INSTITUTE FOR DEFENSE ANALYSES

Quantifying Forecast Uncertainty

John W. Dennis
Evan Miyakawa
James Bishop
Alan B. Gelder

June 2022
Approved for public release;
distribution is unlimited.
IDA Document NS D-33131
Log: H 22-000247

INSTITUTE FOR DEFENSE ANALYSES
730 E. Glebe Road
Alexandria, Virginia 22305

About this Publication
The work was conducted by the Institute for Defense Analyses (IDA) under CRP C6614.

For More Information:
Dr. Alan B. Gelder, Project Leader
agelder@ida.org, 703-845-6879
Dr. Katherine M. Sixt, Director (Interim), SFRD
ksixt@ida.org, 703-575-6695

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-33131

Quantifying Forecast Uncertainty

John W. Dennis
Evan Miyakawa
James Bishop
Alan B. Gelder

This page is intentionally blank.

Executive Summary

Government sponsors are often interested in predicting attrition rates for service members at future time points. The Institute for Defense Analyses developed a tool, the Finite-Interval Forecasting Engine (FIFE), to produce point forecasts of attrition rates using neural networks and gradient boosted trees. Point forecasts are most commonly considered in isolation, but while point forecasts are useful, they only provide partial information regarding the future distribution of the quantity of interest and no information regarding the uncertainty in the forecast.

We discuss methods for quantifying uncertainty in these survival forecasts. Point estimates for future values of interest can be close to the truth, but they are always subject to uncertainty. In some classification prediction problems, future realized data can be classified incorrectly with high probability. While FIFE currently uses advanced approaches for achieving high-quality forecasts, robust methods for quantifying uncertainty in these predictions have not been implemented. Use of distributional forecasts or prediction intervals around point estimates can facilitate understanding of the uncertainty associated with these predictions when used appropriately.

We define relevant terminology in relation to prediction uncertainty and address how these terms differ based on field of study, and we discuss the properties of prediction intervals and describe the specifics of our objective in adding methods to FIFE that can produce these intervals. Our literature review investigates differing approaches to quantifying forecast uncertainty, which includes generic methods and use-specific methods; we implement some of these approaches in FIFE. Among those implemented methods, we find that generic approaches can often be too conservative in the sense that they provide intervals that are too wide, and use-specific methods can be misleading in the sense that they provide intervals that do not attain nominal coverage. In some situations, we find that these intervals can be severely distorted. We discuss the performance of these methods and suggest improvements.

This page is intentionally blank.

IDA

Quantifying Forecast Uncertainty

WEAI 2022

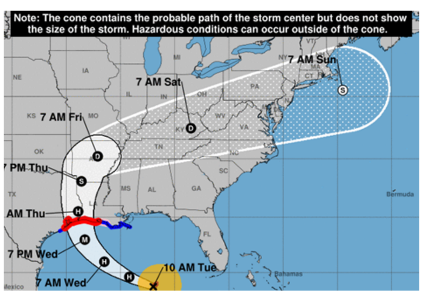
John W. Dennis
jdennis@ida.org

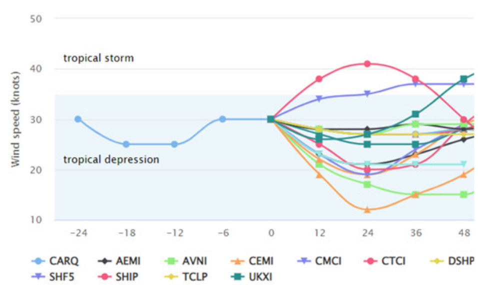
with Evan Miyakawa, Alan Gelder, & James Bishop

June 2022

IDA | Uncertainty?

- All forecasts have uncertainty





Data Uncertainty

$\hat{y}_t = f_t(\hat{\theta}) \pm error$

Approximation Uncertainty

$\hat{y}_{1,t} = f_{1,t}(\hat{\theta})$ $\hat{y}_{2,t} = f_{2,t}(\hat{\theta})$ $\hat{y}_{3,t} = f_{3,t}(\hat{\theta})$

Epistemic Uncertainty: $f \in ???$

2

Introduction - Context

DOD planning requires forecasting

IDA's Finite Interval Forecasting Engine (FIFE) is a tool to forecast categorical outcomes

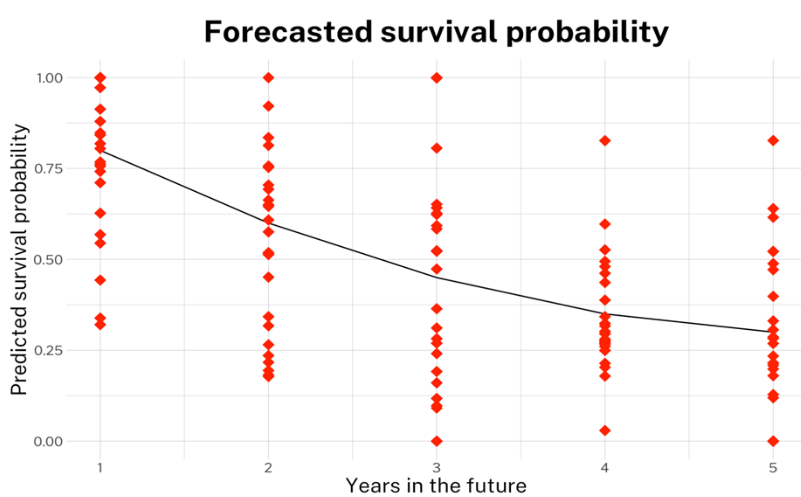
- Often used to predict when an individual leaves service
- Uses gradient boosted tree and neural network modelers
- Primary output: individual level probability of exit
- Can use this to get items of interest like expected count of exits for a group of individuals

- *Quantifying uncertainty in forecasts can help sponsors make better decisions*

3

Introduction - FIFE

How do we quantify uncertainty in FIFE Predictions?



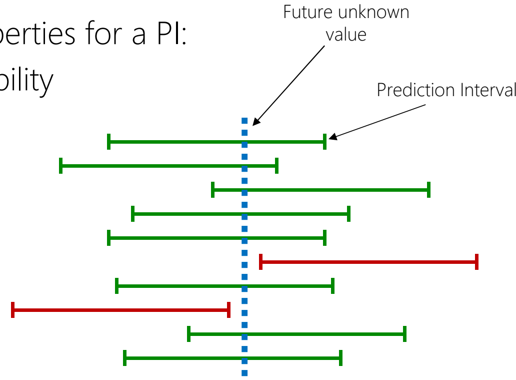
4

Introduction – Prediction Intervals

A **Prediction Interval** (PI) is a set of bounds around a prediction for a future unobserved value

There are two desirable properties for a PI:

1. Bigger Coverage Probability
2. Smaller Interval Width



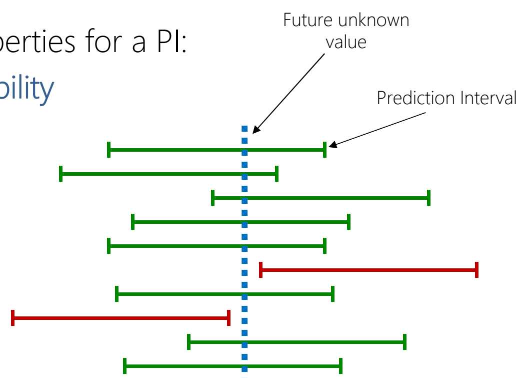
Introduction – Prediction Intervals

A **Prediction Interval** is a set of bounds around a prediction for a future unobserved value.

There are two desirable properties for a PI:

1. Bigger Coverage Probability
2. Smaller Interval Width

Coverage Probability = 8/10



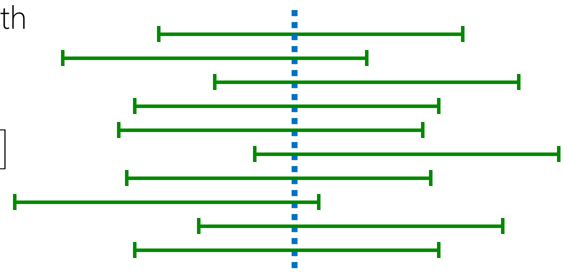
IDA | Introduction – Prediction Intervals

A **Prediction Interval** is a set of bounds around a prediction for a future unobserved value

There are two desirable properties for a PI:

1. Bigger Coverage Probability
2. Smaller Interval Width

Coverage Probability = 10/10



7

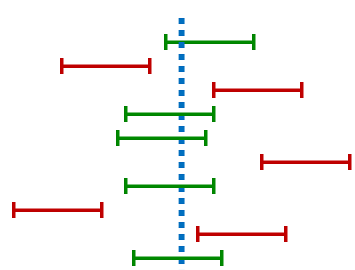
IDA | Introduction – Prediction Intervals

A **Prediction Interval** is a set of bounds around a prediction for a future unobserved value, based on a given confidence level

There are two desirable properties for a PI:

1. Bigger Coverage Probability
2. Smaller Interval Width

Coverage Probability = 5/10



8

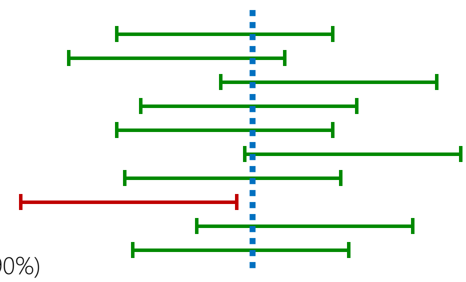
IDA | Introduction – Prediction Intervals

A **Prediction Interval** is a set of bounds around a prediction for a future unobserved value, based on a given confidence level

There are two desirable properties for a PI:

1. Bigger Coverage Probability
2. Smaller Interval Width

Coverage Probability = 9/10



Procedure:

1. Choose a confidence level (90%)
2. Get the tightest intervals that have at least 90% coverage probability

9

Prediction Interval Methods

IDA 10

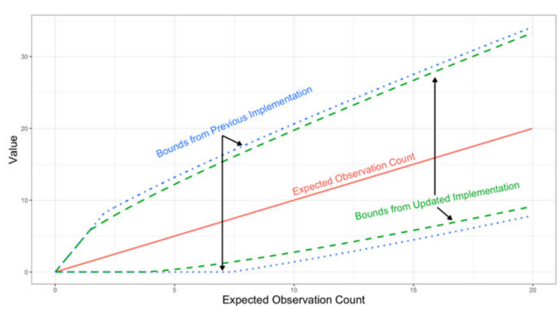
IDA | Prediction Interval Methods - Overview

1. Generic Methods
   - Do not rely on underlying algorithm
   - Can be applied to any predictions with our data type
   - Often conservative

2. Use-Specific Methods
   - Uses underlying modeling framework
   - Vary based on type of algorithm
   - Often tighter intervals

IDA | PI Generic Methods – Chernoff Bounds

- Start with a binomial random variable
- Use Chebyshev's inequality to derive probabilistic bounds
- Can be used for the expected count of exits for a group of individuals at a future time point



- Pointwise, Chebyshev, not using specific information => overly conservative

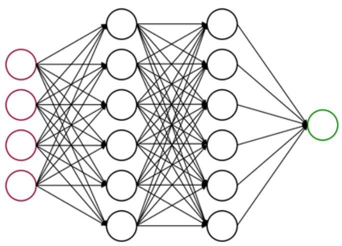
PI Use-Specific Methods

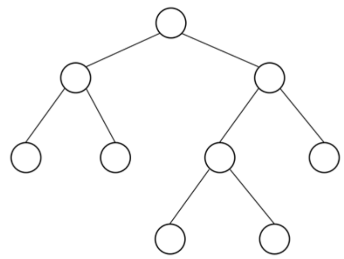
Specific Method = your favorite algorithm or modeling framework

Many algorithms, sometimes many methods per algorithm

Today: cover 2 methods

- Neural Networks: MC Dropout
- Gradient Boosted Trees: Stochastic Gradient Langevin Boosting

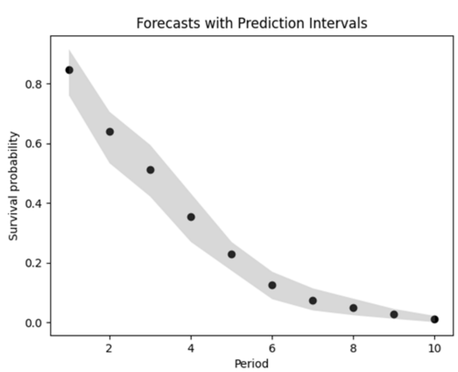




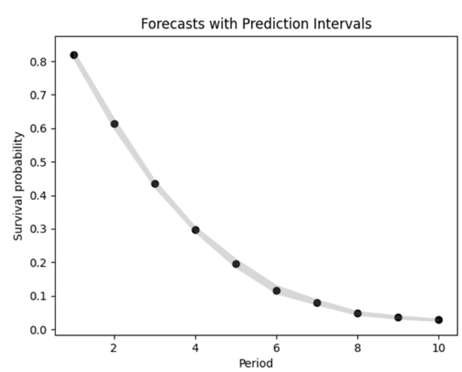
13

Adaptations

Naïve Ensemble 95% PIs



MC Dropout (Neural Network)

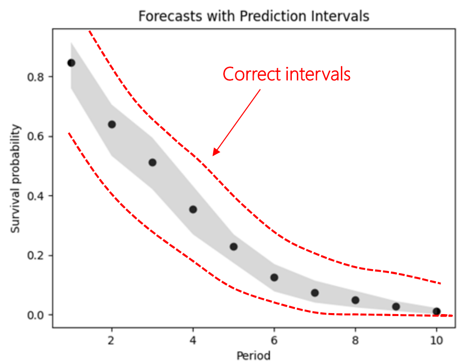


SGLB (GB Tree)

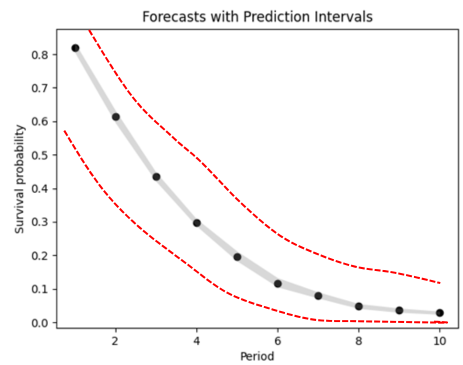
14

IDA | PI Use-Specific Methods

Ensemble PIs needed for 95% Coverage



MC Dropout (Neural Network)
PI Coverage Probability: 48%

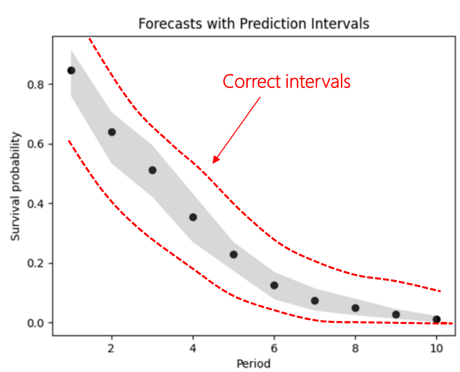


SGLB (GB Tree)
PI Coverage Probability: 2%

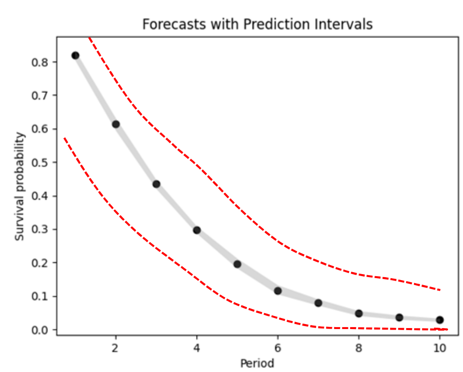
15

IDA | PI Use-Specific Methods

Ensemble PIs needed for 95% Coverage



MC Dropout (Neural Network)
PI Coverage Probability: 48%



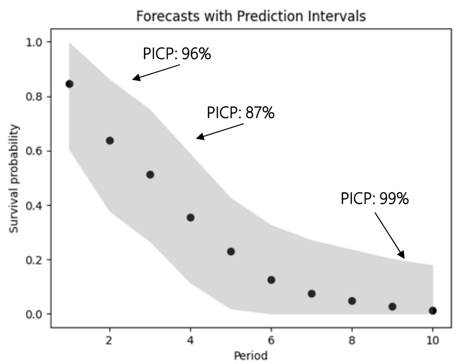
SGLB (GB Tree)
PI Coverage Probability: 2%

Need to Inflate predictive variance to account for the underlying uncertainty

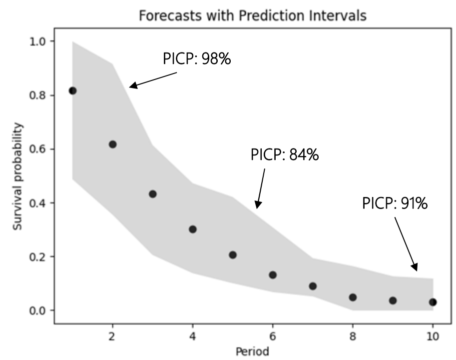
16

IDA | Adaptations

Ensemble PIs with inflated variance



MC Dropout (Neural Network)
PI Coverage Probability: 95%



SGLB (GB Tree)
PI Coverage Probability: 95%

17

Conclusion and Future Work

IDA 18

IDA | Conclusion

DOD forecasts commonly focus only on point estimates

- The uncertainty regarding those forecasts is important
- Many methods are available to quantify that uncertainty

Future work

- Investigate other PI methods
- Update hyperparameter tuning method
- Investigate heterogeneity across forecast horizons

FIFE updates:

- Updated Chernoff Bound PIs for sums of counts
- PIs implemented for Gradient Boosted Tree and Neural Network modelers in FIFE

FIFE Github: https://github.com/IDA-HumanCapital/fife feature/prediction_intervals branch

19

IDA

J. W. Dennis
jdennis@ida.org

References

Chatfield, Chris. 1993. "Calculating Interval Forecasts." *Journal of Business & Economic Statistics* 11 (2): 121–35. https://doi.org/10.2307/1391361.

Gal, Yarin. n.d. "What My Deep Model Doesn't Know... Yarin Gal - Blog Cambridge Machine Learning Group." Yarin Gal. Accessed July 19, 2021. http://mlg.eng.cam.ac.uk/yarin/blog_3d801aa532c1ce.html.

Gal, Yarin, and Zoubin Ghahramani. 2016. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning." International Conference on Machine Learning, 1050–59. PMLR. http://proceedings.mlr.press/v48/gal16.html.

Goemans, Michel. 2015. "Chernoff bounds, and some applications." MIT 18.310 Lecture Notes..https://math.mit.edu/~goemans/18310S15/chernoffnotes.pdf.

Huang, Jing, and Matthew Perry. 2016. "A Semi-Empirical Approach Using Gradient Boosting and k-Nearest Neighbors Regression for GEFCom2014 Probabilistic Solar Power Forecasting." *International Journal of Forecasting* 32 (3): 1081–86. https://doi.org/10.1016/j.ijforecast.2015.11.002.

Hwang, J. T. Gene, and A. Adam Ding. 1997. "Prediction Intervals for Artificial Neural Networks." *Journal of the American Statistical Association* 92 (438): 748–57. https://doi.org/10.1080/01621459.1997.10474027.

Hyndman, Rob. 2013. "The Difference Between Prediction Intervals and Confidence Intervals Rob J Hyndman." Hyndsight Blog. Rob J Hyndman. https://robjhyndman.com/hyndsight/ intervals/.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/ file/6449f44a102fde848669bdd9eb6b76faPaper.pdf.

Khosravi, Abbas, Saeid Nahavandi, Doug Creighton, and Amir F. Atiya. 2011b. "Lower Upper Bound Estimation Method for Construction of Neural Network-Based Prediction Intervals." *IEEE Transactions on Neural Networks* 22 (3): 337–46. https://doi.org/10.1109/TNN.2010.2096824.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. 2017. "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles." arXiv:1612.01474 [Cs, Stat], November. http://arxiv.org/abs/1612.01474.

Malinin, Andrey, Liudmila Prokhorenkova, and Aleksei Ustimenko. 2021. "Uncertainty in Gradient Boosting via Ensembles." arXiv:2006.10562 [Cs, Stat], April. http://arxiv.org/abs/2006.10562.

Miyakawa, Evan T., Dennis, John W., Bishop, James M. and Gelder, Alan B. 2021. *Quantifying and Visualizing Forecast Uncertainty with the FIFE*. IDA Paper NS P-31857. Institute for Defense Analysis. Alexandria, VA:

Ustimenko, Aleksei, and Liudmila Prokhorenkova. 2021. "SGLB: Stochastic Gradient Langevin Boosting." arXiv:2001.07248 [Cs, Stat], July. http://arxiv.org/abs/2001.07248.

21

Appendix

22

PI Methods – Chernoff Bounds

Chernoff bounds are PIs around sums of independent Bernoulli trials, derived using Chebyshev's inequality.

$$P(X \geq (1+\delta_U)\mu) \leq \left(\frac{e^{\delta_U}}{(1+\delta_U)^{1+\delta_U}}\right)^{\mu} \text{ for all } \delta_U > 0$$

A looser version was previously used in FIFE:

$$P(X \geq (1+\delta_U)\mu) \leq \exp\left(-\frac{\delta_U^2}{2+\delta_U}\mu\right) \text{ for all } \delta_U > 0$$

PI Use-Specific Methods – Gal and Malinin

	MC Dropout	Stochastic Gradient Langevin Boosting
Modeler	Neural Network	Gradient Boosting
Source of ensemble variance	Dropout of NN units	Random sub-sampling of trees
Source of added uncertainty	Added model precision τ	Gaussian noise injected in gradient
Restrictions	Must have required choice of L2 regularization parameter	Must have required choice of L2 regularization parameter

Simulation Study Setup

- N = 2000, Time Periods = 50, Avg Exit Prob = 0.4
- First 2/3 of data used for training, last 1/3 for testing

ID	Period	X1	X2	X3	Exit Prob	Exit?
556	3	B	-0.03	0.94	0.17	No
556	4	B	-0.03	0.94	0.30	No
556	5	B	-0.03	0.94	0.48	Yes

- Known underlying survival probabilities used to calculate out-of-sample PI Coverage Probability (PICP)

This page is intentionally blank.