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

Vladik Kreinovich[1], Hung T. Nguyen[2,3],
Songsak Sriboonchitta[3], and Olga Kosheleva[1]
[1]University of Texas at El Paso, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu
[2]Department of Mathematics, New Mexico State University
Las Cuces, New Mexico 88003, USA, hunguyen@nmsu.edu
[3]Faculty of Economics, Chiang Mai University
Chiang Mai 50200 Thailand, songsakecon@gmail.com

1. 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".
- *Examples:* 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.
- *Example:* where a spaceship will be if we apply a certain trajectory correction?
- In both science and engineering, prediction is one of the main objectives.

2. Traditional Statistics Approach to Prediction: Estimate then Predict

- In the traditional statistical approach, we first fix a statistical model with unknown parameters.
- For example, we can assume that the dependence of y on $x_1, \ldots, x_n$ is linear:

$$y = a_0 + \sum_{i=1}^{n} a_i \cdot x_i + \varepsilon, \quad \varepsilon \sim N(0, \sigma).$$

- 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 parameters to make the corresponding predictions.

3. 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.
- Advantage of this approach: a computationally intensive parameter estimation part is performed only once.
- In the past, when computations were much slower than now, this was a big advantage.
- With this advantages, come a potential limitation:
  - hopefully, by tailoring parameter estimation to a specific prediction problem,
  - we may be able to make more accurate predictions.

4. Predictive Approach

- In the past, because of the computer limitations, we had to save on computations.
- Thus, the traditional approach was, in most cases, all we could afford.
- However, now computers have become much faster.
- As a result, it has become possible to perform intensive computations in a short period of time.
- So, we can directly solve the prediction problem.
- In other words:
  - on the intermediate step of estimating the parameters,
  - we can take into account what exactly quantities we need to predict.

5. What We Do in This Talk

- There are many examples of successful use of the predictive approach.
- However, most of these examples remain anecdotal.
- In this talk:
  - 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.

6. Measurement Uncertainty: Reminder

- Data processing starts with values that come from measurements.
- Measurement are not 100% accurate:
  - the measurement result $\widetilde{x}$ is, in general, different from
  - the actual (unknown) value x of the corresponding quantity.
- In other words, in general, we have a non-zero *measurement error* $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$.
- In some situations, we know the probability distribution of the measurement error.
- For example, we often that the Δx is normally distributed, with 0 mean and known st. dev. σ.

7. Robust Interval Uncertainty

- However, often, the only information that we have about Δx is the upper bound Δ: $|\Delta x| \le \Delta$.
- This bound is provided by the manufacturer of the measuring instrument.
- In other words:
  - we only know that the probability distribution of the measurement error Δx is located on $[-\Delta, \Delta]$,
  - but we do not have any other information about the probability distribution.
- Such *interval uncertainty* is a particular case of the general *robust statistics*.
- 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.

8. Robust Interval Uncertainty (cont-d)

- In many practical situations, this is possible; namely:
  - in addition to our 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,
  - and thus, the results of using the standard MI can be taken as the actual values.
- We can then find the prob. distribution for Δx if we measure quantities by both our MI and standard MI.
- In many situations, however, our MI is already state-of-the-art, no more-accurate standard MI is possible.

9. Robust Interval Uncertainty (cont-d)

- For example, in fundamental science, 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.
- 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 standard MIs.
- In view of the practical importance, in this talk, we consider the case of robust interval uncertainty.

10. Analysis of the Problem

- Let y denote the quantity that we would like to predict.
- To predict the desired quantity y, we need to know the relation between y and easier-to-estimate quantities $x_1, \ldots, x_n$.
- Then, to predict y, we:
  - compute estimates $\widetilde{x}_i$ for x_i based on the measurement results, and then
  - use these estimates $\widetilde{x}_i$ and the known relation between $y = f(x_1, \ldots, x_n)$ to get a prediction for y:

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

- Let $v_1, \ldots, v_N$ denote all the quantities whose measurement results are used to estimate the quantities x_i.
- The estimation of x_i is based on the known relation between x_i and v_j: $x_i = g_i(v_1, \ldots, v_N)$.

11. Analysis of the Problem (cont-d)

- The estimation of x_i is based on the known relation between x_i and v_j: $x_i = g_i(v_1, \ldots, v_N)$.
- So, $\widetilde{x}_i = g_i(\widetilde{v}_1, \ldots, \widetilde{v}_N)$.
- 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

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

12. 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: $x_i = g_i(v_1, \ldots, v_N)$, where

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

- 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

$$\widetilde{y} = F(\widetilde{v}_1, \ldots, \widetilde{v}_N).$$

13. How Accurate Are These Estimates?

- Measurements are usually reasonably accurate, so the measurement errors Δv_j are reasonably small.
- Thus, we ignore terms quadratic in Δv_j:

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

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

$$\sum_{i=1}^{N} g_{ij} \cdot \Delta v_j, \text{ where } g_{ij} \stackrel{\text{def}}{=} \frac{\partial g_i}{\partial v_j}.$$

- This 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 is the largest: $\Delta v_j = \Delta_j$.

14. How Accurate Are These Estimates (cont-d)

- 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 is the smallest: $\Delta v_j = -\Delta_j$.
- 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

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

15. How Accurate Are These Estimates (cont-d)

- 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, we can conclude that the possible values of the prediction error lie in the interval $[-\Delta, \Delta]$, where

$$\Delta = \sum_{i=1}^{n} |f_i| \cdot \Delta_i^x, \text{ where } f_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}.$$

- Alternatively, if we use the function $F(v_1, \ldots, v_N)$ to directly predict y, we get $\Delta y \in [-\delta, \delta]$, where

$$\delta = \sum_{j=1}^{N} |F_j| \cdot \Delta_j, \text{ and } F_j \stackrel{\text{def}}{=} \frac{\partial F}{\partial v_j}.$$

16. Comparison of Two Approaches

- *Traditional:* $\Delta = \sum_{i=1}^{n} |f_i| \cdot \Delta_i^x$, where $\Delta_i^x = \sum_{j=1}^{N} |g_{ij}| \cdot \Delta_j$, so $\Delta = \sum_{j=1}^{n} C_j$, where $C_j = \sum_{i=1}^{n} |f_i| \cdot |g_{ij}| \cdot \Delta_j$.

- *Predictive:* $\delta = \sum_{j=1}^{N} |F_j| \cdot \Delta_j$, where $F_j = \sum_{i=1}^{n} f_i \cdot g_{ij}$, so $\delta = \sum_{j=1}^{n} c_j$, where $c_j \stackrel{\text{def}}{=} \left|\sum_{i=1}^{n} f_i \cdot g_{ij}\right| \cdot \Delta_j$.

- So, $\Delta = \sum_{j=1}^{n} C_j$ and $\delta = \sum_{j=1}^{n} c_j$, where $C_j = \sum_{i=1}^{N} |c_{ij}|$ and $c_j = \left|\sum_{i=1}^{N} c_{ij}\right|$.

- $|a + b| \le |a| + |b|$, so $c_j = \left|\sum_{i=1}^{N} c_{ij}\right| \le \sum_{i=1}^{N} |c_{ij}| = C_j$: predictive approach is more accurate.

17. How More Accurate?

- In principle, each term $c_{ij} = f_i \cdot g_{ij} \cdot \Delta_j$ can take any real value, positive and negative.
- 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 distributions of 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, it makes to consider them to be independent.

18. How More Accurate (cont-d)

- Under these assumptions, for large n,
  - the sum $\sum\limits_{i=1}^{n} c_{ij}$ is normally distributed,
  - with 0 mean and variance which is n times larger than $\sigma^2 \stackrel{\text{def}}{=} V[c_{ij}]$.
- Thus, the mean 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 $C_j = \sum\limits_{i=1}^{n} |c_{ij}|$ of n such independent terms is equal to $\mu \cdot n$.
- For large n, we have $\mu \cdot n \gg \sigma \cdot \sqrt{n}$.
- Thus, te predictive approach is $\sqrt{n}$ times more accurate.

19. What We Did: Summary

- In this talk, we compare:
  - the traditional statistical approach, in which:
    * we first use the observations to estimate the values of the parameters
    * and then we 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 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.

20. Conclusion

- Predictive techniques require more computations.
- However, 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.

21. Acknowledgments

- We acknowledge the support of the Center of Excellence in Econometrics, Chiang Mai Univ., Thailand.
- This work was also supported in part:
  - by the National Science Foundation grants HRD-0734825, HRD-1242122, and DUE-0926721, and
  - by an award from Prudential Foundation.