} p_k H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k) \leqq H_n(q_1, \ldots, q_n) \tag{4}$$

a rovnost nastane, když a jen když

$$\pi_{k,l} = p_k q_l \quad (k = 1, \ldots, m \,;\; l = 1, \ldots, n)\,;$$

v tomto případě (3) nabývá tvaru

$$H_{mn}(\bar{\pi}) = H_m(\bar{p}) + H_n(\bar{q}) \,.$$

Všechny tyto vlastnosti mohou být ověřeny elementárním způsobem. Aniž bychom zacházeli do technických detailů, povšimněme si toho, že vlastnosti 1–3 jsou zřejmé, zatímco vlastnost 5 se dá ověřit přímým výpočtem z definice entropie. Konečně, z Jensenovy nerovnosti

$$\sum_{k=1}^{m} a_k f(b_k) \leqq f\Big(\sum_{k=1}^{m} a_k b_k\Big) \,,$$

použité na konkávní funkci $f(x) = -x \ln x$, dostaneme vlastnost 4 volbou $a_k = 1/m$, $b_k = p_k$ $(k = 1, \ldots, m)$ a nerovnost (4) volbou $a_k = p_k$, $b_k = \pi_{k,l}/p_k$ $(k = 1, \ldots, m)$ pro každé $l = 1, \ldots, n$ a nakonec sečtením výsledných n nerovností.

Interpretace výše uvedených vlastností je ve shodě se zdravým rozumem, intuicí a odůvodnitelnými požadavky, které by měla míra neurčitosti splňovat. Vskutku, pravděpodobnostní experiment s pouze jedním možným výsledkem (tzn. striktně deterministický experiment) neobsahuje vůbec žádnou neurčitost; ještě před uskutečněním pokusu víme, co nastane. To je přesně vlastnost 2. Připojíme-li k možným výsledkům pravděpodobnostního experimentu jiný výsledek mající nulovou pravděpodobnost, zůstává množství neurčitosti ohledně průběhu experimentu nezměněné (vlastnost 3). Vlastnost 4 nám říká, že ve třídě všech pravděpodobnostních pokusů majících m možných výsledků je maximální neurčitost obsažena ve speciálním pravděpodobnostním pokusu, jehož výsledky jsou stejně pravděpodobné. Dříve než budeme interpretovat poslední dvě vlastnosti, uvažujme dvě diskrétní náhodné veličiny X a Y, jejichž obory hodnot obsahují m, resp. n hodnot. Při užití téže notace jako u vlastnosti 5, nechť $\bar{\pi}$ je sdružená pravděpodobnostní distribuce dvojice (X, Y) a nechť $\bar{p}$ a $\bar{q}$ jsou marginální pravděpodobnostní rozdělení veličiny X, resp. Y. V tomto případě lze rovnost (3) psát přehledněji jako

(5)
$$H(X, Y) = H(X) + H(Y \mid X),$$

kde

$$H(X, Y) = H_{mn}(\pi_{1,1}, \ldots, \pi_{m,n}),$$
$$H(X) = H_m(p_1, \ldots, p_m)$$

a kde

$$H(Y \mid X) = \sum_{k=1}^{m} p_k H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k)$$

značí podmíněnou entropii Y za podmínky X. Vzhledem k (5) se množství neurčitosti obsažené ve dvojici náhodných veličin (neboli ve složeném či součinovém pravděpodobnostním experimentu) obdrží sečtením množství neurčitosti obsažené v jedné složce (řekněme v X) a neurčitosti obsažené ve druhé složce (Y) podmíněné první z nich (X). Podobně dostáváme pro $H(X, Y)$ rozklad

(6)
$$H(X, Y) = H(Y) + H(X \mid Y),$$

kde

$$H(Y) = H_n(q_1, \ldots, q_n)$$

a

$$H(X \mid Y) = \sum_{l=1}^{n} q_l H_m(\pi_{1,l}/q_l, \ldots, \pi_{m,l}/q_l).$$

Zde

$$H_m(\pi_{1,l}/q_l, \ldots, \pi_{m,l}/q_l)$$

je podmíněná entropie X vzhledem k l-té hodnotě Y_l. H_m je definována pouze pro takové hodnoty l, pro něž $q_l > 0$. Z (5) a (6) dostáváme

$$H(X) - H(X \mid Y) = H(Y) - H(Y \mid X),$$

což je tak zvaná „rovnováha neurčitosti", jediný zákon zachování platný pro entropii.

Konečně vlastnost 6 ukazuje, že nějaké údaje o X mohou pouze zmenšit neurčitost Y, totiž

$$H(Y \mid X) \leqq H(Y) \tag{7}$$

s rovností právě když X a Y jsou nezávislé. Ze vztahů (5) a (7) dostáváme

$$H(X, Y) \leqq H(X) + H(Y)$$

s rovností právě když X a Y jsou nezávislé. Tato nerovnost platí pro libovolný počet složek. Obecněji pro s náhodných veličin s libovolným konečným oborem hodnot můžeme psát

$$H(X_1, \ldots, X_s) \leqq H(X_1) + \ldots + H(X_s)$$

s rovností právě když $X_1, \ldots, X_s$ jsou nezávislé. Tedy výraz

$$W(X_1, \ldots, X_s) = \sum_{i=1}^{s} H(X_i) - H(X_1, \ldots, X_s) \geqq 0$$

měří *globální závislost* náhodných veličin $X_1, \ldots, X_s$, tj. stupeň, v jakém je systém $(X_1, \ldots, X_s)$ „něčím více" než pouhým výčtem svých složek. Speciálně platí, že $W = 0$ právě tehdy, když $X_1, \ldots, X_s$ jsou nezávislé.

Všimněme si, že rozdíl mezi množstvím neurčitosti obsažené ve dvojici (X, Y) a množstvím závislosti mezi složkami X a Y, totiž

$$d(X, Y) = H(X, Y) - W(X, Y)$$

neboli ekvivalentně

$$d(X, Y) = 2H(X, Y) - H(X) - H(Y) = H(X \mid Y) + H(Y \mid X),$$

lze považovat za *vzdálenost* náhodných veličin X a Y, přičemž dvě náhodné veličiny chápeme jako identické, když každá z nich plně předpovídá druhou, tj. když $H(X \mid Y) = 0$ a $H(Y \mid X) = 0$. Tedy „čistá náhodnost" obsažená ve dvojici (X, Y), tj. neurčitost celku minus závislost mezi složkami, je vzdálenost. To dává geometrickou představu o chaosu!

Diskrétní entropie jako míra neurčitosti byla zavedena C. E. Shannonem [12] jako analogie k Boltzmannově H-funkci [1] ve statistické mechanice. Shannon ji také používal jako míru informace, kde informace znamená odstraněnou neurčitost. *Před* provedením pravděpodobnostního experimentu měří entropie množství neurčitosti spojené s možnými výsledky. *Po* experimentu entropie měří množství dodané informace. Zdůrazňujeme, že zde bylo poprvé využito matematické funkce k měření neurčitosti obsažené v pravděpodobnostním experimentu — entity tolik odlišné od měřitelných charakteristik reálného světa, jakými jsou délka, obsah, objem, teplota, tlak, hmotnost, náboj atd.

Je Shannonova entropie jednoznačně určena? Odpověď závisí na zvolených axiómech pro míru neurčitosti. Chinčin [9] ukázal, že vlastnosti 1, 3, 4 a 5 brané jako axiómy (což je z intuitivního hlediska docela přijatelné) implikují vyjádření míry neurčitosti ve tvaru (2), a to jednoznačně až na násobení libovolnou kladnou konstantou. To nám dovoluje zvolit za základ logaritmu libovolné číslo větší než 1, aniž bychom ovlivnili základní vlastnosti míry.

Princip maxima entropie

Vraťme se k vlastnosti 4. Neurčitost je maximální, jsou-li výsledky stejně pravděpodobné. Rovnoměrné rozdělení maximalizuje entropii; rovnoměrné rozdělení obsahuje největší množství neurčitosti. Avšak to je přesně v souladu s Laplaceovým principem scházejícího důvodu, podle něhož není-li důvod preferovat jednu ze dvou či několika možností, pak nejlepší strategií je považovat je za stejně pravděpodobné. V Laplaceově případě samozřejmě šlo o subjektivní hledisko, založené na opatrnosti a zdravém rozumu. Skutečně, i když nic nevíme o pojmu entropie, používáme Laplaceův princip scházejícího důvodu v každodenním životě, dokonce i při analýze nejjednodušších experimentů. Například při házení mincí obyčejně přikládáme stejnou pravděpodobnost oběma možným výsledkům, a to nikoliv až po dlouhé sérii opakování tohoto jednoduchého pokusu a bedlivém zkoumání stability relativních četností možných výsledků, nýbrž jednoduše proto, že používáme Laplaceův princip a jsme si vědomi toho, že není důvod dávat některému výsledku přednost. Jak jsme však již viděli, přijmeme-li Shannonovu entropii za míru neurčitosti, je vlastnost 4 vlastně matematickým zdůvodněním principu maxima entropie, který říká, že není-li na pravděpodobnostní rozdělení kladeno žádné omezení, je entropie maximalizována rozdělením rovnoměrným. V takovém případě nám naše intuice založená na zkušenosti dává správné řešení. Co se však stane, budou-li na pravděpodobnostní rozdělení kladena nějaká omezení?

Před zodpovězením této otázky si nejprve všimněme, jakého druhu tato omezení mohou být. V aplikacích máme velmi často k dispozici jednu či několik středních hodnot jedné či několika náhodných veličin. Protože stavový prostor ve statistické mechanice je pravděpodobnostní, jsou stavové funkce náhodnými veličinami a je možno měřit pouze některé jejich střední hodnoty. Například každému mikrostavu odpovídá dobře definovaná hodnota energie systému. Avšak určit s jistotou v daném okamžiku t skutečný jediný mikroskopický stav systému nemůžeme, proto namísto toho konstruujeme pravděpodobnostní rozdělení možných stavů systému. Potom se energie stává náhodnou veličinou a co měřit skutečně můžeme (ovšem na makroskopické úrovni), je střední hodnota této náhodné veličiny, tj. makroskopická energie. Makroskopická úroveň je úroveň středních hodnot, z nichž některé mohou být změřeny. Avšak my potřebujeme pravděpodobnostní model mikroskopické úrovně, tj. pravděpodobnostní rozdělení možných mikrostavů systému. Obecně existuje mnoho pravděpodobnostních rozdělení (dokonce nekonečně mnoho!), která jsou v souladu s naměřenými středními hodnotami. Odtud vyvstává otázka, jaké pravděpodobnostní rozdělení je „nejlepší" a podle jakého kritéria.

V roce 1957 podal E. T. Jaynes [8] velmi přirozené kritérium výběru zavedením *principu maxima entropie*: z množiny všech pravděpodobnostních rozdělení, jež jsou v souladu s jednou či několika středními hodnotami jedné či několika náhodných veličin, se vybírá takové, které maximalizuje Shannonovu entropii. Takové pravděpodobnostní rozdělení je „nejširší“; předem nevyloučí žádnou možnost a je při daném omezení nejrovnoměrnější. Princip maxima entropie, zavedený k řešení problému statistické mechaniky, se stal široce používaným nástrojem pro sestrojování pravděpodobnostních rozdělení ve statistice, v teorii rozhodování, při rozpoznávání obrazců, v teorii informace a v analýze časových řad, protože ve všech těchto oblastech jsou naše znalosti všeobecně vyjádřeny středními hodnotami jistých náhodných veličin a hledá se pravděpodobnostní rozdělení, které nevylučuje žádnou možnost podřízenou příslušným omezením.

Abychom poznali, jak se tento princip projevuje, uvažujme nejjednodušší možný případ, totiž takový, kdy známe střední hodnotu $E(f)$ náhodné veličiny f, jejíž možné hodnoty jsou $f_1, \ldots, f_m$. Hledáme pravděpodobnostní rozdělení $\bar{p} = (p_1, \ldots, p_m)$,

$$p_k > 0\,, \quad (k = 1, \ldots, m)\,; \quad \sum_{k=1}^{m} p_k = 1 \tag{8}$$

vyhovující omezení

$$E(f) = \sum_{k=1}^{m} f_k p_k\,. \tag{9}$$

V triviálním případě $m = 2$ střední hodnota $E(f)$ jednoznačně definuje odpovídající pravděpodobnostní rozdělení na základě lineární rovnice

$$E(f) = f_1 p_1 + f_2(1 - p_1)\,.$$

Avšak pro každé $m \geqq 3$ existuje nekonečně mnoho pravděpodobnostních rozdělení tvaru (8) splňujících (9). Použitím principu maxima entropie vybíráme pravděpodobnostní rozdělení s největší neurčitostí, tj. pravděpodobnostní rozdělení maximalizující entropii

$$H_m(\bar{p}) = -\sum_{k=1}^{m} p_k \ln p_k$$

při omezeních (8) a (9). Samozřejmě H_m je konkávní a spojitá funkce definovaná v konvexní oblasti charakterizované pomocí (8) a (9). Existuje tedy pouze jeden bod globálního maxima patřící do otevřené množiny

$$\{\bar{p} = (p_1, \ldots, p_m) \mid p_k > 0\,, \quad k = 1, \ldots, m\,; \ \sum_{k=1}^{m} p_k - 1 = 0\,,$$
$$\sum_{k=1}^{m} f_k p_k - E(f) = 0\}\,.$$

Použitím Lagrangeovy funkce

$$L = H_m(p_1, \ldots, p_m) - \alpha\Big(\sum_{k=1}^{m} p_k - 1\Big) - \beta\Big(\sum_{k=1}^{m} f_k p_k - E(f)\Big)\,,$$

kde α a β jsou Lagrangeovy multiplikátory příslušející našim dvěma podmínkám, a položením parciálních derivací prvního řádu rovným nule dostáváme

$$\frac{\partial L}{\partial p_k} = -\ln p_k - 1 - \alpha - \beta f_k = 0\,, \quad (k = 1, \ldots, m)\,,$$

$$\frac{\partial L}{\partial \alpha} = 1 - \sum_{k=1}^{m} p_k = 0\,,$$

$$\frac{\partial L}{\partial \beta} = E(f) - \sum_{k=1}^{m} f_k p_k = 0\,.$$

Odtud řešením je

$$p_k = \frac{e^{-\beta_0 f_k}}{\sum_{r=1}^{m} e^{-\beta_0 f_r}}\,, \qquad (k = 1, \ldots, m)\,, \tag{10}$$

kde β_0 je řešení exponenciální rovnice

$$\sum_{k=1}^{m} [f_k - E(f)]\, e^{-\beta(f_k - E(f))} = 0\,. \tag{11}$$

Je-li náhodná veličina f nedegenerovaná (tj. nabývá-li f alespoň dvou různých hodnot), pak takové řešení existuje a je jediné, poněvadž funkce

$$G(\beta) = \sum_{k=1}^{m} [f_k - E(f)]\, e^{-\beta(f_k - E(f))} \tag{12}$$

je striktně klesající s

$$\lim_{\beta \to -\infty} G(\beta) = +\infty\,, \quad \lim_{\beta \to +\infty} G(\beta) = -\infty\,.$$

Viděli jsme již, že neexistuje-li žádné omezení, je řešením principu maxima entropie rovnoměrné pravděpodobnostní rozdělení. Je-li dána střední hodnota $E(f)$ náhodné veličiny f, pak řešením principu maxima entropie je (10), nebo ekvivalentně

$$p_k = \frac{1}{\Phi(\beta)}\, e^{-\beta_0 f_k} \quad (k = 1, \ldots, m)\,,$$

kde

$$\Phi(\beta) = \sum_{k=1}^{m} e^{-\beta f_k}$$

a β_0 je jediné řešení rovnice

$$\frac{d \ln \Phi(\beta)}{d\beta} = -E(f)\,.$$

To je přesně Gibbsovo neboli kanonické rozdělení, vyskytující se skoro ve všech knihách o statistické mechanice a mnohem nověji v některých knihách teorie rozhodování. Vidíme nyní, proč je kanonické rozdělení v aplikacích užitečné: má největší neurčitost, je nejrovnoměrnější, nevylučuje předem žádnou možnost s přihlédnutím k omezení danému střední hodnotou $E(f)$.

Poněvadž (11) je exponenciální rovnice, může být její řešení β_0 transcendentní číslo. Ovšem skutečnost, že funkce G daná vztahem (12) je striktně klesající, nám umožňuje aproximovat řešení β_0 s velkou přesností.

Například, nechť $m = 3, f_1 = 12, f_2 = 15, f_3 = 20$ a střední hodnota $E(f) = 18.12$. Použitím jednoduchého kapesního kalkulátoru TI-57 můžeme získat řešení rovnice (11) (správné na šest desetinných míst) během několika minut, a to $\beta_0 \approx -0.2364201$. Příslušné řešení principu maxima entropie je $p_1 = 0.1035103$, $p_2 = 0.2103835$, $p_3 = 0.6861062$. Toto je pravděpodobnostní rozdělení o největší neurčitosti, které je ve shodě s danou střední hodnotou. Pro některá jiná omezení mohou být hodnoty Lagrangeových multiplikátorů zavedených k maximalizaci entropie určeny exaktně. Aniž bychom zacházeli do technických detailů, zmíníme se dále o některých pozoruhodných výsledcích týkajících se principu maxima entropie.

a) Je-li f náhodná veličina se spočetným oborem hodnot

$$\{ku \mid u > 0\,,\; k = 0, 1, 2, \ldots\}$$

(což je splněno pro energii v kvantové mechanice — v tomto případě je u kvantum energie — nebo pro mnoho diskrétních funkcí používaných v operačním výzkumu — v tomto případě je u jednotka) a je-li dána střední hodnota $E(f)$, pak pravděpodobnostní rozdělení

$$p_k > 0\,, \quad (k = 0, 1, \ldots)\,, \quad \sum_{k=0}^{\infty} p_k = 1$$

maximalizující spočetnou entropii

$$H = -\sum_{k=0}^{\infty} p_k \ln p_k$$

je tvaru

$$p_k = \frac{u(E(f))^k}{(u + E(f))^{k+1}}\,, \quad k = 0, 1, 2, \ldots$$

Je vidět, že řešení principu maxima entropie je plně určeno veličinou u a střední hodnotou $E(f)$. Důležitost tohoto pravděpodobnostního rozdělení zdůraznil M. Born [2].

Dříve než budeme diskutovat spojitý případ, povšimněme si jedné neobvyklé vlastnosti entropie, která nám dovolí maximalizovat ji dokonce tehdy, má-li být řešením posloupnost splňující

$$p_k > 0\,, \quad \sum_{k=0}^{\infty} p_k = 1\,.$$

V takovém případě namísto výpočtu parciálních derivací vzhledem ke spočetné množině proměnných ($p_0, p_1, \ldots$ a Lagrangeovy multiplikátory příslušející omezením) postačí vzít v úvahu jednoduchou rovnost

$$t \ln t = (t - 1) + \frac{1}{2\tau}(t - 1)^2\,,$$

platnou pro *všechna* $t > 0$, přičemž *kladné* číslo τ závisí na t a leží někde mezi 1 a t (tato rovnost se obdrží z Taylorova rozvoje funkce $t \ln t$ v bodě 1). Jejím použitím za podmínky tvaru

$$E(f) = \sum_{k=0}^{\infty} kup_k < \infty$$

máme pro $\alpha > 0$, $\beta > 0$

$$H - \alpha \, . \, 1 - \beta E(f) = - \sum_{k=0}^{\infty} p_k \ln \left(p_k \, \mathrm{e}^{\alpha + \beta k u} \right) =$$

$$= - \sum_{k=0}^{\infty} \mathrm{e}^{-\alpha - \beta k u} \left(p_k \, \mathrm{e}^{\alpha + \beta k u} \right) \ln \left(p_k \, \mathrm{e}^{\alpha + \beta k u} \right) \leqq - \sum_{k=0}^{\infty} \mathrm{e}^{-\alpha - \beta k u} \left(p_k \, \mathrm{e}^{\alpha + \beta k u} - 1 \right) =$$

$$= -1 + \sum_{k=0}^{\infty} \mathrm{e}^{-\alpha - \beta k u} \, ;$$

ohraničení shora je zde nezávislé na pravděpodobnostním rozdělení $\{p_k, k = 0, 1, \ldots\}$ a rovnost máme právě tehdy, když

$$p_k = \mathrm{e}^{-\alpha - \beta k u} , \quad k = 0, 1, \ldots$$

Z první podmínky

$$\sum_{k=0}^{\infty} \mathrm{e}^{-\alpha - \beta k u} = 1$$

obdržíme

$$\mathrm{e}^{-\alpha} = 1 - \mathrm{e}^{-\beta u}$$

a z druhé podmínky

$$E(f) = \sum_{k=0}^{\infty} k u (1 - \mathrm{e}^{-\beta u}) \, \mathrm{e}^{-\beta k u}$$

obdržíme řešení

$$p_k = \frac{u (E(f))^k}{(u + E(f))^{k+1}} , \quad k = 0, 1, \ldots$$

b) Ve spojitém případě předpokládejme, že známe střední hodnotu μ nezáporné náhodné veličiny, jejíž hustota pravděpodobnosti δ je integrovatelná se čtvercem. V tomto případě je spojitá entropie

$$H(\delta) = - \int_{-\infty}^{+\infty} \delta(x) \ln \delta(x) \, \mathrm{d}x \tag{13}$$

maximalizována při

$$\delta(x) = \begin{cases} \dfrac{1}{\mu} \, \mathrm{e}^{-x/\mu} & \text{pokud} \quad x > 0 \\ 0 \quad \text{jinak,} & \end{cases}$$

což je dobře známá hustota pravděpodobnosti exponenciálního rozdělení. Nyní máme oprávnění k obvyklému předpokladu v teorii hromadné obsluhy, že totiž doba mezi událostmi má exponenciální rozdělení. Takové pravděpodobnostní rozdělení má největší neurčitost, je nejopatrnější a nevylučuje žádnou možnost zachovávající střední dobu μ mezi dvěma událostmi.

c) Samozřejmě je možno klást více podmínek. Předpokládejme ve spojitém případě, že je známa střední hodnota μ i rozptyl σ^2 spojité náhodné veličiny, jejíž hustota pravděpodobnosti je integrovatelná se čtvercem. Je příjemným překvapením, že v tomto případě se spojitá entropie (13) maximalizuje právě při

$$\delta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad (-\infty < x < +\infty),$$

což je hustota pravděpodobnosti normálního rozdělení $N(\mu, \sigma^2)$. Nyní tedy vidíme, proč se toto pravděpodobnostní rozdělení hojně používá v aplikacích statistických závěrů a proč si zasluhuje adjektivum „normální"; v nekonečné množině hustot pravděpodobnosti na reálné přímce integrovatelných se čtvercem se střední hodnotou μ a rozptylem σ^2 je normální rozdělení (neboli Moivrovo-Laplacovo-Gaussovo rozdělení) rozdělením s největší neurčitostí a maximalizujícím entropii. Entropie by musela být vynalezena už jen proto, aby demonstrovala tuto variační vlastnost normálního rozdělení.

Skutečnost, že princip maxima entropie umožňuje jednotný variační přístup k některým dobře známým pravděpodobnostním rozdělením, je právě jedním z důvodů jeho důležitosti. Tutéž strategii (tj. maximalizaci entropie) lze použít i v případě četnějších a důmyslnějších omezení, jako je předepsaný velký počet momentů (řádu většího než 2) několika náhodných veličin. Použitím principu maxima entropie obdržíme jako řešení dosud nevídaná pravděpodobnostní rozdělení.

Na závěr několik poznámek objasňujících poměr subjektivního a objektivního při používání principu maxima entropie. Sám variační problém (maximalizace entropie za daných podmínek pro střední hodnoty nějakých náhodných veličin) je stejně objektivní jako každý jiný optimalizační matematický problém. Přijetí pravděpodobnostní entropie za míru neurčitosti a interpretace řešení z hlediska množství obsažené neurčitosti však je navzdory „přirozenosti" výše uvedených vlastností 1–6 postojem subjektivním. Ovšem skutečnost, že některá důležitá pravděpodobnostní rozdělení statistických závěrů (exponenciální, kanonické, rovnoměrné a především normální rozdělení, které je nejdůležitější) jsou jeho řešeními, nás opravňuje k prohlášení, že princip maxima entropie je více než pouhá jednoduchá konvence.

Princip maxima entropie v sobě zahrnuje jak některé další entropické variační problémy (minimalizace Kullbackovy-Leiblerovy divergence, minimalizace vzájemné závislosti), tak i mnohé nové aplikace (například jeho nové použití v analýze časových řad a entropický algoritmus při rozpoznávání obrazců, který má nejmenší střední délku), což je ovšem již jiná záležitost.

Přeložil M. Krutina

Literatura

[1] L. Boltzmann: *Vorlesungen über Gastheorie.* J. A. Barth, Leipzig, 1896.

[2] M. Born: *Atomic Physics.* Blakie & Son Ltd., London—Glasgow, 1969.

[3] L. L. Campbell: *Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities.* Ann. Math. Statist. *41* (1970), 1011—1015.

[4] J. W. Gibbs: *Elementary principles in statistical mechanics developed with especial reference to the rational foundation of thermodynamics.* Yale University Press, New Haven—Conn, 1902.
[5] S. Guiasu: *Information theory with applications.* McGraw-Hill, New York—London—Düsseldorf, 1977.
[6] S. Guiasu, T. Nguyen Ky: *On the mean length of the entropic algorithm of pattern-recognition.* J. Combinatorics, Information & System Sciences 7 (1982), 203—211.
[7] S. Guiasu, R. Leblanc, C. Reischer: *On the principle of minimum interdependence.* J. Information & Optimization Sciences *3* (1982), 149—172.
[8] E. T. Jaynes: *Information theory and statistical mechanics.* Phys. Rev. *106* (1957), 620—630, *108* (1957), 171—182.
[9] A. I. Khinchin: *Mathematical foundations of information theory.* Dover Publications, New York, 1957.
[10] S. Kullback: *Information theory and statistics.* Wiley, New York, Chapman & Hall, London, 1959.
[11] E. Parzen: *Maximum entropy interpretation of autoregressive spectral densities.* Statistics & Probability Letters *1* (1982), 7—11.
[12] C. E. Shannon: *A mathematical theory of communication.* Bell Syst. Techn. Journal *27* (1948), 379—423, 623—656.
[13] S. Watanabe: *Knowing and guessing.* Wiley, New York, 1969.

vyučování

FYZIKA A MATEMATIKA V ETIOPII

Josef Jelen, Praha

Tento příspěvek je určěn k přehledné informaci o současné situaci výuky matematiky a fyziky na univerzitě v Addis Abebě. Obecněji však může posloužit též jako příklad postavení fyziky a matematiky i v řadě jiných rozvojových zemích. Snad alespoň někteří čtenáři usoudí, že informace v něm obsažené nejsou zcela nezajímavé a stojí za to je přečíst. Článek je založen na zkušenostech z let 1981 až 1984, kdy autor působil na uvedené univerzitě a byl vedoucím tamní katedry fyziky. Proto také věnuje poněkud větší pozornost stavu fyziky než matematice.

Univerzita v Addis Abebě byla založena r. 1950. Od r. 1962 existují na její přírodovědecké fakultě samostatné katedry matematiky a fyziky. V systému a organizaci výuky jsou dosud zřetelně patrné původní americké a britské vlivy. Vyučovacím jazykem je angličtina. Jinak tomu ostatně dosud ani nemůže být, neboť vlastní vědecká terminologie v amharštině se teprve začíná vytvářet. Původně bylo možno na této univerzitě získat jen hodnost B. Sc. (Bachelor of Science). Za vyšším vzděláním museli studenti odcházet do zahraničí. Od r. 1981, kdy studium bylo upraveno a rozšířeno, je možno zde získat i hodnost M. Sc. (Master of Science). O nejvyšším stupni studia, poskytujícím hodnost Ph.D. (Doctor of Philosophy), se uvažuje pro budoucnost.

Základní program, vedoucí k hodnosti B. Sc., je čtyřletý. Většina jeho absolventů po jeho ukončení působí jako učitelé na