Livestock Breeding and Genomics

Peter von Rohr

2018-09-20

Contents

Preface

Livestock Breeding and Genomics is a masters course taught in the Institute of Agricultural Sciences at ETH Zurich. The aim of this course is to introduce the basic concepts of livestock breeding and genomics to the students.

Prerequisites

There are no explicit prerequisites for this course. All required concepts are introduced during this course. Although basic knowledge in the following areas is definitely helpful for this course.

- **Linear Algebra**: Basic properties of and basic operations with vectors and matrices
- **Statistics**: Simple linear models and least squares
- **R**: Basics of how to work with data and how to do simple computations

Goals

The following goals are taken from the teaching system at ETH. The students are able to set up design matrices, the additive genetic relationship matrix and its inverse as well as the mixed model equations to estimate BLUP breeding values. The concept of BLUP breeding values are also extended to genomic selection. All concepts introduced during this course are not only presented during the lecture, but students are expected to work with the concepts on small example problems.

Chapter 1

Introduction

1.1 Terminology

The terms

- **Livestock Breeding** and
- **Animal Breeding**

are used interchangeably in this course. Although, the latter is a more general term, whereas the former focuses on breeding activities in livestock species, i.e. animals which are typically present on a farm such as cattle, pig, goat and sheep. Animal breeding in general could also include pets such as dogs or cats or even zoo animals. But the very interesting topic of breeding such species is outside of the scope of this lecture and is therefore not covered in these course notes.

The term `Livestock Breeding` is sometimes understood ambiguously. In general, most people do not differentiate between Livestock Breeding and animal husbandry. From a scientific point of view a `Livestock Breeder` is a person who owns a number of animals from which he or she selects parent animals and uses a designed mating scheme to achieve a certain goal with the offspring animals. Most commonly known are breeders of pet animals such as dogs, cats or birds which follow individual breeding goals which focus on specific phenotypic appearance or on special behavioral traits. Around the end of the 19^{th} century, livestock breeders have realized that they have to work together in breeding associations to be able to effectively select parent animals from a large breeding population and thereby are able to achieve more robust selection responses in shorter amounts of time. This development is nicely documented by the archive of Swiss agricultural history available at https://www.histoirerurale.ch/afaahr/.

1.2 History

Livestock Breeding and Genomics are two scientific areas which have seen some quite dramatic changes in the last few years. As already mentioned in the previous section, Livestock Breeding started over 100 years ago and is a much older discipline than Genomics.

1.2.1 Livestock Breeding

In principle, Livestock Breeding did exist for a very long time in a rather unsystematic form. Individual breeders always made choices about which animals they want to select as parents for the next generation of their livestock herds. Accounts that such early activities in Livestock Breeding happened as early as the Middle Ages are given in (Duerst, 1931) and (Arndorfer et al., 2010). But to the best of my knowledge it was only in the second half of the 20th century that the area of Livestock Breeding made some ground-breaking progress which spread all over the world. This progress was initiated by a researcher called **Charles R. Henderson**. He and his team developed a solid methodology that is still applied up to the current day.

1.2.2 Genomics

Genomics started with the Human Genome Project. The publication of the first draft of the complete Human genome (Venter et al., 2001) and a publication by (Meuwissen et al., 2001) that appeared in the same year made it possible to include information of complete genomes into statistical analyses.

The introduction of genomics methodologies in the area of livestock breeding caused a shift of paradigm. In large livestock breeding populations associations between certain genetic variants and desirable expression forms of phenotypic traits can be estimated using older breeding animals. The obtained estimation results can be used to assess the genetic potential of young animals which do not have any phenotypic observations available. This procedure is termed `genomic selection` and it allows for selecting breeding animals at a much younger age which shortens the generation interval.

The basic principle of how animals are selected as parents of future generations did not change, but the availability of different types of information and the amount of information that can be used to assess the genetic potential of a selection candidate changed dramatically since the invention of genomic technologies. Despite these rapid developments of new technologies, Livestock Breeders are still facing the following two fundamental questions.

1.3 Fundamental Questions

In Livestock Breeding and Genomics, we are interested in addressing two fundamental questions that bothered breeders for a very long time. For this course, we put these two fundamental questions into the following form.

1. What is the *best* animal?
2. What can breeders do to obtain the *best* animal?

The term `best` is relative, because there is no `best` animal for all situations. Animals that show high performances in one environment, may be able to produce as much in a different environment. One example for that might be Holstein cows in Europe or North America are able to produce a lot of milk, but they have difficulties to survive in Africa. Knowing that the environment plays an important role for livestock animals, we will be assuming that the animals that we are selecting, are more or less adapted to their environment.

Animals are usually described or characterized in terms of appearance or performance or a combination of both. In any case, we will be talking about **traits** where any trait is an observable or measurable characteristic of an animal. Examples of *observable* traits are

- coat color
- size
- muscling
- leg set
- udder conformation and many more.

Observable traits are mostly used to describe the appearance of an animal. In contrast to that *measurable* traits are mostly used to describe the performance of an animal. Examples of measurable traits are

- body weight
- milk production
- protein and fat yield.

Note, it is important to distinguish between the observed or measured values of a trait which might be `red` coat color or 343 kg of body weight and the traits themselves which are just coat color or body weight. The observed or measured values of a trait are also called **phenotypes**.

1.3.1 Genotypes and Phenotypes

In Livestock Breeding we are mainly concerned with changing animal populations at the genetic level. The reason why we are interested in changing a population genetically is because parents do not pass their phenotypes to their offspring. Parents pass a random sample of their genes to their offspring. For each offspring every parent does transmit a different sample of their genes. From a genetic point of view, we want to know not only the most desirable phenotype, but also the most desirable **genotypes**. From the central dogma of molecular biology (https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology), it follows that an animal's genotype provides the genetic background of phenotypes. The relationship between phenotypes (P) and genotypes (G) can be summarized by the following equation (1.1)

$$P = G + E \tag{1.1}$$

where E represents the **environmental effects**. Because we want to change our populations at a genetic level, we are interested in the effect (G) of every genotype. In most cases, we are not able to directly observe or measure G. But we will see later on how we can estimate G based on measurements and observations of P and based on estimates of E. The estimates of G will later be called **breeding values** and those estimates will be used by breeders as information for their tools to improve animal populations. Those tools are being described in the following section.

1.4 Improvement of Animal Populations

The purpose of Livestock Breeding is to improve animal populations. Once an animal is conceived, the genotype is fixed[1] and cannot be improved anymore. Breeders can improve populations at the genetic level using the following two tools

1. selection
2. mating

1.4.1 Selection

Selection is the process to determine which individuals of a current population become parents of the next generation. The application of selection in a certain population over a certain time changes the animals in that population at the genetic level. The most familiar form of selection is **natural selection** which occurs in natural and wildlife populations. Natural selection is one of the great forces of evolution and it also affects domestic animals. All animals with lethal genetic defects are naturally selected against, i.e., they never become parents.

Although natural selection cannot be ignored for livestock species, what is most important for animal breeders is **artificial selection**. The idea behind artificial selection is simple. For a given trait all animals in a population are ranked according to their breeding value. From this list the animals ranking top are used as parents for the next generation. In most livestock populations, animal breeders are interested to improve their animals with respect to more than just one trait. When considering more than one trait, the question is how to come up with the ranking for the animals that are selected as potential parents. Several strategies to produce such a ranking based on a number of traits. It has been shown that using a weighted mean of the breeding values of all traits which is called **selection index** to rank all animals is an optimal procedure to be used as selection criterion (Hazel, 1943).

---

[1] Here we do not take into account new technologies such as gene editing.

1.4.2 Mating

The second tool we have available as animal breeders is **mating**. In a mating scheme, we decide which of the selected male animals are bred to which selected female animals. There are a number of different rules that can be followed. The application of a given set of rules are summarized as mating system. There are three reasons for using a mating system.

1. producing offspring with extreme breeding values. When parents with extreme breeding values (high or low) are mated, offspring with extreme phenotypes can be expected. This is mostly used when a given trait is to be changed in one direction
2. make use of complementarity in parental traits. When neither of the parents is optimal, a mix of traits can be desirable. In such a case parental genotypes can be quite different. When parents of different breeds are mated, then this is called **crossbreeding**.
3. obtain positive effects due to heterosis. Hybrid vigor or heterosis in crossbreeding occurs when offspring performance exceeds the performance of the pure-breds.

There might also be other aspects that influence a mating system, e.g. to restrict the level of inbreeding or to consider optimum genetic contribution theory (Meuwissen, 1997).

1.5 Statistics

Several authors such as (Schaeffer, 2013) and (Gianola and Rosa, 2015) have reviewed the development of statistical methods in the area of animal breeding. Both authors mention that statistical methodology plays an important role in animal breeding. Most animal breeders are concerned with estimating or predicting breeding values. This is still done using a set of methods resulted from the theory developed by Charles Henderson and his team ((Henderson, 1953) and (Henderson, 1975)). These methods are known under the name of **BLUP**. BLUP shows some important regularisation properties. These properties allow us to estimate or to predict many more unknown parameters than we have observations. In animal breeding, breeding values of all animals in a population can be predicted also for those animals for which we do not have observations. This is particularly important for traits which can only be observed in animals of one sex.

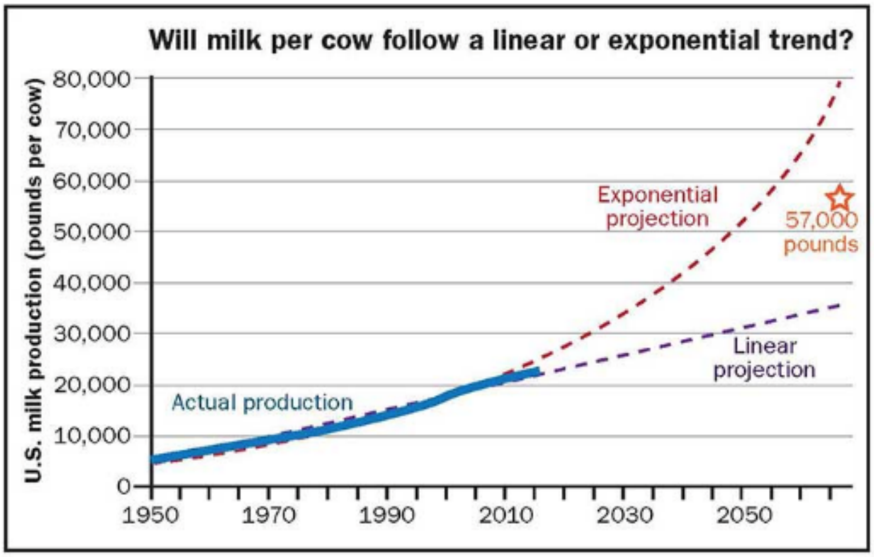
There are more methods with regularisation properties. The so-called **Bayesian methods** are one example. Bayesian methods use the so-called Bayes theorem ((Bayes and Price, 1763) and https://en.wikipedia.org/wiki/Bayes%27_theorem) to come up with parameter estimates. Although Bayesian methods are much older than other methods such as BLUP, they were only introduced into practical animals breeding in the early 1990's. Important pioneering papers for the use of Bayesian methods in animal breeding are (Gianola and Foulley, 1982) and (Gianola and Fernando, 1986). The reasons for the late adoption of Bayesian methods are certainly related to development of cheap computing infrastructure. This is described in more detail in subsection 1.6.

1.6 Computer Science

The development of computing power is often summarized by the so-called **Moore's Law** ((Moore, 1965) and https://en.wikipedia.org/wiki/Moore%27s_law). Moore's law is not a law in the sense of mathematics or physics, but it is a prediction that Gordon Moore[2] made as early as 1965. He predicted that the number of components that could be placed on a certain integrated circuit would double roughly every 18 months between 1959 and 1975. This prediction was generalized into a statement that the general computing performance could be doubled every 18 months. In retro-spect this was more or less true for the last 50 years. This considerable increase in computing performance had also a dramatic impact in the costs of a certain computation.

---

[2] One of the co-founders and a director of Intel

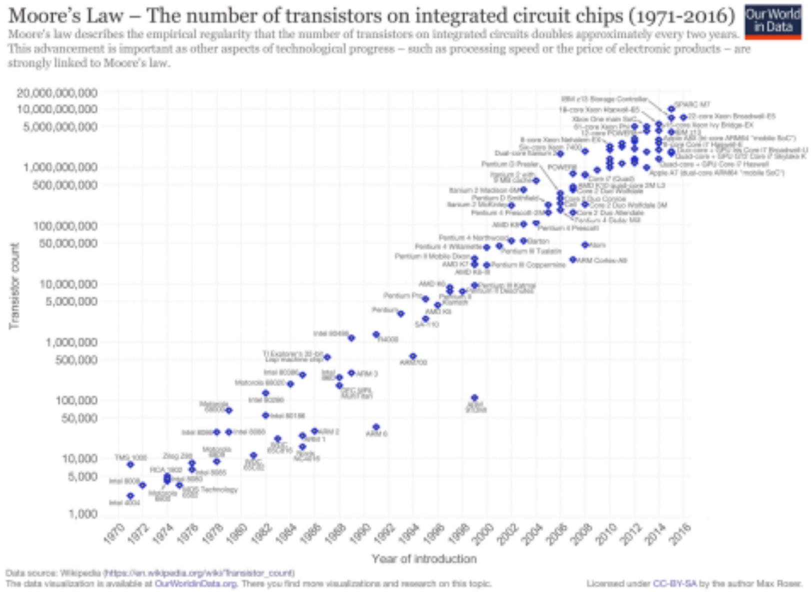


Milk Performance per Cow
(Source: https://hoards.com/article-20808-what-will-dairy-cows-and-farms-look-like-in-50-years.html)

Figure 1.1: Yearly Milk Yield per Cow in the USA

When comparing the development of computing performance with the performance of livestock animals, there is an obvious relation between the two. This means the performance increase of livestock animals was in part facilitated by the development of cheap computing power. The two figures 1.1 and 1.2 compare the two developments. The first diagram shows the annual milk production per cow.

The Figure 1.2 below shows the development of computing power according to Moore's law.


Moore's Law – The number of transistors on integrated circuit chips (1971-2016)

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.

Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

Source: https://en.wikipedia.org/wiki/Moore%27s_law

Figure 1.2: Computing Performance According To Moore's Law

Bibliography

Arndorfer, M., Blauensteiner, P., Berger, B., Falschlunger, G., Fischerleitner, F., Froihofer, W., Gaber, R., Holzner, W., Kajtna, B., Koller, B., Kriechbaum, M., Machatschek, M., Neunteufel, M., Spitzenberger, F., Michael, S., Stekovics, E., Vogl, C., Vogl-Lukasser, B., Wallner, R. M., Wilfing, A., and Zipser, P. (2010). *Wie viele Arten braucht der Mensch?* Böhlau, Wien, Köln, Weimar.

Bayes, T. and Price, R. (1763). An essay towards solving a problem in the doctrine of chance. by the late rev. mr. bayes, communicated by mr. price, in a letter to john canton. *Philosophical Transactions of the Royal Society of London*, 53:370—418.

Duerst, J. U. (1931). *Grundlagen der Rinderzucht*. Springer, Berlin; Heidelberg.

Gianola, D. and Fernando, R. L. (1986). Bayesian methods in animal breeding theory. *J. Anim. Sci.*, 63:217—244.

Gianola, D. and Foulley, J. L. (1982). Non linear prediction of latent genetic liability with binary expression: An empirical bayes approach. *Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain*, 7:293––303.

Gianola, D. and Rosa, G. J. M. (2015). One hundred years of statistical developments in animal breeding. *Annual Review of Animal Biosciences*, 3:13.1—13.38.

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, 28:476–490.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9:226—252.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31:423—449.

Meuwissen, T. H. E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75:934—940.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157:1819–1829.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, 38.

Schaeffer, L. R. (2013). *History of Genetic Evaluation Methods in Dairy Cattle*. self.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., Mckusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V. D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., ru Ji, R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., Mccawley, S., Mcintosh, T., Mcmullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., hui Rogers, Y., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Blick, L., Caminha, M., Carnes-stine, J., Caulk, P., hui Chiang, Y., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., Mcdaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The Sequence of the Human Genome. *Science*, 291(5507):1–38.