УДК 550.34

ON THE MEANING OF STATISTICS IN GEOPHYSICS

H. Moritz

Technical University, Graz, Austria

The paper discusses basic concepts such as objective and subjective probability, ergodicity etc. in the context of geophysics, with special emphasis on the statistics of gravity anomalies.

О СМЫСЛЕ СТАТИСТИКИ В ГЕОФИЗИКЕ

Г. Мориц

Технический университет в Граце, Австрия

В статье обсуждаются такие фундаментальные понятия, как объективная и субъективная вероятность, эргодичность и другие в их применении к геофизике, к особенности в статистике аномалий силы тяжести.

Introduction

All concepts of probability share the same *structure* of the well–known axioms of Kolmogorov. Their *interpretation*, however, is a matter of hot disagreement, especially if one takes an extreme standpoint (a standpoint is one's horizon if its radius goes to zero (ascribed to D. Hilbert)).

Two interpretations are customary. The *objective interpretation* tries to deal with actual phenomena which are so irregular that they can only be described statistically. The *subjective interpretation* asserts that probability is a statement not about nature, but about our subjective knowledge.

For most scientists an objective interpretation comes natural: histograms of natural phenomena are standard. The subjective interpretation arises in cases of prediction; the accuracy of prediction is considered to express our trust in the validity of the prediction. This is particularly evident in *Bayes estimation* where one assumes that the quantity to be predicted has a certain a priori probability which we assign subjectively.

In the view of importance of Bayes estimation, some mathematicians and even one famous geophysicist, Sir Harold Jeffreys, advocate a subjective interpretation throughout [1, 2].

I think the majority of scientists, however, is rather pragmatic (usually automatically without much reflection). To them, objective probability is used in most cases, in agreement with their feeling that science should tell us something objective about nature rather than about our mental state. At the same time, they nonchalantly use subjective probability in prediction, including Bayes' theorem, cf. [3].

This seems to be alright if, wherever available, we try to use objective probabilities for our subjective assumptions.

1. Earth Statistics

Take population statistics. We speak of average income, average occurence of diseases of the world etc. One can (and does) as well publish statistical data for countries like Austria. Thus we have global and regional data which perfectly make sense and provide very useful information.

Gravity anomalies Δg are too irregular to be modeled by simple mathematical functions. The statistical treatment of Δg works pretty well, however. Histograms of Δg distributions, globally and regionally, are frequently remarkable like Gaussian (normal) distributions. Why? Poincaré said that physicists believe in the validity of the Gaussian distribution because they think it is a mathematical law, whereas mathematicians believe in it because they think it is a law of nature. In any case, it usually works well, and the official explanation is the Central Limit Theorem of probability.

2. True Randomness?

An eminent colleague from physics and geodesy told me that the statistical interpretation of the gravity field cannot be right because this field cannot be considered "truly random" but is perfectly determined by the mass distribution of the Earth [4]. He is of course right. My answer is that also the topographic masses are rather irregular, especially if we try to remove trends to the best possible extent. At any rate, gravity anomalies are as regular and as random as, e.g., population statistics. The boundaries between determinism and randomness are rather fluid, as modern chaos theory shows.

3. Ergodicity

A very clever attack against gravity statistics was leveled by Lauritzen [5]. In a rather complicated way he proved that a gravity distribution over the sphere cannot be at the same time Gaussian and ergodic.

What does ergodicity mean? Statistical mechanics teaches us that a mechanical system is embedded in an infinite ensemble of fictitious similar systems; this ensemble is called *phase space* and serves as a probability space for a probabilistic study of the system which, as is well known, leads to thermodynamics.

If the phase-space average is the same as the average over the original system (in thermodynamics, if the time average equals the phase average) we speak of *ergodicity*. This is usually assumed in statistical mechanics because it offers an easy way to introduce probabilistic methods.

We may, of course, imagine a fictitious infinity of other Earths, but this seems to us highly artificial and unnatural. We have enough problems with our natural Earth; the thought of other possible Earths hurts our national feelings (this is a good nationalism because our nation is the whole Earth!).

Meaningful averaging of assumed homogeneous and isotropic functions over the sphere is about all points of the sphere (homogeneity) and all directions (isotropy). In mathematical terms, it is averaging or integration over *rotation group space*.

Concerning the ensemble, a simple dirty but efficient trick is to let the other members of the ensemble consist of all imaginable rotations of the original function. Then the phase space is again the rotation group and we have the same case as the above-mentioned averaging of the original function over all points and directions. Hence the individual average automatically equals the phase-space average: the functions, considered as random functions in phase space, are *trivially ergodic*.

Ergodic, but not Gaussian, as the following compactness argument shows.

A normal distribution always goes from $-\infty$ to ∞, that is, arbitrarily great values must be allowed. Since any continuous function on the sphere (or on any compact surface) is bounded, there can be no arbitrarily large function values. Thus it cannot be subject to a normal distribution.

However, it may be asymptotically normally distributed if we approximate the sphere, in a certain region, by a plane (which is noncompact). Thus, for many purposes, it is possible practically to assume normal distribution after all, and this is well confirmed by empirical histograms.

Similar considerations may be applied to the irregular terrain, after removing the trends, e.g., by an isostatic reduction. The remaining topography may be considered, with some reservations, as homogeneous and isotropic.

4. Statistics Without Stochastics

Our principal reason for introducing stochastic processes has been to avail ourselves of the corresponding mathematical apparatus and the statistical terminology, such as the concept of covariance functions; this is very convenient, useful and of considerable heuristic value. If it is possible to retain this apparatus while at the same time working with one Earth only, then there is little reason why we should not take "Occam's razor" and cut away all the other fictitious "sample Earths". (Occam: "*Entia non sunt multiplicanda praeter necessitatem*".)

We are thus back to the problem mentioned in Sect. 2, returning to the global statistics of the human population. There are random variations from one human individual to the other. We have regular trends, such as the racial and cultural background, but there are genuinely irregular features left, distributed over the human population and thus over the Earth's surface. This is not completely unlike the surface distribution of gravity anomalies (although the analogy, if pushed too hard, quickly becomes nonsense).

Is it permitted to study the global population statistics at a given time and to calculate various statistical distributions? Everyone will answer this question in the affirmative, although there is only *one* global population. All statistical distributions are simply calculated on the basis of this population.

This is the subject of *descriptive statistics*, which computes relative frequencies, histograms, distribution functions, mean values, variances and covariances only on the basis of the available population. This has the character of a classification of the data and does not necessarily presuppose a "stochastic" behavior.

It appears that the statistics of the gravitational field should be handled similarly. We simply must take seriously the fact that there is only *one* gravitational field, and compute the whole statistics from this one field only.

The appropriate mathematical apparatus for studying the "second–order statistics" (variances and covariances) of the gravitational field is thus Norbert Wiener's "covariance analysis of individual functions" from 1930. This model is implicit in almost all geodetic work in this field.; explicitly it is formulated in [6], to which we refer for mathematical details.

It is now of basic importance that, *formally*, the results of this theory can also be interpreted within the framework of ergodic stochastic processes, as we have seen above. This is, of course, independent of the question whether the anomalous gravitational field is "really" a stochastic process in some physical sense. Probability theory then simply serves to provide a convenient mathematical formalism.

The deeper reason why this formalism can be applied is that, mathematically, both probabilities and relative frequencies satisfy the axioms of *measure theory* according to Kolmogorov. Therefore, if we wish to avoid the term "probability", we might simply speak of "measure". However, probabilistic terminology possesses an attractive intuitive flavor and is thus frequently preferred.

Finally, the present arguments may thus not be irrelevant to earthquake theory and prediction, but I do not know this topic as thoroughly as Volodya Keilis–Borok and his colleagues do.

REFERENCES

1. *Jeffreys H.* Theory of Probability. 3rd ed. Oxford Univ. Press, 1961. 447 p.
2. *Jeffreys H.* Scientific Inference. 3rd ed. Cambridge Univ. Press, 1973. 273 p.
3. *Cressie N.* Statistics for Spatial Data. N. Y.: Wiley, 1993. 900 p.
4. *Moritz H., Sansò F.* A dialogue on collocation. // Boll. di Geodesia e Scienze Affini, 1980. Vol.39. P.49–51.
5. *Lauritzen S.* The probabilistic background of some statistical methods in physical geodesy. Publ. N.48, Copenhagen: Danish Geodetic Inst., 1973. 99 p.
6. *Moritz H.* Advanced Physical Geodesy. Karlsruhe: Wichmann, 1980. 500 p.