Machine Learning for Industrial Anomaly Detection using Bivariate Gaussian Probability Density Function: A Case Study

Chinnaraji Annamalai
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur, India
Email: anna@iitkgp.ac.in
https://orcid.org/0000-0002-0992-2584

Abstract: In the field of Machine Learning (ML), transitioning from univariate monitoring to a multivariate framework allows for a sophisticated analysis of complex industrial systems. While a univariate model analyzes a single feature, such as temperature, in isolation, the Bivariate Gaussian Probability Density Function (PDF) enables the simultaneous evaluation of two related sensors, such as temperature and humidity, to establish a dynamic mathematical baseline for operational normality. This research presents a paradigm shift in industrial monitoring, specifically within server room environments, by transitioning from rigid, static threshold systems to a framework of probabilistic intelligence. By integrating multidimensional statistical baselines and correlation coefficients, the model can distinguish between benign stochastic fluctuations and genuine early-stage hardware malfunctions with high granularity. The study demonstrates that by understanding how sensors dance together through covariance, systems can identify contextual anomalies that individually appear normal but statistically violate the learned operational relationship.

Keywords: Automated Monitoring**,** Probabilistic Intelligence, Statistical Baseline, stochasticity

1. Introduction

Traditional industrial monitoring architectures often rely on rigid, binary limits that trigger false alarms or fail to detect subtle degradations. In complex environments like server rooms, variables such as temperature and humidity are not independent; they are governed by physical processes that create a predictable relationship [1-8]. This research presents a paradigm shift in industrial monitoring by transitioning from static threshold systems to a framework of probabilistic intelligence. By integrating bivariate Gaussian distributions into Machine Learning (ML) frameworks [9-13], the system moves from simple range checking to contextual anomaly detection.

2. The Mathematical Foundation

The core of this framework is based on the joint probability of two random variables, x and y.

2.1. Theorem of Bivariate Normal Distribution

Two random variables x and y follow the standard bivariate normal distribution with a correlation coefficient ρ if their joint Probability Density Function (PDF) is defined as:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{\left(y-\mu_y\right)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)\left(y-\mu_y\right)}{\sigma_x\sigma_y}\right)\right]$$

where $\rho \in [-1, 1]$ or $|\rho| \leq 1$, where $\rho = 1$ a perfect positive linear relationship, $\rho = -1$ indicates a perfect negative linear relationship, and $\rho = 0$ indicates no linear correlation.

- **Positive Correlation ($\rho > 0$):** As x increases, y tends to increase.
- **Negative Correlation ($\rho < 0$):** As x increases, y tends to decrease.
- **Independence ($\rho = 0$):** The variables x and y are independent.

Sample covariance (Bessel's Correction) is the standard measurement used throughout Statistics, Data Science, and scientific experiments,

$$Cov(x, y) = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{n - 1}$$

Population covariance is primarily utilized in Probability Theory and Census data analysis,

$$Cov(x, y) = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{n}$$

Where n represents the sample size or the total number of paired observations (e.g., simultaneous readings of temperature and humidity) collected during a specific period.

The Correlation Coefficient (ρ) is essentially a normalized version of covariance. It is defined as:

$$\rho = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y} \Rightarrow Cov(x, y) = \rho \sigma_x \sigma_y$$

2.2. Derivation of Bivariate Gaussian PDF

The bivariate Gaussian probability density function [10-13] in general is:

$$f(\boldsymbol{x}) = \frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

Covariance Matrix and Determinant:

Given the covariance Matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x, y) \\ \mathrm{Cov}(x, y) & \mathrm{Var}(y) \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{pmatrix}$$

The determinant $|\boldsymbol{\Sigma}|$ is:

$$|\boldsymbol{\Sigma}| = (\sigma_x^2)(\sigma_y^2) - (\rho \sigma_x \sigma_y)^2 = \sigma_x^2 \sigma_y^2 (1 - \rho^2)$$

Then, the normalization constant is:

$$\frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} = \frac{1}{2\pi \sqrt{\sigma_x^2 \sigma_y^2 (1 - \rho^2)}} = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}}$$

Inverse Covariance Matrix:

The inverse matrix is

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{|\boldsymbol{\Sigma}|} \begin{pmatrix} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{pmatrix} = \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)} \begin{pmatrix} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{pmatrix} \text{ or}$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} {}^{1}\!/_{\sigma_x^2} & {}^{-\rho}\!/_{\sigma_x \sigma_y} \\ {}^{-\rho}\!/_{\sigma_x \sigma_y} & {}^{1}\!/_{\sigma_y^2} \end{pmatrix}.$$

Quadratic Form Expansion:

Let $\Delta x = (x - \mu_x)$ and $\Delta y = (y - \mu_y)$. Then, $(\boldsymbol{x} - \boldsymbol{\mu}) = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$ and $(\boldsymbol{x} - \boldsymbol{\mu})^T = (\Delta x, \Delta y)$.

The quadratic form $(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$ expands to

$$(\Delta x, \Delta y)\frac{1}{1-\rho^2}\begin{pmatrix} {}^{1}/_{\sigma_x^2} & {}^{-\rho}/_{\sigma_x\sigma_y} \\ {}^{-\rho}/_{\sigma_x\sigma_y} & {}^{1}/_{\sigma_y^2} \end{pmatrix}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}$$

$$= \frac{1}{1-\rho^2}(\Delta x, \Delta y)\begin{pmatrix} \left({}^{1}/_{\sigma_x^2}.\Delta x\right) + \left({}^{-\rho}/_{\sigma_x\sigma_y}.\Delta y\right) \\ \left({}^{-\rho}/_{\sigma_x\sigma_y}.\Delta x\right) + \left({}^{1}/_{\sigma_y^2}.\Delta y\right) \end{pmatrix}$$

$$= \frac{1}{1-\rho^2}\left(\frac{(\Delta x)^2}{\sigma_x^2} - \frac{\rho\Delta x\Delta y}{\sigma_x\sigma_y} - \frac{\rho\Delta x\Delta y}{\sigma_x\sigma_y} + \frac{(\Delta y)^2}{\sigma_y^2}\right)$$

$$\therefore (\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \frac{1}{1-\rho^2}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\frac{\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)$$

The final PDF:

The general Bivariate Normal Distribution formula is

$$f(\boldsymbol{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

Substituting the determinant and the quadratic form back into the formula, the bivariate normal PDF is:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right]$$

Hence, theorem is proved.

3. Case Study: Environmental Monitoring of a Server Room

To demonstrate the practical utility of this model, we apply the Bivariate Gaussian PDF to the monitoring of temperature (x) and humidity (y) within an industrial server room.

3.1. Parameter Estimation (Training Phase)

During a stable operational period, the model analyzes historical sensor data to establish a baseline of normality. Using Maximum Likelihood Estimation (MLE) [14-16], the following parameters are derived:

- **Mean Temperature** ($\mu_x = 22^{\circ}$C) and **Mean Humidity** ($\mu_y = 50\%$)
- **Temperature Variance** ($\sigma_x^2 = 2$) and **Humidity Variance** ($\sigma_y^2 = 25$)
- **Standard deviations**: $\sigma_x = \sqrt{2} \approx 1.414$; $\sigma_y = \sqrt{25} = 5$; $\sigma_x\sigma_y = 7.07$

Assume we have collected $n = 10$ pairs of sensor readings (Temperature x_i and Humidity y_i) from the server room and calculated the sum of the products of their deviations as 44.55.

$$Cov(x,y) = \frac{\sum_{i=1}^{10}(x_i-\mu_x)(y_i-\mu_y)}{10-1} = \frac{44.55}{9} = 4.95$$

$$\rho = \frac{\mathrm{Cov}(x,y)}{\sigma_x\sigma_y} = \frac{4.95}{7.07} = 0.7$$

$$1-\rho^2 = 1-0.49 = 0.51\ and\ \sqrt{1-\rho^2} = \sqrt{0.51} \approx 0.714$$

To calculate the anomaly detection threshold, epsilon (ε), using the Three-Sigma Rule (**68-95-99.7 rule**), we must determine the range of values that cover 99.7% of the normal operational distribution. In a Bivariate Gaussian framework, this range is represented as an elliptical boundary rather than a simple linear range. However, calculating the threshold for each individual variable (x and y) based on $\mu \pm 3\sigma$ allows us to define the "band of normality"

$$\mu_x - 3\sigma_x = 22 - 3(1.414) = 17.758^{\circ}\mathrm{C} \text{ and } \mu_x + 3\sigma_x = 22 + 3(1.414) = 26.242^{\circ}\mathrm{C}$$

$$\mu_y - 3\sigma_y = 50 - 3(5) = 35\% \text{ and } \mu_y + 3\sigma_y = 50 + 3(5) = 65\%$$

$$\varepsilon = f(17.758, 35) = f(26.242, 65) \approx 0.000315$$

The system compares the sensor reading $f(x, y)$ against the statistically grounded threshold, epsilon ($\varepsilon = 0.000315$). This value (ε) acts as the "Score of Suspicion" limit.

- **Normal operational noise:** Any sensor reading that results in $f(x, y) \geq 0.000315$ is accepted as standard.
- **Anomaly:** Any reading where $f(x, y) < 0.000315$ is flagged as a genuine hardware malfunction because it falls outside the 99.7% band of normality.

3.2. Contextual Anomaly Detection

The power of the bivariate approach lies in its ability to identify "impossible combinations":

For example, let's consider the current reading: $x = 24, y = 58$.

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right]$$

$$x - \mu_x = 24 - 22 = 2 \text{ and } y - \mu_y = 58 - 50 = 8$$

$$f(24, 58) = \frac{1}{2\pi(7.07)(0.714)} \exp\left[-\frac{1}{2(0.51)}\left(\frac{(2)^2}{(1.414)^2} + \frac{(8)^2}{(5)^2} - \frac{2(0.7)(16)}{7.07}\right)\right] \approx 0.008$$

$$f(24, 58) \approx 0.008.$$

Benign Stochasticity: A reading of $24^{\circ}\mathrm{C}$ and 58% humidity yields a PDF value $f(24, 58) \approx 0.008$. Since this follows the positive correlation, it is treated as normal operational noise.

Contextual Anomaly: A reading of $19^{\circ}\mathrm{C}$ and 65% humidity might individually seem safe, but because the relationship is broken (low temp should not have high humidity in this setup), $f(x, y)$ drops below ε and triggers an alert.

4. Conclusion

Integrating Multivariate Gaussian PDFs into ML architectures provides a precise, adaptive, and scalable lens for industrial monitoring. By shifting from binary limits to continuous probabilistic inference, systems can distinguish between standard operational noise and genuine hardware failures with high granularity.

References

[1] Annamalai, C. (2023) Factorial Theorem: An Alternative to Gamma Function. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4376857.

[2] Annamalai, C. (2022) Factorials, Integers, and Factorial Theorems for Computing and Cryptography. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4189241.

[3] Annamalai, C. (2023) Factorial Theorem for Computing the Factorial of Positive Real Number. *COE, Cambridge University Press.* https://dx.doi.org/10.33774/coe-2023-7p472.

[4] Annamalai, C. (2022) Combinatorial Theorems in Factorials with Multinomial Computation. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4185700.

[5] Annamalai, C. (2023) Review on the Gamma Function and Error Correction. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4397585.

[6] Annamalai, C. (2023) Computation of the Gamma Function of Half using the Gaussian Integral. *Zenodo*. https://doi.org/10.5281/zenodo.13849817.

[7] Annamalai, C. (2023) Gamma Function derived from Factorial Function based-Pi Function. *COE, Cambridge University Press*. https://dx.doi.org/10.33774/coe-2023-wv4rc.

[8] Annamalai, C. (2023) Gamma Function derived from Natural Logarithm and Pi Function. *COE, Cambridge University Press*. https://dx.doi.org/10.33774/coe-2023-wv4rc-v3.

[9] Annamalai, C. (2023) An Alternative Method for the Gamma Function derived from Natural Logarithm and Pi Function. *COE, Cambridge University Press*. https://doi.org/10.33774/coe-2023-wv4rc-v4.

[10] Annamalai, C. (2004) The Gaussian Integral. *The ResearchGate*. https://doi.org/10.13140/RG.2.2.10969.33122.

[11] Annamalai, C. (2026) The Application of Gaussian Probability Density Functions in Machine Learning for Industrial Anomaly Detection. *ResearchGate*. http://doi.org/10.13140/RG.2.2.16407.94889.

[12] Annamalai, C. (2026) Computation and Prediction of Gaussian Probability Density Functions in Machine Learning for Industrial Anomaly Detection. *Zenodo*. http://doi.org/10.5281/zenodo.19699189.

[13] Annamalai, C. (2026) Machine Learning for Industrial Anomaly Detection: A Probabilistic Intelligence Approach. *COE, Cambridge University Press*. https://doi.org/10.33774/coe-2026-3bzkc.

[14] Annamalai, C. (2026) Multivariate Gaussian Probability Density Function in Machine Learning for Industrial Anomaly Detection. *COE, Cambridge University Press*. https://doi.org/10.33774/coe-2026-lgqh5.

[15] Annamalai, C. (2026) The Multi-Dimensional Role of the Softmax Function in Computational Systems. *ResearchGate*. http://doi.org/10.13140/RG.2.2.32234.89285.

[16] Annamalai, C. (2026) The Mathematical Convergence of Softmax and Bayes' Theorem in Predictive Modeling. *ResearchGate*. http://doi.org/10.13140/RG.2.2.16611.11045.

[17] Annamalai, C. (2026) Stochastic Foundations of Large Language Models: Bridging Neural Networks and Frequentist Inference. *ResearchGate*. http://doi.org/10.13140/RG.2.2.25025.93280.