Antonín Lukš; Stanislav Komenda

Variance of estimator of a discrete parameter

Kybernetika, Vol. 16 (1980), No. 3, (273)--284

Persistent URL: http://dml.cz/dmlcz/125115

KYBERNETIKA — VOLUME 16 (1980), NUMBER 3

Variance of Estimator of a Discrete Parameter

ANTONÍN LUKŠ, STANISLAV KOMENDA

Studying statistical properties of a simple probabilistic model of the school achievement test authors faced the problem of expressing the amount of information about a parameter contained in an empirically observable quantity in the case when the parametric space is to be considered as discrete. In this paper a measure of information is proposed, which can be applied in this situation.

1. A PROBABILISTIC MODEL OF THE SCHOOL ACHIEVEMENT TEST

It is supposed that the matter the student is obliged to master represents a measurable space $[\Omega, \mathcal{A}]$, where Ω is an abstract set and $\mathcal{A}$ is a σ-algebra of its subsets. After a period of education this universe can be partitioned into two classes $A, A' \in \mathcal{A}$. A is a set of the facts mastered by the student and A' a set of the facts unmastered by him. We suppose that the measurable space $[\Omega, \mathcal{A}]$ can be turned into a measure space $[\Omega, \mathcal{A}, P]$ in such a way that it is meaningful to speak about the proportion $\pi = P(A')$ of the matter unmastered by the student. The measure space $[\Omega, \mathcal{A}, P]$ is identified in a natural way with the Kolmogorov probability space.

The student is put to a school achievement test

$$\left(\begin{array}{c|ccc} s_1 & r_{11} & \dots & r_{1q} \\ \dots & \dots & \dots & \dots \\ s_n & r_{n1} & \dots & r_{nq} \end{array}\right),$$

which consists of questions s_i, $i = 1, \dots, n$, and some offered answers r_{ij}, $i = 1, \dots, n$, $j = 1, \dots, q$ $(q \geqq 2)$ to these questions. For every i it holds that the concept represented by the answer r_{ij} to the question s_i is a fact for just one j, i.e. just one offered answer to a given question is correct. We assume that the school achievement test is made and used in such a way that a student, having unmastered π

of the matter required, knows the correct answer to a given question with the probability $1 - \pi$, and does not know it with the probability π. See [2, 3].

The number Y of questions of the test the student does not know is, accordingly, a random quantity which has the binomial distribution,

$$P(y \mid \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}.$$

(See Fig. 1.) However, the student can guess the correct answer to a given question from q offered answers. Thus the probability p of the student not answering the given question correctly, is given by the formula

$$p = \frac{q - 1}{q} \pi .$$

The number X of questions of the test the student does not answer correctly, is, accordingly, a random quantity, which has the binomial distribution,

$$P(x \mid p) = \binom{n}{x} p^x (1 - p)^{n-x} ,$$

or

$$P(x \mid \pi) = \binom{n}{x} \left(\frac{q - 1}{q} \pi\right)^x \left(1 - \frac{q - 1}{q} \pi\right)^{n-x}.$$

We state yet the conditional distribution of the number X of questions of the test the student does not answer correctly, given the number y of questions of the test the student does not know,

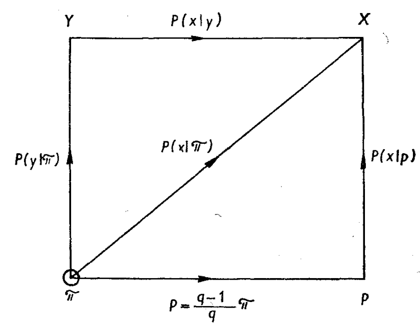

Fig. 1. Scheme of conditioning, dependence on parameters and forming the mixed model.

$$(1) \qquad P(x \mid y) = \binom{y}{x}\left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{y-x}.$$

(For the statistical inference about parameters, see Fig. 2.) If a value x of X is known, x/n is an estimate of the quantity p. The estimator X/n is the unique unbiased estimator for p, as can be proved. The relation $\pi = [q/(q-1)]\, p$ implies that $[q/(q-1)]\,(X/n)$ is a reasonable estimator for the parameter π, although for $x \geqq (q-1)/q$ it takes on values greater than 1. The estimator $[q/(q-1)]\,(X/n)$ is unbiased.

The quantity y unknown by us can be considered a parameter as well. As will be proved in what follows, $[q/(q-1)]\, X$ is the unique unbiased estimator for this quantity. The situation is rather unusual in that this parameter is discrete.

Since Y/n would be the unique unbiased estimator for the parameter π, if Y were observable, we conclude that the estimator $(1/n)\,[q/(q-1)]\, X$ will be a reasonable estimator for the parameter π. But

$$\frac{1}{n} \cdot \frac{q}{q-1} X = \frac{q}{q-1} \cdot \frac{X}{n},$$

so that the two ways of reasoning lead to the same estimator. (See Fig. 2.)

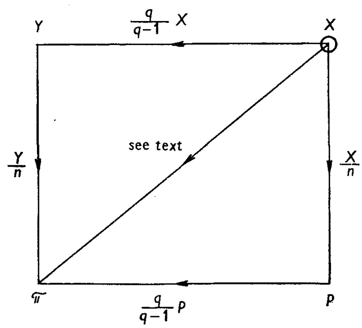

Fig. 2. Scheme of inference. For the estimator of π in the mixed model see text.

We can find easily that Fisher's measure of information about the parameter π yielded by the quantity X [4] is given by the formula

$$\frac{n}{\pi\left(\dfrac{q}{q-1} - \pi\right)}.$$

As for the rest of the problems of the estimation theory, as indicated above, see Fig. 3. As for the estimation of the discrete parameter y, Fisher's measure of information cannot be applied. The purpose of this paper is to introduce a simple measure of information, suitable as a substitute in this situation for that commonly used.

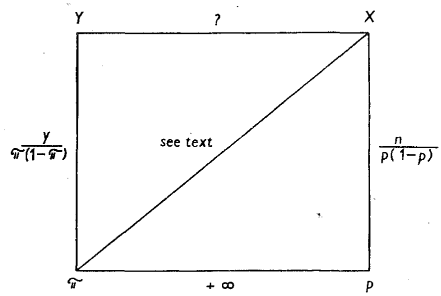

Fig. 3. Information scheme. Our problem is indicated by a question mark. For Fisher's measure of information in the mixed model see text.

Let us show that in the case of the distribution

$$P(x \mid y) = \binom{y}{x}\left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{y-x}$$

$[q/(q-1)]\,X$ is the unique unbiased estimator of y.

Proof. (a) Above all, it is unbiased, for

$$E\left(\frac{q}{q-1}X|y\right) = \frac{q}{q-1}E(X|y) = \frac{q}{q-1}\,y\,\frac{q-1}{q} = y\,.$$

(b) If any estimator $t(X)$ is unbiased, then it satisfies the relation

$$\sum_{x=0}^{\infty} t(x)\binom{y}{x}\left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{y-x} = y \tag{2}$$

for every y. We shall provide an inductive proof that it is necessary for this estimator to satisfy the relation

$$t(x) = [q/(q-1)]\,x \tag{3}$$

for every $x = 0, 1, 2, \ldots$. Substituting 0 for y in (2) we obtain that $t(0) = 0$ or that

(3) is valid for $x = 0$. Let $\bar{x} = 0, 1, 2, \ldots$. If (3) holds for every integer x, $0 \leqq x \leqq \bar{x}$, then substituting $\bar{x} + 1$ for y in (2) we have

$$\sum_{x=0}^{\bar{x}+1} t(x) \binom{\bar{x}+1}{x} \left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{\bar{x}+1-x} = \bar{x} + 1 ,$$

or

$$\sum_{x=0}^{\bar{x}+1} t(x) \binom{\bar{x}+1}{x} \left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{\bar{x}+1-x} =$$

$$= \sum_{x=0}^{\bar{x}+1} \frac{q}{q-1} x \binom{\bar{x}+1}{x} \left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{\bar{x}+1-x} ,$$

which yields the following equation

$$\sum_{x=0}^{\bar{x}} \frac{q}{q-1} x \binom{\bar{x}+1}{x} \left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{\bar{x}+1-x} + t(\bar{x}+1) \left(\frac{q-1}{q}\right)^{\bar{x}+1} =$$

$$= \sum_{x=0}^{\bar{x}} \frac{q}{q-1} x \binom{\bar{x}+1}{x} \left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{\bar{x}+1-x} + \frac{q}{q-1} (\bar{x}+1) \left(\frac{q-1}{q}\right)^{\bar{x}+1}$$

whence, necessarily,

$$t(\bar{x}+1) = \frac{q}{q-1} (\bar{x}+1) .$$

So (3) holds for x, $0 \leqq x \leqq \bar{x} + 1$, which proves the proposition.

2. A MEASURE OF INFORMATION

It is well known [4] that the ability of a random quantity to yield information about the value of a parameter can be judged on what change of distribution of the quantity will be brought about by a change of the value of the parameter. In the case of a continuous parameter, when infinitesimal change is possible, using various metrics [4, 5] one always arrives at the same, Fisher's measure of information. It cannot be expected that this will be preserved in the case of a discrete parameter.

Now, we shall apply some formulae from [5] for the mean information for discrimination between hypotheses to assess the change of distribution (2) induced by augmenting y by 1. Denote μ_y the probability measure determined by the distribution (1). We shall see to the fact that $\mu_{y-1} \ll \mu_y$ (read: μ_{y-1} is *absolutely continuous* with respect to μ_y), but not $\mu_y \ll \mu_{y-1}$. The distribution (1) is discrete, $\mu_y \ll \lambda$, where λ is the counting measure of the countable set $\{0, 1, 2, \ldots\}$.

The validity of $\mu_{y-1} \ll \mu_y$ suffices for the quantity

$$I(y-1 : y) = \sum_{P(x|y-1)>0} P(x \mid y-1) \log \frac{P(x \mid y-1)}{P(x \mid y)} ,$$

$y = 1, 2, \ldots$, to be well-defined. This quantity can be interpreted, but solely for illustrative use in this section, as the information $J(y)$ yielded by the random quantity X about a value of the parameter y. Unfortunately, this interpretation is ambiguous.

We can take

$$J(y) = I(y - 1 : y), \quad y = 1, 2, \ldots$$

as well as

$$J(y) = I(y : y + 1), \quad y = 0, 1, 2, \ldots$$

(For this "ambiguity" and numerical values, see Tab. I.)

Table I.

	$q = 2$	$q = 3$	$q = 4$	$q = 5$
$I(0 : 1)$	0.6932	1.0986	1.3863	1.6094
$I(1 : 2)$	0.3466	0.6365	0.8664	1.0549
$I(2 : 3)$	0.2158	0.4301	0.6163	0.7766
$I(3 : 4)$	0.1521	0.3159	0.4686	0.6059
$I(4 : 5)$	0.1161	0.2453	0.3723	0.4907
$I(5 : 6)$	0.0937	0.1985	0.3055	0.4084
$I(6 : 7)$	0.0785	0.1658	0.2572	0.3472
$I(7 : 8)$	0.0676	0.1420	0.2209	0.3003
$I(8 : 9)$	0.0594	0.1240	0.1930	0.2635
$I(9 : 10)$	0.0530	0.1100	0.1710	0.2341

3. AN ANALOGY OF THE RAO-CRAMÉR INEQUALITY

The theory of the efficiency of estimation is known only for a continuous parameter. It concentrates around the Rao-Cramér theorem, furnishing Fisher's definition of the measure of information with a solid basis. We should like to demonstrate an analogy of this theory for a discrete parameter, which is very transparent, includes an analogy of the Rao-Cramér inequality and is a basis, too, we believe, that a definition of the measure of information can rest on.

Let (a) X be a discrete random quantity which can assume nonnegative integral values $x = 0, 1, 2, \ldots$. Consider a discrete system of distributions of the random quantity X dependent on a discrete parameter $y = 0, 1, \ldots, N$. This system of distributions of the random quantity X is described by the probabilities

$$P(x \mid y), \quad x = 0, 1, 2, \ldots, \quad y = 0, 1, \ldots, N,$$

where

$$\sum_{x=0}^{\infty} P(x \mid y) = 1 , \quad y = 0, 1, \ldots, N .$$

Suppose (b) $P(x \mid y - 1) > 0 \Rightarrow P(x \mid y) > 0$ for every x, $y = 1, 2, 3, \ldots$. Denote

$$J(y) = \sum_{\substack{x \\ P(x/y) > 0}} \left[1 - \frac{P(x \mid y - 1)}{P(x \mid y)}\right]^2 P(x \mid y) ,$$

$y = 1, 2, \ldots, N$, supposing (c) that the right-hand sums are all finite and positive.

Example 1. (*School achievement test.*) Let the random quantity X have the binomial distribution,

$$P(x \mid y) = \binom{y}{x} \left(\frac{q-1}{q}\right)^x \left(\frac{1}{q}\right)^{y-x} ,$$

where q is a known integer, $q = 2, 3, 4, \ldots$.

Note that $P(x \mid y) > 0$ if and only if $0 \leqq x \leqq y$. Now,

$$\frac{P(x \mid y - 1)}{P(x \mid y)} = \frac{y - x}{y} q = q - \frac{q}{y} x ,$$

$$1 - \frac{P(y \mid y - 1)}{P(x \mid y)} = \frac{q}{y} x - (q - 1) = \frac{q}{y}\left(x - \frac{q-1}{q} y\right),$$

$$J(y) = \sum_{x=0}^{y} \left(\frac{q}{y}\right)^2 \left(x - \frac{q-1}{q} y\right)^2 P(x \mid y) =$$

$$= \left(\frac{q}{y}\right)^2 \mathrm{D}(X \mid y) = \left(\frac{q}{y}\right)^2 y \frac{q-1}{q}\frac{1}{q} = \frac{q-1}{y}, \quad y = 1, 2, \ldots, N .$$

Thus,

(4) $$J(y) = \frac{q-1}{y}, \quad y = 1, 2, \ldots, N .$$

Note that this quantity takes on only finite positive values.

Example 2. Let a random quantity X have the hypergeometric distribution

$$P(x \mid y) = \frac{\binom{y}{x}\binom{N-y}{n-x}}{\binom{N}{n}} ,$$

$0 \leqq x \leqq n$, $n \leqq y \leqq N - n$, where n and N are known integers. The algebraic expression for $J(y)$ is found to be irreducible, appropriate only to numerical computations.

Theorem. Let $T = t(X)$ be an estimator for y that has finite second moment for $y = 1, 2, \ldots, N$. Let $b(y) = E(T) - y$, $y = 0, 1, \ldots, N$, be the bias of the estimator T. Let the assumptions (a), (b), (c) hold. Then

$$E(T - y)^2 \geqq \frac{[1 + b(y) - b(y-1)]^2}{J(y)}\,, \tag{5}$$

$y = 1, 2, \ldots, N$.

Proof. The function $b(y)$ is defined implicitely by the formula

$$\sum_{x=0}^{\infty} t(x)\, P(x \mid y) = y + b(y)\,,$$

$y = 0, 1, \ldots, N$. Therefore

$$\sum_{x=0}^{\infty} t(x)\,[P(x \mid y) - P(x \mid y-1)] = 1 + b(y) - b(y-1)\,, \tag{6}$$

$y = 1, 2, \ldots, N$. Moreover,

$$\sum_{x=0}^{\infty} y[P(x \mid y) - P(x \mid y-1)] = 0\,, \tag{7}$$

$y = 1, 2, \ldots, N$. Subtracting (7) from (6) and after slight modification we obtain

$$\sum_{\substack{x \\ P(x|y)>0}} [t(x) - y]\left[1 - \frac{P(x \mid y-1)}{P(x \mid y)}\right] P(x \mid y) = 1 + b(y) - b(y-1)\,.$$

On the left-hand side there is essentially a mean value of a product of two measurable functions. According to the Schwarz inequality [1],

$$\begin{aligned}&[1 + b(y) - b(y-1)]^2 \leqq \\ &\leqq \sum_{\substack{x \\ P(x|y)>0}} [t(x) - y]^2\, P(x \mid y) \sum_{\substack{x \\ P(x|y)>0}} \left[1 - \frac{P(x \mid y-1)}{P(x \mid y)}\right]^2 P(x \mid y)\,.\end{aligned} \tag{8}$$

It proves the inequality

$$[1 + b(y) - b(y-1)]^2 \leqq E(T - y)^2\, J(y)\,.$$

It is equivalent to the assertion (5) by virtue of the assumption that the value of $J(y)$, for $y = 1, 2, \ldots, N$, be finite and positive. Q.E.D.

Now, we shall investigate into when, under the assumptions of the theorem, the equality in (5) is attained. As can be seen, this equality is attained if and only if the equality in (8) is valid. But the Schwarz inequality turns into an equality, if either

$$\sum_{x=0}^{\infty} [t(x) - y]^2\, P(x \mid y) = 0\,, \tag{9}$$

or there exists such a function $K(y)$, $y = 1, 2, \ldots, N$, that

$$1 - \frac{P(x \mid y-1)}{P(x \mid y)} = K(y)\,[t(x) - y] \tag{10}$$

for every x, y, $P(x \mid y) > 0$.

The first case is an absurd case of an exact estimator T of the parameter y, which need not be taken into considerations. (By the way, this case can be eliminated on a purely logical basis, too. [1].)

In the second case, we constrain ourselves to x, y's such that $P(x \mid y) > 0$. If x, y's are such that, moreover, $P(x \mid y-1) > 0$, then the equality (10) can be modified to the form

$$P(x \mid y) = \frac{1}{1 + K(y)\,[y - t(x)]}\, P(x \mid y-1)\,. \tag{11}$$

If y is the least of the values for which $P(x \mid y-1) > 0$ then $P(x \mid y)$ is not determined by the relation (10) and can be chosen arbitrarily, but $\sum P(x \mid y) = 1$ for $y = 0, 1, 2, \ldots$. So the recurrent relation (11) characterizes essentially the type of distribution for which there exists such estimator $T = t(X)$ for y for which the equality in (5) holds.

Example 3. Let us return to the case we have treated in Example 1. Here $P(x \mid y) > 0$ if and only if $x \leqq y$. The equality (11) runs as follows:

$$P(x \mid y) = \frac{y}{y-x}\,\frac{1}{q}\, P(x \mid y-1)\,.$$

We put

$$1 + K(y)\,[y - t(x)] = \frac{(y-x)\,q}{y}$$

$$[y - t(x)]\,K(y) = \frac{qy - qx - y}{y} = \frac{(q-1)\,y - qx}{y} = \frac{q-1}{y}\left(y - \frac{q}{q-1}\,x\right),$$

i.e.

$$t(x) = \frac{q}{q-1}\,x\,, \quad K(y) = \frac{q-1}{y}\,.$$

Accordingly, the equality in (5) is attained for the estimator $T = [q/(q-1)]\,X$ of the parameter y.

Example 4. Let us return to the case treated in Example 2. In a similar way as in Example 3 we arrive at the conclusion that the equality (11) runs here as follows:

$$P(x \mid y) = \frac{y}{y - x} \frac{N - n + 1 - (y - x)}{N + 1 - y} P(x \mid y - 1).$$

Put

$$1 + K(y)\left[y - t(x)\right] = \frac{y - x}{y} \frac{N + 1 - y}{N - n + 1 - (y - x)},$$

$$\left[y - t(x)\right] K(y) = \frac{(N + 1)(y - x) - (N - n + 1) y}{(N - n + 1) y - (y - x) y}.$$

Apparently, it is not possible to find $K(y)$ and $t(x)$ such that this equality be satisfied. The equality in (5) is, accordingly, not attained for any estimator T of the parameter y.

From the theorem it follows immediately that for every unbiased estimator T with finite variance of the parameter y it holds

$$D(T) \geqq \frac{1}{J(y)},$$

for $y = 1, 2, \ldots, N$.

The *efficiency* e of an unbiased estimator T with finite variance is defined as

$$e = \frac{1}{J(y)\, D(T)}.$$

Notice that this is, in fact, a system of N numbers. It holds, of course, $0 \leqq e \leqq 1$. In the case when $e = 1$ for every $y = 1, 2, \ldots, N$, call the estimator T *efficient.*

Example 5. Let us continue the treatment we have begun in Examples 1 and 3. Since

$$E(X \mid y) = y \frac{q - 1}{q},$$

the estimator

$$T = \frac{q}{q - 1} X$$

of the parameter y is unbiased. We know that the equality in (5) is attained for it, or that it is efficient. Independently of it we can compute its variance as follows:

$$D(T) = \left(\frac{q}{q - 1}\right)^2 D(T) = \left(\frac{q}{q - 1}\right)^2 y \frac{q - 1}{q} \frac{1}{q} = \frac{y}{q - 1}.$$

Then

$$e = \frac{1}{J(y)\, D(T)} = \frac{1}{\dfrac{q - 1}{y} \dfrac{y}{q - 1}} = 1.$$

If one believes that an estimator T is efficient, he can define $J(y)$ by the equality $J(y) = 1/D(T)$. This has led one of the authors (A. L.) to all the theory, rather trivial perhaps, for the estimator treated is the only unbiased one.

Example 6. In the case we have treated in Examples 2 and 4, apparently there exists no efficient estimator of the parameter y.

4. THE MEASURE OF INFORMATION

The function $J(y)$ defined by the following formula

$$J(y) = \sum_{\substack{x \\ P(x|y)>0}} \left[1 - \frac{P(x \mid y - 1)}{P(x \mid y)}\right]^2 P(x \mid y),$$

$y = 1, 2, \ldots, N$, is now considered as a measure of information about the parameter y, contained in a random quantity X. In practice, we need not stick to the condition $0 < J(y) < \infty$ and can define the value $J(0)$, too.

Example 7. The definition of the measure of information, which is expressed by the formula (4), can be extended to $y = 0$, $J(0) = +\infty$.

The formula

$$J(y) = \sum_{\substack{x \\ P(x|y)>0}} \frac{[P(x \mid y) - P(x \mid y - 1)]^2}{P(x \mid y)}$$

is equivalent to the foregoing one.

5. DISCUSSION

In our paper, we have been motivated primarily by the effort to find a suitable solution in the case of school achievement test introduced above. However, the analogy of the Rao-Cramér formula presented here, can be generalized in many directions. First of all, one need assume that the univariate random quantity is discrete; it can be continuous, too. Into the formulae, some generalized densities can be introduced, similarly as in Kullback's theory [5]. Further, even the assumption that the parameter takes on positive integral values is not necessary, although appropriate.

Thus, we may derive an inequality for discrimination between two states of our knowledge, between two hypotheses, which can be applied even in the continuous case. Here, however, the equality may be a simple consequence of the well-known Rao-Cramér inequality.

The analysis of the effect of change of parameter on the distribution of a random quantity leads to analogies of Fisher's measure of information, the appropriate choice of metric being decisive in the discrete case. The two metrics treated here are statistically related to the two types of tests of good fit: the likelihood-ratio test (Kullback's metric) and the χ^2-test (our metric), the latter having proved to be the handier one.

(Received April 18, 1979.)

REFERENCES

[1] J. Anděl: Matematická statistika. SNTL, Praha 1978.
[2] S. Komenda, J. Klementa: Modelování a analýza vlivu náhody v didaktickém testu. Matematika a fyzika ve škole *9* (1978/79), 311—318.
[3] S. Komenda, J. Klementa: Ein wahrscheinlichkeitstheoretisches Modell für einen didaktischen Test und seine statistische Analyse. Wiss. Z. (Halle/S.) XVI (1978), 3—9.
[4] C. R. Rao: Lineární metody statistické indukce a jejich aplikace. Academia, Praha 1978.
[5] S. Kullback: Teorija informacii i statistika. Moskva 1967.

RNDr. Antonín Lukš, Ústav pro výzkum vyšší nervové činnosti Univerzity Palackého v Olomouci (Institute of Higher Nervous Activity — Palacký University), I. P. Pavlova 13, 775 20 Olomouc. Czechoslovakia.

RNDr. Stanislav Komenda, CSc., Katedra lékařské biofyziky a nukleární medicíny Lékařské fakulty University Palackého v Olomouci (Institute of Medical Biophysics and Nuclear Medicine — Palacký University), Dr. S. Allende 3, 775 15 Olomouc. Czechoslovakia.