

USE OF ENTROPIC AND INFORMATION ANALYSIS OF LIVING WEIGHT OF DAIRY COWS FOR PRODUCTIVITY

Elena KARATIEIEVA, Irina GALUSHKO, Elena KRAVCHENKO, Michael GILL

Mikolaev National Agrarian University, 9 George Gongadze St., 54020, Nikolaev, Ukraine

Corresponding author email: karateeva1207@gmail.com

Abstract

In the article, due to the use of entropy-informational analysis, the statistical parameters of the biological system, the degree of its entropy of absolute and relative organization, which is represented by the live weight of dairy cows, have been studied. It has been established that biological systems, regardless of the breed factor, become more ordered at the age of nine and fifteen months, as well as at their birth. It has also been proven that cows of the Ukrainian black-and-white dairy breed have lower values of unconditional entropy, that is, entropy tends to zero, and the system becomes more orderly. Thus, animals with high levels of ordering of systems in terms of live weight, respectively, will have a high degree of ordering of systems represented by the main indicators of milk productivity. This can serve as a kind of marker when predicting it.

Key words: body weight, dairy cattle, entropy, entropy and information analysis (EIA).

INTRODUCTION

The processes of system development in define direction can be modeled by the exploring information transfer mechanism. Whereas it gives opportunity to clarify the mechanisms of system progress inclusive of its amplification, orderliness and organization degree increase (Gill, 2010; García-Garibay, 2010; Merkyryeva, 1989).

Entropy in general is the degree of uncertainty, any system disorderliness, it is some degree of chaos, disorder expression. Entropy in the context of selection is a capability to show variability in time under the influence of probable factors (Shannon, 1963).

Compared to immunogenotypic analysis that allow to estimate only by the alleles of blood types and protein polymorphism types, EIA takes into account also heterozygosity or homozygosity by the basic features of selection. It gives an opportunity for deeper evaluation of the population variability that is useful while predicting of productivity – selection planning (Gill, 2010; Halushko, 2009; Karateeva, 2013; Pidpala et al., 2018).

In recent years, many statistical approaches have been proposed for detecting gene-gene (G x G) interactions, including numerous methods based on information theory, inspired by the concept of entropy.

They are considered to be especially powerful and, because of their non-linearity, they are better able to capture non-linear relationships between genetic variants and / or variables. In addition, the presented estimates based on entropy are fundamentally different in their design and even in the basic definition of interactions (Ferrario & König, 2018).

Kwon et al. (2014) used IGEN, iNteraction method based on GENome information theory. IGEN is an efficient algorithm for detecting genome-wide gene-gene interactions (GGI) and gene-environment interactions (GEI). And it was found that entropy-based gene interaction (GGI) analysis reveals much of the unexplained inheritance of complex traits.

Malten & König (2020) modified an efficient method for the simultaneous detection of the main effects and effects of gene interaction which is the entropy-based IGEN method. This modification is based on conditional mutual information, subject to the equilibrium of links. The modified estimate is investigated in complex modeling based on five models of genetic interaction using entropy analysis.

Stock-breeder's system concept has to include complete analysis of entropy qualities changes of biological systems of any complexity as they accommodate define number of information that indicative to particular population. The systems are bounded within these populations,

and for such systems the features entropy is growing or remaining the same (Gill, 2010; Kramarenko, 2005).

Bio-computational gene grouping facilitates genetic analysis, sequencing, and structural analysis. To calculate the Shannon's entropy of DNA sequences of genes in the extraction of cluster genes that control milk production in dairy cows, Dehghanzadeh et al. (2017) used the Kulback-Leibler (KL) divergence equation which is based on the similarities and differences of nucleotides and different orders of relative entropy. The research results showed that exons with the maximum entropy value are suitable for genotype analysis using molecular markers, and both coding and non-coding sequences have low or high complexity of system organization.

KL divergence can be used to cluster large gene sets of dairy cattle with other methods to group biologically significant gene sets (Dehghanzadeh et al., 2017).

The relationship between indicators of milk production and genes associated with it was also studied by Dehghanzadeh et al. (2020). Entropy is a measure of the uncertainty of a set of information. In his study (Dehghanzadeh et al., 2020), which is based on the relative entropy of genes and exons, the Kullback-Leibler divergence was calculated using the clustering of genes responsible for the milk production of cows. As a result of research on the study of metabolic pathways of genes based on gene annotations, it was found that the proposed clustering method gives correct, logical and quick results. At the same time, this method did not have the disadvantages associated with alignment, made it possible to take into account genes of actual length and content, and also did not require high memory for long sequences. Therefore, Dehghanzadeh et al. (2020) believe that the performance of their proposed method can be used with other competitive gene clustering techniques for grouping a biologically significant set of genes. Therefore, the proposed method can be considered as a method for predicting genes associated with performance indicators and genes with weak genomic annotations (Dehghanzadeh et al., 2020).

Using entropy information analysis, Ruiz-Marín et al. (2010) studied the etiology of

complex diseases caused by a combination of genetic and environmental factors. Using symbolic dynamics and symbolic entropy as a measure of gene dependence Ruiz-Marín et al. (2010) developed a new, simple, consistent and powerful test for detecting the genetic association of biallelic / SNP markers. This test is based on entropy measures and avoids smoothed nonparametric estimates and is more efficient than Fisher's analysis, especially for a large number of markers (Ruiz-Marín et al., 2010).

Borowska et al. (2018) used information theory as an alternative statistical approach to identify regions of the genome and candidate genes associated with economically useful traits of livestock. The following sperm quality variables were analyzed: CASA sperm kinematics (total motility, mean path speed, straight-line speed, curvilinear speed, amplitude of lateral head displacement, lateral beat frequency, straightness, linearity), sperm membrane integrity (plasmolem, mitochondrial function), the content of ATP in sperm. Entropy and conditional entropy were estimated for each SNP. Conditional entropy quantifies the remaining uncertainty about the values of a variable, taking into account SNPs. The most informative SNPs for each variable were determined. The results of the study showed that important regions of the genome and candidate genes that determine the variable qualities of bovine semen are located on several chromosomes. Scientists have proved the reliability of the effect of SNPs on some variables in the quality of Holstein-Friesian bovine sperm using entropy analysis (Borowska et al., 2018).

Based on the assessment of the results of entropy-information analysis of the signs of milk productivity of Holstein cows, Pidpaloi et al. (2018), it was established the degree of organization and information content for Holstein cows of German and Ukrainian breeding of adjacent generations based on the characteristics of productivity, reproductive and adaptive ability of animals. The following traits were characterized by a high level of determinism: the content of fat and protein in milk in both Holstein cows of German and Ukrainian breeding (Pidpala et al., 2018).

Thus, the use of entropy analysis to study the state of a specific biological system will

provide objective data on the influence of various factors on the level of performance indicators of animals. This, in turn, will improve the accuracy of assessing animals and will make it possible to predict various options for selecting animals for their further use in selection and breeding work.

MATERIALS AND METHODS

The object of research was the full grown cows of three breeds what are specific for South region of Ukraine (n = 189): Red Steppe Breed (n = 88), Ukrainian Black-and-white Dairy Breed (n = 52), Ukrainian Red Dairy Breed (n = 49). These cows belong to two lead farms in Nikolayev region: Red Steppe and Ukrainian Black-and-white Dairy cows to GP «PR Stepnoy», and Ukrainian Red Dairy cows to PSPH «Kozyrscoe». Body weight of newborn cows and three-, six-, nine-, 12-, 15-, and 18-months cows was the subject to study. Entropy-information manipulation of data was made by the generally accepted procedure in S.S Kramarenko version (Kramarenko, 2005). The measure of intrapopulation unconditional entropy of quantitative character was calculated using a formula:

$$H = -\sum_{i=1}^k (p_i \cdot \log_2 p_i)$$

where:

H - entropy of concrete statistical system;
 p_i - probability (or frequency) of characteristic variability by gradations of variation series;
 k - quantity of probable system variants (characters).

Maximum possible theoretically determinate entropy for this system stage is calculated by the formula:

$$H_{\max} = \log_2 k$$

where:

H_{\max} - the degree of complexity or maximum system indeterminateness;
 k - maximum number of system conditions of the characteristic.

The level of absolute system orderliness is calculated using a formula:

$$O = H_{\max} - H$$

The level of relative system orderliness is calculated using a formula:

$$R = 1 - H / H_{\max}$$

Entropy zero level demonstrates the highest orderliness. In determine systems level of relative entropy is high and reaches one. $R = 0$ in completely disorganize systems.

Two-factor analysis of variance was used for establishment of factors effect on the system organization.

RESULTS AND DISCUSSIONS

Ouma et al. (2007) used maximum entropy in search theory with an emphasis on heterogeneity among animals in terms of their live weight and its impact on adaptation to specific conditions of livestock management technology. The data obtained indicate the existence of heterogeneity in live weight and their adaptation to certain conditions. Thus, cattle with higher live weight showed better adaptation to industrial technologies than animals with lower live weight (Ouma et al., 2007).

Based on the entropy of Fukuda et al. (2013), an RBF neural network model was built to predict the weight of pigs based on the growth parameters of Landrace sows. The results showed that the RBF neural network modeling method using entropy analysis was an effective way to build a pig weight prediction model. Entropy removed the collinearity of the explanatory variables in linear regression analysis and allowed for predicting the live weight of pigs better than the linear regression model (Fukuda et al., 2013).

At the same time, there are no data on the study of live weight of cattle using entropy and its main indicators. This technique will improve the accuracy of assessing animals and will make it possible at an early age to predict various options for selecting animals for their further use in their selection. This was the goal of our research.

It should be noted that entropy level is variable within 1.597 3.228 bit by using body weight entropy-information analysis of cows at birth and in the age of three, six and nine months, and also in the age of 12, 15, and 18 months. It means that range of variability is rather wide (Table 1). Absolute entropy by the body weight of Red Steppe Breed cows within the prescribed periods comes up to 2.165 3.228 bit, Ukrainian Red Dairy Breed 2.024 3.171 bit and Ukrainian Black-and-white

Dairy Breed - 1.597 3.179 bit. It shows Ukrainian Black-and-white Dairy Breed system organization high level exactly at representatives.

Table 1. EIA of body weight variability (kg) of cows in the Southern Ukraine

Cows age	n	Body mass entropy parameters of cows, bit			
		$H \pm SEH$	H_{max}	O	R
Newborn	88	2.165±0.076	3.322	1.157	0.348
3 months	88	3.228±0.039		0.094	0.028
6 months	88	2.241±0.114		1.081	0.325
9 months	88	2.222±0.111		1.110	0.331
12 months	88	3.108±0.057		0.214	0.065
15 months	88	2.733±0.086		0.589	0.117
18 months	88	3.138±0.054		0.184	0.055
Newborn	49	2.024±0.080	3.322	1.298	0.391
3 months	49	3.044±0.061		0.278	0.084
6 months	49	3.103±0.045		0.218	0.066
9 months	49	3.090±0.049		0.232	0.070
12 months	49	3.171±0.060		0.151	0.045
15 months	49	3.073±0.052		0.248	0.075
18 months	49	3.083±0.049		0.239	0.072
Newborn	52	1.597±0.096	3.322	1.725	0.519
3 months	52	3.084±0.071		0.237	0.071
6 months	52	3.179±0.063		0.143	0.043
9 months	52	2.814±0.010		0.508	0.153
12 months	52	2.848±0.064		0.474	0.143
15 months	52	2.991±0.066		0.330	0.099
18 months	52	2.966±0.072		0.356	0.101

Absolute and relative entropy in this group increases: O from 0.143 to 1.725 bit and R from 0.043 to 0.519 bit inclusive. Animals with high indexes of the system orderliness by the body weight respectively will have high level of the system orderliness by main characters of milk productivity. It can be kind of marker in milk productivity forecasting.

It has been noted that the tendency of orderliness degree increases at birth, and at the age of nine and fifteen months that is not depend on the breed, it is demonstrated by entropy low values of Red Steppe Breed (2.0165; 2.222; 2.273 bit), Ukrainian Red Dairy Breed (2.024; 3.090; 3.073 bit), Ukrainian Black-and-white Dairy Breed (1.597; 2.814;

2.991 bit). The system becomes more organize in the end of growing period.

As a result, presented biological systems by cows body weight generally by Antomonov classification (Antomonov, 1977) are found as the stochastic quasideterministic systems, as their relative orderliness is not more than 0.1, that points at their high separability during the stock breeding.

Performed analysis of variance established that age factor or body weight period formation to 61% ($P \leq 0.01$) have essential influence on the dynamics of orderliness degree display during the whole cows growing period. It is not depend on the breed only ($\eta^2 = 1\%$), random factors have impact (38%) (Table 2).

Table 2. Factorial conditionality of the entropy level of cows body weight

Impact factors	SS	Df	MS	F	p	η^2
A - breed	0.224649	2	0.112324	1.293	0.310	1.0
B - age	3.047053	6	0.507842	5.847	0.005	61.0
Random factors	1.042297	12	0.086858			38.0
Total variability	4.313999	20				

CONCLUSIONS

It is proved, that Biological systems that are presented by body weight of Red Steppe, Ukrainian Red Dairy and Ukrainian Black-and-white Dairy cows belong to stochastic quasideterministic systems.

It was found out that orderliness degree increases at birth, and at the age of nine and fifteen months. It is appropriate to use received results of different systems as accessory parameters in the selective stock breeding and in the milk productivity forecasting.

It was deduced statistically significant factors influence on body weight formation during development: body weight formation period (age) dominates (to 61%) in comparison with breed factor - 1.0%.

Thus, the obtained results make it possible to assert that the use of entropy-informational analysis in the selection of animals can be used as an additional indicator of their assessment for the main economically useful traits, in particular, their live weight. That will allow you to get a more accurate and complete assessment and predict their future live weight at an early stage of development.

ACKNOWLEDGEMENTS

The authors are grateful to Vladimir Pogorelov, Director of the State Enterprise «Plemproductor Stepnoy» (Steppe Breeding Reproducer), Bytun Semen Demjanovich, Director of the private agricultural enterprise «Kozyrskoe» for the assistance provided to conduct this research on the basis of the above-mentioned enterprises.

We thank to Sergey Kramarenko, Doctor of Biological Sciences, for the author's method of calculating the entropy-information analysis used for biological systems.

We are grateful to one of the co-authors and scientific leader, Doctor of Agricultural Sciences, Professor Mikhail Ivanovich Gill for the idea of using this technique in our research.

We are grateful to Tetiana Kushchova, the teacher of the foreign languages department, for the assistance in the translation of the article.

REFERENCES

- Antamonov, Y.G. (1977). Modelirovanie biologicheskikh sistem: Spravochnik. Kyiv: *Naukova dumka*, 260.
- Borowska, A., Szwaczkowski, T., Kamiński, S., Hering, D. M., Kordan, W., & Leczewicz, M. (2018). Identification of genome regions determining semen quality in Holstein-Friesian bulls using information theory. *Animal reproduction science*, 192, 206-215.
- Dehghanzadeh, H., Mirhoseini, S.Z., Ghaderi-Zefrehei, M., Tavakoli, H., & Esmailkhaniyan, S. (2017). Clustering dairy cattle genes by Kullback-Leibler divergence. *Online Journal of Veterinary Research*, 21(9), 600-615.
- Dehghanzadeh, H., Ghaderi-Zefrehei, M., Mirhoseini, S.Z., Esmailkhaniyan, S., Haruna, I.L., & Amirpour, N.H. (2020). A new DNA sequence entropy-based Kullback-Leibler algorithm for gene clustering. *Journal of applied genetics*, 61(2), 231.
- Fan, R., Zhong, M., Wang, S., Zhang, Y., Andrew, A., Karagas, M., & Moore, J.H. (2011). Entropy-based information gain approaches to detect and to characterize gene-gene and gene-environment interactions/correlations of complex diseases. *Genetic epidemiology*, 35(7), 706-721.
- Ferrario, P.G., & König, I.R. (2018). Transferring entropy to the realm of GxG interactions. *Briefings in bioinformatics*, 19(1), 136-147.
- Fukuda, O., Nabeoka, N., & Miyajima, T. (2013). Estimation of marbling score in live cattle based on ICA and a neural network. *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 1622-1627.
- Galushko, I.A. (2009). Selektiino-henetychna otsinka produktyvnykh oznak koriv holshtynskoi porody zarubizhnoi selektzii. *Dysertatsiia na zdobuttia naukovooho stupenia kandydata silskohospodarskykh nauk*, 6(01).
- Garcia-Garibay, M.A. (2010). The entropic enlightenment of organic photochemistry: strategic modifications of intrinsic decay pathways using an information-based approach. *Photochemical & Photobiological Sciences*, 9(12), 1574-1588.
- Gil M.I., & Kovalenko, V.V. (2010). Vykorystannia entropiinoho analizu v otsintsi molochnoi produktyvnosti khudoby riznoi intensyvnosti formuvannia orhanizmu. *Ahrarnoi nauky prykhornomoria Vypusk*, 1 (52), 2, 12.
- Golan, A., Perloff, J.M., & Shen, E.Z. (2001). Estimating a demand system with nonnegativity constraints: Mexican meat demand. *Review of Economics and Statistics*, 83(3), 541-550.
- Karateeva, O.I. (2013). *Vplyv intensyvnosti formuvannia koriv riznykh porid v yikh rannomu postnatalnomu ontogenezi na produktyvnist*. Doctoral dissertation, stupenia kand. s.-h. nauk: spets. 06.02. 01 "Rozvedennia ta selektsiia"/Olena Ivanivna Karatieieva-Mykolaiv, 24 s.
- Kramarenko, S.S. (2005). Metod ispolzovaniya entropiyno-informatsionnogo analiza dlya kolichestvennykh priznakov. *Izvestiya Samarского nauchnogo tsentra Rossiyskoy akademii nauk*, 7(1).

- Khvostyk, V.P. (2012). Vykorystannia entropiinoho analizu dlia kharakterystyky systemy za chastotoiu alelei ovoproteinovykh lokusiv u miaso-yaiechnykh kurei riznoho henezysu. *Zbirnyk naukovykh prats. Kamianets-Podilskyi: Podilskyi ahrarno-tekhnichnyi universytet*, 284-285.
- Kwon, M.S., Park, M., & Park, T. (2014). IGENT: efficient entropy based algorithm for genome-wide gene-gene interaction analysis. *BMC medical genomics*, 7(1), 1-11.
- Malten, J., & König, I.R. (2020). Modified entropy-based procedure detects gene-gene-interactions in unconventional genetic models. *BMC Medical Genomics*, 13(1), 65-65.
- Merkyryeva, E.K., Bertazin, A.B. (1989). Primenenie jentropijnogo analiza i kojeficienta informativnosti pri ocenke selekcionnykh priznakov v molochnom skotovodstve. *Doklady VASHNIL*, 2, 21-23.
- Musani, S.K., Shriner, D., Liu, N., Feng, R., Coffey, C. S., Yi, N., & Allison, D.B. (2007). Detection of gene \times gene interactions in genome-wide association studies of human population data. *Human heredity*, 63(2), 67-84.
- Ouma, E., Abdulai, A., & Drucker, A. (2007). Measuring heterogeneous preferences for cattle traits among cattle-keeping households in East Africa. *American Journal of Agricultural Economics*, 89(4), 1005-1019.
- Patreva, L.S. (2005). Informacijno-statistichnij metod analizu oznak u populacijah kachok. *Tvarinnictvo Ukraïni*, 10, 16-17.
- Pidpala, T.V., Kramarenko, O.S., & Zaitsev, Y.M. (2018). Vykorystannia entropiinoho analizu dlia otsinky rozvytku oznak molochnoi khudoby holshtynskoi porody. Naukovyi visnyk Lvivskoho natsionalnoho universytetu veterynarnoi medytsyny ta biotekhnologii imeni SZ Ghzhtskoho. Serii: Silskohospodarski nauky, 20(84), 3-8.
- Ritchie, M.D., & Van Steen, K. (2018). The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. *Annals of translational medicine*, 6(8).
- Ruiz-Marín, M., Matilla-García, M., Cordoba, J.A.G., Susillo-González, J.L., Romo-Astorga, A., González-Pérez, A., & Gayán, J. (2010). An entropy test for single-locus genetic association analysis. *BMC genetics*, 11(1), 1-15.
- Shannon, C.E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.
- Shannon, K. (1983). Raboty po teorii informatsii i kibernetiki: per. s angl. / Shenon K. Moskva, 831.
- Smetana, O. Y. (2011). Seleksiyno-henetychna otsinka produktyvnykh oznak koriv holshtynskoy porody za umov diyi stabilizuyuchoho vidboru. *Dysertatsiia na zdobuttia naukovoho stupenia kandydata silskohospodarskykh nauk*, 6(01).
- Van Lishout, F., John, J.M.M., Gusareva, E.S., Urrea, V., Cleynen, I., Théâtre, E., & Van Steen, K. (2013). An efficient algorithm to perform multiple testing in epistasis screening. *BMC bioinformatics*, 14(1), 1-10.