Jaroslav Hájek

An inequality concerning random linear functionals on a linear space with a random norm and its statistical application

Czechoslovak Mathematical Journal, Vol. 12 (1962), No. 4, 486–491

Persistent URL: http://dml.cz/dmlcz/100534

AN INEQUALITY CONCERNING RANDOM LINEAR FUNCTIONALS ON A LINEAR SPACE WITH A RANDOM NORM AND ITS STATISTICAL APPLICATION

JAROSLAV HÁJEK, Praha

(Received August 23, 1960)

It is shown that the norm of a mixed (averaged) linear functional with respect to a mixed (averaged) norm in a linear space cannot exceed the mean square norm of the random functional with respect to the random norm in the linear space. Applications are given concerning projection (prediction etc.) of random variables and linear estimation of regression coefficients in stochastic processes.

1. THEORY

Let us have a complex linear space $M = \{x\}$ and a probability space $(\Lambda, \mathscr{B}, F)$, where $\mathscr{B}$ is a Borel field of subsets of Λ, and $F(B)$, $B \in \mathscr{B}$, is a probability measure. Let $s_{\lambda x}$ be a real function on $\Lambda \times M$, representing for every fixed λ a norm on M and for every fixed x a measurable and quadratically integrable function of $\lambda \in \Lambda$.

Further, let $f_{\lambda x}$ be a function on $\Lambda \times M$, representing for every fixed λ a linear functional on M, if M is normed by $s_{\lambda x}$, and for every fixed x a measurable function of λ. Let $n_\lambda(f_\lambda)$ be the norm of $f_{\lambda x}$, if M is normed by $s_{\lambda x}$, $\lambda \in \Lambda$. We shall assume that $n_\lambda(f_\lambda)$ is a measurable and quadratically integrable function of $\lambda \in \Lambda$.

Put

$$\sigma_x = \sqrt{(E s_{\lambda x}^2)} \quad (x \in M)\,, \tag{1}$$

$$\varphi_x = E f_{\lambda x} \quad (x \in M) \tag{2}$$

where $E(.) = \int (.)\,\mathrm{d}F(\lambda)$. Since $|f_{\lambda x}| \leqq n_\lambda(f_\lambda)\, s_{\lambda x}$, the integrability of $f_{\lambda x}$ follows from quadratic integrability of $n_\lambda(f_\lambda)$ and $s_{\lambda x}$.

Theorem 1. *The function σ_x given by (1) is a norm on M and φ_x given by (2) is a linear functional on M, if M is normed by σ_x. Also*

$$\nu(\varphi) \leqq \sqrt{[E n_\lambda^2(f_\lambda)]} \tag{3}$$

where $\nu(\varphi)$ denotes the norm of φ, if M is normed by σ_x.

If $f_{\lambda x} = \varphi_x$, $\lambda \in \Lambda$, $x \in M$, *we have the following inequality*

(4) $$\nu(\varphi) \leqq \frac{1}{\sqrt{\left[E\left(\frac{1}{n_\lambda^2(\varphi)}\right)\right]}}\,.$$

Proof. We have

$$\sigma_{x+y}^2 = Es_{\lambda,x+y}^2 \leqq Es_{\lambda,x}^2 + Es_{\lambda y}^2 + 2Es_{\lambda x}s_{\lambda y} \leqq$$
$$\leqq Es_{\lambda x}^2 + Es_{\lambda y}^2 + 2\sqrt{(Es_{\lambda x}^2 Es_{\lambda y}^2)} =$$
$$= \sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y = (\sigma_x + \sigma_y)^2\,,$$

which shows that σ_x fulfills the triangle inequality. If $\sigma_x = 0$, then $s_{\lambda x} = 0$ for at least one λ, and hence $x = 0$. Finally, $\sigma_{ax} = |a|\,\sigma_x$ follows directly from (1).

Now, additivity and homogeneity of φ_x follows from (2) and from the same property of $f_{\lambda x}$. Boundedness of φ as well as (3) are implied by the following inequality:

$$|\varphi_x| \leqq E|f_{\lambda x}| \leqq En_\lambda(f_\lambda)\, s_{\lambda x} \leqq \sqrt{[En_\lambda^2(f_\lambda)\, Es_{\lambda x}^2]} = \sigma_x\sqrt{[En_\lambda^2(f_\lambda)]}\,.$$

If $f_{\lambda x} = \varphi_x$, then

$$\sigma_x^2 = Es_{\lambda x}^2 \geqq E\,\frac{|\varphi_x|^2}{n_\lambda^2(\varphi)} = |\varphi_x|^2\, E\,\frac{1}{n_\lambda^2(\varphi)}$$

which is equivalent to (4). The proof is completed.

Now we shall assume that the norms $s_{\lambda x}$ are defined by an inner product $s_{\lambda xy}$, $x, y \in M$, $\lambda \in \Lambda$, *i.e.* that

(5) $$s_{\lambda x} = \sqrt{s_{\lambda xx}} \quad (x \in M,\ \lambda \in \Lambda)\,.$$

Then, provided that $s_{\lambda x}$ are quadratically integrable, $s_{\lambda xy}$ are integrable and

(6) $$\sigma_{xy} = Es_{\lambda xy}$$

is an inner product on M.

Suppose that the linear functionals f_λ and φ admit the representation

(7) $$f_{\lambda x} = s_{\lambda x v(\lambda)} \quad (x \in M,\ \lambda \in \Lambda)\,,$$

(8) $$\varphi_x = \sigma_{xh}\,.$$

The elements $v(\lambda) \in M$, $h \in M$, if they exist, are unique and will be called generating elements of the linear functionals f_λ and φ, respectively. For the existence of such elements it is sufficient that M be complete with norms $s_{\lambda x}$ and σ_x, *i.e.* that M be a Hilbert space. As is well known,

(9) $$n_\lambda(f_\lambda) = s_{\lambda v(\lambda)} \quad (\lambda \in \Lambda)\,,$$

(10) $$\nu(\varphi) = \sigma_h\,.$$

In view of (7) and (8), the relation (2) may be written thus

(11) $$\sigma_{xh} = Es_{\lambda x v(\lambda)} \quad (x \in M)\,.$$

Now we can improve (3) as follows:

Theorem 2. *If the norms* $s_{\lambda x}$ *are defined by inner products* $s_{\lambda xy}$, *and* (7) *and* (8) *hold, then*

(12) $$v^2(\varphi) = En_\lambda^2(f_\lambda) - Es_{\lambda,h-v(\lambda)}^2$$

or, equivalently,

(13) $$\sigma_h^2 = Es_{\lambda v(\lambda)}^2 - Es_{\lambda,h-v(\lambda)}^2 .$$

Proof. We have

(14) $$Es_{\lambda,h-v(\lambda)}^2 = Es_{\lambda v(\lambda)}^2 + Es_{\lambda h}^2 - Es_{\lambda hv(\lambda)} - E\overline{s_{\lambda hv(\lambda)}} .$$

Now, in view of (1) and (11), $Es_{\lambda h}^2 = \sigma_h^2$ and $Es_{\lambda hv(\lambda)} = \sigma_{hh} = \sigma_h^2$, and therefore, also $E\overline{s_{\lambda hv(\lambda)}} = \sigma_h^2$. So (14) is equivalent to (13), which is equivalent to (12). The theorem is proved.

2. APPLICATIONS

In what follows, the space M will consists of all finite linear combinations $\sum c_v x_{t_v}$, $t_v \in T$, of values of a stochastic process x_t, $t \in T$. We shall consider a system of second moments

$$s_{\lambda xy} = \int x\bar{y}\, \mathrm{d}P_\lambda \quad (\lambda \in \Lambda)$$

generated by a system of probability measures P_λ, $\lambda \in \Lambda$, and assume that $\int x\, \mathrm{d}P_\lambda = 0$, $x \in M$, $\lambda \in \Lambda$. The inner product $s_{\lambda xy}$ will be defined on the set K consisting of all random variables having finite second moments with respect to all measures P_λ, $\lambda \in \Lambda$. We suppose that $M \subset K$. We shall not complete the space M by adding limit points, because the set of limit points may depend on $\lambda \in \Lambda$, which causes complications unnecessary in this context.

Now we shall apply the derived inequalities in three situations:

2.1. Projections. If $z \in K$, then

(15) $$f_{\lambda x} = s_{\lambda xz}$$

is a linear functional on M normed by $s_{\lambda x}$, and $d_{\lambda z}$ given by

(16) $$d_{\lambda z}^2 = s_{\lambda z}^2 - n_\lambda^2(f_\lambda)$$

represents the distance of z from M. If M is complete, then $d_{\lambda z}$ is the distance of z from its projection onto M.

Now consider a probability space $(\Lambda, \mathscr{B}, F)$, suppose that all required measurability and integrability conditions are satisfied, and put

(17) $$\delta_z^2 = \sigma_z^2 - v^2(\varphi)$$

where σ and φ are given by (1) and (2). Thus δ_z represents the distance of z from M, when K is normed by σ_x. From (3) it follows that

(18) $$\delta_z^2 \geqq Ed_{\lambda z}^2 .$$

If the operation of taking means with respect to dF is called mixing, then (18) may be expressed thus: the square distance of z from M normed by mixed second moments is not less than the mixed square distance.

Example 1. If M consist of all linear combinations of past values of a stationary process $\{x_t, t \leqq 0\}$ and $z = x_\tau$, $\tau > 0$, then (17) determines the minimum square error of prediction. If we have

$$s_{\lambda x_t x_s} = b^2 r\left(\frac{t-s}{\lambda}\right) \quad 0 < \lambda < \infty \tag{19}$$

where b^2 denotes the variance, and

$$r(u) = 1 - |u|, \text{ if } |u| \leqq 1\,, \quad = 0 \text{ otherwise}\,, \tag{20}$$

then (see [2]) we have

$$d^2_{\lambda x_\tau} = b^2(1 - r(\tau/\lambda))\,. \tag{21}$$

To any distribution function $F(\lambda)$ with $F(0) = 0$ there corresponds a mixed correlation function

$$R(u) = \int_0^\infty r\left(\frac{u}{\lambda}\right) \mathrm{d}F(\lambda) \tag{22}$$

for which the inequality (18) takes on the following form:

$$\delta^2_{x_\tau} \geqq b^2(1 - R(\tau))\,. \tag{23}$$

As any convex correlation function with $R(\infty) = 0$ may be represented in the form (22), the inequality (23) remains true for any convex correlation function. This is shown in the paper [2].

Example 2. If in (19) we take, instead of (20),

$$r(u) = e^{-|u|} \tag{24}$$

then, as is well-known,

$$d^2_{\lambda x_\tau} = b^2(1 - e^{-2\tau/\lambda})\,. \tag{25}$$

To any distribution function $F(\lambda)$ with $F(0) = 0$ there corresponds a mixed correlation function

$$R(u) = \int_0^\infty e^{-|u|/\lambda}\, \mathrm{d}F(\lambda) \tag{26}$$

for which the inequality (18) takes on the form

$$\delta^2_{x_\tau} \geqq b^2(1 - R(2\tau))\,. \tag{27}$$

Thus (27) holds for any correlation function which is obtained by mixing exponential correlation functions $e^{-|u|/\lambda}$.

2.2. Linear estimation of regression coefficients. As is shown in [3], $1/n^2(f)$ is the variance of the best unbiased linear estimate $\hat{\alpha}$ of the parameter α in a linear model, where covariance of x_t and x_s is (independently of α) $s_{x_t x_s}$ and the mean value of x_t is αf_{x_t}. So, adjoining an index λ, we have

$$(28) \qquad \operatorname{var}(\hat{\alpha}_\lambda) = \frac{1}{n_\lambda^2(f_\lambda)}$$

where var (.) denotes the variance, $\operatorname{var}(\hat{\alpha}_\lambda) = s^2_{\lambda\hat{\alpha}_\lambda}$. If $f_{\lambda x} = f_x = \varphi_x$, the inequality (4) yields

$$(29) \qquad \operatorname{var} \hat{\alpha} \geqq E \operatorname{var} \hat{\alpha}_\lambda$$

where $\hat{\alpha}$ corresponds to σ_x given by (1). If $f_{\lambda x_t} = \varphi_{x_t} = 1$, $t \in T$, then α denotes the mean value.

Example 3. The inequality (5.10) in [1].

Example 4. If x_t, $0 \leqq t \leqq T_0$, is a stationary process with correlation function $b^2 e^{-|u|/\lambda}$, and α denotes the mean value, then, as is well-known,

$$(30) \qquad \operatorname{var} \hat{\alpha}_\lambda = \frac{2\lambda b^2}{2\lambda + T_0} \quad (0 < \lambda < \infty).$$

Thus (29) gives

$$(31) \qquad \operatorname{var} \hat{\alpha} \geqq b^2 \int_0^\infty \frac{2\lambda}{2\lambda + T_0} \, \mathrm{d}F(\lambda)$$

for any correlation function of form (26).

2.3. The general case. Let us have m linear functionals $\varphi_1, \ldots, \varphi_m$ on M, and denote by M_m the subset of M consisting of those elements for which

$$(31) \qquad \varphi_{\nu x} = c_\nu \quad (\nu = 1, \ldots, m)$$

where $c_1, \ldots, c_m$ are fixed constants. Now let $M_m - x_0$ consist of elements $x' = x - x_0$, where x_0 is a fixed element from M_m and x is any element of M_m. Then, clearly, $M_m - x_0$ is a linear space, and the distance of any element $z \in K$ from M_m equals the distance of $z - x_0$ from $M_m - x_0$. Thus the inequality (18) is applicable in this general case also.

The scheme just described has the following statistical interpretation: We have a process $\{x_t, t \in T\}$ with a class of covariance functions $s_{\lambda x_t x_s}$, $\lambda \in \Lambda$, and a class of mean values $\sum_{\nu=1}^{m} \alpha_\nu \varphi_{\nu x_t}$, where $-\infty < \alpha_1, \ldots, \alpha_m < \infty$ are unknown constants and $\varphi_{\nu t} = \varphi_{\nu x_t}$, $1 \leqq \nu \leqq m$, are known functions. Observe that covariances do not depend of $\alpha_1, \ldots, \alpha_m$ and mean values do not depend of $\lambda \in \Lambda$. Let us assume that the functions $\varphi_{\nu x_t}$ can be extended to the whole space as linear functionals, which we denote by $\varphi_{\nu x}$, $x \in M$, $1 \leqq \nu \leqq m$ (then none of the $\alpha_\nu' s$ can be linearly estimated with zero variance). Now, let us have a random variable z such that its mean value equals

$\sum_{\nu=1}^{m} \alpha_\nu c_\nu$ and that the covariances $s_{\lambda z x_t}$ do not depend on $\alpha_1, \ldots, \alpha_m$. Then M_m defined by (31) consists of unbiased linear estimates of z (*i.e.* of elements of M having the same mean value as z for any $\alpha_1, \ldots, \alpha_m$) and the square distance of z from M_m denotes the minimum possible residual variance in linear unbiased estimation of z by elements from M, or from the closure of M. Especially, if z is a constant, $z = \sum_{\nu=1}^{m} \alpha_\nu c_\nu$, then the square distance denotes the minimum possible variance of linear unbiased estimates of $\sum_{\nu=1}^{m} \alpha_\nu c_\nu$. In all such cases the residual variance (or simply variance, if z is a constant) satisfies the relation (18). For details see [4].

References

[1] *J. Hájek:* Линейная оценка средней стационарного случайного процесса с выпуклой корреляционной функцией. Czech. Math. Journ., *6 (81)*, 1956, 94—117.

[2] *J. Hájek:* Predicting a Stationary Process when the Correlation Function is Convex. Czech. Math. Journ., *8 (83)*, 1958, 150—154.

[3] *J. Hájek:* On a Simple Linear Model in Gaussian Processes. Trans. Second Prague Conf. Inf. Theory, etc., Praha, 1960.

[4] *J. Hájek:* On Linear Estimation Theory for an infinite Number of Observations. Teorija Verojatnostej, *VI*, 1961, 182—193.

Резюме

ОБ ОДНОМ НЕРАВЕНСТВЕ, КАСАЮЩЕМСЯ СЛУЧАЙНЫХ ЛИНЕЙНЫХ ФУНКЦИОНАЛОВ НА ЛИНЕЙНОМ ПРОСТРАНСТВЕ СО СЛУЧАЙНОЙ НОРМОЙ, И О ЕГО ПРИМЕНЕНИЯХ В СТАТИСТИКЕ

ЯРОСЛАВ ГАЕК (Jaroslav Hájek), Прага

Рассматривается система норм $s_{\lambda x}$ и линейных функционалов $f_{\lambda x}(\lambda \in \Lambda)$, определенных на линейном пространстве $M = \{x\}$. Уравнениями (1) и (2) определяется средняя норма σ_x и средний линейный функционал φ_x. Доказывается неравенство (3), где $\nu(\varphi)$ — норма φ, если в M введена норма σ, а $n_\lambda(f_\lambda)$ означает норму f_λ, если в M введена норма s_λ. Если $f_{\lambda x} = \varphi_x$ не зависит от λ, то справедливо более сильное неравенство (4). Если норма $s_{\lambda x}$ дана скалярным произведением $s_{\lambda xy}$, то справедливы соотношения (12) и (13). Эти результаты применяются к предикции стационарного процесса, функция корреляции которого является смесью функций корреляции $R_\lambda(u) = \max [0, 1 - |u|/\lambda]$ и, соответственно, $R_\lambda(u) = e^{-|u|/\lambda}$, а затем к оценке коэффициентов регрессии в линейных моделях.