# Empirical Study of MC-Dropout in Various Astronomical Observing Conditions

Noble Kennamer
University of California, Irvine
nkenname@uci.edu

Alex Ihler
University of California, Irvine
secondauthor@i2.org

David Kirkby
University of California, Irvine
secondauthor@i2.org

## Abstract

*The analysis of large astronomical surveys increasingly incorporates machine learning models to handle a diverse set of tasks. It is important for the scientific analysis of these surveys that the uncertainty of the models be well understood and the predictions properly calibrated. Here we present an empirical study of MC-Dropout for a core prediction problem in astronomy emphasizing how the modeled uncertainty is influenced by changes in observing conditions. We will show that while MC-Dropout results in improved accuracy and better calibrated predictions there is still an underestimation of uncertainty that needs to be addressed.*

## 1. Introduction

We focus our experiments around the Large Synoptic Survey Telescope (LSST)[7] a major ground based astronomical survey. LSST will start to collect data in the early 2020s and operate for over ten years imaging tens of billions of galaxies and allowing domain scientists to explore phenomena from near earth asteroids to the large scale structure of the universe. Ground based surveys enable a large acquisition of data and its processing, but comes with the significant complication of having to factor in atmospheric effects. The effects of the atmosphere can be modeled as the convolution of a blurring function with the source. Due to atmospheric turbulence this blurring functions can change rapidly. In addition to the atmosphere the noise introduced by the instrumentation needs be accounted for. This can be successfully modeled an additive poisson noise.

A crucial inference problems in astronomical surveys is the detection of objects in a given cutout image of the sky. The outputs of a detection model are used in many downstream tasks. It is important to be able to model uncertainty and understand how it is characterized in different observing conditions. Poorly calibrated results can bias downstream analysis and impact the science goals of the survey. Gal and Ghahramani[3] showed that dropout and other stochastic regularization techniques are equivalent to performing variational inference over a bayesian neural network for some family of priors. This technique has been applied and extended with strong empirical performance in a diverse set of computer vision tasks [6]. We have empirically studied how this method performs as observing conditions change and its practical use for astronomical applications.

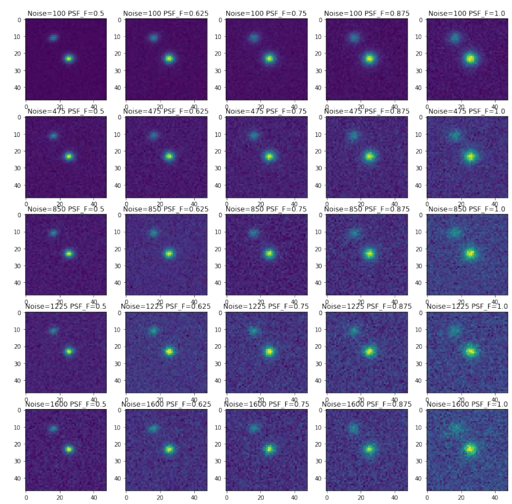


Figure 1. Example of the same point in the sky with two galaxies under the 25 different observing conditions. Five different values for the FWHM of point spread function and five different values for the rate of the additive poisson noise.

## 2. Experiments

We use Galsim an astromical image simulation package[8] to simulate LSST like images with a variable number of galaxies per image at 25 different observing conditions from very good to poor. The number of galaxies per image is poisson distributed with a rate of of about 1 galaxies per image (39.2 galaxies per square arc-minute. We use two parameters to control the quality of the observing conditions: the size of the atmospheric point spread function, specified by its full width at half the maximum and the rate of the additive poisson noise. Figure 1 shows an example of the same point in the sky containing two galaxies under the 25 different observing conditions. We train a simple convolutional neural network with dropout to predict the number of galaxies in each image. Our training set consists of 50000 images with a random assignment for the possible observing conditions. The validation set consists of 10000 images again with a random assignment for the observing conditions. The test set consists of 10000 images, but each image is generated for all possible observing conditions and we study the uncertainty produced by MC-Dropout as the observing conditions change.

Our network consists of two convolutional layers with 32 feature maps and 5x5 kernel size with a max pooling layer of size 2x2 in between. Followed by 3 fully connected layers of size 120, 84 and 9 (the maximum number of galaxies in an image). A relu nonlinearity and dropout is applied after all layers except the last. At prediction time we use 100 MC-Dropout trials. Figures 2 and 3 show accuracy and three different measures of uncertainty: entropy, mutual information and variation ratios as defined by [2]. As expected accuracy decreases and uncertainty decreases with poorer observing conditions. Figure 4 shows a calibration plot as defined by [5] comparing the calibration of using MC-Dropout to a standard dropout approach for a single observing condition. MC-Dropout results in improved calibration. This observation is consistent across all observing conditions.

We can also study uncertainty by using the predictive distribution of each data point to infer a distribution of a given property of our entire dataset. Sampling from the average distribution of the mc trials for each data point we can construct an MC estimate of the number of galaxies per arc minute as predicted by our network. We can compare this distribution with the true number, 39.2, controlled by our simulations. Figure 5 shows this distribution for each observing condition. As we can see the true value is not included posterior distribution. It is unclear if this underestimation of uncertainty is due general properties of variational inference or is specific to MC-Dropout. Further there appears to be some conflict with figures 4 and 5 with one showing strong calibration but the other indicating the method is underestimating uncertainty.

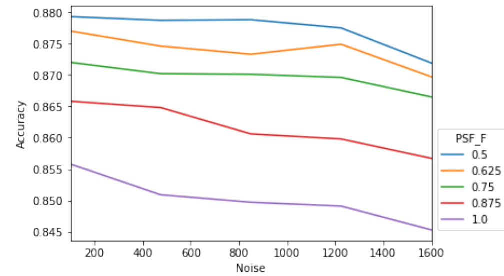


Figure 2. As expected accuracy decreases as the size of the psf increases and an increase in the rate parameter of the additive noise.

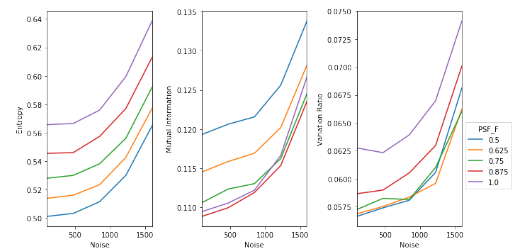


Figure 3. We use entropy, mutual information and variation ratios to quantify uncertainty. As expected uncertainty increases with poorer observing conditions.

## 3. Future Work

We have demonstrated that MC-Dropout can provide useful measures of uncertainty and improve calibration for a core astronomical problem. However it is also clear that in some sense the model is underestimating uncertainty. Understanding the source of this underestimation and how if it is possible to be resolved is an important direction for future work. It is also important that any downstream analysis which uses these estimates are aware of the underestimation. Further future work will involve extending this more complicated problems in astronomy and studying how it will perform in specific hard cases such as when two objects are blended (overlapping). This later problem of blending is especially with certain cosmological probes such as weak lensing[1]. Further study will also focus on more advanced objection detection architectures like R-CNN[4] style models and how uncertainty is characterized in these networks

## Acknowledgements

We would like to thank the MAPS program at UCI supported by the NSF award 1633631, the US Department of Energy award DE-SC0009920, the NSF grant IIS 1254071 and DARPA W911NF-18-C-0015 for supporting this work.

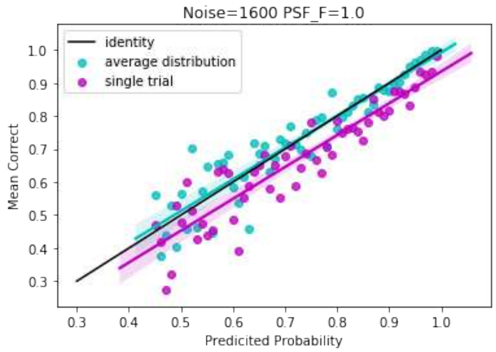


Figure 4. An example calibration plot of a single observing condition comparing calibration from 100 mc trials to 1 mc trial (standard dropout). Using MC-Dropout improves calibration.

## References


[1] G. M. Bernstein and M. Jarvis. Shapes and Shears, Stars and Smears: Optimal Measurements for Weak Lensing. *The Astronomical Journal*, 123:583–618, Feb. 2002. 2

[2] Yarin Gal. *Uncertainty in deep learning*. PhD thesis, PhD thesis, University of Cambridge, 2016. 2

[3] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Insights and applications. In *Deep Learning Workshop, ICML*, volume 1, page 2, 2015. 1

[4] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014. 2

[5] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1321–1330. JMLR. org, 2017. 2

[6] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pages 5574–5584, 2017. 1

[7] LSST Science Collaboration, P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, and et al. LSST Science Book, Version 2.0. *ArXiv e-prints*, Dec. 2009. 1

[8] B. T. P. Rowe, M. Jarvis, R. Mandelbaum, G. M. Bernstein, J. Bosch, M. Simet, J. E. Meyers, T. Kacprzak, R. Nakajima, J. Zuntz, H. Miyatake, J. P. Dietrich, R. Armstrong, P. Melchior, and M. S. S. Gill. GALSIM: The modular galaxy image simulation toolkit. *Astronomy and Computing*, 10:121–150, Apr. 2015. 1

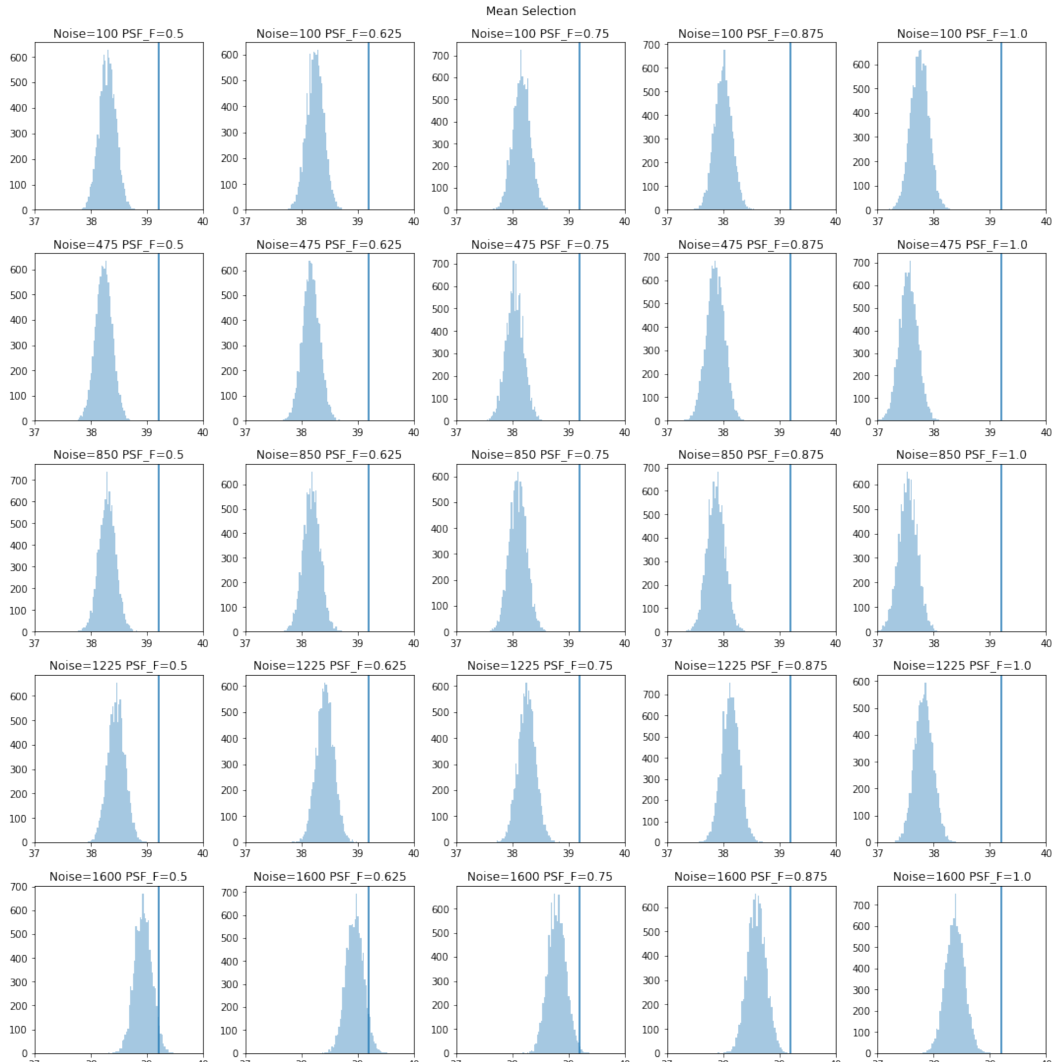


Figure 5. Example of a short caption, which should be centered.