Jaroslav Hájek

A property of J-divergences of marginal probability distributions

Czechoslovak Mathematical Journal, Vol. 8 (1958), No. 3, 460–463

Persistent URL: http://dml.cz/dmlcz/100316

Чехословацкий математический журнал, т. 8 (83) 1958, Прага

A PROPERTY OF J-DIVERGENCES OF MARGINAL PROBABILITY DISTRIBUTIONS

JAROSLAV HÁJEK, Praha

(Received August 2, 1957)

It is proved that the J-divergence of any two probability distributions of any stochastic process equals the supremum of J-divergences of finite-dimensional marginal distributions. If this supremum is finite then the distributions are absolutely continuous with respect to each other.

Let us have two arbitrary probability distributions, P nad Q, on a Borel field $\mathscr{F}$ of subsets Λ of a space $\Omega = \{\omega\}$. Let P_a and Q_a, $a \in A$, be corresponding "marginal" distributions on Borel sub-fields $\mathscr{F}_a \subset \mathscr{F}$, defined by $P_a(\Lambda) = P(\Lambda)$ and $Q_a(\Lambda) = Q(\Lambda)$ for $\Lambda \in \mathscr{F}_a$, $a \in A$.

Definition.[1]) *J-divergence J_a between distributions P and Q on the Borel field $\mathscr{F}_a \subset \mathscr{F}$ is the number*

$$J_a = \int \left(\frac{p_a}{q_a} - 1\right) \log \frac{p_a}{q_a} \, \mathrm{d}Q \quad \text{if} \quad P_a \equiv Q_a\,, \tag{1}$$

and

$$J_a = \infty\,, \quad \text{if} \quad P_a \not\equiv Q_a\,, \tag{2}$$

where $P_a \equiv Q_a$ denotes that $[Q(\Lambda) = 0] \Leftrightarrow [P(\Lambda) = 0]$ for $\Lambda \in \mathscr{F}_a$, and $\frac{p_a}{q_a} = \frac{\mathrm{d}P_a}{\mathrm{d}Q_a}$ is the likelihood ratio (Radon-Nikodym's derivative) of P_a w. r. t. Q_a, i. e. such a function of ω that

$$P(\Lambda) = \int_\Lambda \frac{p_a(\omega)}{q_a(\omega)} \, \mathrm{d}Q\,, \quad \Lambda \in \mathscr{F}_a\,. \tag{3}$$

Divergence J_a is symmetrical in P and Q and possesses certain valuable properties. It may be easily shown, that $J_a \geqq 0$, where the sign of equality holds if and only if $P_a = Q_a$. Furthermore, $\mathscr{F}_b \subset \mathscr{F}_a$ implies $J_b \leqq J_a$, where

[1]) See [2], page 158, and [3]. Our definition is that of [3] extended to the case when $P_a \not\equiv Q_a$.

the sign of equality holds if and only if either $J_b = \infty$ or $\frac{p_a}{q_a} = \frac{p_b}{q_b}$ $[P]$; in the latter case F_b is a sufficient Borel field for distinguishing between P_a and Q_a.[2])

Theorem 1. *Let $J_1 \leqq J_2 \leqq \ldots \leqq J_\infty$ be a sequence of J-divergences (see definition) between distributions P and Q on the Borel fields $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \subset \mathscr{F}_\infty$, where $\mathscr{F}_\infty$ is the smallest Borel field containing $\bigcup_1^\infty \mathscr{F}_n$. Then*

$$J_\infty = \lim_{n\to\infty} J_n . \tag{4}$$

If $\lim_{n\to\infty} J_n < \infty$, then $P_\infty \equiv Q_\infty$.

Proof. If $\lim_{n\to\infty} J_n = \infty$, then (4) follows from $J_n \leqq J_\infty$, $n \geqq 1$. Hence we may restrict ourselves to the case when

$$\lim_{n\to\infty} J_n < \infty . \tag{5}$$

First let us prove that (5) implies $P_\infty \equiv Q_\infty$. We shall suppose that $P_\infty \not\equiv Q_\infty$ and deduce a contradiction. If, for example, $P_\infty \ll Q_\infty$ does not hold, then there exists an event $A \in \mathscr{F}_\infty$ such that $P_\infty(A) = \varepsilon > 0$ and $Q_\infty(A) = 0$. Consequently (P. R. Halmos, Exercise 8, § 13), to each $k \geqq 1$ we may choose $n_k \geqq 1$ such that there exists an event A_k in $\mathscr{F}_{n_k}$ satisfying the inequalities

$$\frac{3}{4}\varepsilon < P(A_k) , \quad Q(A_k) < \frac{\varepsilon}{4(k-1)} . \tag{6}$$

Bearing (6) in mind and denoting

$$A_k^* = A_k \cap \left\{\omega : \frac{p_{nk}}{q_{nk}} \geqq k\right\} \tag{7}$$

we may write

$$\frac{\varepsilon}{2} < P(A_k) - Q(A_k) = \int_{A_k} \left(\frac{p_{nk}}{q_{nk}} - 1\right) \mathrm{d}Q = \int_{A_k^*} \left(\frac{p_{nk}}{q_{nk}} - 1\right) \mathrm{d}Q +$$

$$+ \int_{A_k - A_k^*} \left(\frac{p_{nk}}{q_{nk}} - 1\right) \mathrm{d}Q \leqq \int_{A_k^*} \left(\frac{p_{nk}}{q_{nk}} - 1\right) \mathrm{d}Q + (k-1)\, Q(A_k - A_k^*) \leqq$$

$$\leqq \int_{A_k^*} \left(\frac{p_{nk}}{q_{nk}} - 1\right) \mathrm{d}Q + \frac{\varepsilon}{4} ,$$

[2]) It may be shown, however, that J-divergence does not possess the triangle property of a metric: Let us consider three normal distributions on the real line having variances $\sigma_1^2 = 0.1$, $\sigma_2^2 = 1$, $\sigma_3^2 = 2$ and mean values $\mu_1 = \mu_2 = \mu_3 = 0$. The J-divergence of any two of them turns out to be $J_{ik} = \frac{\sigma_i^2}{\sigma_k^2}\left(\frac{\sigma_k^2}{\sigma_i^2} - 1\right)^2$, $1 \leqq i \neq k \leqq 3$, from which we get $J_{12} = 8.1$, $J_{13} = 18.05$, $J_{23} = 0.5$, i. e. $J_{12} + J_{23} < J_{13}$.

i. e.

$$\int\limits_{A_k^*} \left(\frac{p_{n_k}}{q_{n_k}} - 1\right) dQ \geqq \frac{\varepsilon}{4} \,. \tag{8}$$

From (7) and (8) it follows that

$$J_{n_k} = \int \left(\frac{p_{n_k}}{q_{n_k}} - 1\right) \lg \frac{p_{n_k}}{q_{n_k}} dQ \geqq \int\limits_{A_k^*} \left(\frac{p_{n_k}}{q_{n_k}} - 1\right) \lg \frac{p_{n_k}}{q_{n_k}} dQ \geqq$$

$$\geqq \lg k \int\limits_{A_{n_k}^*} \left(\frac{p_{n_k}}{q_{n_k}} - 1\right) dQ \geqq \frac{\varepsilon}{4} \lg k \,, \quad k = 1, 2, \ldots.$$

This last inequality contradicts the supposition (5) and thereby proves $P_\infty \equiv Q_\infty$.

Now, from $P_\infty \equiv Q_\infty$ it follows that there exists $\frac{p_\infty}{q_\infty}$ and

$$J_\infty = \int \left(\frac{p_\infty}{q_\infty} - 1\right) \lg \frac{p_\infty}{q_\infty} dQ \,.$$

Moreover, $\frac{p_n}{q_n} = M\left\{\frac{p_\infty}{q_\infty} \mid \mathscr{F}_n\right\}$, which implies (J. L. Doob, Theorem 4.3, Ch. VII) that

$$\frac{p_\infty}{q_\infty} = \lim_{n\to\infty} \frac{p_n}{q_n} \,. \qquad [Q] \tag{9}$$

By means of (9) and using Fatou's lemma we get from (1) that

$$\lim_{n\to\infty} J_n \geqq \int \left(\frac{p_\infty}{q_\infty} - 1\right) \lg \frac{p_\infty}{q_\infty} dQ = J_\infty \,. \tag{10}$$

Inequality (10) combined with the obvious opposite inequality, $\lim J_n \leqq J_\infty$, gives (4). The theorem is thus proved.[3])

The following version of Theorem 1 is useful for stochastic processes.

Theorem 2. *Let $\{x_t, t \in T\}$ be an arbitrary system of random variables. Let J_K be the J-divergence between distributions P and Q on the Borel field $\mathscr{F}$ generated by a sub-system $\{x_t, K \subset T\}$. Then*

$$J_T = \sup_{K \in \mathscr{K}} J_K \,, \tag{11}$$

where $\mathscr{K}$ is the class of all finite subsets of T. If $\sup_{K\in\mathscr{K}} J_K < \infty$, then $P_T \equiv Q_T$.

---

[3]) We may write $J_n = - H_n(P, Q) - H_n(Q, P)$, where $H_n(P, Q) = - \int \frac{p_n}{q_n} \lg \frac{p_n}{q_n} dQ$ is the entropie of P w. r. t. Q on $\mathscr{F}_n$, $P_n \ll Q_n$, (see [5]). The corresponding theorem for entropies, namely that (i) $H_n(P, Q) \to H_\infty(P, Q)$ and (ii) $\lim H_n(P, Q) > -\infty \Rightarrow P_\infty \ll Q_\infty$, could be proved without any essential change in our method.

A related but considerably weaker result is contained in [5], theorem 7, part (ii), where $H_n(P, Q) \to H_\infty(P, Q)$ is proved under supposition that $\lim_{n\to\infty} H_n(P, Q) \gg -\infty$ and $P_\infty \ll Q_\infty$; the supposition $P_\infty \ll Q_\infty$, being implied by $\lim H_n(P, Q) > -\infty$, is superfluous.

Proof. If T is countable, then it is possible to choose finite subsets $K_1 \subset \subset K_2 \subset \ldots$ such that $\mathscr{F}_T$ is the smallest Borel field containing $\bigcup_{n=1}^{\infty} \mathscr{F}_{K_n}$, and the theorem is reduced to theorem 1.

If T fails to be countable, it may be easily shown, that $J_T = J_S$ for some countable subset $S \subset T$: When $P_T \not\equiv Q_T$ then $P_S \not\equiv Q_S$ for at least one countable $S \subset T$, so that $J_T = J_S = \infty$. When $P_T \equiv Q_T$, then there exists $\frac{p_T}{q_T}$ which is measurable with respect to $\mathscr{F}_T$. However, we know, that every function measurable w. r. t. $\mathscr{F}_T$ is measurable w. r. t. $\mathscr{F}_S$ for at least one countable $S \subset T$, so that $\frac{p_T}{q_T} = \frac{p_S}{q_S}$, which implies $J_T = J_S$.

REFERENCES

[1] *J. L. Doob:* "Stochastic processes", London, New York, 1953.

[2] *H. Jeffreys:* "Theory of probability", Oxford, 1948.

[3] *S. Kullback, R. A. Leibler:* "On information and sufficiency", Ann. Math. Stat. *22* (1951), 79—86.

[4] *P. R. Halmos:* "Measure theory", New York, 1950.

[5] *A. Pérez:* „Notions generalisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales", Transactions of the First Prague Conference on information theory, statistical decision functions, random processes, Prague, 1957.

Резюме

ОБ ОДНОМ СВОЙСТВЕ J-ОТЛИЧИЙ МАРГИНАЛЬНЫХ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ

ЯРОСЛАВ ГАЕК (Jaroslav Hájek), Прага

(Поступило в редакцию 2/VIII 1957 г.)

Доказывается, что J-отличие двух произвольных распределений вероятностей любого стохастического процесса равно верхней грани J-отличий конечно-мерных маргинальных распределений. Если эта верхняя грань конечна, то распределения абсолютно непрерывны одно по отношению к другому.

J-отличие J_a между распределениями P и Q на борелевском поле $\mathscr{F}_a \subset \mathscr{F}$ является числом, определенным следующим образом:

$$J_a = \int \left(\frac{p_a}{q_a} - 1\right) \lg \frac{p_a}{q_a} \,\mathrm{d}Q\,, \quad \text{если} \quad P_a \equiv Q_a\,, \quad J_a = \infty\,, \quad \text{если} \quad P_a \not\equiv Q_a\,,$$

где $P_a \equiv Q_a$ означает, что $[Q(\Lambda) = 0] \Leftrightarrow [P(\Lambda) = 0]$ для всех событий $\Lambda \in \mathscr{F}_a$, а $\frac{p_a}{q_a}$ есть отношение правдоподобия (производная Радон-Никодима) P по Q относительно борелевского поля $\mathscr{F}_a \subset \mathscr{F}$.