arize

USING STATISTICAL DISTANCES FOR

Machine Learning Observability

Introduction

Statistical Distances are used to quantify the distance between two distributions and are extremely useful in ML observability. This blog post will go into **statistical distance** measures and how they are used to detect common machine learning model failure modes.

Table of contents

Why Statistical Distance Checks?

Data problems in Machine Learning can come in a wide variety that range from sudden data pipeline failures to long-term drift in feature inputs. Statistical distance measures give teams an indication of changes in the data affecting a model and insights for troubleshooting. In the real world post model-deployment, these data distribution changes can come in a myriad of different ways and cause model performance issues.

Here are some real-world data issues that we've seen in practice.

Incorrect data indexing mistake - breaks upstream mapping of data

Software engineering changes the meaning of a field

3rd Party data source makes a change dropping a feature, changing format, or moving data

Newly deployed code changes an item in a feature vector

Outside world drastically changes (e.g., the COVID-19 pandemic) and every feature shifts

Periodic daily collection of data fails, causing missing values or lack of file

Presumption of valid format that changes and is suddenly not valid

3rd Party Library Functionality Changes

Date string changes format

Bad text handling - causes new tokens model has never seen

- Mistakes Handling Case
- Problems with New Text String

System naturally evolves and feature shifts

Drastic increase in volume skews statistics

Different sources of features with different coordinates or indexing

These are examples of data issues that can be caught using statistical distance checks.

Where To Use Statistical Distance Checks

Statistical distances can be used to analyze

- **Model Inputs:** Changes in inputs into a model, especially critical most important features or features that might be the output of an upstream model.
- **Model Outputs:** Changes in outputs of a model
- **Actuals:** Changes in actuals (ground truth received). In some cases, the ground truth might not be available within a short time horizon after prediction. In these cases, teams often use proxy metrics/data.

These checks are extremely insightful for model performance troubleshooting and they allow teams to get in front of major model issues before these problems affect business outcomes. In this image below, there are statistical checks that can be done on model inputs (features) and model outputs (predictions).

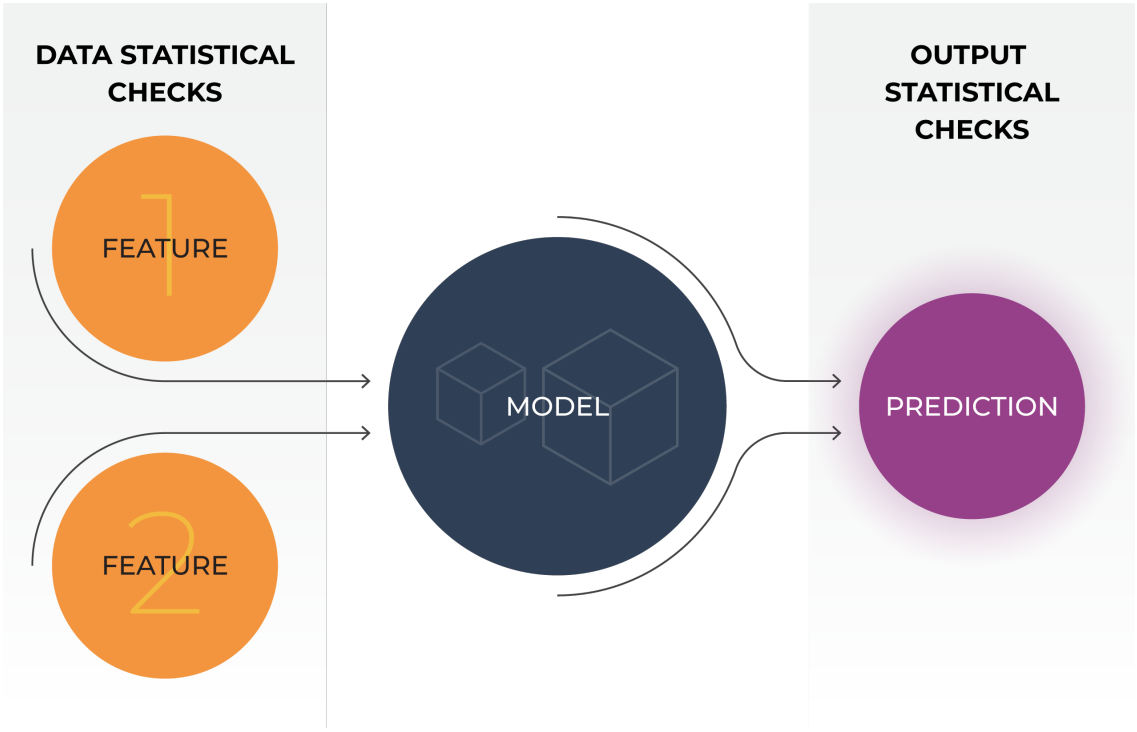


Image by Author. Statistical Checks: Inputs, Outputs and Actuals

How to Use Statistical Distance Checks

Statistical distance measures are defined between two distributions - distribution A and distribution B. One of these distributions is commonly referred to as the reference distribution (we will refer to this as distribution A) - this is what you are comparing against. The other distribution is typically the current state of the system that you are comparing to the reference distribution (we will refer to this as distribution B).

What is a Reference Distribution?

In the context of ML Observability, the reference distribution can be a number of different options. The first distinction to make is that the reference distribution can be a distribution across a fixed time window (distribution doesn't change) or a moving time window (distribution can change).

In the image below, you can see that the examples of the fixed reference distribution include a snapshot of training distribution, initial model deployment distribution (or a time when the distribution was thought to be stable), and validation/ test set distribution. An example of a statistical distance using a fixed reference distribution is to compare the model's prediction distribution made from a training environment (distribution A) to the model's prediction distribution from a production environment (distribution B).

The reference distribution can also be a moving window. In the image below, there are two examples where the reference distribution is a moving window - last week and A/B testing use cases. For the first example, one might want to compare the distribution of last week's predictions to this week's prediction. In this case, each week

there will be a new reference distribution. If you are A/B testing two different model versions in production, you can compare if the prediction distribution from your champion model is similar to the prediction distribution from the challenger model.

The options for what you set as the reference distribution will depend on what you are trying to catch and we will dive into common statistical distance checks to set up for models.

Image by Author. Example Reference Distributions for Distance Checks

Common Use Cases for Statistical Distance Checks

MODEL INPUTS

Model inputs can drift suddenly, gradually, or on a recurring basis depending on what is set as the reference distribution. As the old adage goes, “garbage in, garbage out.” In other words, a model is only as good as the data flowing into the model. If the input data changes and is drastically different than what the model has observed previously or has been trained on, it can be indicative that the model performance can have issues.

Here are a few statistical distance checks to set up on model inputs:

1. Feature Distribution in Training vs Feature Distribution in Production

The distribution of a feature can drift over time in production. It is important to know if a feature distribution has changed over time in production and if this is impacting the model. In this setup, the reference distribution (distribution A) is the feature distribution in training. The current window (distribution B) can be set to the feature distribution over a certain time window (ex: a day, a week, a month, etc). If the feature distribution is highly variant, it can be helpful to set a longer lookback window so the statistical distance check can be less noisy.

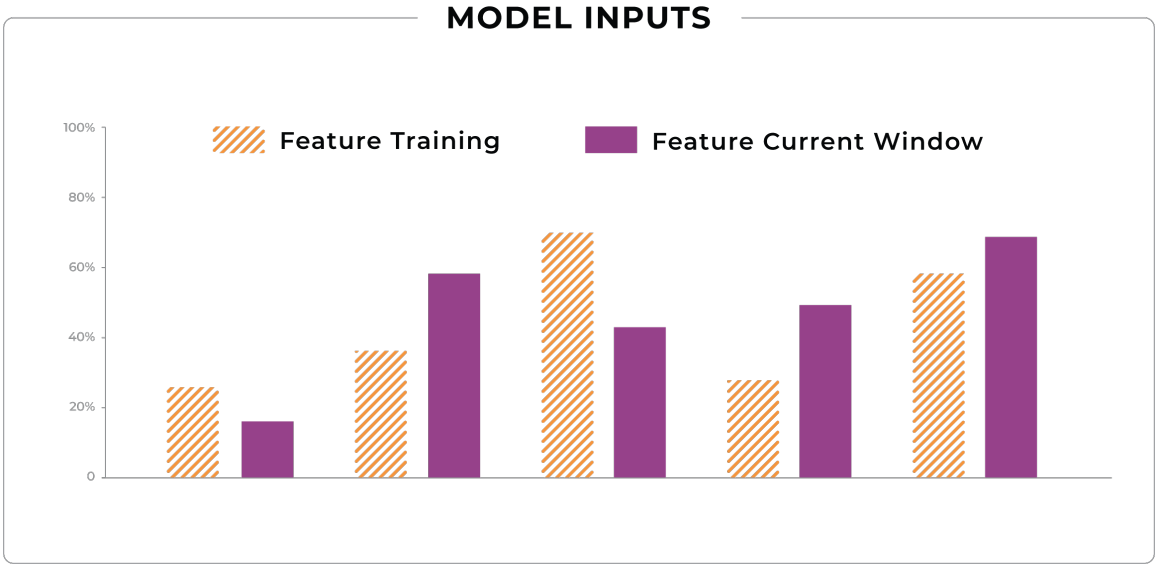


Image by Author. Feature (Training) vs Feature (Production)

2. Feature Distribution in Production Time Window A vs Feature Distribution in Production Time Window B

It can also be useful to set up distribution checks on a feature at two different intervals in production. This distribution check can focus on more short-term distribution changes compared to the training vs production check. If setting the training distribution as the reference distribution, setting a short production time window can be noisy if there are any fluctuations (ex: traffic patterns, seasonal changes, etc). Setting up a statistical distance check against last week vs the current week can give an indication of any sudden outliers or anomalies in the feature values. These can also be extremely useful to identify any data quality issues that might get masked by a larger time window.

Identifying if there has been a distribution change in the feature can give early indications of model performance regressions or if that feature can be dropped if it's not impacting the model performance. It can lead to model retraining if there are significant impacts to the model performance. While a feature distribution change should be investigated, it does not always mean that there will be a correlated performance issue. If the feature was less important to the model and didn't have much impact on the model predictions, then the feature distribution change might be more of an indication it can be dropped.

Teams in the real world use model input checks to determine when models are growing stale, when to retrain models, and slices of features that might indicate performance issues. One team I spoke to that uses a financial model for underwriting generates analysis on feature stability by comparing the feature in production to training to make sure the model decisions are still valid.

"While a feature distribution change should be investigated, it does not always mean that there will be a correlated performance issue."

MODEL OUTPUTS

Just like model inputs can drift over time, model outputs distributions can also change over time. Setting statistical distance checks on model predictions make certain the outputs of the model are not drastically different than their reference distributions.

3. Prediction Distribution in Training vs Prediction Distribution in Production

The goal of output drift is to detect large changes in the way the model is working relative to training. While these are extremely important to ensure that models are acting within the boundaries previously tested and approved, this does not guarantee that there is a performance issue. Similar to how a feature distribution change does not necessarily mean there is a performance issue, prediction distribution changes doesn't guarantee there is a performance issue. A common example is if a model is deployed to a new market, there can be distribution changes in some model inputs and also the model output.

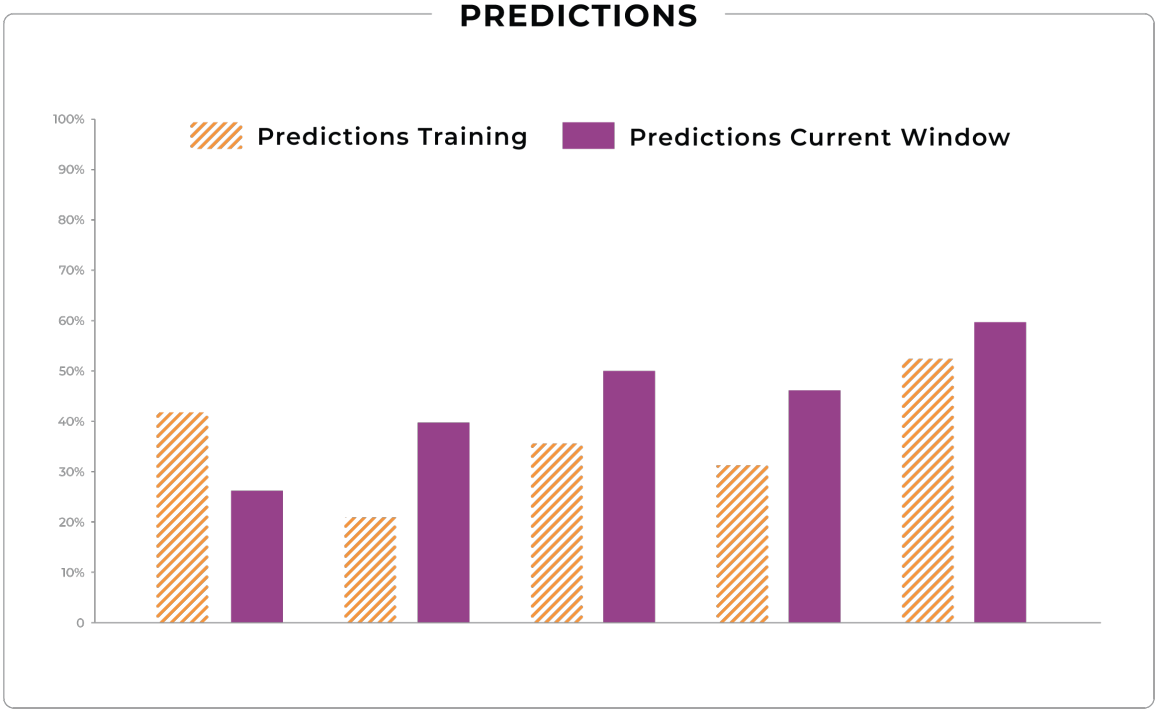


Image by Author. Predictions (training) versus Prediction (production)

4. Prediction Distribution at Production Time Window A vs Prediction Distribution in Production Time Window B

Similar to model inputs, the prediction distribution can also be monitored to another time window in production. One team we talked with, who was evaluating a spam filter model, uses the distribution of the output of the model versus a fixed time frame to surface changes in attack patterns that might be getting through the model. The reference distribution here can either be a moving time window or a fixed time frame. A common fixed time frame we hear is using the initial model launch window.

5. Prediction Distribution for Model Version A vs Prediction Distribution for Model Version B at Same Time Window

Teams that have support for canary model deployment can set up statistical distance checks on the prediction distributions for different model versions. While A/B testing two different models in production with each model receiving a certain amount of traffic or backtesting a model on historical data, comparing the prediction distribution gives insight into how one model performs over another.

MODEL ACTUALS

6. Actuals Distribution at Training vs Actuals Distribution in Production

Actuals data might not always be within a short-term horizon after the model inferences have been made. However, statistical distance checks on actual distributions help identify if the structure learned from the training data is no longer valid. A prime example of this is the Covid-19 pandemic causing everything from traffic, shopping, demand, etc patterns to be vastly different today from what the models in production had learned before the pandemic began. Apart from just large-scale shifts, knowing if the actuals distribution between training data vs production for certain cohorts can identify if there

MODEL PREDICTIONS VS ACTUALS

7. Predictions Distribution at Production vs Actuals Distribution at Production

This statistical distance check is comparing production distribution of predictions vs actuals. This can help catch performance issues by pinpointing specific cohorts of predictions that have the biggest difference from their actuals. These checks can sometimes catch issues that are masked in averages such as MAE or MAPE.

Common Statistical Distance Measures

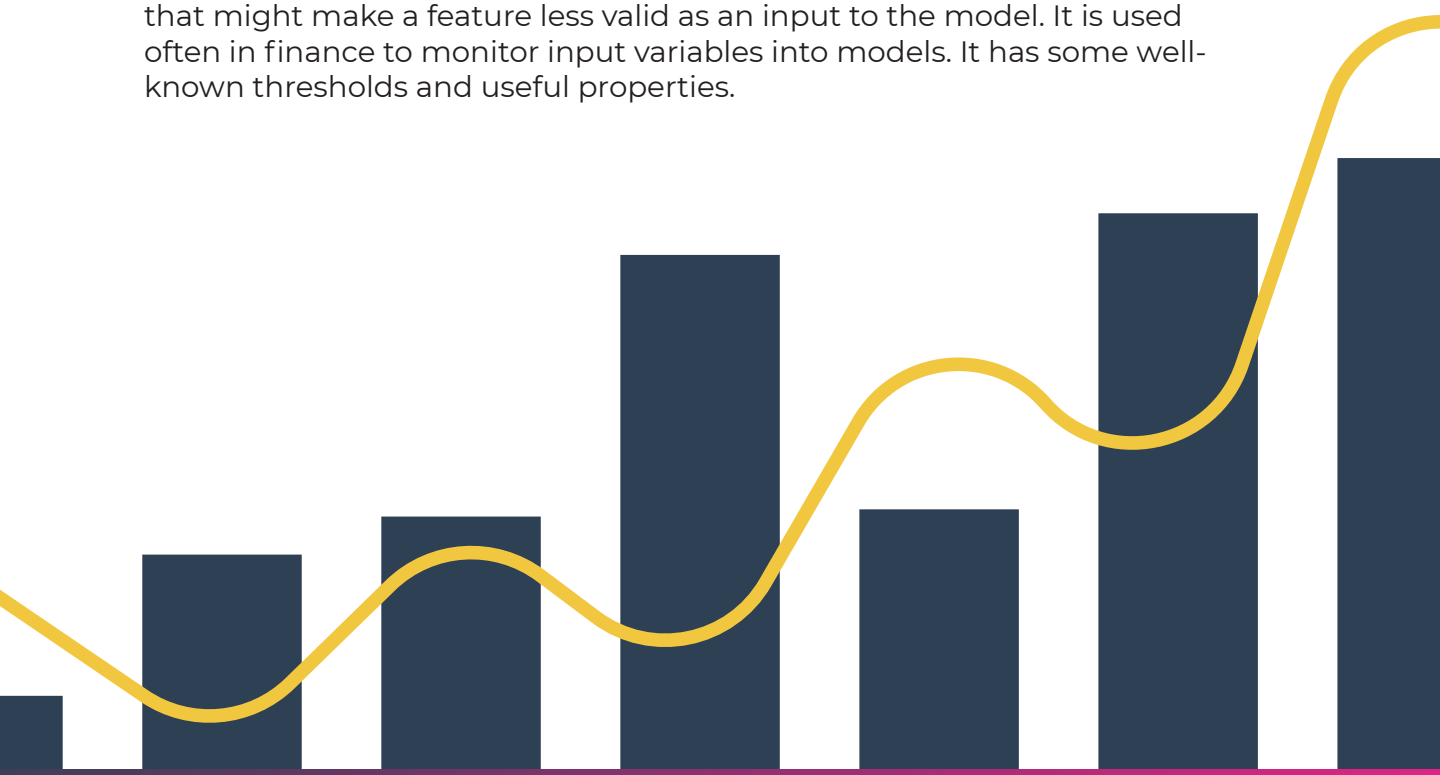
We just covered common use cases for statistical distances. There a number of different statistical distance measures that quantify change between distributions. Different types of distance checks are valuable for catching different types of issues. In this blog post, we will cover the following 4 distance measures and when each can be most insightful.

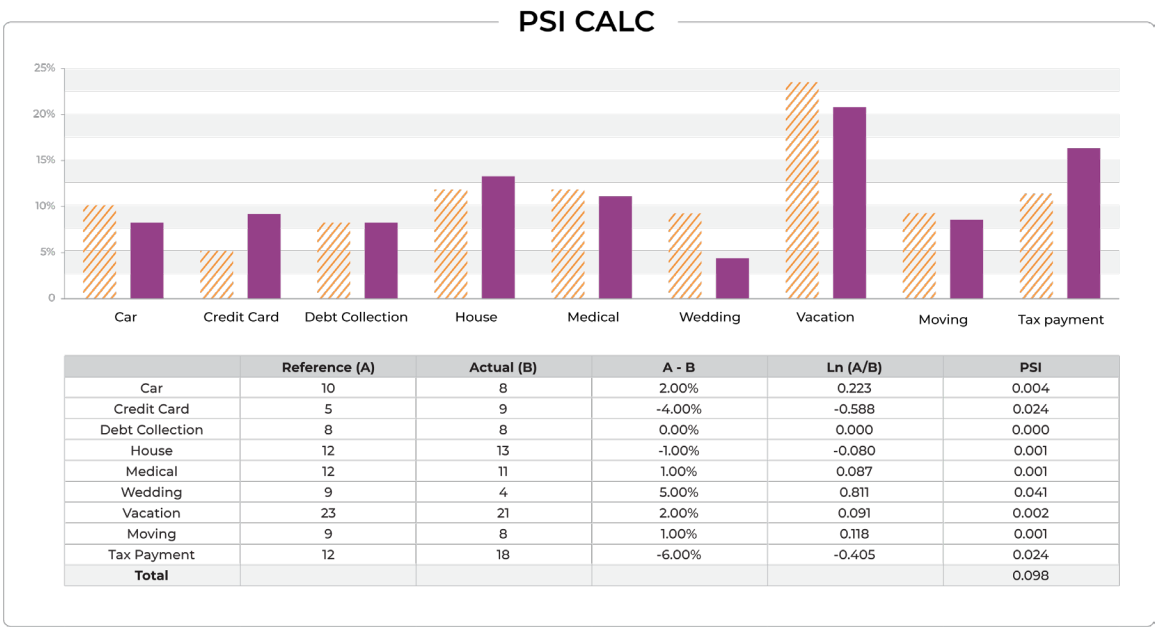
- Population Stability Index (PSI)
- Kullback–Leibler divergence (KL-Divergence)
- Jensen–Shannon divergence (JS-Divergence)
- Earth Mover's Distance (EMD)

PSI

The PSI metric has many real-world applications in the finance industry. It is a great metric for both numeric and categorical features where the distributions are fairly stable.

Equation:
*PSI = ∑(Pa - Pb) * ln(Pa/Pb)*
PSI is an ideal distribution check to detect changes in the distributions that might make a feature less valid as an input to the model. It is used often in finance to monitor input variables into models. It has some well-known thresholds and useful properties.



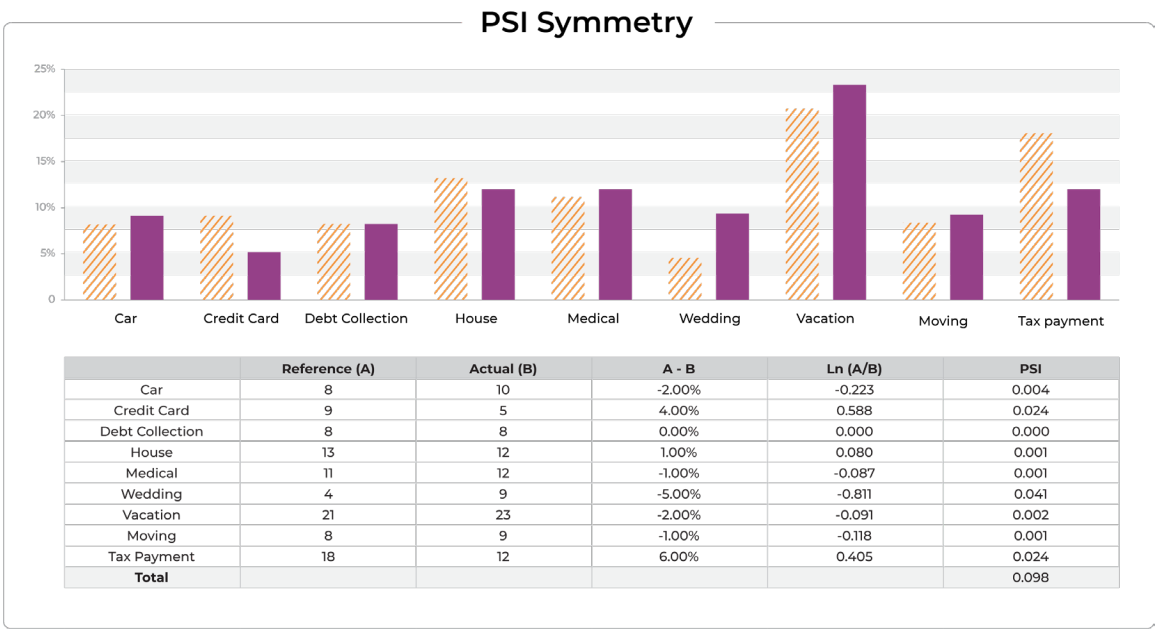


	Reference (A)	Actual (B)	A - B	Ln (A/B)	PSI
Car	10	8	2.00%	0.223	0.004
Credit Card	5	9	-4.00%	-0.588	0.024
Debt Collection	8	8	0.00%	0.000	0.000
House	12	13	-1.00%	-0.080	0.001
Medical	12	11	1.00%	0.087	0.001
Wedding	9	4	5.00%	0.811	0.041
Vacation	23	21	2.00%	0.091	0.002
Moving	9	8	1.00%	0.118	0.001
Tax Payment	12	18	-6.00%	-0.405	0.024
Total					0.098

Image by Author. Calculation of the Raw Values of PSI from Distribution

Going through the PSI values above, one can see:

- Any change in the distribution will add to the PSI total -- whether the change is positive or negative. This means it doesn't matter if the distribution grows or shrinks, any change increases PSI.
- The ln(Pa/Pb) term implies that a large change in a bin that represents a small percentage of a distribution will have a larger impact (on PSI) than a large change in a bin with a large percentage of the distribution.
- An increase in a distribution bin from 4% to 9% has almost double the PSI affect than a move from 12% to 18%
- The example distribution above which includes a number of small percentage changes (less than 6 percent) where none individually generate a PSI term over 0.1 which is a rule-of-thumb benchmark for model investigation. The point is small changes will not move the needle relative to industry benchmarks.
- Setting of threshold we recommend either common finance industry benchmarks or basing on days/hours of previous samples of PSI for that feature/prediction/actual.
- Industry benchmarks of 0.1-0.25 in finance will typically catch moves of around 10% between bins.



	Reference (A)	Actual (B)	A - B	Ln (A/B)	PSI
Car	8	10	-2.00%	-0.223	0.004
Credit Card	9	5	4.00%	0.588	0.024
Debt Collection	8	8	0.00%	0.000	0.000
House	13	12	1.00%	0.080	0.001
Medical	11	12	-1.00%	-0.087	0.001
Wedding	4	9	-5.00%	-0.811	0.041
Vacation	21	23	-2.00%	-0.091	0.002
Moving	8	9	-1.00%	-0.118	0.001
Tax Payment	18	12	6.00%	0.405	0.024
Total					0.098

Image by Author. PSI Calculation Switch Distribution: Symmetry

The PSI is symmetric -- that is, if you reverse the distributions, the PSI value is the same. In the above example we have switched the purple graph with the yellow graph of previous examples, the value of 0.98 is the same as the previous reversed distribution.

The example below uses the Population Stability Index (PSI) on an important feature. The check is run periodically, trading off how quickly you want to be alerted on change, versus the type of change you are trying to detect.

When the check falls below a well-defined threshold, the change needs to be investigated and could indicate a model performance issue.

FPO - graphic

The example below uses the Population Stability Index (PSI) on an important feature. The check is run periodically, trading off how quickly you want to be alerted on change, versus the type of change you are trying to detect.

When the check falls below a well-defined threshold, the change needs to be investigated and could indicate a model performance issue.

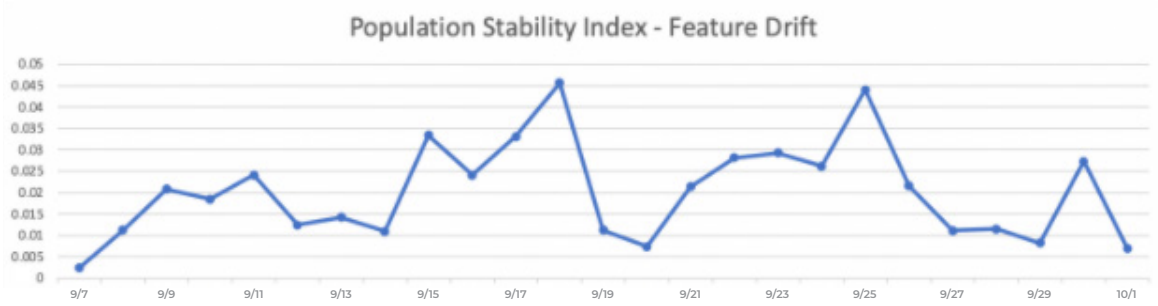


Image by Author. PSI - By Day On Feature

The following shows a live feature where the stability index is far below the 0.15 limit that was set (0.1-0.25 finance industry standard range). On setup, we recommend looking at a multi-day window of statistics for setting the detection range.

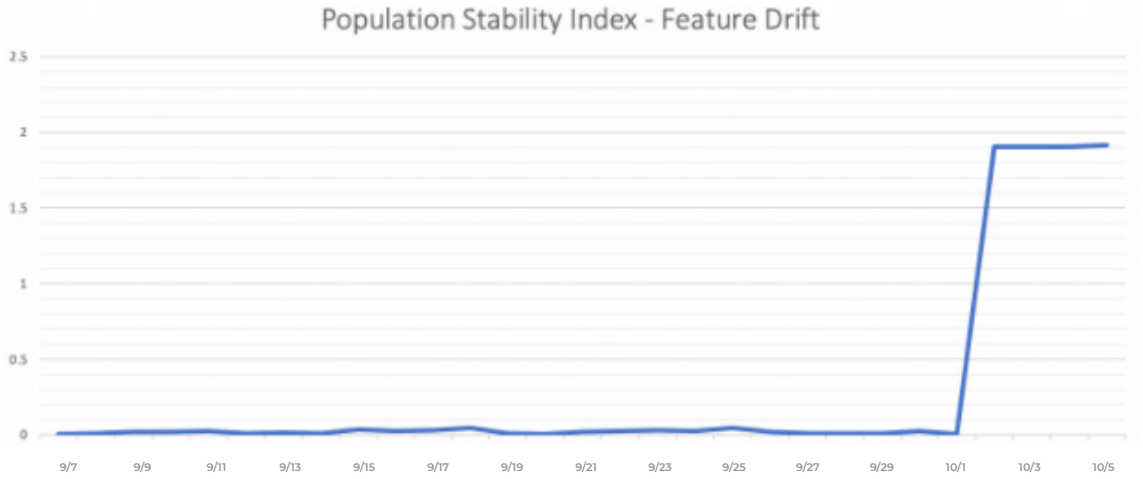


Image by Author. PSI (Large Change) - By Day On Feature

The above daily PSI distribution change is a true change on a measured feature where a new categorical feature was introduced.

KL Divergence

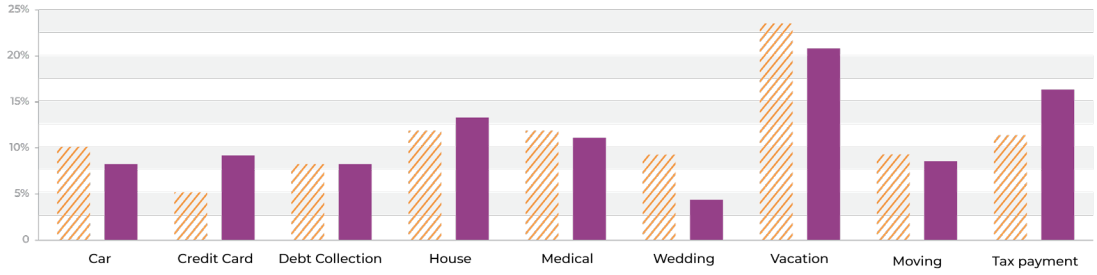
The KL divergence statistic is useful if one distribution has a high variance relative to another or small sample size.

Equation:
KLdiv = Ea[ln(Pa/Pb)] = ∑ (Pa)ln(Pa/Pb)

KL Divergence is a well-known metric that can be thought of as the relative entropy between a sample distribution and a reference (prior) distribution. Like PSI, KL Divergence is also useful in catching changes between distributions. Also similar to PSI, it has its basis in information theory.

One important difference from PSI is that KL Divergence is not symmetric. A reversed distribution will have a different value - you will get different values going from A -> B then B -> A. There are a number of reasons that having a non-symmetric metric is not ideal for distribution monitoring in that you get different values, when you switch what is the reference versus compared distribution. This can come across as non-intuitive to users of monitoring.

KL 1



	Reference (A)	Actual (B)	A	Ln (A/B)	KL
Car	10	8	10.00%	0.223	0.022
Credit Card	5	9	5.00%	-0.588	-0.029
Debt Collection	8	8	8.00%	0.000	0.000
House	12	13	12.00%	-0.080	-0.010
Medical	12	11	12.00%	0.087	0.010
Wedding	9	4	9.00%	0.811	0.073
Vacation	23	21	23.00%	0.091	0.021
Moving	9	8	9.00%	0.118	0.011
Tax Payment	12	18	12.00%	-0.405	-0.049
Total					0.050

Image by Author. KL-Divergence Example Calculation

JS Divergence

JS Div(P, Q) = ½ KL-DIV(P,M) + ½ KL-DIV(Q,M)

Reference = M (mixture distribution) = ½ (P + Q)

JS Divergence has some useful properties. Firstly, it's always finite, so there are no divide-by-zero issues. Divide by zero issues come about when one distribution has values in regions the other does not. Secondly, unlike KL-Divergence, it is symmetric.

The JS divergence uses a mixture of the two distributions as the reference. There are challenges with this approach for moving window checks; the mixture-reference changes based on the changes in the moving window distribution. Since the moving window is changing each period, the mixture-reference is changing, and the absolute value of the metric in each period can not be directly compared to the previous periods without thoughtful handling. There are workarounds but not as ideal for moving windows.

Image by Author. Issues with JS Divergence (Unstable Reference)

The moving window changes each period for every distribution check. It represents a sample of the current periods distribution. The JS Distribution has a unique issue with a moving window, in that the mixture will change with each window of time you are comparing. This causes the meaning of the value returned by JS Divergence to shift on a periodic basis, making comparing different time frames on a different basis, which is not what you want.

The PSI and JS are both symmetric and have potential to be used for metric monitoring. There are adjustments to PSI that we recommend versus JS as a distance measure for moving windows used for alerts.

Earth Mover's Distance (Wasserstein metric)

$EMD_0=0$

$EMD_{i+1}=(Ai+EMDi) - B_i$

Total Distance = $\sum|EMD_i|$

The Earth Mover's Distance measures the distance between two probability distributions over a given region. The Earth Movers Distance is useful for statistics on non-overlapping numerical distribution moves and higher dimensional spaces (images, for example)

Using both PSI & KL calculations above a Bin0 is compared to Bin0, Bin1 to Bin1, etc... as part of the distribution check. The Bin0 is never compared to Bin1, the calculation fixes the bin comparisons.

The following statistical distance checks do not have locked bins as part of the calculation. The Bin number is irrelevant; what matters more is the distance between distributions.

The Earth Mover's Distance is a fairly old calculation -- it was formulated in 1781. In the case of a one-dimensional distribution, it captures how much the shape and distance to the mean of a distribution is retained in moving one distribution to the other.

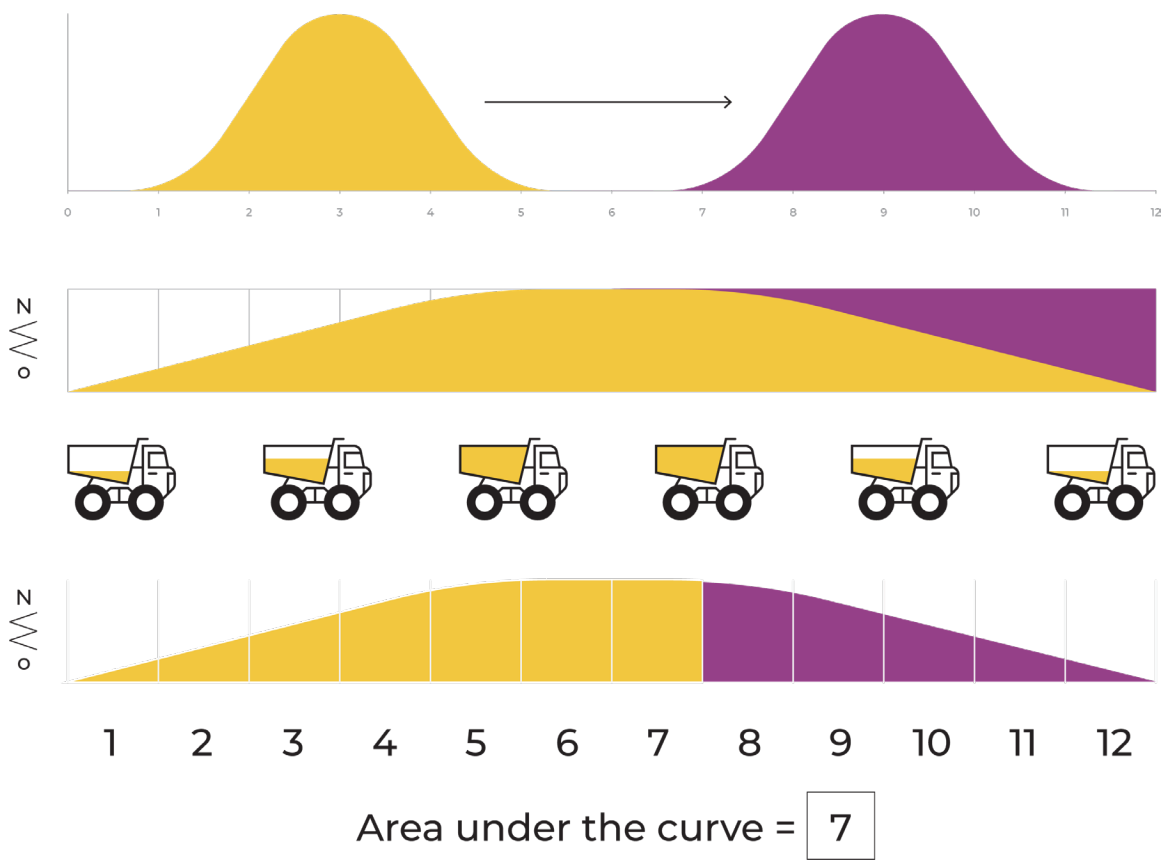


Image by Author. Visualization of 1-Dimensional EMD

The Earth Mover's Distance can be simply demonstrated using a one-dimensional case such as that which is illustrated above (originally the EMD algorithm was designed to solve a problem around moving dirt). The Earth Mover's Distance here can be thought of as the work needed to move one pile of dirt into another pile of dirt. The dirt is filled up by a truck along a straight road (the X-axis) by putting the dirt into the truck. The work needed to move the dirt is calculated by each unit along the X-axis, as well as how much dirt is in the truck, and how many units of dirt that the truck can transport. The truck empties the dirt into the other distribution. The further away the means of the distributions, the larger the Earth Mover's Distance because the truck will transport the dirt farther to get from one mean to the other. The more spread out and overlapping the distributions are, the smaller the number.

Compared to KL divergence, EMD handles naturally non-overlapping distributions where KL/PSI need modifications.

Using Statistical Distances along with Model Performance Metrics

Statistical distance checks can be extremely powerful for model observability. There are many setup configurations that can help identify different model issues (drift, data distribution changes, data quality issues, model performance regressions, etc). These statistical distance measures are best analyzed alongside model performance metrics.

Using performance metrics with distributions changes enables teams to identify slices of predictions that might be bringing down overall performance of the model. In this example below, the overall model accuracy is 71%. Using statistical distance measures, we can see there is significant movement in 2 bins. What is the performance for those bins? Did this movement cause any global performance issues in the model? In this example, we can see that those bins have poorer performance than the overall accuracy.

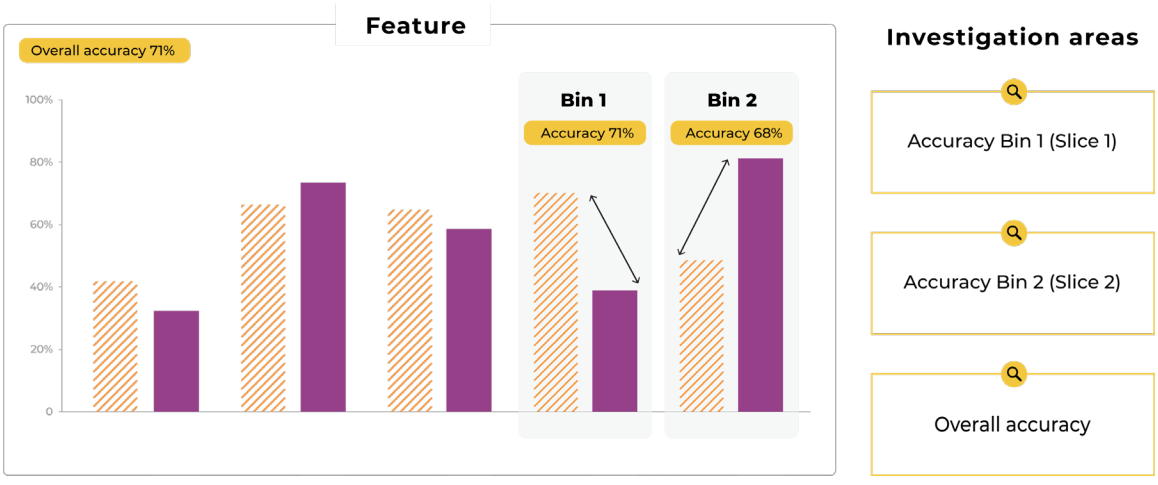


Image by Author. Performance Analysis of a Feature

Connecting a statistical distance check to a broader model troubleshooting workflow can help identify different model failure modes. What we recommend is:

- Set statistical distance checks on features, predictions, and actuals
- When a statistical distance metric has gone above comfortable thresholds, determine whether this is model performance impacting
- When looking at model performance, compare with performance from a training/validation set
- Look at performance for specific slices related to the change

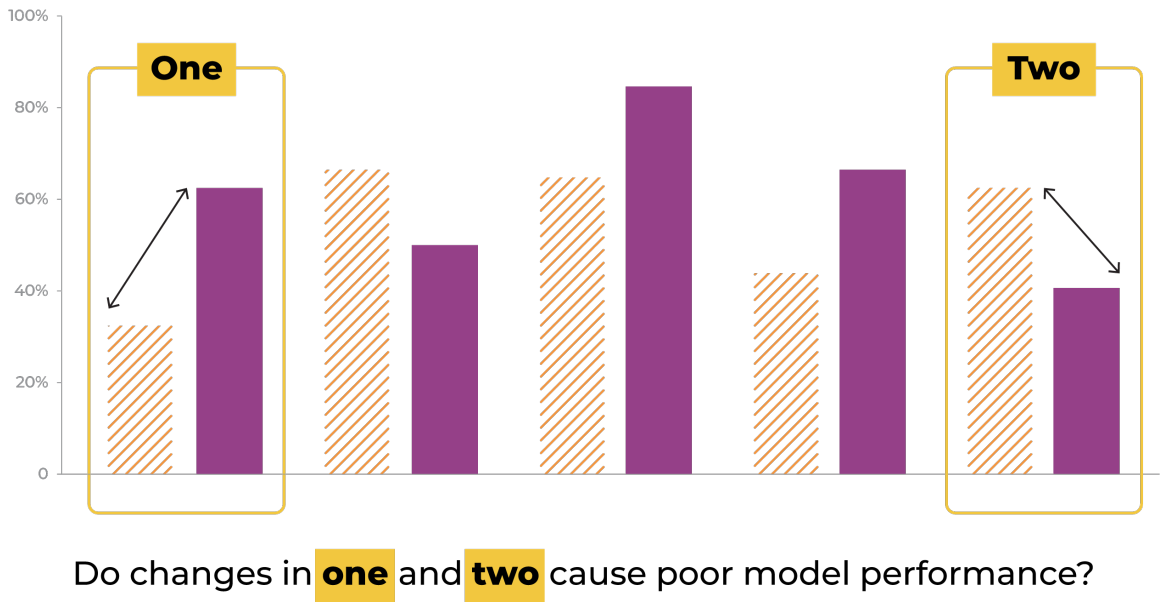


Image by Author. Distribution Changes May Not Always Mean Performance Issue

The changes in a distribution may or may not cause large downstream issues. The point is that no change should be looked at in a vacuum, or investigated just because something changed. The changes should be filtered against other system performance metrics to investigate the ones that matter.

FPO - graphic

Conclusion

Here are final thoughts/recommendations we will conclude on:

- Statistical distance checks can be immensely useful in tracking changes to inputs, outputs, and actuals of models allowing teams to catch issues before business impact.
- PSI should be used to monitor top features for feature drift
- PSI should be used to monitor prediction output and actuals for concept drift
- KL Divergence should be used when one distribution is much smaller in sample numbers and has a large variance
- JS Divergence can be used for feature drift but since the mixture (reference) is changing you should look at absolute numbers versus day by day
- Zero value comparisons, when one bin has a value and another does not need to be only handled by setting a prior based on data. The standard approach of adding a small value does not work well. Please reach out for more detail.
- When a significant distribution change has occurred, we recommend looking alongside performance metrics and investigating if retraining can be a solution.

Contact Us

If this blog post caught your attention and you're eager to learn more, follow us on Twitter and Medium! If you'd like to hear more about what we're doing at Arize AI, reach out to us at contacts@arize.com. If you're interested in joining a fun, rockstar engineering crew to help make models successful in production, reach out to us at jobs@arize.com!

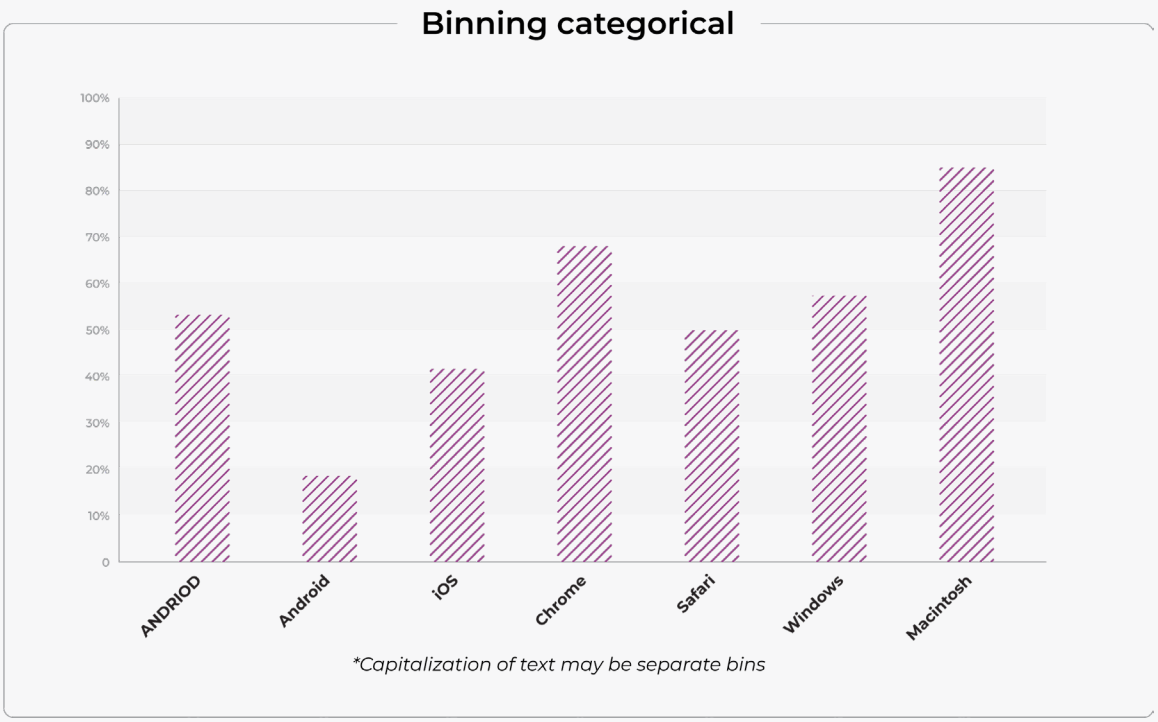
APPENDIX

There are a bunch of details we could dive into on statistical distance checks! These details are extremely important when trying to set up distance checks and are crucial to making the metrics make sense in real-world applications.

Types of Bins:

CATEGORICAL:

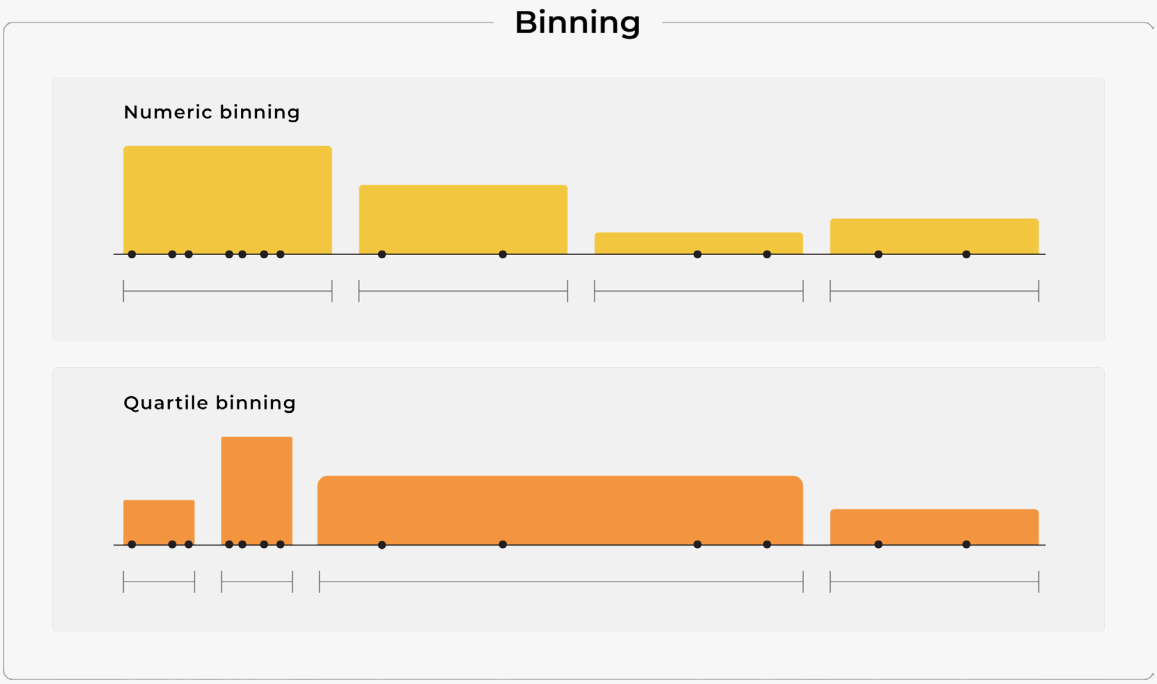
The binning of categorical variables occurs by binning on the value itself, based on inputs before 1-hot encoding. The text string represents the bin. Depending on how the system handles capitalization, a capitalized word might or might not be binned separately based on how the feature pipeline handles capitalization.



BINNING A NUMERICAL DISTRIBUTION:

We recommend binning for all variables in statistical distance checks in addition to their use in graphical analysis. There are a number of reasons to do this but mainly it comes down to making alerts more useful. Binning allows for easier troubleshooting of issues by providing a subspace of the model input space to investigate. The bin is an easy system filter to slice on in order to analyze model performance and tie a change in a feature/model output TO issues such as Accuracy or RMSE.

The binning of a numeric feature is not absolutely required to get a metric but it's very helpful for visualization and debugging.



As a numerical input to the model changes it will move between bins, for example moving from bin 1.0-4.0 (decreases) to bin 4.0-8.0 (increases). As you evaluate the change you can slice performance metrics (Accuracy, RMSE,etc) by those bins to see if the model itself has any issues with the new distribution.

NUMERIC:

Numeric data can be binned using fixed distance between points(knots), custom points or quintiles each with different trade offs.

FIXED DISTANCE:

The fixed distance is easy to set up and easy to analyze. It works best for data that doesn't have a lot of variation in a small area relative to the entire distribution. Data that is more evenly distributed over a range.

QUINTILES

The quintiles can be used for data that is not evenly distributed. The quintiles are taken from a single distribution, say reference, and then used to define the knots that all distributions use. It helps ensure each bin region has a similar amount of data. The differing points or knots between distributions can make visual comparisons harder for regions with fewer samples on secondary distributions.

QUINTILE BREAK POINTS:

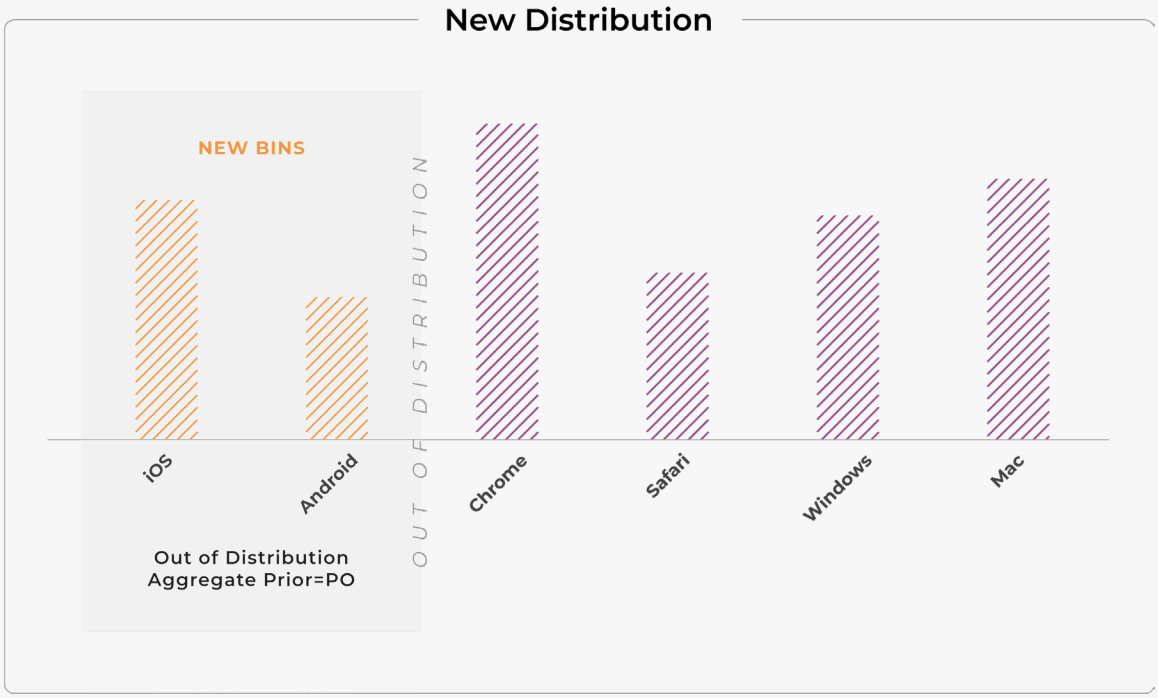
The Quintiles determine breakpoints to have a certain % of the distribution included. If you evenly space quintiles 10% / 20% / 30% you put an even number of samples into each bin. Or you can have quintiles that are more spaced as a Normal distribution cut offs 0.1%, 13.6%, ...

CUSTOM:

If you know your data well and have common breakpoints for data OR you want to capture movements between well defined regions/bins you can break up your data with custom breakpoints.

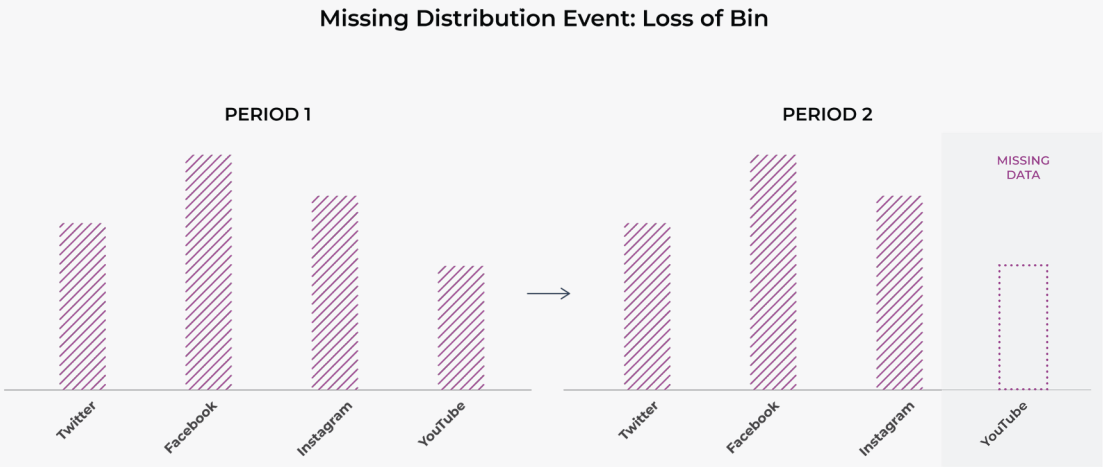
OUT OF DISTRIBUTION EVENTS, MOVEMENT OF BINS AND OUTLIERS:

In a future distribution sample, events can occur outside of the range of distributions seen when the analysis was set up. Events that fall outside of the distributions used to set up the analysis, we label as out-of-distribution events.



In order to handle out-of-distribution events typically certain bins are defined with an infinity edge. There are some well understood criteria of how mappings occur from out-of-distribution to a specific bin designed to handle those events.

In addition to Out-of-Distribution bins there is another grouping of bins we call Movement-out-of bins. This applies the same concept as out-of-distribution but symmetrically back to the reference distribution. In an Out-of-distribution bin/event, you have a bin that exists in the compared-to distribution versus is empty in reference distribution. In a movement-out-of bin you have a bin that is empty in the compared-to distribution that has values in the reference distribution.



COMPARE PERIODS: OUT OF DISTRIBUTION EVENTS

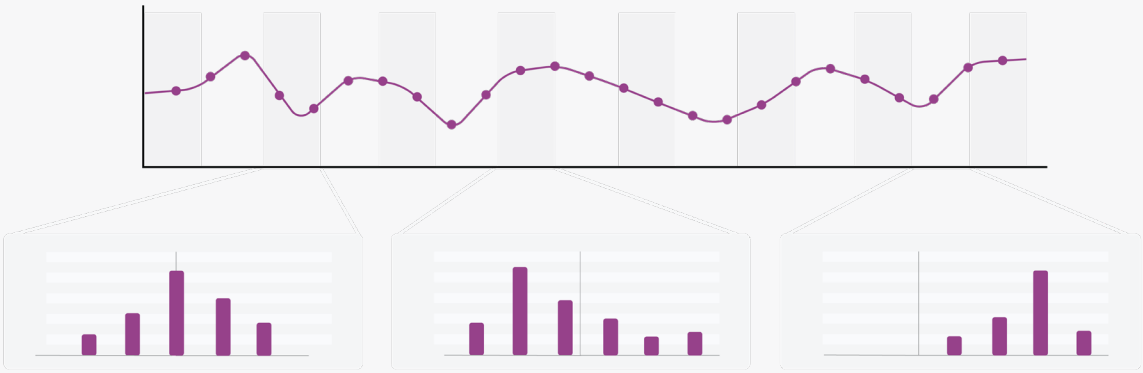
In traditional statistical distance checks, you might have all the statistics on both distributions, at one place & time, so you can define bins in a hand-crafted fashion. In the case of moving windows, you need to define bins in an automated fashion that will handle future unique distribution samples.

Another unique challenge with moving windows is that you want to define bins that don't change, so you truly have a reference, but you need to do it based on the distributions you have initially. New distributions that show up in the future need to be covered by the previous bins you had created.

Lastly, gracefully handling zeros in statistical distance analysis is extremely important. The zero means that either the prior or posterior is statistically impossible, yet that is often not the case for out-of-distribution events.

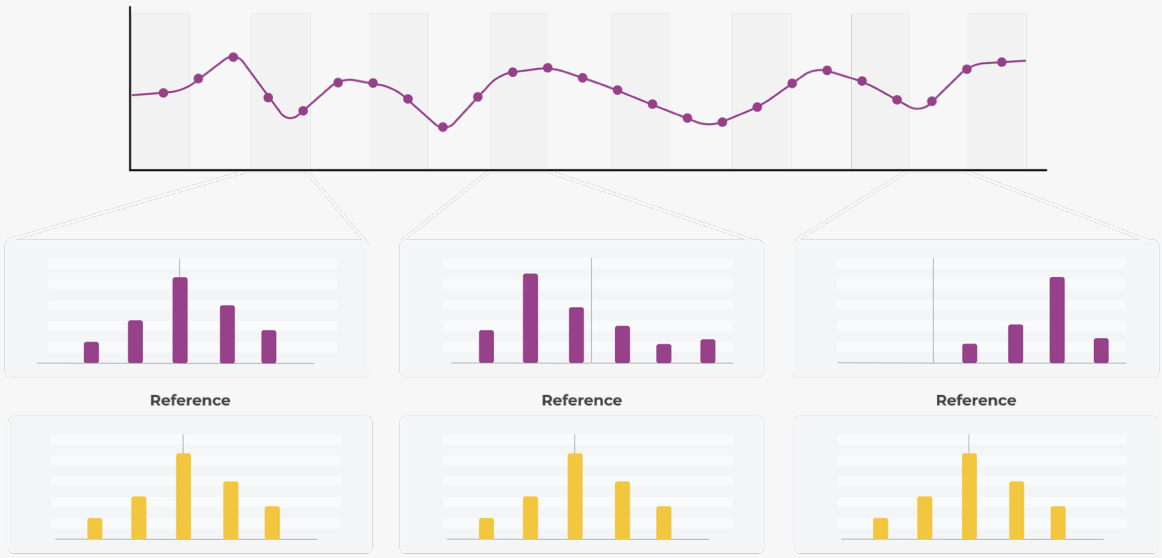
UNIQUE CHALLENGES WITH MOVING WINDOWS:

One unique challenge with looking at distributions over moving windows is that the moving window distribution you compare to can change drastically and have very different distribution points than all the previous samples.



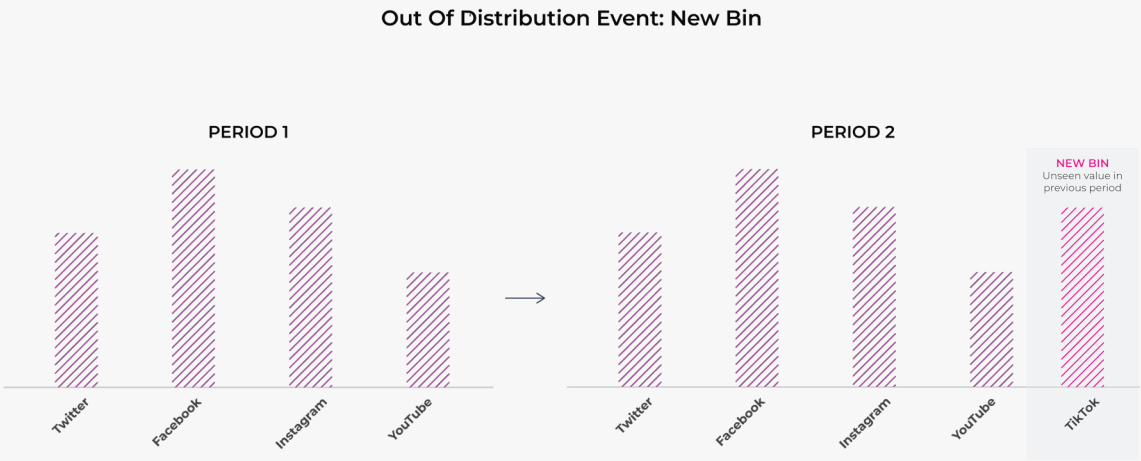
Distribution Checks: Moving Windows

The moving distributions are compared with a reference distribution that is supposed to be a stable larger sample from training.



Distribution Checks: Against Stable Reference

The movement requires solutions to choosing reference distribution bins that handle outliers and out-of-distribution future events. It also implies that, even if a bin has values for something in the reference distribution, in a future distribution those events may no longer be in a bin.

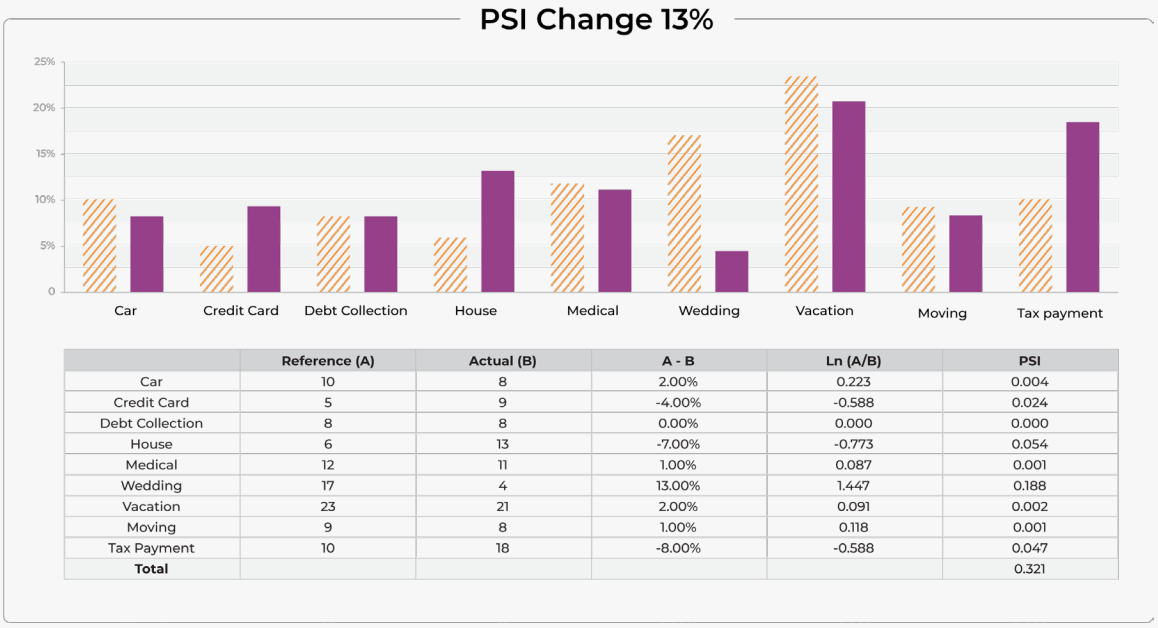


Out of Distribution Event: New Bin

EXAMPLE MOVEMENTS AND STATISTICS OF PSI:

These examples run through how a change in a statistical distance relates to the change in the distribution. The goal is to give teams some intuition as to numbers.

The example below might be the purpose of the loan for a fictitious business. One can imagine post-covid loans to cover Weddings might jump a large percentage versus previous periods. This example shows how that change would be caught by PSI with a setting of 0.25.



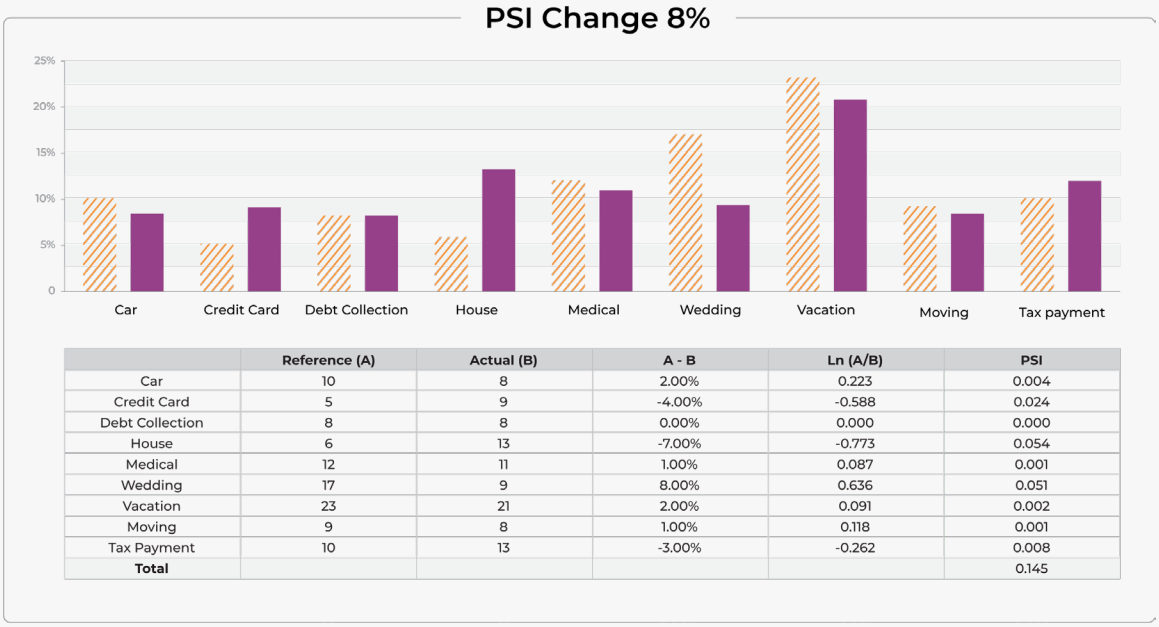
	Reference (A)	Actual (B)	A - B	Ln (A/B)	PSI
Car	10	8	2.00%	0.223	0.004
Credit Card	5	9	-4.00%	-0.588	0.024
Debt Collection	8	8	0.00%	0.000	0.000
House	6	13	-7.00%	-0.773	0.054
Medical	12	11	1.00%	0.087	0.001
Wedding	17	4	13.00%	1.447	0.188
Vacation	23	21	2.00%	0.091	0.002
Moving	9	8	1.00%	0.118	0.001
Tax Payment	10	18	-8.00%	-0.588	0.047
Total					0.321

PSI Example with 13% Change

The example above is meant to show how big of a change in a distribution is needed, to hit the rule of thumb threshold used to benchmark the finance industry.

The above changes generate a PSI over 0.25 which, in financial industries, would require a model rebuild:

- Delta change of one bin in this case is 13% versus an initial bin value of 4%
- There are a three large movements (greater than 5%) between bins of this distribution



	Reference (A)	Actual (B)	A - B	Ln (A/B)	PSI
Car	10	8	2.00%	0.223	0.004
Credit Card	5	9	-4.00%	-0.588	0.024
Debt Collection	8	8	0.00%	0.000	0.000
House	6	13	-7.00%	-0.773	0.054
Medical	12	11	1.00%	0.087	0.001
Wedding	17	9	8.00%	0.636	0.051
Vacation	23	21	2.00%	0.091	0.002
Moving	9	8	1.00%	0.118	0.001
Tax Payment	10	13	-3.00%	-0.262	0.008
Total					0.145

PSI Example with 8% Change

The example above with a smaller change, where the maximum delta is 8% and a PSI value above 0.1 but below 0.25. This example would fall into the finance industries range requiring an investigation of the model, but not above the 0.25 requiring a new model or retrain.