Covariance and Correlation in AI and Anomaly Detection

Shameek Bhattacharjee (Asst. Prof. WMU, Dept. of CS)

Scribe: Shourav Das (UG Student, WMU, Dept. of CS)

This material is part of NSF grant OAC-2017389

Introduction

- Covariance and Correlation measures give important clues about the underlying structure of a data or system behavior.

- Anomalies that are a result of attacks, faults, unusual events disturb the underlying structure of data

- Therefore, a deeper understanding of physical meaning of covariance and correlation is essential to solve real-world anomaly detection problems using Statistical AI

Section Goals

- Understand Physical Meaning of Covariance and Correlation
- Why they are important for anomaly detection
- Different types of Correlation and Covariance Measures
- When to use which correlation measure

Simple Covariance

Module 1 DRAFT Version

Understanding covariance is key because correlation coefficient is just a normalized version of covariance

Normalization of covariance helps to compare regardless of the range of values for the pair of random variables.

Depending on the correlation measure the process of normalization as well covariance quantification will vary.

Physical Meaning

Covariance has a Magnitude and a Sign

Magnitude indicates strength of dependence (proportional) Between r.v. *x* and *y*. The closer it is to zero the weaker the dependence

Sign → indicates whether the dependence (however strong or weak) is direct or inversely related to the other r.v.

Covariance Sign

Positive(+)
If both variables tend to be on the *same side from their respective means* μ_x and μ_y

Negative(-)
If variables tend to be on the *opposite sides from their respective means* μ_x and μ_y

Bottom Line: if both variables tend to increase/decrease at the same sampling round, they have positive covariance
if one variable's increase show decrease of the r.v. (and vice-versa), they have negative covariance

Most basic correlation measure

Pearson's Correlation Coefficient

This is the most basic measure of correlation that is found in all books

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x.\sigma_y}$$

Covariance between random variables x and y

Standard Deviation of random variable x

Standard Deviation of random variable y

- If x and y are independent
  ρ_{xy} = 0
- If x and y are perfectly positively correlated
  ρ_{xy} = +1
- If x and y are perfectly negatively correlated
  ρ_{xy} = -1

Limitation of Pearson's Correlation Coefficient:

Pearson's Correlation Coefficient *only captures linear dependence* between x and y

If two variables are non-linearly dependent, then a zero correlation may be observed.
However, zero correlation does not imply that x and y are not dependent on each other.

Bottom Line: Cannot rely on Pearson's Coefficient if you expect complex relationships (e.g. climate, human behavioral data)

Other Covariance measures Module 1 DRAFT Version

Anomalies caused by triggers from a fraction of components of a system

Anomalies do not agree with the fraction that remains unaffected by the anomaly trigger

This disturbs the space time covariance structure during normal mode of operation

For non-linear systems, we need to capture the covariance structure mathematically but not with pearson's Correlation coefficient

Alternatives to Pearson's Correlation Coefficient

- Distance Correlation
- Maximal Information Coefficient (MIC)

We do not cover Spearman's' or Kendall's Tau coefficient

Distance Correlation

Module 1 DRAFT Version

Physical Meaning

Replace how two variables vary with respect to their respective means,
with how two variables vary with respect to all other points in the dataset

How to calculate?

Both R and Python have libraries for Distance Correlation. Check energy package in R.

$$Dist\rho_{xy} = \frac{Cov_{(d)}(x,y)}{\sigma_{(d)_x} \cdot \sigma_{(d)_y}}$$

Distance Covariance between random variables x and y

Self Distance Covariance

Limitation?

May not work accurately for discrete r.v.s

Essence of Mutual Information

$$MI\ (x, y) = D_{KL}\ [\ P(x, y)\ ||\ P(x).P(y)\ ]$$

Mutual Information is *the* *Kullback Leibler Divergence* between the joint distribution and product of marginal distributions

Joint Distribution

Product of Marginal Distributions

Kullback Leibler divergence (also called relative entropy)

is one of the popular ways to quantify the difference between any two distributions A and B defined over the probability space s. Mathematically,

$$D_{KL}(\mathrm{A}\ ||\ \mathrm{B}) = \sum_{A \in s} p(A)\ (\log\ (\frac{p(A)}{p(B)}))$$

Maximal Information Coefficient (MIC)

Module 1 DRAFT Version

For discrete random variables, the distance correlation may not work accurately.

We need a third measure that works for both discrete (and continuous r.v.) and capture non linear dependence.

Maximal Information Coefficient (MIC)

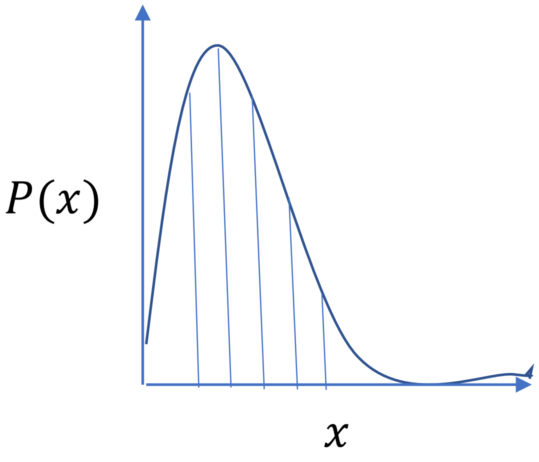
captures non-linear dependence between discrete and continuous random variables

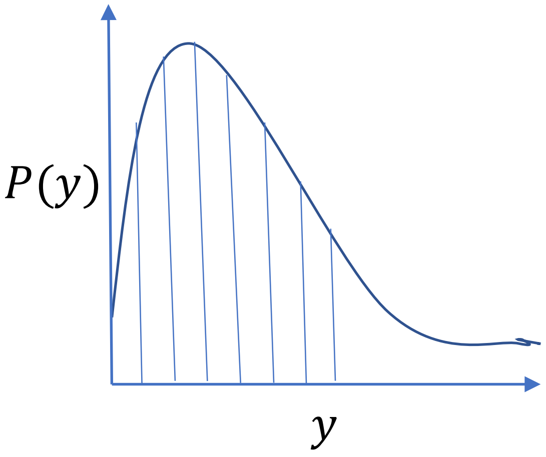
$$MIC_{xy} = \frac{MI(x,y)}{log_2(\min(b_x, b_y))}$$

Is the Mutual Information (MI) between x and y

MI → see next slide

b_x and b_y are Number of levels (or bins) in prob. distribution of x and y





MIC and Covariance

If x and y are truly independent, then it is well known that joint distribution equals the product of the marginals

$$P(x, y) = P(x).P(y)$$

If not, then $P(x, y) > \ P(x).P(y)$

Application: An otherwise provably correlated system when disturbed due to threats should see a drop/change in the MIC under certain assumptions.

Bigger the difference between joint distribution and the product of the marginal distributions → indicates more strong dependence of x with y

Hence if $P(x, y) > \ P(x).P(y)$ then $MI(x, y) \neq 0$

The denominator of MIC_{xy} is just a normalizing term, but $b_x \ * \ b_y \ \leq (SampleSize^{0.6})$

The larger the MIC, the stronger the correlation.

Limitation: Does not work well under low sample sizes in the distributions involved