**Likelihood Ratios for Out-of-Distribution Detection**

Jie Ren*, Peter J. Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark A. DePristo, Joshua V. Dillon, Balaji Lakshminarayanan*

New benchmark dataset + code is available at https://github.com/google-research/google-research/tree/master/genomics_ood

*Contact: jjren,peterjliu,balajiln@google.com

---

1. **Motivation**

- Bacteria identification based on genomic sequences
  - AGCTAAACC...GCGCTC ⇒ label
  - Promising for early detection of disease
- Classifier can achieve high accuracy on known classes, but perform poorly in real world:
  - 60-80% of real-world test inputs belong to as yet unknown bacteria
  - Ideally, say “I don’t know” on OOD inputs yet

2. **Generative Models Can Assign Higher Likelihood to OOD Inputs**

- Generative models for OOD detection:
  - do not require labeled data
  - model the input distribution $p_{\text{data}}(x)$ and evaluate the likelihood of new inputs.
- Prior work [Nalisnick et al., 2018, Choi et al. 2019] observed failure modes of generative models:
  - Higher likelihoods for OOD than in-dist. e.g. Fashion-MNIST (in-dist.) vs. MNIST (OOD)
- We observe a similar failure mode on models trained on genomic sequences.

3. **Explaining Why Density Models Fail At OOD Detection**

- $p(x) = p(x|y)p(y|x)$ can be dominant
- $p(x|y)$ has two components:
  - Semantic (i.e. semantic contrastive score)
  - Background
- Humans ignore background and focus primarily on semantics for OOD

4. **Proposed Solution: Likelihood Ratios For OOD Detection**

- How do we automatically extract the semantic component of $p(x)$?
- We propose training a background model on perturbed inputs and computing the likelihood ratio.
- $LR(x) = \log \frac{p_{\text{data}}(x)}{p_{\text{data}}(x)} = \log \frac{p(x|y)p(y|x)}{p(x|y)p(y|x)} = \log \frac{p(x|y)p(y|x)}{p(x|y)p(y|x)}$
- Likelihood ratio is highly correlated with the background
  - proportion of zeros in an image
  - GC-content in genomic sequence

5. **OOD Detection For Images**

- Investigate auto-regressive models: which pixels contribute the most to the likelihood ratio?
- Fashion-MNIST (in-dist.) vs. MNIST (OOD). PixelCNN++ model is trained on Fashion-MNIST.
- Likelihood ratio focuses on the semantic pixels ⇒ LRR(Fashion-MNIST) > LRR(MNIST)

6. **OOD Detection For Genomic Sequences**

- LSTM model is trained using sequences from in-distribution classes
- Likelihood Ratio significantly improves OOD Detection
- Effect of background GC-content is corrected
- OOD detection correlates with its distance to in-distribution

---

**Summary**

- Create a realistic benchmark dataset for OOD detection (and open-set classification) in genomics
- Show that the likelihood from deep generative models can be confounded by background statistics
- Propose a likelihood ratio method for unsupervised OOD detection, outperforming the raw likelihood
- Our method performs well on images and achieves SOTA performance on genomic dataset.