ISSN 1973-2201 STATISTICA, anno LXXXIV, n. 2, 2024
DOI: https://doi.org/10.60923/issn.1973-2201/22077

REJOINDER

Laurie Davies[1]
University of Duisburg-Essen, Essen, Germany

GENERALITIES

Here is my translation of Müller (1974) which contains Müller's concept of "distanced rationality":

> *... distanced rationality. By this we mean an attitude to the given, which is not governed by any possible or imputed immanent laws, but which confronts it with draft constructs of the mind in the form of models, hypotheses, working hypotheses, definitions, conclusions, alternatives, analogies, so to speak from a distance, in the manner of partial, provisional, approximate knowledge.*

Müller was a mathematics Professor at Heidelberg, Lutz Dümbgen was a student of his. I once asked Müller why he chose statistics, his reply was that it was the only subject he did not understand. I had the same experience. The statistics course in the Maths Tripos at Cambridge was given by John Kingman. It was the only course I did not understand, in spite of Kingman's excellent didactic qualities. It was based on a previous course by David Lindley, but was, fortunately, non-Bayesian, otherwise I would have been completely lost.

Treating a model as true is the opposite of distanced rationality, far from being distanced, one immerses oneself in it. The Gaussian covariate approach is more distanced. Instead of immersion, the data is confronted with the question as to whether there exist valid linear approximations. The estimation of the unique true true parameter values as accurately as possible is replaced by a search for multiple valid approximations.

With respect to Dümbgen and Davies (2023), I have come to the conclusion that the use of "model-free" is not correct. The concept of approximation in Davies (2014) was largely based on the ability to simulate under the model which, in the case of linear regression, requires a model for the errors. Thus "model-free" means, more precisely, "model-free errors". The errors for the Boston housing data have not, as far as I know,

[1] Corresponding Author. E-mail: laurie.davies@uni-due.de

been modelled. Nevertheless the linear regression is a model as it specifies a linear dependence of the dependent variable on the covariates. Even without a model for the residuals, the F distribution P-values, which are all false, give information about the covariates: covariate 13, median value of owner occupies homes in 1000, has the lowest P-value and therefore the largest reduction in the sum of squared residuals. In Chapter 5 of Huber (2011) Huber points out that "approximate model" is a pleonasm, the two words mean the same thing.

Model selection is important. Statistics offers several procedures for selecting models of which AIC, BIC and MDL are the best known, but, being based on likelihood, they are fundamentally flawed. An approximation is to the data, not to the truth which is much too complicated to be approximated: two samples, whose individuals values differ only by a small amount, should lead to similar conclusions. Consequently the topology of data analysis is weak with few open sets. For real data the topology is generated by the Kolmogorov metric, more generally by Vapnik-Cervonenkis sets. Given a model simulations under the model are generated using the distribution function, for one-dimensional data $F^{-1}(U)$ with $U \sim unif(0,1)$. If F has a density f, then $F(x) = \int_{-\infty}^{x}$, in the other direction $f = D(F)$, where D is the pathologically discontinuous differential operator. Transferred to likelihood, it follows that likelihood is a pathologically discontinuous function of the distribution. For examples see Chapters 1.3 and 11.6 of Davies (2014).

The equality of standard F P-values and Gaussian P-Values is a rather unexpected coincidence which, in a sense, "saves" F P-values; it is irrelevant for Gaussian P-values. Take the Boston housing data. All the standard F P-values are false as the residuals are clearly not i.i.d Gaussian, but, if they are reinterpreted as Gaussian P-values, they are all correct. It goes further. Take any subset and run a regression for this subset. Even if the errors are i.i.d. Gaussian for the full model, all the standard F P-values for this subset are false as some covariates have been omitted. Reinterpret them as Gaussian P-values and they are again all correct. This is very useful as many software packages calculate standard F P-values as a matter of course.

Christian Hennig

I have considerable difficulties understanding Christian Hennig's CH contribution. He may have the same problem with mine. He places great importance on the Data Generating Process DGP behind the data. Take the riboflavin data. The data generating process involve bacteria, Bacillus subtilis. It is extremely complicated and involves the genes of the bacteria, it cannot be modelled. It seems reasonable that certain genes are more actively involved in the production of riboflavin so 4000 are chosen and their gene expression measured during the production of riboflavin. A linear relationship will hopefully indicate which genes are most active even if the relationship is not linear. This is not modelling the data generating process.

CH states that a method, which works well when based on model assumptions, can still be used when the assumptions do not hold. Take the easiest example, the mean and the normal model. What are the assumptions required before calculating the mean? See Chapter 5 of Davies (2014) for a discussion of the location-scale problem. Problems start when statistical inference enters and confidence intervals are calculated and hypotheses tested. CH writes that such theory has been very stimulating. It was for me but in the opposite sense: I was stimulated to replace it by an approximation approach.

Good performance under the model may mislead, and the better the performance the more misleading. This is the case for high dimensional regression. The model used for the lasso is still the basic model for high dimensional regressions, nothing has changed in the last 30 years. Simulations under the model are performed with success, Hastie *et al.* (2020) is a perfect example, and this good performance justifies its application to real data. Not only are the simulations under the model good, the asymptotics are excellent. Under the model the desparsifide lasso is asymptotically optimal if the inverse of the covariance matrix is sparse. All this results in complete failure.

Larry Wasserman

Larry Wasserman's LW approach to approximate models is almost the opposite of mine. There is indeed a true distribution P which generated the data but it is not assumed that P is in the family of distributions $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$. Using a discrepancy D the idea is to find that θ_0 which minimizes $D(P, P_\theta)$, that is the θ for which P_θ is closest to the truth P. Various discrepancies are considered including the L_2 discrepancy defined in terms of Lebesgue densities. In Davies (2014) there are no true models, the topology is that defined by the Kolmogorov metric and the statistical analysis is always based on distribution functions. The two approaches are orthogonal. LW cites Buja *et al.* (2019b) which in turn cites Davies (2014). In the third of the three papers Buja *et al.* (2019a), refers to Davies (2014) as an "excellent book".

Pietro Coretto

I never thought of a Gaussian covariate as a random perturbation of the data. The idea was to compare the covariate of interest with a covariate known to be irrelevant. I suppose the standard calculation of F P-values could be called a perturbation of the data, but I don't think this is done.

I don't like the use of the phrase "null hypothesis". The Gaussian P-values are just calculated, there is no hypothesis testing and consequently none of the problems it causes. I am not quite sure what you mean by "the null random model is an adequate representation of the observed data. There is a concept of a valid approximation which is an approximation with all Gaussian P-values less than the chosen threshold $p0$.

I have never given prediction any thought but I will now do so. Thank you for mentioning it.

Moritz Herrmann and Michael Herrmann

Indoctrination from The Oxford English Dictionary

1.a To imbue with learning, to teach, first reference 1626

1.b To instruct *in* a subject, principle etc, 1656

1.c To imbue *with* a doctrine, idea or opinion *spec* To imbue with Communist ideas 1833

1.d To bring *into* knowledge of something 1841

Google search "indoctrination Cambridge Dictionary"
the process of repeating an idea or belief to someone until they accept it without criticism or question:

1.c and the Cambridge University meaning apply to Statistic Courses.

I suspect that the Catholic Church indoctrinates its priests without bad intent. I think the same applies to statisticians. I thank you for the Breiman reference.

Efthymios Costa and Ioanna Papatsouma

The two articles you cite Zrnic and Fithian (2024) are truth based approaches, confidence regions make no sense, there are no true parameter values or functions to cover. Confidence regions can be replaced by approximation regions, see Chapter 3.8 of Davies (2014). For the time series Example 1 of Zrnic and Fithian (2024) consider the sunspot data. Run

```
library(gausscov)
data(snspt)
x<-fgentrig(3253,500)[[1]]
a<-f1st(log(1+snspt/5),x)
snsptapprx<-5*exp(log(1+snspt/5)-a[[2]])
plot(snspt,col="grey")
lines(snsptapprx,col="red")
```

snsptapprx is an approximation, a confidence band makes no sense, there is no true sunspot function to cover.

On philosophy, "distanced rationality" has influenced me a great deal, especially in statistics.

CLINTIN P. DAVIS-STOBER AND RICHARD D. MOREY

There are parts of this contribution I do not understand. We read " a small p value ... because obviously the parameter does not exist". If I run a regression and a covariate has a small p value with a negative parameter value, then it seems to me that I can deduce that the direction of the effect. This does not require that the model holds. Take the Boston housing data. Then covariate 6 has a P-values of less than 2e-16 and the coefficient 3.18 is positive. I deduce that this covariate has a strong positive effect on house prices, and this despite the fact that the residuals are nowhere near standard Gaussian noise. The covariate is the average number of rooms per dwelling. However one must be careful. Judging by P-values, covariate 12, the proportion of blacks per town, has only the sixth lowest P-value and may be regarded as not important. Run

```
library(gausscov)
data(boston)
bostintr<-fgeninter(boston[,1:13],8) # all interactions of order 8 and less
a<-f1st(boston[,14],bostintr[[1]])
a[[1]]
       [,1]          [,2]          [,3]          [,4]
[1,]  20184 -2.507666e-03  9.177633e-09  4.510093e-14
[2,]  21876  1.778293e-07  9.556210e-13  4.696134e-18
[3,]  66133 -1.533748e-03  6.826126e-08  3.354510e-13
[4,] 181203 -1.000746e-04  8.199808e-06  4.029584e-11
[5,] 190771 -9.388218e-09  6.533263e-65  3.210590e-70
[6,] 191110  2.607536e-05 1.271655e-147 6.249197e-153
[7,]      0  1.675045e+01 9.079230e-142 9.079230e-142

The covariate 191110 has by far the smallest P-value
decode it
bostintr[[2]][191110,]
[1]  6  6  6  6 12  0  0  0
```

This indicates a very strong interaction between the covariates 6 and 12. So covariate 12 turns out to be very important after all.

I would never use Kullback-Leibler which is pathologically discontinuous in the case of Lebesque densities. Robust procedures are always to be used where possible. Chapter 6.1 of Davies (2014) considers the problem of comparing location parameters for different data sets. It is based on the M-functionals of Chapter 5.4 and would seem to me to be more appropriate than the non-robust means.

ALEXANDRE G. PATRIOTA AND ANDREY B. SARMENTO

As far as I understand it, this response is about an inconsistency in the "assumed (revealed?) truth" approach. The Gaussian covariate approach looks only for valid subsets

and here there is no inconsistency. For the example you gave there are no valid subsets if $p0 = 0.01, 0.05$ and only one, x_3, for $p0 = 0.1$.

References

A. Buja, L. Brown, A. Kuchibhotla, R. Berk, E. George, L. Zhao (2019a). *Models as approximations –rejoinder*. Statistical Science, 34, no. 4, pp. 606–620.

A. Buja, L. Brown, A. Kuchibhotla, R. Berk, E. George, L. Zhao (2019b). *Models as approximations I: Consequences illustrated with linear regression*. Statistical Science, 34, no. 4, pp. 523–544.

L. Davies (2014). *Data analysis and approximate models*, vol. 133 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL.

L. Dümbgen, L. Davies (2023). *Connecting model-based and model-free approaches to linear least squares regression*. arXiv e-prints arXiv:1807.09633v4.

T. Hastie, R. Tibshirani, R. Tibshirani (2020). *Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons*. Statistical Science, 35, no. 4, pp. 579–592.

P. J. Huber (2011). *Data Analysis*. Wiley, New Jersey.

D. W. Müller (1974). *Thesen zur Didaktik der Mathematik*. Mathematisch-Physikalische Semesterberichte, Neue Folge, 21, pp. 164–169.

T. Zrnic, W. Fithian (2024). *Locally simultaneous inference*. The Annals of Statistics, 52, no. 3, pp. 1227–1253.