6-2017

How Better Are Predictive Models: Analysis on the Practically Important Example of Robust Interval Uncertainty

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Hung T. Nguyen
New Mexico State University - Main Campus, hunguyen@nmsu.edu

Songsak Sriboonchitta
Chiang Mai University, songsakecon@gmail.com

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Comments:
Technical Report: UTEP-CS-17-46

How Better Are Predictive Models: Analysis on the Practically Important Example of Robust Interval Uncertainty

Vladik Kreinovich, Hung T. Nguyen, Songsak Sriboonchitta, and Olga Kosheleva

Abstract One of the main applications of science and engineering is to predict future value of different quantities of interest. In the traditional statistical approach, we first use observations to estimate the parameters of an appropriate model, and then use the resulting estimates to make predictions. Recently, a relatively new *predictive* approach has been actively promoted, the approach where we make predictions directly from observations. It is known that in general, while the predictive approach requires more computations, it leads to more accurate predictions. In this paper, on the practically important example of robust interval uncertainty, we analyze how more accurate is the predictive approach. Our analysis shows that predictive models are indeed much more accurate: asymptotically, they lead to estimates which are $\sqrt{n}$ more accurate, where n is the number of estimated parameters.

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

Hung T. Nguyen
Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA
and
Faculty of Economics, Chiang Mai University, Chiang Mai 50200 Thailand, e-mail: hunguyen@nmsu.edu

Songsak Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai 50200 Thailand, e-mail: songsakecon@gmail.com

Olga Kosheleva
Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

1 Formulation of the Problem

Predictions are important. One of the main applications of science and engineering is to predict what will happen in the future:

- In science, we are most interesting in predicting what will happen "by itself" – e.g., where the Moon will be a year from now.
- In engineering, we are more interested in what will happen if we apply a certain control strategy – e.g., where a spaceship will be if we apply a certain trajectory correction.

In both science and engineering, prediction is one of the main objectives.

Traditional statistics approach to prediction: estimate then predict. The traditional statistical approach to prediction problems (see, e.g., [6]) is as follows:

- First, we fix a statistical model with unknown parameters. For example, we can assume that the dependence of some quantity y on the quantities $x_1, \ldots, x_n$ is described by a linear dependence $y = a_0 + \sum_{i=1}^{n} a_i \cdot x_i + \varepsilon$, where ε is normally distributed with 0 mean and some standard deviation σ. In this case, the parameters are a_0, a_1, ..., a_n, and σ.
- Then, we use the observations to confirm this model and estimate the values of these parameters.
- After that, we use the model with the estimated values of the parameters to make the corresponding predictions.

Traditional statistical approach to prediction: advantages and limitations. In the traditional approach, when we perform estimations, we do not take into account what exactly characteristic we plan to predict. In the above example, the same estimates for the parameters a_i and σ are used, whether we are trying to predict the future value of the quantity y or whether we are trying to predict a different quantity z that depends on y and on several other quantities.

A natural advantage of this approach is that a computationally intensive parameter estimation part is performed only once, and the resulting estimates can then be used to solve many different prediction problems. In the past, when computations were much slower than now, this was a big advantage: by using pre-computer estimates for the values of the corresponding parameters, we can perform many different predictions fast, without the need to re-do time-consuming parameter estimation part.

With this advantages, come a potential limitation: hopefully, by tailoring parameter estimation to a specific prediction problem, we may able to make more accurate predictions.

Predictive approach. In the past, because of the computer limitations, we had to save on computations, and thus, the traditional approach was, in most cases, all we could afford. However, now computers have become much faster. As a result, in

many practical situations, it has become possible to perform intensive computations in a short period of time.

As a result, taking into account the above disadvantage of the traditional approach, many researchers now advocate to use *predictive* approach to statistics, in which we directly solve the prediction problem – i.e., in other words, on the intermediate step of estimating the parameters, we take into account what exactly quantities we need to predict; see, e.g., [1, 2, 3].

What we do in this paper. There are many examples of successful use of the predictive approach. However, most of these examples remain anecdotal.

In this paper, on a practically important simple example of robust interval uncertainty, we prove a general result showing that predictive models indeed lead to more accurate predictions. Moreover, we provide a numerical measure of accuracy improvement.

2 Robust Interval Uncertainty: A Brief Reminder

Measurement uncertainty. Data processing starts with values that come from measurement or from an expert estimate. Expert estimates are often important, but of course, a measuring instrument provides much more data than an expert. As a result, the overwhelming majority of data values come from measurements.

With the exception of simplest cases like counting number of people in a small group, measurement are not 100% accurate: the measurement result $\widetilde{x}$ is, in general, different from the actual (unknown) value of the corresponding quantity. In other words, in general, we have a non-zero *measurement error* $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$.

What do we know about measurement uncertainty: case when we know the probability distribution and case of robust interval uncertainty. In some situations, we know the probability distribution of the measurement error; for example, in many practical cases, we know that the measurement error is normally distributed, with 0 mean and known standard deviation σ.

However, in many practical situations, the only information that we have about the measurement error Δx is the upper bound Δ on its absolute value – the bound provided by the manufacturer of the measuring instrument; see, e.g., [5].

In other words, we only know that the probability distribution of the measurement error Δx is located on the interval $[-\Delta, \Delta]$, but we do not have any other information about the probability distribution. Such *interval uncertainty* is a particular case of the general *robust statistics*; see, e.g., [4].

Why cannot we always get this additional information? To get information about $\Delta x = \widetilde{x} - x$, we need to have information about the actual value x. In many practical situations, this is possible. Namely, in addition to the current measuring instrument (MI), we often also have a much more accurate ("standard") MI, so much more accurate that the corresponding measurement error can be safely ignored in com-

parison with the measurement error of our MI, and thus, the results of using the standard MI can be taken as the actual values.

In such a situation, we can find the probability distribution for the measurement error Δx if, for each of several quantities, we measure this quantity both by using the current MI and by using the standard MI. The difference between the two measurement results is a good approximation to the corresponding measurement error. Thus, the collection of such differences is a sample from the desired probability distribution for Δx. Based on this sample, we can find the corresponding probability distribution.

In many situations, however, our MI is already state-of-the-art, no more-accurate standard MI is possible. For example, in fundamental science, when we perform state-of-the-art measurements, we use state-of-the-art measuring instruments. For a billion-dollar project like space telescope or particle super-collider, the best MI are used. In this case, it is not possible to apply the above technique, so the best we can do is to use the bound Δ on the measurement error.

Another frequent case when we have to use Δ is the case of routine manufacturing. In this case, theoretically, we can calibrate every sensor, but sensors are cheap and calibrating them costs a lot – since it means using expensive state-of-the-art standard MIs. In routine manufacturing, such a calibration is just not financially possible – and not needed. For example, a simple thermometer for measuring a body temperature is reasonable cheap. If we had to calibrate each thermometer, it would become an order of magnitude more expensive – and what is the purpose? Honestly, all we need to know is whether a patient has a fever and, if yes, how severe, but the difference between, say 38.1 and 38.2 will not result in any changes in medical diagnosis or treatment.

Robust interval uncertainty is what we consider in this paper. In view of the practical importance, in this paper, we consider the case of robust interval uncertainty.

3 Comparing Predictive and Traditional Statistics on the Example of Robust Interval Uncertainty: Analysis of the Problem

Let us describe the traditional approach in precise terms. Let y denote the quantity that we would like to predict.

To predict a quantity, we need to know the relation between this future quantity y and certain "estimate-able" quantities $x_1, \ldots, x_n$. Then, to predict y, we:

- estimate the quantities $x_1, \ldots, x_n$ based on the measurement results, and then
- use these estimates and the known relation between y and x_i to predict the desired future value y.

Let us denote the corresponding relation between y and x_i by $y = f(x_1, \ldots, x_n)$. Let us denote the measurement-based estimates for the quantities x_i by $\widetilde{x}_i$. In these terms,

after generating these estimates, we get the following prediction for y:

$$\widetilde{y} \stackrel{\text{def}}{=} f(\widetilde{x}_1, \ldots, \widetilde{x}_n).$$

The quantities x_i are estimated based on measurement results. Let $v_1, \ldots, v_N$ denote all the quantities whose measurement results are used to estimate the quantities x_i. This estimation is based on the known relation between x_i and v_j. Let us denote this relation by $x_i = g_i(v_1, \ldots, v_N)$, and let us denote the result of measuring each quantity v_j by $\widetilde{v}_j$. In these terms, the process of computing estimates $\widetilde{x}_i$ for the quantities x_i consists of the following two steps:

- first, we measure the quantities $v_1, \ldots, v_N$;
- then, the results $\widetilde{v}_1, \ldots, \widetilde{v}_N$ of measuring these quantities are used to produce the estimates $\widetilde{x}_i = g_i(\widetilde{v}_1, \ldots, \widetilde{v}_N)$.

Overall, the traditional approach takes the following form:

- first, we measure the quantities $v_1, \ldots, v_N$;
- then, the results $\widetilde{v}_1, \ldots, \widetilde{v}_N$ of measuring these quantities are used to produce the estimates $\widetilde{x}_i = g_i(\widetilde{v}_1, \ldots, \widetilde{v}_N)$;
- finally, we use the estimates $\widetilde{x}_i$ to compute the corresponding prediction

$$\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n).$$

How will predictive approach look in these terms. The predictive approach means that, instead of first estimating the parameters x_i and then using these parameters to predict y, we predict y based directly on the measurement results v_j.

To make such a prediction, we need to know the relation between the predicted quantity y and the measurement results. Since we know that $y = f(x_1, \ldots, x_n)$ and that $x_i = g_i(v_1, \ldots, v_N)$, we thus conclude that $y = F(v_1, \ldots, v_N)$, where we denoted

$$F(v_1, \ldots, v_N) \stackrel{\text{def}}{=} f(g_1(v_1, \ldots, v_N), \ldots, g_n(v_1, \ldots, v_N)).$$

In these terms, the predictive approach to statistics takes the following form:

- first, we measure the quantities $v_1, \ldots, v_N$;
- then, the results $\widetilde{v}_1, \ldots, \widetilde{x}_N$ of measuring these quantities are used to produce the prediction $\widetilde{y} = F(\widetilde{v}_1, \ldots, \widetilde{v}_N)$.

How accurate are these estimates and predictions? We are interested in the accuracy of the corresponding estimates and predictions.

For each estimated quantity x_i, the estimation error Δx_i is naturally defined as the difference $\widetilde{x}_i - x_i$ between the estimate $\widetilde{x}_i = g_i(\widetilde{v}_1, \ldots, \widetilde{v}_N)$ and the actual value $x_i = g_i(v_1, \ldots, v_N)$, i.e., the value that we would have got if we knew the exact values v_j of the measured quantities v_j. Similarly, for the prediction, the prediction error Δx_i is naturally defined as the difference $\widetilde{y} - y$ between the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ and the actual value $y = g_i(x_1, \ldots, x_n)$, i.e., the value that we would have got if we knew the exact values x_i of the estimated quantities x_i.

Measurements are usually reasonably accurate, which means that the measurement errors Δv_j are reasonably small. So, we can substitute the formula $v_j = \widetilde{v}_j - \Delta v_j$, expand the resulting expression for

$$\Delta x_i = g_i(\widetilde{v}_1, \ldots, \widetilde{v}_N) - g_i(v_1, \ldots, v_N) =$$

$$g_i(\widetilde{v}_1, \ldots, \widetilde{v}_N) - g_i(\widetilde{v}_1 - \Delta v_1, \ldots, \widetilde{v}_N - \Delta v_N)$$

in Taylor series, and keep only linear terms in this expansion. As a result, we get the following formula:

$$\Delta x_i = \sum_{i=1}^{N} g_{ij} \cdot \Delta v_j,$$

where we denoted $g_{ij} \stackrel{\text{def}}{=} \dfrac{\partial g_i}{\partial v_j}$.

What can we conclude about the value Δx_i? The only thing we know about each of the measurement errors Δv_j is that this measurement error can take any value from the interval $[-\Delta_j, \Delta_j]$. The above sum attains its largest possible value when each of the terms attains its largest value.

- when $g_{ij} \geq 0$, the term $g_{ij} \cdot \Delta v_j$ is an increasing function of Δv_j, so its maximum is attained when Δv_j attains its largest possible value $\Delta v_j = \Delta_j$; the resulting largest value of this term is $g_{ij} \cdot \Delta_j$;
- when $g_{ij} < 0$, the term $g_{ij} \cdot \Delta v_j$ is a decreasing function of Δv_j, so its maximum is attained when Δv_j attains its smallest possible value $\Delta v_j = -\Delta_j$; the resulting largest value of this term is $-g_{ij} \cdot \Delta_j$.

In both cases, the largest possible value of the term is equal to $|g_{ij}| \cdot \Delta_j$. Thus, the largest possible value Δ_i^x of Δx_i is equal to

$$\Delta_i^x = \sum_{j=1}^{N} |g_{ij}| \cdot \Delta_j. \tag{1}$$

One can easily check that the smallest possible value of Δx_i is equal to $-\Delta_i^x$. Thus, possible values of Δx_i form an interval $[-\Delta_i^x, \Delta_i^x]$.

Similarly, based on the estimates $\widetilde{x}_i$ and bounds Δ_i^x on the estimation errors, we can conclude that the possible values of the prediction error lie in the interval $[-\Delta, \Delta]$, where

$$\Delta = \sum_{i=1}^{n} |f_i| \cdot \Delta_i^x, \tag{2}$$

and we denoted $f_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$.

Alternatively, if we use the function $F(v_1, \ldots, v_N)$ to directly predict the value y from the measurement results, we conclude that the possible value of the prediction error lie in the interval $[-\delta, \delta]$, where

$$\delta = \sum_{j=1}^{N} |F_j| \cdot \Delta_j, \tag{3}$$

and we denoted $F_j \stackrel{\text{def}}{=} \dfrac{\partial F}{\partial v_j}$.

Preliminary conclusion. Depending on whether we consider the traditional statistical approach or the predictive approach, we get the same estimate $\widetilde{y}$ for the predicted quantity y. However, for the accuracy Δy, we have, in general, different bounds:

- if we use the traditional approach, then we get the bound Δ as described by the formulas (1) and (2);
- alternatively, if we use the predictive approach, we get the bound δ as described by the formula (3).

Comparing the two bounds. One can see that δ is the actual bound: in principle, the value δ can be attained if we take appropriate values of $\Delta v_j = \Delta_j \cdot \text{sign}(F_j)$.

Since all possible values of Δy also lie in the interval $[-\Delta, \Delta]$, the value δ also lies in this interval, thus the estimate δ coming from the predictive approach is smaller than or equal to the traditional estimate Δ. But is is better? Let us compare the two expressions.

If we substitute the expression (1) into the formula (2), we conclude that

$$\Delta = \sum_{i=1}^{n} \left(|f_i| \cdot \left(\sum_{j=1}^{N} |g_{ij}| \cdot \Delta_j \right) \right),$$

i.e., equivalently,

$$\Delta = \sum_{j=1}^{n} C_j, \tag{4}$$

where we denoted

$$C_j \stackrel{\text{def}}{=} \sum_{i=1}^{n} |f_i| \cdot |g_{ij}| \cdot \Delta_j.$$

Since $|a \cdot b| = |a| \cdot |b|$ and $\Delta_j > 0$, we can thus conclude that

$$C_j = \sum_{i=1}^{N} |c_{ij}|, \tag{5}$$

where we denoted $c_{ij} \stackrel{\text{def}}{=} f_i \cdot g_{ij} \cdot \Delta_j$.

On the other hand, the formula (3) takes the form

$$\delta = \sum_{j=1}^{n} c_j, \tag{6}$$

where $c_j \stackrel{\text{def}}{=} |F_j| \cdot \Delta_j$. By using the chain rule, we conclude that the derivative F_j of the composition function $F(v_1, \ldots, v_N)$ takes the form $F_j = \sum_{i=1}^{n} f_i \cdot g_{ij}$. Thus, the coefficient c_j in the formula (6) has the form

$$c_j = \left| \sum_{i=1}^{n} f_i \cdot g_{ij} \right| \cdot \Delta_j,$$

i.e., equivalently,

$$c_j = \left| \sum_{i=1}^{N} c_{ij} \right|. \tag{7}$$

By comparing formulas (4)-(5) with formulas (6)-(7), we can see that indeed $\delta \le \Delta$: indeed, since $|a+b| \le |a| + |b|$, we have

$$c_j = \left| \sum_{i=1}^{n} c_{ij} \right| \le \sum_{i=1}^{n} |c_{ij}| = C_j$$

and thus, indeed

$$\delta = \sum_{j=1}^{N} c_j \le \sum_{j=1}^{N} C_j = \Delta.$$

Is δ smaller? If yes, how smaller? To answer these equations, let us take into account that, in principle, each term $c_{ij} = f_i \cdot g_{ij} \cdot \Delta_j$ can take any real value, positive and negative. A priori, we do not have have any reason to believe that positive values will be more frequent than negative ones, so it is reasonable to assume that the mean value of each such term is 0. Again, there is no reason to assume that the values c_{ij} are different, so it makes sense to assume that all these values are identically distributed. Finally, there is no reason to believe that there is correlation between different values, so its makes to consider them to be independent.

Under these assumptions, for large n, the sum $\sum_{i=1}^{n} c_{ij}$ is normally distributed, with 0 mean and variance which is n times larger than the variance σ^2 of the original distribution of c_{ij}. Thus, the means value of the absolute value c_j of this sum is proportional to its standard deviation $\sigma \cdot \sqrt{n}$.

On the other hand, the expected value μ of each term $|c_{ij}|$ is positive, thus, the expected value of the sum $C_j = \sum_{i=1}^{n} |c_{ij}|$ of n such independent terms is equal to $\mu \cdot n$.

For large n, $\mu \cdot n \gg \sigma \cdot \sqrt{n}$. Thus, we arrive at the following conclusion.

4 Conclusion

In this paper, we compare:

- the traditional statistical approach, in which we first use the observations to estimate the values of the parameters and then use these estimates for prediction, and
- the predictive approach to statistics, in which we make predictions directly from observations.

We make this comparison on the example of the practically important case of robust interval uncertainty, when the only information that we have about the corresponding measurement error is the upper bound provided by the manufacturer of the corresponding measurement instrument.

It turns out that while predictive techniques require more computations, they result in much more accurate estimates: asymptotically, $\sqrt{n}$ times more accurate, where n is the total number of parameters estimated in the traditional approach.

Acknowledgments

We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand. This work was also supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

References

1. W. Briggs, *Uncertainty: The Soul of Modeling, Probability & Statistics*, Springer Verlag, Cham, Switzerland, 2016.
2. J. Dutta, "On predictive evaluation of econometric models", *International Economic Review*, 1980, Vol. 21, No. 2, pp. 379–390.
3. T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilsitic forecasts, calibration, and sharpness', *Journal of the Royal Statistical Society, Part B*, 2007, Vol. 69, Part 2, pp. 243–268.
4. P. J. Huber and E. M. Ronchetti, Robust Statistics, Wiley, Hoboken, New Jersey, 2009.
5. S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.
6. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2011.