Comprehensive Description of Uncertainty in Measurement for Representation and Propagation with Scalable Precision

Ali Darijani[1,2], Jürgen Beyerer[1,2], Zahra Sadat Hajseyed Nasrollah[1,2], Luisa Hoffmann[3], and Michael Heizmann[3]

[1]Fraunhofer IOSB
[2]KIT, IES
[3]KIT, IIT

October 30, 2026

Probability theory has become the predominant framework for quantifying uncertainty across scientific and engineering disciplines, with a particular focus on measurement and control systems. However, the widespread reliance on simple Gaussian assumptions—particularly in control theory, manufacturing, and measurement systems—can result in incomplete representations and multistage lossy approximations of complex phenomena, including inaccurate propagation of uncertainty through multi stage processes.

This work proposes a comprehensive yet computationally tractable framework for representing and propagating quantitative attributes arising in measurement systems using *Probability Density Function*s (PDFs). Recognizing the constraints imposed by finite memory in software systems, we advocate for the use of *Gaussian Mixture Model*s (GMMs), a principled extension of the familiar Gaussian framework, as they are universal approximators of PDFs whose complexity can be tuned to trade off approximation accuracy against memory and computation. From both mathematical and computational perspectives, GMMs enable high performance and, in many cases, closed form solutions of essential operations in control and measurement.

The paper presents practical applications within manufacturing and measurement contexts especially circular factory, demonstrating how the GMMs framework supports accurate representation and propagation of measurement uncertainty and offers improved accuracy–compared to the traditional Gaussian framework–while keeping the computations tractable.

1. Introduction

In this paper, we present a consistent, generic framework to efficiently represent and propagate measurement uncertainties in industrial systems with scalable accuracy. When measuring quantitative attributes, it is not sufficient to just report a numerical value for the measurement result. Only together with a quantitative description of the measurement uncertainty about the measured attribute value does the result become usable. According to the state of the art in metrology, uncertainty is described on the basis of a number with the semantics of a standard deviation or multiples thereof [21]. To reduce stochastic errors, the measurement result usually emanates from an averaging calculation on the observed data. From a probabilistic point of view, with the measurement result and the uncertainty in form of a mean value and a standard deviation, respectively, a Gaussian distribution for perturbations is implicitly assumed regarding the formation of observable data based on the underlying measurand. In the Guide to the Expression of Uncertainty in Measurement [21], however, this methodology is used not only to describe aleatory uncertainties (due to randomness), but also to represent epistemic uncertainties (due to lack of knowledge), e.g. to quantify systematic but unknown errors. The interpretation of probability used then corresponds to a *Degree of Belief* (DoB), which allows non-random uncertainties and expert knowledge to be represented and treated mathematically consistently using the calculus of probability. The DoB is the more general interpretation of probability that is compatible with Kolmogorov's axioms and allows a unified quantitative description of aleatory and epistemic uncertainties.

The reason why in this paper, as well as by the metrological community, uncertainties are regularly described by means of probabilities and alternative calculus for describing uncertainty do not play a role is explained in Section 3.1 and Section 3.2. In practice, there are many cases where a Gaussian assumption (implicitly or explicitly) is not adequate; see Section 4.2. In particular, in the propagation of measurement results and their uncertainties through multi-stage systems, errors accumulate along the calculations due to the simple assumption of Gaussian distributions. Consequently, the uncertainty would then ideally have to be described by means of a complete PDF, but this is generally mathematically very challenging in practice; see Section 4.2.2. In this paper, we propose to describe PDFs using GMMs. These are universal approximators for PDFs [30][28][18], where the precision of the approximation can be specifically adjusted by the number of components of the GMM; see Section 5.1.1. This leads to a finitely dimensional, usually even low-dimensional description of the PDFs using the parameter tuple GMM.

We even suggest that all uncertain quantitative attributes that play a role in a measurement process should be described as PDF in the form of GMM, i.e. GMM should be regarded as the "native data type" of such quantities. Many practically relevant operations, such as the addition of quantities, Bayesian fusion of different information with respect to a quantity of interest, the mixing of similar quantities from different sources, the focusing on individual quantities from a set of different quantities, and the exploitation of observed quantities for inference with respect to unobserved quantities, can be performed by algebraic operations on the parameter tuples of the GMMs with simple formulas, very efficiently and accurately; see Table 1. For operations between quantities for which an algebraic calculation of the parameters of the GMM of the result is not possible, a mapping to a description using GMMs can be accomplished by efficient Monte Carlo sampling and subsequent fitting of a GMM using the *Expectation Maximization* (EM) algorithm; see Section 5.1.4 and Section 5.1.2. Explanatory examples for the operations in Table 1 are discussed in Section 6.

In addition to the uniform description of uncertain quantities using GMMs, measuring instruments and systems can also be described in a scalable and probabilistic manner also by using GMMs; see Section 7. Based on this representation, the change in the uncertainty of a quantity of interest due to a measurement operation can also be calculated purely on the basis of algebraic operations on the parameter tuples of GMMs. From a Bayesian point of view, prior knowledge about the quantity of interest can be embodied as GMM. The observation of measurement data converts the GMM of the measurement system into the likelihood function in the form of a GMM, which can then be fused algebraically using Bayes' theorem with the prior knowledge to achieve the posterior PDF in the form of a GMM as the final measurement result. Hence, measurement processes can be described by means of algebraic operations on the parameter tuples of GMMs.

At first glance, the formula apparatus in Table 1, together with the mathematical details listed in Section A, could be a deterrent for practical use. In terms of software, the representations and operations can be encapsulated in detail and the user is shielded from them. The user can handle the quantities in the usual manner and only in the background the quantities are treated and calculated as GMMs. In multi-stage systems, this can be used to calculate more accurately internally, whereby a reversal to the usual description by means of a mean value and a standard deviation is possible at any time and can be calculated algebraically from the GMM without any problems; see Section 5.1.7.

2. Mastering Uncertainty within the Circular Factory

In the Circular Factory, full-fledged products are made mixed from new and used parts. Since used parts have different histories and usage durations behind them, it does not make sense to want to represent the ensemble of a certain component type as a statistical population with only one single PDF with regard to its quality-relevant and function-relevant properties (attributes). Rather, an individual probabilistic description of the relevant properties is required for each component instance. For continuous attributes, this can be achieved with scalable accuracy and efficient handling by means of component-specific GMMs. These GMMs are determined in such a way that they reflect the individual level of information about the attributes of a specific component, including the associated uncertainties and correlations.

In the Circular Factory probability is consistently interpreted as DoB, so that probabilities based on statistical evidence as well as subjective expert knowledge and opinions can be described and handled uniformly. Each instance of a component is therefore linked to an individual data structure that reflects the current level of information about the quantitative attributes relevant to function and quality in the form of GMM parameter tuples.
If further information about a component instance is to be added, e.g. through a measurement, the stored GMM is understood as the prior distribution and the additional information as a likelihood. Both are fused in the Bayesian sense to the posterior distribution in the form of an updated GMM.

Overall, this approach allows all relevant component attributes, including their aleatoric as well as epistemic uncertainties, to be accurately represented in an instance-individually adjustable manner. In addition, all important operations regarding uncertainties and their propagation can be carried out very efficiently in the Circular Factory, mainly by means of closed calculations and consistently based on GMMs.

3. Uncertainty Representation and the Probability Framework

Probabilities provide a consistent and powerful framework for representing uncertainty and the subsequent decision making and reasoning under it. Its adaptation by both practitioners and theoreticians comes from its historical success and the math that powers it.

3.1 Philosophical Aspects and Historical Success

In modern era much of the philosophical aspect success of the probability is due to [12]. Finetti showed by a specific reward function based on a decision that is to be made under uncertainty, one must adhere to probability to maximize their reward. Later Lindley relaxed the assumptions on the reward function which ascertained the place of probability in uncertainty quantification [26]. Interestingly, he also challenged those proposing other nonprobabilistic frameworks–mainly Lotfi Zadeh and Arthur Dempster– arguing that everything that can be done within their frameworks can be done better using probability [27].

Within probability there are different interpretations–mainly classical [29], frequentist [37], Bayesian [2]–which lead to different formulations and all having their pros and cons. Classical interpretation is fairly limited and can only answer simple queries such as "What is the probability that the outcome of a dice roll is an odd number given that we have a fair dice?" which requires idealization of the experiment as hinted by the assumption that the dice is fair which is also known as the equally likely assumption and quite strong symmetry. The frequentist interpretation, also known as the asymptotic interpretation, defines probability as the long run relative frequency of an event in repeated identical trials which is limited as one cannot hope to have identical trials even for two experiments let alone quite a large number of them.

The Bayesian interpretation or also known as the subjective interpretation formulates the DoB one might have about the outcome of an experiment as a PDF and would update the PDF given new experimental results or new data through the Bayes' theorem. Two main advantages of the Bayesian interpretation are that it can draw conclusion based on zero number of data points and it is suitable for online learning and online inference. The challenge is that the decisions drawn from the Bayesian interpretation when having small amount of data depends on the prior knowledge or the expert knowledge, formulated as a DoB PDF, and can only become better by having a better expert or by acquiring new data. In the context of scientific machine learning, measurement systems, and physical systems there is a strong inclination to use the Bayesian interpretation due to lack of data and the expensive cost of performing experiments.

3.2 Mathematical Aspects

Probability theory does not only have historical success in solving real world problems, but also has solid mathematical foundations. It possesses a well defined calculus–Kolmogorov's formulation [24]–based on measure theory and by extension analysis [34]. All the interpretations described in Section 3.1 are compatible with Kolmogorov's axiomatic foundations of probability theory and with the corresponding probability calculus [4]. This foundation forms the basis of modern statistical inference and probabilistic machine learning.

4. Uncertainty in Measurement, the Gaussian Framework and its Insufficiency

Measurement uncertainty [15] is usually being handled using the Gaussian framework. While the Gaussian framework has been a powerful tool it has inherent limitations in nonlinear systems and multimodal situations. According to [20], measurement uncertainty is determined, represented, and propagated based on standard deviations (root of variance). Restricting uncertainty to the notion of standard deviation implicitly induces a Gaussian representation by virtue of the principle of maximum entropy [9].

4.1 The Gaussian Framework

Under the Gaussian assumption, measurement uncertainty is represented by a normal distribution characterized by its mean and standard deviation. This model implies that most measurements cluster around the mean, with decreasing probabilities for values farther away. Consequently, the uncertainty can be quantified and visualized effectively using confidence intervals.

4.1.1 The Advantages of the Gaussian Framework

Assuming a Gaussian PDF offers theoretical and computational benefits. Given an expected value and a variance a Gaussian is the representation that stays impartial or uncertain the most [9] which for example in the context of reliability models means they become safer to use as they would overestimate the danger rather than underestimate it. If only the expected value m and the variance σ^2 are known about a *Random Variable* (RV) X, a Gaussian $\mathcal{N}_X(x|m, \sigma^2)$ is that PDF with these parameters with the maximum entropy [9]. Therefore, describing measurement results based on a mean value and measurement uncertainty by a variance implicitly assumes a Gaussian distribution. Gaussian formulation has many nice properties [36][31][14] that make decision making and inference faster and more robust.

4.2 The Insufficiency of the Gaussian Framework

Gaussian models assume symmetry, light tails, linearity, and additive, homoscedastic noise. Real world data often exhibit skewed bahavior, heavy tails, multimodality, constraints, discontinuities, outliers, and nonlinear dynamics, violating these assumptions and therefore leading to miscalculation of risk and other values of interest, if Gaussian distributions are assumed.

4.2.1 Inherently Non Gaussian Physics

While at the macro scale and low velocity most of the physical phenomena that exhibit randomness are modeled using the Gaussian framework, in smaller scales and high velocity the Gaussian framework fails. To name a few Relativistic Breit-Wigner distribution [6], quantum mechanical operators [33], electromagnetic operators [19] cannot be handled using the Gaussian framework.

4.2.2 Nonlinear Transformation

Nonlinear transformation of inputs that are distributed according to a Gaussian in general are non Gaussian. The sensitivity analysis which is based on the first order Taylor series is insufficient to propagate the uncertainty forward and a loss of information will occur. The canonical example is that if two variables are normally distributed around zero, their ratio is Cauchy distributed and as the expectation value and the variance are not defined for the Cauchy distribution, therefore it is impossible to approximate it using a Gaussian distributed variable.

4.2.3 Mixture of Sources

If we have different sources for observing values for one and the same quantity X, and if the sources are independent and active with different probabilities, a common PDF for the pool of these sources can be established in the form of the convex linear combination of the PDFs of the individual sources. In general, a linear combination of Gaussian PDFs does not result in a Gaussian PDF, thus, the Gaussian framework is not sufficient anymore in this case [7][35].

4.2.4 Constraint Handling

As an example, modeling a positive physical quantity like length within the Gaussian framework might cause problems when propagated using expansion methods or Monte Carlo methods if the mean is close to zero and the standard deviation is quite large. In general hard constraints cannot be directly incorporated into the Gaussian framework due to the infinite support of the domain of the Gaussian variables.

4.2.5 Abstract Non Gaussian Distributions

Beyond physical quantities, there exist abstract non Gaussian probability distributions that have diverse applications in different disciplines. These distributions play a crucial role in various fields, offering tools for modeling complex phenomena that do not conform to traditional Gaussian assumptions. Non Gaussian phenomena in finance [22] is probably the one that cannot be categorized under any physical phenomena and still exhibit non Gaussian behaviour.

5. Gaussian Mixture Model and Its Advantages

GMMs are the natural extension of the Gaussian framework which are simply the weighted sum of finitely many Gaussian PDFs. A list of nice properties of GMMs is presented in the Table 1. Doing statistical inference under the GMM assumption provides mathematical and computational benefits. It is worth giving an exact mathematical definition of a GMM before discussing its advantages as it provides a solid ground for the subsequent discussions.

Definition 5.1. *Let*

$$
\begin{aligned}
&d, K \in \mathbb{N} \setminus \{0, +\infty\} \\
&x \in \mathbb{R}^d, m_i \in \mathbb{R}^d \text{ for } i = 1, \ldots, K \\
&\{\Sigma_i\}_{i=1,\ldots,K} \subset \{A \in \mathbb{R}^{d\times d} \colon x^\top \cdot A \cdot x > 0 \text{ for } x \neq 0\} \\
&\pi_i \in \mathbb{R}_+ \text{ for } i = 1, \ldots, K \wedge \sum_{i=1}^{K} \pi_i = 1,
\end{aligned}
\tag{1}
$$

then the mapping

$$
\mathcal{G}_X(x|K, \{\pi_i, m_i, \Sigma_i\}_{i=1,\ldots,K}) \coloneqq \sum_{i=1}^{K} \pi_i \mathcal{N}_X(x|m_i, \Sigma_i), \tag{2}
$$

is called a Gaussian Mixture or a GMM.

Remark 5.1 (Notation)**.** *If we encapsulate all the parameters required to define a GMM then:*

$$
\begin{aligned}
&g \coloneqq (K, \{\pi_i, m_i, \Sigma_i\}_{i=1,\ldots,K}) \\
&\mathcal{G}_X(x|g) \coloneqq \mathcal{G}_X(x|K, \{\pi_i, m_i, \Sigma_i\}_{i=1,\ldots,K}).
\end{aligned}
\tag{3}
$$

Remark 5.2 (Notation)**.** *When the Random Variable X is Gaussian Mixture distributed, we write $X \sim \mathcal{G}_X(x|g)$.*

Doing statistical inference under the GMM assumption provides mathematical and computational benefits.

5.1 Mathematical Arguments

5.1.1 Universal Density Approximation

GMM is a universal PDF approximator. For a target PDF $p_X(x)$ having enough regularity and an appropriate dissimilarity measure $\mathrm{d}(\cdot, \cdot)$, a series of GMMs $\mathcal{G}_X(x|K, \{\pi_i, m_i, \Sigma_i\}_{i=1,\ldots,K})$ exist such that $\mathcal{G}_X(x|g)$ converges to $p_X(x)$: $\lim_{K\to+\infty} \mathrm{d}(p_X(x), \mathcal{G}_X(x|g))$. i.e., by increasing the number of components, a GMM can approximate the target PDF $p_X(x)$ with any required precision [28][30].

5.1.2 Fitting a Gaussian Mixture Model

One of the standard ways to fit a GMM to data is through an iterative algorithm called EM [11]. EM algorithm is highly efficient in the case of a GMM as it provides closed form updates at each iteration [5]. The convergence is also guaranteed due to the celebrated Banach fixed point theorem.

5.1.3 Closure Under Important Operations

GMMs are closed under fundamental operations that are of interest in statistics. Affine transformations preserve mixture form, allowing propagation through linear mappings without approximation. Marginals and conditionals of a joint GMM remain a GMM. Exact formulas in Proposition B.1 and Proposition B.2 enable exact uncertainty quantification of the variables. The product of two GMMs is an unnormalized GMM which enables closed form calculation of some dissimilarity measures and more importantly the fusion based on Bayes' theorem. Convolution of two GMM is again a GMM with the explicit formula as seen in Proposition A.5 in Section A.

5.1.4 High Quality Generative Property

Sampling from a GMM is a straightforward process. Using composition method [32] and the Box-Muller [32] transform high quality samples can be drawn from a GMM. The composition method selects the component at random and then a sample is drawn from the corresponding Gaussian making a two step sequential process requiring $d + 1$ RVs. This makes GMM an attractive model as a generative model which has applications in data augmentation, data completion and data imputation.

5.1.5 Finite Dimensional

Statistical inference over general probability distributions is inherently an infinite-dimensional problem, as it involves reasoning over function spaces with no finite representation. Gaussian mixture framework provides a way to represent distributions with a finite set of parameters. Having a finite set of parameters by itself makes it easier to do inference since calculations in infinite dimensional spaces are inherently more difficult and theoretically involved [8].

5.1.6 High Dimensional Integrals

Many statistical properties and operations–such as marginalization, conditioning, expectations, and entropy–are fundamentally defined in terms of integrals of functions of PDFs [23] over usually infinite domains. In high dimensional spaces, these integrals are often intractable for general distributions and suffer from the curse of high dimensionality [38][39]. However, when working within the Gaussian mixture framework, a remarkable number of these operations including but not limited to marginalization, conditioning, expectation, and variance admit closed form or highly efficient analytic approximations.

5.1.7 Gaussian Fallback

GMMs provide a powerful yet flexible framework for representing complex distributions while preserving compatibility with the Gaussian framework. When configured with a single component, the Gaussian mixture framework reduces to the Gaussian framework, allowing users to opt for simplicity when needed. A Gaussian distribution can approximate a Gaussian mixture up to the second central moment. The expectation and the variance of a mixture can be calculated in closed form [1] and used as parameters of an equivalent Gaussian. This enables seamless fallback to the familiar Gaussian framework. The expectation and the variance of the correspondent Gaussian are given in the following propositions.

Proposition 5.1 (GMM Expectation)**.**

$$\mathbb{E}(X \sim \mathcal{G}_X(x|g)) = \sum_{i=1}^{K} \pi_i m_i \tag{4}$$

Proposition 5.2 (GMM Covariance Matrix)**.**

$$\mathbb{V}(X \sim \mathcal{G}_X(x|g)) = \sum_{i=1}^{K} \pi_i (\Sigma_i + m_i \cdot {m_i}^\top) - (\sum_{i=1}^{K} \pi_i m_i) \cdot (\sum_{i=1}^{K} \pi_i m_i)^\top \tag{5}$$

5.2 Computational Arguments

5.2.1 Parallelizability of Gaussian Mixture Fitting

EM [11] is often used to fit a GMM to data which innately is parallelizable. The E-step is what the high performance computing community calls "embarrassingly parallelizable" [16] which roughly means that if you have N computing resources the runtime is going to be $1/N$ times the runtime it would taken using only 1 computing resource. The M-step is also embarrassingly parallelizable as it only uses sum reduction. This results in a linear speedup within the iteration step. Streaming EM enable online updates making it invaluable for real time scenarios. Implementations fit well with distributed computing which in turn result in scaling being feasible.

5.2.2 Small Storage Footprint

On a computer one needs 2 integers, the number of components K and the dimension d, and $K-1+Kd+1/2Kd(d+1)$ [4] floating point variables to fully define a Gaussian mixture. This allows for a compact and efficient representation and access of the model both at fitting and inference time. Due to the additive nature of the Gaussian mixture it is even possible to split the model and store it on different nodes for again both at fitting and inference time.

5.3 Closed Form Calculations

By representing attributes and their uncertainties by GMMs, many practical relevant operations with attributes can be calculated algebraically on the parameter tuples g. Table 1 lists those operations together with the closed form formulas.

E.g. if two attributed with the same physical meaning are added and if attributes are treated as stochastically independent RVs, the PDFs in form of GMMs are convolved. This results again in a GMM, the parameters of which can be calculated in closed form from the parameters of the two GMMs; see Equation (8) in Table 1.

Table 1: This table provides a compact overview of the use of GMMs for the representation and combination of uncertain quantitative attributes and their PDFs. Column 1 lists the most important operations for the practical handling of attributes in real-world applications. Column 2 explains the associated mathematical meaning. The computation type (CT) in column 3 indicates whether the calculations can be accomplished algebraically on the parameter tuples of the GMMs (CF: Closed Form) or whether a numerical calculation (N) is required. Finally, in column 4, the closed formulas for the algebraic operations on the GMM parameters are given compactly. $X \perp Y$ means that the RVs X and Y are stochastically independent. Additional details to these formulas are presented in Section A. For each row of the table there is an explanatory concrete example in Section 6.

Practical Meaning	**Mathematical Expression**	**CT**	**Formulas/Algorithms**
Traditional measurement result; value m	Expectation of RVs; see Equation (4) in Proposition 5.1	CF	$m = \sum_{i=1}^{K} \pi_i m_i$ (6)
Traditional measurement uncertainty $\pm k\sigma$; σ^2 = variance	Variance of RVs; see Equation (5) in Proposition 5.2	CF	$\Sigma = \sum_{i=1}^{K} \pi_i(\Sigma_i + m_i \cdot {m_i}^\top) - (\sum_{i=1}^{K} \pi_i m_i) \cdot (\sum_{i=1}^{K} \pi_i m_i)^\top$ (7)
Empirical addition of physical quantities of the same type	Addition of independent RVs corresponds with: Convolution of PDFs; see Proposition B.5 in Section A	CF	$X + Y = Z, \quad X \perp Y$ $K_Z = K_X \cdot K_Y, \quad \pi_{Zij} = \pi_{Xi} \cdot \pi_{Yj}$ $m_{Zij} = m_{Xi} + m_{Yj}, \quad \Sigma_{Zij} = \Sigma_{Xi} + \Sigma_{Yj}$ (8)
Focusing on variables of interest	Partial marginalization of RVs corresponds with: Integrating out nuisance RVs; see Proposition B.1 in Section A	CF	$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \in \mathbb{R}^d, \quad Z \sim \mathcal{G}_Z(z\|g_Z)$ $m_{Zi} = \begin{bmatrix} m_{ZiX} \\ m_{ZiY} \end{bmatrix}, \quad \Sigma_{Zi} = \begin{bmatrix} \Sigma_{ZiXX} & \Sigma_{ZiXY} \\ \Sigma_{YiYX} & \Sigma_{ZiYY} \end{bmatrix}$ $X \sim \mathcal{G}_X(x\|g_X), \quad Y \sim \mathcal{G}_Y(x\|g_Y)$ $K_X = K_Y = K_Z$ $\pi_{Xi} = \pi_{Yi} = \pi_{Zi}$ $m_{Xi} = m_{ZiX}, \quad m_{Yi} = m_{ZiY},$ $\Sigma_{Xi} = \Sigma_{ZiXX}, \quad \Sigma_{Yi} = \Sigma_{ZiYY}$ (9)

Continued on next page

Practical Meaning	Mathematical Expression	CT	Formulas/Algorithms
Exploiting evidence about a set of RVs	Partial conditioning of RVs regarding the observed RVs; see Proposition B.2 in Section A	CF	$X\|Y \sim \mathcal{G}_{X\|Y}(x\|y, g_{X\|Y})$ $K_{X\|Y} = K_Z$ $m_{X\|Y\,i} = m_{iX} + \Sigma_{iXY} \cdot {\Sigma_{iYY}}^{-1} \cdot (y - m_{iY})$ $\Sigma_{X\|Y\,i} = \Sigma_{iXX} - \Sigma_{iXY} \cdot {\Sigma_{iYY}}^{-1} \cdot \Sigma_{iYX}$ $\pi_{X\|Y\,i} = \frac{\pi_i \mathcal{N}(y\|m_{Y\,i}, \Sigma_{Y\,i})}{\sum_{j=1}^{K_Y} \pi_j \mathcal{N}(y\|m_{Y\,j}, \Sigma_{Y\,j})}$ $Y\|X \sim \mathcal{G}_{Y\|X}(y\|x, g_{Y\|X})$ $K_{Y\|X} = K_Z$ $m_{Y\|X\,i} = m_{iY} + \Sigma_{iYX} \cdot {\Sigma_{iXX}}^{-1} \cdot (x - m_{ix})$ $\Sigma_{Y\|X\,i} = \Sigma_{iYY} - \Sigma_{iYX} \cdot {\Sigma_{iXX}}^{-1} \cdot \Sigma_{iXY}$ $\pi_{Y\|X\,i} = \frac{\pi_i \mathcal{N}(x\|m_{X\,i}, \Sigma_{X\,i})}{\sum_{j=1}^{K_X} \pi_j \mathcal{N}(x\|m_{X\,j}, \Sigma_{X\,j})}$ (10)
Mix of Sources	Convex linear combination of PDFs; see Proposition B.4 in Section A	CF	$X^j \sim \mathcal{G}_{X^j}(x\|g_X^j)$ for $j = 1, \ldots, J$ $X \sim \mathcal{G}_X(x\|g_X) = \sum_{j=1}^{J} w^j \mathcal{G}_X^j(x\|g_X^j) =$ $\sum_{j=1}^{J} w^j \sum_{i=1}^{K_j} \pi_i^j \mathcal{N}_X(x\|m_i^j, \Sigma_i^j) =$ $\sum_{j=1}^{J} \sum_{i=1}^{K_j} w^j \pi_i^j \mathcal{N}_X(x\|m_i^j, \Sigma_i^j).$ (11)
Fusion of independent information contributions	Bayesian fusion corresponds with: Multiplication of PDFs; see Proposition B.3 in Section A	CF	$\mathcal{G}_X(x\|g_a) \cdot \mathcal{G}_X(x\|g_b) = \sum_{i=1}^{K_a} \sum_{j=1}^{K_b} c_{a_i b_j} \pi_{a\,i} \pi_{b\,j} \mathcal{N}_X(x\|m_{a_i b_j}, \Sigma_{a_i b_j})$ $c_{a_i b_j} = \frac{\exp\left[-\frac{1}{2}(m_{a_i} - m_{b_j})^\top \cdot (\Sigma_{a_i} + \Sigma_{b_j})^{-1} \cdot (m_{a_i} - m_{b_j})\right]}{\sqrt{\det\left(2\pi(\Sigma_{a_i} + \Sigma_{b_j})\right)}}$ $m_{a_i b_j} = (\Sigma_{a_i}^{-1} + \Sigma_{b_j}^{-1})^{-1} \cdot (\Sigma_{a_i}^{-1} \cdot m_{a_i} + \Sigma_{b_j}^{-1} \cdot m_{b_j})$ $\Sigma_{a_i b_j} = (\Sigma_{a_i}^{-1} + \Sigma_{b_j}^{-1})^{-1}$ (13)
Seeing the outcome of a GMM distributed RV	Drawing samples from a GMM distributed RV	N	Randomly select a component using the mixture weights and sample from its Gaussian distribution. See Proposition A.8 and Proposition B.6 in Section A.

Continued on next page

Practical Meaning	Mathematical Expression	CT	Formulas/Algorithms
Reducing the number of components	Reducing the number of parameters of GMMs based on $\|\cdot\|_{L^2(\mathbb{R}^d)}$.	N	$\operatorname{argmin}_{g_b} \|\mathcal{G}_X(x\|g_a) - \mathcal{G}_X(x\|g_b)\|_{L^2(\mathbb{R}^d)}$ with $K_a > K_b$. $\operatorname{argmin}_{g_b} \|\mathcal{G}_X(x\|g_a) - \mathcal{G}_X(x\|g_b)\|_{L^2(\mathbb{R}^d)} =$ $\operatorname{argmin}_{g_b} \sqrt{\int_{\mathbb{R}^d} (\mathcal{G}_X(x\|g_a) - \mathcal{G}_X(x\|g_b))^2 \mathrm{d}x} =$ $\operatorname{argmin}_{g_b} \sqrt{I_{aa} - 2I_{ab} + I_{bb}}$ $I_{aa} = \sum_{i=1}^{K_a} \sum_{j=1}^{K_a} \pi_{a_i} \pi_{a_j} c_{a_i a_j}$ $I_{ab} = \sum_{i=1}^{K_a} \sum_{j=1}^{K_b} \pi_{a_i} \pi_{b_j} c_{a_i b_j}$ (14) $I_{bb} = \sum_{i=1}^{K_b} \sum_{j=1}^{K_b} \pi_{b_i} \pi_{b_j} c_{b_i b_j}$ $c_{uv} = \frac{\exp\left[-\frac{1}{2}(m_u - m_v)^\top \cdot (\Sigma_u + \Sigma_v)^{-1} \cdot (m_u - m_v)\right]}{\sqrt{\det(2\pi(\Sigma_u + \Sigma_v))}}$ $m_{uv} = (\Sigma_u^{-1} + \Sigma_v^{-1})^{-1} \cdot (\Sigma_u^{-1} \cdot m_u + \Sigma_v^{-1} \cdot m_v)$ $\Sigma_{uv} = (\Sigma_u^{-1} + \Sigma_v^{-1})^{-1}$ $u, v \in \{a_i\}_{i=1,\ldots,K_a} \cup \{b_j\}_{j=1,\ldots,K_b}$

6. Examples

Now we present some examples to demonstrate the practical usefulness of representing quantitative attributes by the means of GMMs and to calculate with them according to the formulas of Table 1.

6.1 Mechanical Assembly

In assembly situations parts are put together in a way that it results in an additive behavior. The lengths are modeled as RVs, when the part A with length $X \sim \mathcal{G}_X(x|g_X)$ and the part B with length $Y \sim \mathcal{G}_Y(y|g_Y)$ are aligned according to Figure 1 to make up the total length Z. In general for stochastically independent RVs $X \sim p_X(x)$ and $Y \sim p_Y(y)$, the PDF of the total length Z is given by the convolution of the PDFs of X and Y:

$$Z = X + Y \text{ and } X \perp Y \implies p_Z(z) = \int_\Omega p_X(x)\, p_Y(z-x)\, \mathrm{d}x = p_X(x) * p_Y(y). \tag{15}$$

However, when the PDFs of the RVs are distributed according to GMMs the total length can be calculated according to Equation (8) in Table 1 in an arithmetical fashion without any approximation on the PDF of the RV Z.

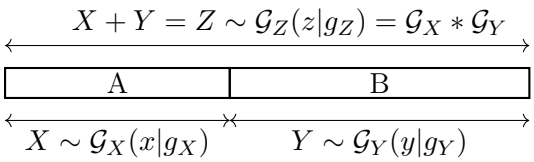


Figure 1: Length X and length Y make up the total length Z. $\mathcal{G}_Z$ can be calculated using Equation (8) in Table 1. The symbol $*$ denotes the convolution operator.

As a numerical example consider two six components GMMs for the two RVs X and Y according to Figure 1. It can be seen in Figure 2 that the closed form arithmetic form PDF assigned to the RV Z coincides with the histogram resulted from samples achieved using the forward Monte Carlo calculation [21].

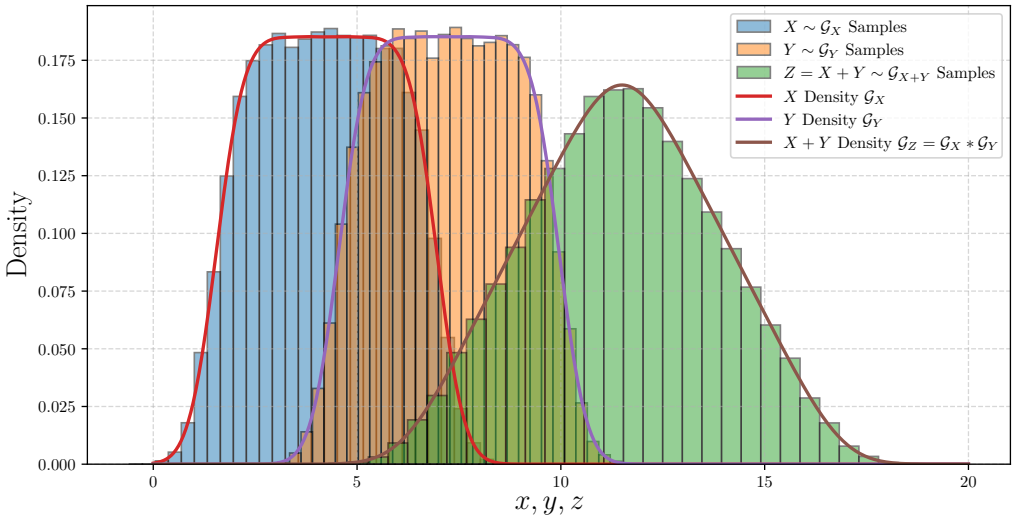


Figure 2: Analytical calculation of the resulting PDF vs forward Monte Carlo according to [21] using the ad hoc sampler in Section 5.1.4

Remark 6.1. *It is possible to apply Equation* (8) *successively to treat the countable summation cases for example Figure 3 similarly.*

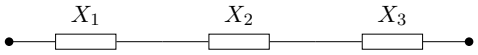


Figure 3: Three independent resistances X_1, X_2 and X_3 make up the total resistance $Z = X_1 + X_2 + X_3$. Then for $X_i \sim \mathcal{G}_{X_i}, i = 1, 2, 3$ the total resistance $Z \sim \mathcal{G}_Z = \mathcal{G}_{X_1} * \mathcal{G}_{X_2} * \mathcal{G}_{X_3}$ which can be calculated by applying Equation (8) successively.

6.2 Mechanical Disassembly

In disassembly situations parts are taken apart in a way that it would result in a subtractive behavior. The lengths are modeled as RVs, when the part A with length $Y \sim \mathcal{G}_Y(y|g_Y)$ and the part B with length $Z \sim \mathcal{G}_Z(z|g_Z)$ are aligned according to Figure 4 to make up the total length X. In general for stochastically independent RVs $X \sim p_X(x)$ and $Y \sim p_Y(y)$, the PDF of the subtracted length Z is given by the convolution of the PDFs of X and $-Y$. In the case of GMMs when $Y \sim \mathcal{G}_Y(y|g_Y)$ then $-Y \sim \mathcal{G}_{-Y}(y|g_{-Y})$ where g_{-Y} is the same as g_Y except that the means are negated.

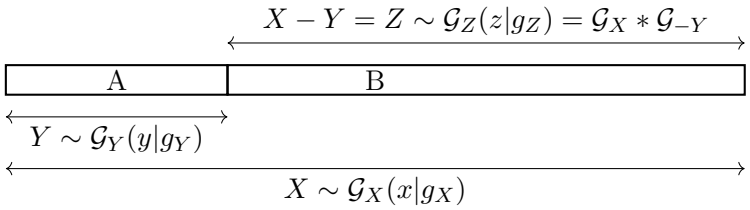


Figure 4: Length Y and length Z make up the total length X. $\mathcal{G}_Z$ can be calculated using Equation (8) in Table 1. The symbol $*$ denotes the convolution operator.

6.3 Metal Sheet

The mass of a rectangle sheet of metal in Figure 5 is proportional to the area of the rectangle having side lengths $X \sim \mathcal{G}_X(x|g_X)$ and $Y \sim \mathcal{G}_Y(x|g_Y)$. The area is the result of a multiplication $Z = X \cdot Y$.

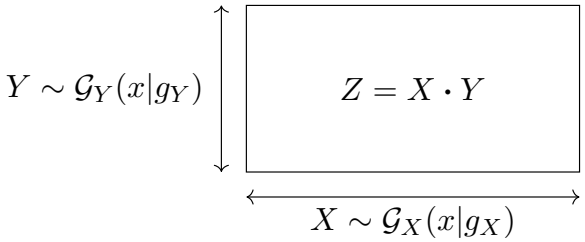


Figure 5: A rectangle with side lengths $X \sim \mathcal{G}_X(x|g_X)$, $Y \sim \mathcal{G}_Y(x|g_Y)$, and area $Z = X \cdot Y$.

In this case the PDF of Z cannot be calculated in closed form even if X and Y are distributed according to simple PDFs [10]. In general if $X \sim p(x)$ and $Y \sim p(y)$ and $Z \sim p(z)$:

$$Z = X \cdot Y \text{ and } X \perp Y \implies p_Z(z) = \int_\Omega \frac{1}{|x|} p_X(x) p_Y\left(\frac{z}{x}\right) \mathrm{d}x. \tag{16}$$

As it can be deduced from the equation, the parameters of $Z \sim \mathcal{G}_Z(z|g_Z)$ cannot be calculated using an arithmetical procedure. The PDF of Z is a complicated distribution in terms of the PDFs of X and Y; called "the product distribution" [10]. However, it is still possible to fit a GMM to Z using the samples achieved from the forward Monte Carlo procedure [21]. For the sampler one can use the ad hoc sampler discussed in Section 5.1.4 or a general sampler like [17].

As a numerical example consider two six components GMMs for the two RVs X and Y according to Figure 6. The EM algorithm needs the number of components that is to be fit to be fixed. Here a plausible heuristic would be $K = 36$ as in the special case where all the standard deviations were zero the resulting product would only assume at maximum 36 different distinct values. It can be seen in Figure 6 that the PDF assigned to the RV Z coincides with the histogram of the samples achieved using the forward Monte Carlo [21].

6.4 Mix of Sources

In many production and manufacturing processes outsourcing must be done. Assume parts are to be delivered by two suppliers A and B. Assume the part is a shaft and the value of interest to the main manufacturer is the diameter of the shaft X and he is interested to model the

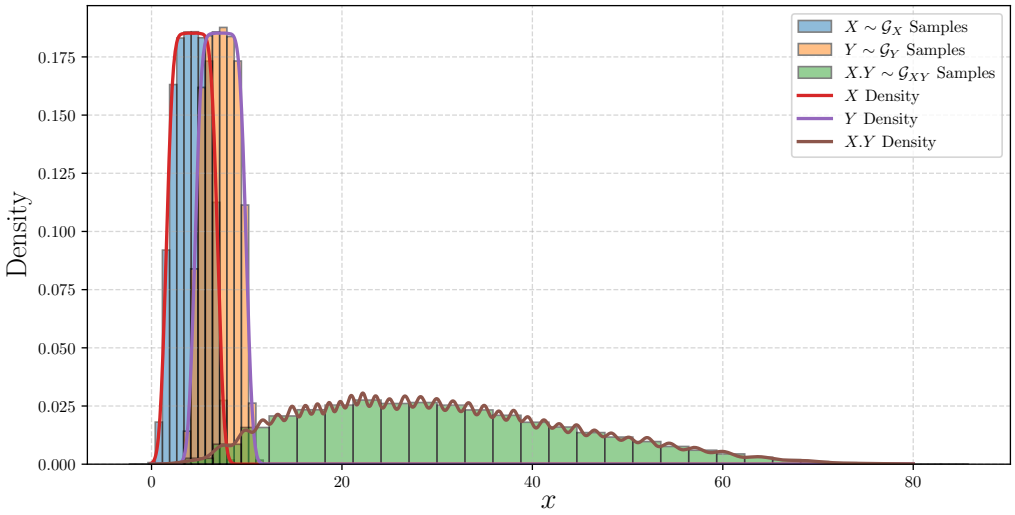


Figure 6: Numerical calculation of the resulting PDF using the EM algorithm based on the samples produced using the forward Monte Carlo according to [21] using the ad hoc sampler in Section 5.1.4

ensemble of shafts they are receiving by a RV and assign a PDF in the form of a GMM to it. Due to different machineries of the suppliers one cannot assume symmetry and must come up with a systematic way of assigning the GMM. If supplier A delivers $r_A \in [0, 100]\%$ of the shafts and supplier B supplies $r_B \in [0, 100]\%$ of the shafts and the diameter of the shafts supplied by A is to be modelled by the RV $X \sim \mathcal{G}_X(x|g_{X_A})$ and the diameter of the shafts supplied by B is to be modelled by the RV $X \sim \mathcal{G}_X(x|g_{X_B})$ then the diameter of the shafts supplied to the main manufacturer can be modelled by the RV $X \sim \mathcal{G}_X(x|g_{X_S})$ where:

$$\mathcal{G}_X(x|g_{X_S}) = \frac{r_A}{r_A + r_B}\mathcal{G}_X(x|g_{X_A}) + \frac{r_B}{r_A + r_B}\mathcal{G}_X(x|g_{X_B}). \tag{17}$$

The resulting distribution is again a Gaussian mixture model with $K_S = K_A + K_B$ number of components. Using a single combined index k, the resulting mixture can be written as

$$\mathcal{G}_{X_S}(x|g_{X_S}) = \sum_{k=1}^{K_S} \pi_{Sk}\mathcal{N}(x|m_{Sk}, \Sigma_{Sk}). \tag{18}$$

The parameters of the resulting mixture are obtained as follows:

$$\begin{aligned} \pi_{Sk} &= \frac{r_A}{r_A + r_B}\pi_{Ak}, \quad m_{Sk} = m_{Ak}, \quad \Sigma_{Sk} = \Sigma_{Ak} \text{ for } k = 1, \ldots, K_A \\ \pi_{Sk} &= \frac{r_B}{r_A + r_B}\pi_{Bk-K_A}, \quad m_{Sk} = m_{Bk-K_A}, \quad \Sigma_{Sk} = \Sigma_{Bk-K_A} \text{ for } k = K_A + 1, \ldots, K_S. \end{aligned} \tag{19}$$

Thus, an implementation only requires concatenating the Gaussian components of both mixtures and rescaling the mixture weights according to the supplier proportions.

This is consistent both with the sampling procedure of Monte Carlo methods [32] and also the functional analysis point of view of the probability theory [23]. The beauty of it is that

the mixture of two GMMs is again a GMM and the parameters of the resulting GMM can be calculated in a arithmetical procedure and closed form; see Proposition B.4 in Section A.

Remark 6.2. *Circular factories use partly old parts and products degraded varyingly from different generations and different durations of use. Therefore, the mixing of sources is an inevitable operation in circular factories and must be handled properly.*

6.5 Probabilistic Quality Control

Let the vector of the product attributes be denoted by $x \in \mathbb{R}^d$ which is the realization of the RV $X \sim \mathcal{G}_X(x|g)$. The quality engineer specifies an acceptable region in form of a set $Q \subset \mathbb{R}^d$ according to the Figure 7 in the attribute space, representing the set that the features are allowed to fall in. The probability that the attribute vector of the product to fall into the allowed set Q is:

$$\mathbb{P}(x \in Q) = \int_Q \mathcal{G}_X(x|g)\mathrm{d}x. \tag{20}$$

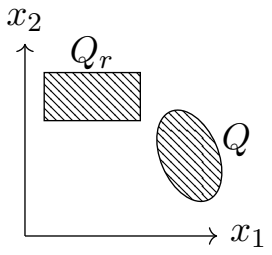


Figure 7: If quality is defined by the set Q, the probability of the attributes of the product instance to fall into set Q can be calculated by integrating the PDF over the set Q. Tolerances are usually given as intervals for the components x_i of the product instance attribute vector x, are a special case of Q, where Q becomes the hyperrectangle Q_r. If the tolerances of the components of the product instance attribute vector x, are correlated (not independent), Q can assume a general shape.

As sampling from a GMM is straightforward; Monte Carlo methods [32] can be used to estimate the probability of the attribute vector of the product instance to fall into the set Q.

6.6 Dependent Variables

Assume the inner and outer radii of the pipe in Figure 8 are modeled as stochastically dependent RVs and their relation to each other is modeled by the RV $Z = \left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right] \in \mathbb{R}^2$, $Z \sim \mathcal{G}_Z(z|g_Z)$. Two common questions could arise:

1. One might be interested in having an estimate of the inner radius X without observing, measuring, or looking at the outer radius. Then Y is a nuisance variable which is integrated out (marginalization) using Proposition B.1 in Section A. For GMMs this means that the nuisance variable is simply dropped in the joint GMM of X and Y, i.e. in $\mathcal{G}_Z$ to obtain the PDF of X.

2. One might be interested in having an estimate of the inner radius X after observing the outer radius, mathematically speaking putting $Y = y^*$ (conditioning), which can be handled using Proposition B.2 in Section A.

Both cases can be treated in closed form without any approximation and in an arithmetical fashion.

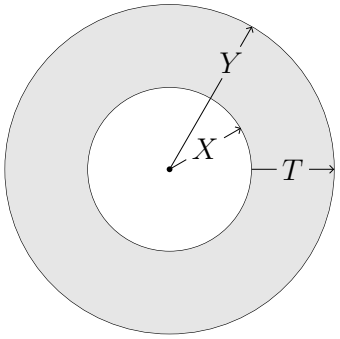


Figure 8: A pipe with the inner diameter X and the outer diameter Y. The thickness of the pipe is T. The variables are modeled as RVs.

6.7 Fusion of Information from Multiple Measurement Devices

Let $X \sim \mathcal{G}_X(x|g_A)$ and $X \sim \mathcal{G}_X(x|g_B)$ be the measurement results of a physical quantity x of interest done by two stochastically independent apparatuses A and B respectively. To fuse these information one must multiply the two PDFs according to the Bayes' theorem, and normalize the result to arrive at a valid PDF. Equation (44) gives a closed form formula to calculate the normalization constant. The derivation does not comprise any approximations and is done in an arithmetical fashion.

6.8 Reduction of Gaussian Mixture Models

In GMMs, the number of parameters increases rapidly under operations such as Bayesian fusion (multiplying two mixtures) or convolution, since each pairwise interaction between components typically produces new mixture components. Multiplying or convolving two GMMs with K and L components result in a GMM with $M = K \cdot L$ components. With respect to the number of such operations, the number of components of GMM grows exponentially. Therefore, it is essential to perform mixture reduction to maintain tractability. A key advantage of working with Gaussian mixtures is that the L^2 distance between mixtures admits a closed-form expression [18], because products and integrals of Gaussian functions can be calculated analytically. As a result, instead of solving an infinite dimensional functional optimization problem over arbitrary densities, we can reformulate mixture reduction as a finite dimensional optimization problem over the mixture weights, means, and covariances. This significantly improves computational feasibility while preserving a principled approximation framework. The closed form formula for the L^2 distance between two Gaussian mixtures is:

$$\|\mathcal{G}_X(x|g_a) - \mathcal{G}_X(x|g_b)\|_{L^2(\mathbb{R}^d)} = \sqrt{\int_{\mathbb{R}^d} (\mathcal{G}_X(x|g_a) - \mathcal{G}_X(x|g_b))^2 \mathrm{d}x} = \sqrt{I_{aa} - 2I_{ab} + I_{bb}} \tag{21}$$

where

$$
\begin{aligned}
I_{aa} &= \sum_{i=1}^{K_a} \sum_{j=1}^{K_a} \pi_{a_i} \pi_{a_j} c_{a_i a_j} \\
I_{ab} &= \sum_{i=1}^{K_a} \sum_{j=1}^{K_b} \pi_{a_i} \pi_{b_j} c_{a_i b_j} \\
I_{bb} &= \sum_{i=1}^{K_b} \sum_{j=1}^{K_b} \pi_{b_i} \pi_{b_j} c_{b_i b_j}
\end{aligned}
\tag{22}
$$

and

$$
\begin{aligned}
c_{uv} &= \frac{\exp\left[-\frac{1}{2}(m_u - m_v)^\top \cdot (\Sigma_u + \Sigma_v)^{-1} \cdot (m_u - m_v)\right]}{\sqrt{\det\left(2\pi(\Sigma_u + \Sigma_v)\right)}} \\
m_{uv} &= (\Sigma_u^{-1} + \Sigma_v^{-1})^{-1} \cdot (\Sigma_u^{-1} \cdot m_u + \Sigma_v^{-1} \cdot m_v) \\
\Sigma_{uv} &= (\Sigma_u^{-1} + \Sigma_v^{-1})^{-1} \\
u, v &\in \{a_i\}_{i=1,\ldots,K_a} \cup \{b_j\}_{j=1,\ldots,K_b}.
\end{aligned}
\tag{23}
$$

For a given GMM $\mathcal{G}_X(x|g_a)$ with K_a components the reduction to a GMM with $K_b < K_a$ components can be accomplished by solving the optimization problem:

$$
\text{argmin}_{g_b} \, \|\mathcal{G}_X(x|g_a) - \mathcal{G}_X(x|g_b)\|_{L^2(\mathbb{R}^d)} \text{ with } K_a > K_b. \tag{24}
$$

Figure 9 shows an illustration on how a GMM having four components can be approximated with a GMM having two components. According to Section 5.2.2 and since $d = 2$ now we have 11 parameters instead of 23 parameters after the reduction.

An overview of common reduction methods for GMMs can be found in [13][18].

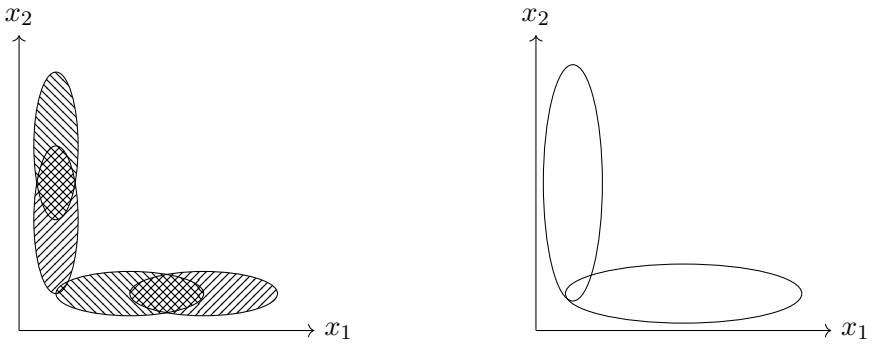


Figure 9: Reduction of a GMM with four components (of the left) into a GMM with two components (on the right); illustration.

7. Modeling of Measurement Systems with GMMs

Abstractly, a measuring instrument can be described probabilistically by a joint PDF $p(x, y)$ over the quantities to be measured x (measurands) and the observable data y (observables).

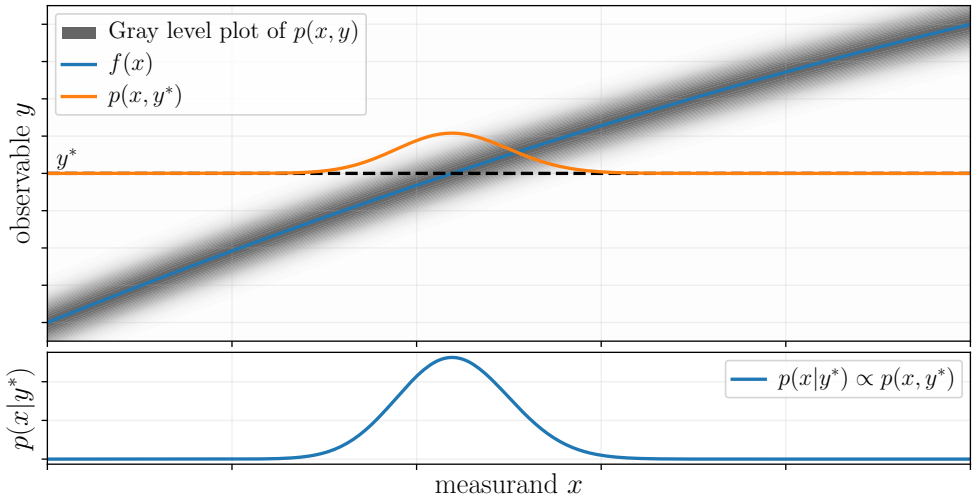


Figure 10: Schematic illustration of the conditioning process to obtain the posterior PDF $p(x|y^*)$ of the measurand based on an observation of the measurement device.

$p(y|x) = \frac{p(x,y)}{p(x)}$ is the so-called forward model, which describes the PDF of the observables y depending on the underlying value of the measurand. If concrete data y^* are observed and inserted into $p(y|x)$, this becomes the likelihood function $p(y^*|x)$, which is now interpreted as a function of x. According to Bayes' formula, the posterior PDF of x can be calculated in the form of $p(x|y^*) = \frac{p(y^*|x)p(x)}{p(y^*)}$ [3]. Measuring has the goal to infer the underlying value of x based on concretely observed data y^*. The question is: What should be reported from the data y^* in the best possible way with respect to x? If one wants to keep the loss of information to a minimum in the sense of Shannon's information theory, it is optimal to calculate the posterior PDF $p(x|y^*)$ and report it as a result of the processing of y^* [25][40]. If $p(x, y)$ is modeled in the form of a GMM, which is possible with arbitrary precision due to the universal approximation property of GMMs [30][28], then $p(x|y^*) \propto p(x, y^*)$ is itself a GMM, whose parameters can be algebraically calculated from the parameters of the $p(x, y)$ GMM. If one also assumes an unbiased measuring device for which one can assume the prior PDFs $p(x)$ is approximately uniformly distributed over the Range(x).

Note that for measurement systems that have, up to noise, an approximately linear characteristic between y and x, the marginal PDF $p(y)$ is approximately uniformly distributed as well. Then $p(x|y^*) \propto p(y^*|x)$, i.e. proportional to the likelihood function $p(y^*|x)$. For illustration see Figure 10.

Formulas for conditioning of GMMs can be found in Proposition B.2 in Section A. Thus, not only continuous attributes itself can be comprehensively described by means of GMM in the probabilistic sense, but also measurement systems in general as well as the processing of observed (read off) data $y*$ in order to calculate the measurement result in the form of the posterior PDF $p(x|y*)$. Interestingly, all required descriptions and calculations can be accomplished within the function set of GMMs. The following example is intended to illustrate the interrelations.

7.1 Model Selection

When fitting a GMM, the number of mixture components K is typically unknown and must be estimated from the data. Two widely used criteria for this purpose are the *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC) [5]. Both criteria balance goodness of fit and model complexity. For a candidate model with K components, we first fit the GMM and compute the maximized log-likelihood $\log L$. The criteria are then calculated as

$$\mathrm{AIC} = -2\log L + 2p, \qquad \mathrm{BIC} = -2\log L + p\log N, \qquad L = \prod_{n=1}^{N} \mathcal{G}_X(x_n|g_X) \tag{25}$$

where p is the number of free parameters in the mixture model and N is the number of observations. Increasing K typically improves the likelihood but also increases p, which both criteria penalize. By fitting models for several values of K and selecting the one that minimizes AIC or BIC, we obtain a principled estimate of the number of mixture components. In practice, BIC often favors simpler models because its penalty grows with sample size N, whereas AIC may select slightly larger mixtures.

7.2 Example

Assume the simplest case where a scalar input x is to be measured by a measurement device and the output is a scalar observable y which is corrupted by additive Gaussian noise. The physical laws governing the measurement device will manifest as the measurement curve $f(x)$. In short:

$$\begin{aligned} &f\colon \mathbb{R} \to \mathbb{R} \qquad x \mapsto f(x) \\ &\widetilde{f}\colon \mathbb{R}^2 \to \mathbb{R} \quad x, w \mapsto y = f(x) + w \\ &W \sim \mathcal{N}(w|0, \sigma^2). \end{aligned} \tag{26}$$

Now, we assume that we have observed N pairs of $\{(x_i, y_i)\}_{i=1,\dots,N}$ where X is sampled uniformly over its Range(x). Based on the data a GMM is fitted.

$$X, Y \sim \mathcal{G}_{XY}(x, y|g_{XY}) \tag{27}$$

that represents the measurement curve as well as the superposed noise. If the GMM accurately represents the measurement device one must have:

$$\begin{aligned} \|\mathbb{E}(Y|X) - f(x)\|_{L^2(\mathbb{R})} &= \sqrt{\int_{\mathbb{R}} \left(\mathbb{E}(Y|X) - f(x)\right)^2 \mathrm{d}x} = \mathbb{E}(W) = 0 \\ \|\mathbb{V}(Y|X) - \sigma^2\|_{L^2(\mathbb{R})} &= \sqrt{\int_{\mathbb{R}} \left(\mathbb{V}(Y|X) - \sigma^2\right)^2 \mathrm{d}x} = 0. \end{aligned} \tag{28}$$

For the case:

$$\begin{aligned} &f\colon [0,1] \to [0, 0.8] \qquad x \mapsto x - 0.2x^2 \\ &\widetilde{f}\colon [0,1] \times (-\infty, +\infty) \to \mathbb{R} \quad x, w \mapsto y = x - 0.2x^2 + w \\ &W \sim \mathcal{N}(w|0, 0.0025) \end{aligned} \tag{29}$$

a Gaussian mixture model was obtained using the EM algorithm. As stated BIC or AIC are commonly used to find the optimal number of components K.

$N = 1000$ simulated data points and $K = 10$ were used to obtain the Gaussian mixture and:

$$\begin{aligned} \|\mathbb{E}(Y|X) - f(x)\|_{L^2[0,1]} &= \sqrt{\int_0^1 \left(\mathbb{E}(Y|X) - f(x)\right)^2 \mathrm{d}x} = 5.24 \times 10^{-7} \\ \|\mathbb{V}(Y|X) - 0.0025\|_{L^2[0,1]} &= \sqrt{\int_0^1 \left(\mathbb{V}(Y|X) - 0.0025\right)^2 \mathrm{d}x} = 6.20 \times 10^{-6}. \end{aligned} \tag{30}$$

The metrics indicate that the obtained Gaussian mixture can accurately represent the measurement device; see Figure 11.

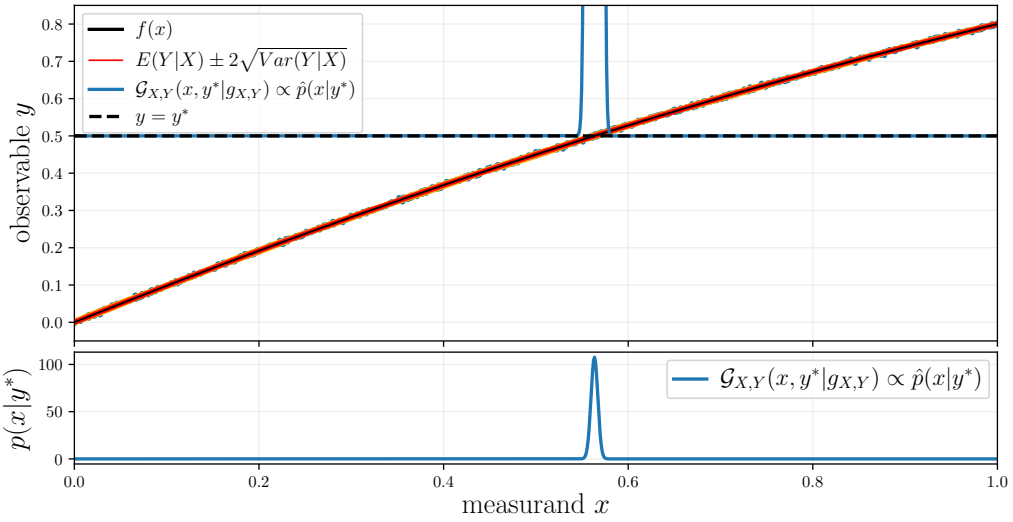


Figure 11: Modeling of a measurement system using GMM; low noise case.

To be able to see the conditioning with the naked eye assume now the noise is distributed according to $\mathcal{N}(w|0, 0.03)$. The results are shown in Figure 12. The blue dots represent the simulated measurement data, the radii of the orange ellipses represent twice the square root of the eigenvalues of the covariance matrices of the Gaussian components of the GMM, the black curve represents the $f(x)$ and the red curves represent $\mathbb{E}(Y|X) \pm 2\sqrt{\mathbb{V}(Y|X)}$. About 95% of the data lie between the red curves. After the observable is read of the measurement device as $y^* = 0.5$ a PDF in the form of a GMM can be assigned to the measurand x using conditioning formulas found in Equation (10) in Table 1 and alternatively in Proposition B.2 in Section A.

8. Conclusion

In this paper, we have presented a consistent and scalable probabilistic framework for the representation, propagation, and processing of measurement uncertainty in industrial and engineering systems, circular factories in particular. Departing from the traditional restriction of

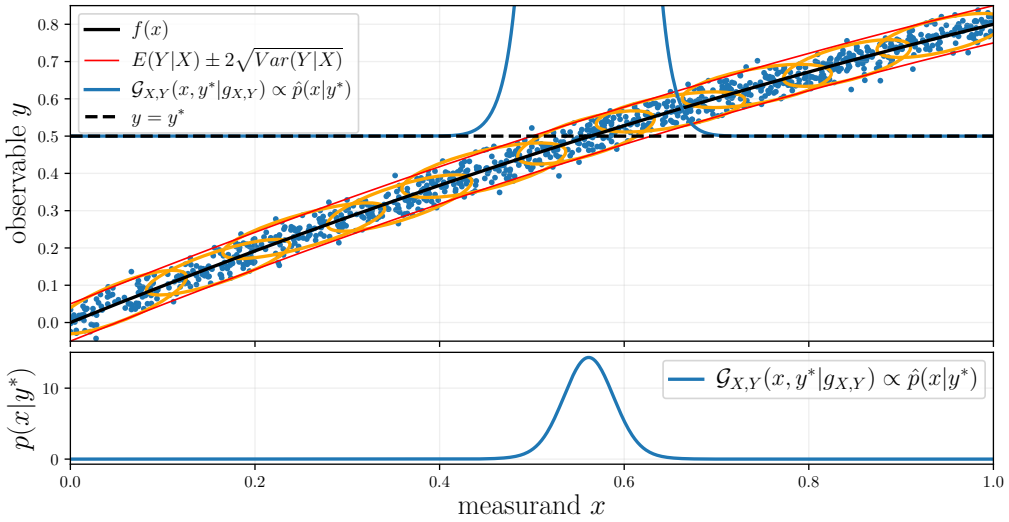


Figure 12: Modeling of a measurement system using GMM; high noise case.

uncertainty representation to Gaussian distributions characterized solely by mean and variance, we have argued that many practically relevant measurement scenarios violate the assumptions underlying the Gaussian framework, particularly in the presence of nonlinear transformations, multimodal effects, constraints, and source uncertainty fusion. To address these limitations, we proposed GMMs as a unifying and expressive representation for uncertain quantitative attributes. Owing to their universal approximation capability, finite-dimensional parameterization, and closure under a wide range of algebraic operations, GMMs enable uncertainty to be propagated and fused with high precision while remaining computationally tractable. We demonstrated that many operations of practical relevance–such as affine transformations, convolution, marginalization, conditioning, and Bayesian fusion–can be performed exactly and efficiently within the Gaussian mixture framework all in closed form. For cases where closed form propagation is not available, efficient ad hoc samplers combined with EM based fitting provides an effective and scalable approximation strategy. Beyond the representation of uncertain quantities themselves, we showed that entire measurement systems can be modeled probabilistically using GMMs. This enables a principled Bayesian treatment of measurement processes, where prior knowledge and posterior distributions are all expressed within the same functional class. As a result, the complete measurement chain–from raw observables to final measurement results–can be described by EM fitting and algebraic operations on Gaussian mixture parameters, minimizing information loss and avoiding multistage Gaussian approximations. Importantly, the framework remains compatible with established metrological practice through an explicit Gaussian fallback, allowing results to be reduced to mean and standard deviation whenever required. The examples presented throughout the paper illustrate that the proposed approach is not merely of theoretical interest but is well suited for real world applications, including tolerance analysis, quality control, uncertainty propagation in multi stage systems, and the fusion of heterogeneous information sources. From a software and computational perspective, the encapsulation of GMM based operations allows users to benefit from improved uncertainty handling without increased conceptual or computational burden. In sum-

mary, GMM provide a robust, expressive, and computationally efficient foundation for modern uncertainty quantification in measurement science. By treating GMMs as a native data type for uncertain quantities, the proposed framework enables more accurate uncertainty propagation, better risk assessment, and a unified probabilistic treatment of aleatory and epistemic uncertainties by interpreting probabilitis as DoB. Future work may focus on automated model selection, real time adaptive mixture refinement, and the integration of the framework into standardized metrological software and digital twin architectures.

Acknowledgments

Funded by:
- Project Name: SFB 1574, A Circular Factory for the Perpetual Product
- Funding Agency: Deutsche Forschungsgemeinschaft (DFG)
- Project ID: 471687386

A. Appendix

Here we give a terse theorem like exposition blended with explanatory text to the basic results on GMM that are used throughout the paper. We define the Gaussians, express the key operations on them, and then lift these operations whenever possible to GMMs. The sequence in which the results are presented are based on mathematical construction and to avoid circularity which is simple terms is not to have logical flaws in the proofs even though the proofs are not included. As we do not present proofs but rather present the results it might be a bit more involved to convince yourself on the correctness of the order of presentation. We hope it enables the avid reader to build their personal library based of GMMs and make use of their powerful representation and reasoning in their respective field.

A.1 Theorem Style Constructs

We make heavy use of four theorem like constructs namely **Definitions**, **Propositions**, **Corollaries** and **Remarks**. Definitions are used to formally define new concepts and objects. Propositions are used to state important properties or results that are not as general or significant as theorems and usually do not have names. Corollaries are used to state results that follow directly from a theorem or proposition. Remarks are used to provide additional insights, clarifications, notational conventions or observations related to the preceding content.

A.2 Gaussians

As the Gaussian PDFs are the atomic building blocks; it is of great importance to precisely define them and their parameters and the spaces that the said parameters belongs to. This will introduce a clear notation and from the computational and software perspective give a clear way to define the class of Gaussian PDFs if desired.

Definition A.1 (Gaussian PDF)**.** *Let*

$$\begin{aligned} &d \in \mathbb{N} \setminus \{0, +\infty\} \\ &x \in \mathbb{R}^d, m \in \mathbb{R}^d \\ &\Sigma \in \{A \in \mathbb{R}^{d \times d} \colon x^\top \cdot A \cdot x > 0 \text{ for } x \neq 0\} \end{aligned} \tag{31}$$

then the mapping

$$\mathcal{N}_X(x|m, \Sigma) \coloneqq \frac{\exp\left[-\frac{1}{2}(x-m)^\top \cdot \Sigma^{-1} \cdot (x-m)\right]}{\sqrt{\det\left(2\pi\Sigma\right)}} \tag{32}$$

is called the Gauss function or the Gaussian PDF.

It is common in probability theory and statistics to have a notation for "When the RV is distributed according to the PDF $p(x)$" and also as the Gaussian PDF is so ubiquitous in the literature there exist a notation its PDF too which brings us to the following remark.

Remark A.1 (Notation)**.** *When the Random Variable X is Gaussian distributed, we write $X \sim \mathcal{N}_X(x|m, \Sigma)$.*

Expectation value and the variance/covariance are two of the most important statistics of a PDF and possibly the first two properties one would think of when facing a new PDF. Expectation can be thought of as the average of the data that the PDF represents and the variance can be thought of as the deviation of the data around their average. If one has a good grasp on how they are calculated in closed form in the case of Gaussians, they can be easily extended to GMMs as we will see later. Next you will find the familiar results in the case of Gaussians presented as propositions.

Proposition A.1 (Expectation of Gaussians)**.**

$$\mathbb{E}(X \sim \mathcal{N}_X(x|m, \Sigma)) = m \tag{33}$$

Proposition A.2 (Covariance Matrix of Gaussians)**.**

$$\mathbb{V}(X \sim \mathcal{N}_X(x|m, \Sigma)) = \Sigma \tag{34}$$

A.2.1 Marginalization

In probability and statistics, **marginalization** is the process of obtaining the probability distribution of a subset of variables by summing or integrating over the other variables. Suppose we have a joint probability distribution for two random variables, X and Y. To find the marginal distribution of X, we "marginalize out" Y by integrating over all possible values of Y. The result is the marginal probability of X alone. Marginalization is fundamental in Bayesian inference, expectation calculations, and simplifying complex probabilistic models by focusing only on variables of interest. Thanks to the properties of the Gaussian distribution, marginalization can be performed in closed form and the result is also a Gaussian distribution.

Proposition A.3 (Gaussian Marginalization)**.** *Let* $Z = \begin{bmatrix} X \\ Y \end{bmatrix} \in \mathbb{R}^d$, $Z \sim \mathcal{N}_Z(z|m_Z, \Sigma_Z)$

$$m_Z = \begin{bmatrix} m_{ZX} \\ m_{ZY} \end{bmatrix}, \quad \Sigma_Z = \begin{bmatrix} \Sigma_{ZXX} & \Sigma_{ZXY} \\ \Sigma_{ZYX} & \Sigma_{ZYY}, \end{bmatrix}$$

then:

$$\begin{aligned} &X \sim \mathcal{N}_X(x|m_X, \Sigma_X), \quad Y \sim \mathcal{N}_Y(y|m_Y, \Sigma_Y) \\ &m_X = m_{ZX}, \quad m_Y = m_{ZY}, \quad \Sigma_X = \Sigma_{ZXX}, \quad \Sigma_Y = \Sigma_{ZYY} \end{aligned} \tag{35}$$

A.2.2 Conditioning

In probability and statistics, **conditioning** in the case of PDFs refers to finding the conditional density of one continuous random variable given another. If X and Y have a joint PDF, the conditional PDF of X given $Y = y^*$ is obtained by dividing the joint PDF by the marginal PDF of Y at y^*. This produces a new density that describes the distribution of X when Y is fixed at a specific value. The conditional PDF integrates to one over X and reflects how knowledge of Y influences the likelihood of different values of X. Thanks to the properties of the Gaussian distribution, conditioning can be performed in closed form and the result is also a Gaussian distribution.

Proposition A.4 (Gaussian Conditioning)**.**

$$\begin{aligned}
&X|Y \sim \mathcal{N}_{X|Y}(x|y, m_{X|Y}, \Sigma_{X|Y}) \\
&m_{X|Y} = m_X + \Sigma_{XY} \cdot {\Sigma_{YY}}^{-1} \cdot (y - m_Y) \\
&\Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY} \cdot {\Sigma_{YY}}^{-1} \cdot \Sigma_{YX} \\
&Y|X \sim \mathcal{N}_{Y|X}(y|x, m_{Y|X}, \Sigma_{Y|X}) \\
&m_{Y|X} = m_Y + \Sigma_{YX} \cdot {\Sigma_{XX}}^{-1} \cdot (x - m_X) \\
&\Sigma_{Y|X} = \Sigma_{YY} - \Sigma_{YX} \cdot {\Sigma_{XX}}^{-1} \cdot \Sigma_{XY}
\end{aligned} \tag{36}$$

Proposition A.5 (Convolution of Gaussians)**.** *If* $X \sim \mathcal{N}_X(x|m_X, \Sigma_X)$ *and* $Y \sim \mathcal{N}_Y(y|m_Y, \Sigma_Y)$ *are stochastically independent, then* $Z = X+Y \sim \mathcal{N}_Z(z|m_Z, \Sigma_Z) = \mathcal{N}_X(x|m_X, \Sigma_X) * \mathcal{N}_Y(y|m_Y, \Sigma_Y)$ *where* $*$ *denotes the convolution operator and:*

$$\begin{aligned}
m_Z =& m_X + m_Y \\
\Sigma_Z =& \Sigma_X + \Sigma_Y.
\end{aligned} \tag{37}$$

A.2.3 Fusion

Two independent Gaussian PDFs about the same quantity can be combined into one through the Bayes' theorem. Each Gaussian may come from a sensor, model, or prior knowledge. The result is another Gaussian whose mean is a weighted average, giving more weight to the more reliable (lower uncertainty) source, and whose uncertainty is reduced. This is fundamental in sensor fusion, estimation, and tracking (e.g., Kalman filters). Its importance lies in providing an optimal way to merge information. Having a closed-form solution avoids numerical integration, making computation fast, stable, and scalable for real-time engineering systems.

Proposition A.6 (Bayesian Fusion of Gaussians)**.** *Let* $\mathcal{N}_X(x|m_s, \Sigma_s)$ *denote a Gaussian density of* X*, then:*

$$\mathcal{N}_X(x|m_a, \Sigma_a) \cdot \mathcal{N}_X(x|m_b, \Sigma_b) = c_{ab} \mathcal{N}_X(x|m_{ab}, \Sigma_{ab}) \tag{38}$$

where:

$$\begin{aligned}
c_{ab} &= \frac{\exp\left[-\frac{1}{2}(m_a - m_b)^\top \cdot (\Sigma_a + \Sigma_b)^{-1} \cdot (m_a - m_b)\right]}{\sqrt{\det\left(2\pi(\Sigma_a + \Sigma_b)\right)}} \\
m_{ab} &= (\Sigma_a^{-1} + \Sigma_b^{-1})^{-1} \cdot (\Sigma_a^{-1} \cdot m_a + \Sigma_b^{-1} \cdot m_b) \\
\Sigma_{ab} &= (\Sigma_a^{-1} + \Sigma_b^{-1})^{-1}
\end{aligned} \tag{39}$$

Knowing the normalization constant of the product of two Gaussians is important for several reasons. First, it allows us to compute the exact posterior distribution in Bayesian inference when both the prior and likelihood are Gaussian. This is crucial for applications like Kalman filtering, where we need to update our beliefs based on new measurements. Second, it provides a way to evaluate metric to compare different Gaussian distributions, which is useful in model selection and hypothesis testing.

Corollary A.1 (Normalization Constant of Bayesian Fusion of Gaussians)**.**

$$\int_\Omega \mathcal{N}_X(x|m_a, \Sigma_a) \cdot \mathcal{N}_X(x|m_b, \Sigma_b) \mathrm{d}x = c_{ab} \tag{40}$$

A.2.4 Sampling

While sampling from a general PDF and performing a Monte Carlo simulation is not a straight forward process; in the case of Gaussian PDFs it is easy to generate random numbers using only uniform random numbers, which are easy for computers to produce. By applying simple mathematical transformations, two independent standard Gaussian samples are obtained. This method is important because many engineering simulations, noise models, and Monte Carlo algorithms rely on Gaussian randomness. The Box-Muller transformation coupled with an affine transformation provides an exact, closed-form way to convert samples drawn from a uniform distribution to samples drawn from the desired multivariate PDF. Its simplicity, efficiency, and reliability makes Monte Carlo simulations and stochastic modeling with Gaussian PDFs feasible and widely used in engineering applications.

Proposition A.7 (Box-Muller Transformation)**.** *Let U and V be independent samples drawn from the $\mathcal{U}(x|0,1)$ PDF. Then for $R = \sqrt{-2\log U}, \Theta = 2\pi V$ the samples $X = R\cos\Theta, Y = R\sin\Theta$ are independent samples drawn from the standard Gaussian PDF $\mathcal{N}(x|0,1)$.*

Proposition A.8 (Multivariate Gaussian Sampling)**.** *Let $Z \in \mathbb{R}^d$ such that Z_i is a sample drawn from the PDF $\mathcal{N}(z_i|0,1)$ for $i = 1,\dots,d$. For the full rank matrix A and the vector $m \in \mathbb{R}^d$, the sample $X = m + AZ$ will be a sample drawn from the PDF $\mathcal{N}_X(x|m, AA^\top)$.*

B. GMMs

As a reminder as seen in Definition 5.1 The definition of GMMs is a direct extension of the definition of Gaussians, where we have a weighted sum of Gaussians instead of just one Gaussian. Now let us express the key operations on GMMs.
Using the Proposition A.3 the **marginalization** of GMMs can be easily derived. The result is a GMM with the same number of components as the original joint GMM,

Proposition B.1 (GMM Marginalization)**.** *Let $Z = \begin{bmatrix} X \\ Y \end{bmatrix} \in \mathbb{R}^d$, $Z \sim \mathcal{G}_Z(z|g_Z)$*

$$m_{Zi} = \begin{bmatrix} m_{ZiX} \\ m_{ZiY} \end{bmatrix}, \quad \Sigma_{Zi} = \begin{bmatrix} \Sigma_{ZiXX} & \Sigma_{ZiXY} \\ \Sigma_{YiYX} & \Sigma_{ZiYY} \end{bmatrix}$$

then:

$$\begin{aligned} &X \sim \mathcal{G}_X(x|g_X), \quad Y \sim \mathcal{G}_Y(x|g_Y) \\ &K_X = K_Y = K_Z \\ &\pi_{Xi} = \pi_{Yi} = \pi_{Zi} \\ &m_{Xi} = m_{ZiX}, \quad m_{Yi} = m_{ZiY}, \\ &\Sigma_{Xi} = \Sigma_{ZiXX}, \quad \Sigma_{Yi} = \Sigma_{ZiYY} \end{aligned} \tag{41}$$

where the subscript i denotes the i-th component of the resulting GMM.

Using the Proposition A.4 and normalization of the weights the **conditioning** of GMMs can be easily derived. The result is a GMM with the same number of components as the original joint GMM,

Proposition B.2 (GMM Conditioning)**.**

$$
\begin{aligned}
&X|Y \sim \mathcal{G}_{X|Y}(x|y, g_{X|Y}) \\
&K_{X|Y} = K_Z \\
&m_{X|Y\,i} = m_{iX} + \Sigma_{iXY} \cdot {\Sigma_{iYY}}^{-1} \cdot (y - m_{iY}) \\
&\Sigma_{X|Y\,i} = \Sigma_{iXX} - \Sigma_{iXY} \cdot {\Sigma_{iYY}}^{-1} \cdot \Sigma_{iYX} \\
&\pi_{X|Y\,i} = \frac{\pi_i \mathcal{N}(y|m_{Y\,i}, \Sigma_{Y\,i})}{\sum_{j=1}^{K_Y} \pi_j \mathcal{N}(y|m_{Y\,j}, \Sigma_{Y\,j})} \\
&Y|X \sim \mathcal{G}_{Y|X}(y|x, g_{Y|X}) \\
&K_{Y|X} = K_Z \\
&m_{Y|X\,i} = m_{iY} + \Sigma_{iYX} \cdot {\Sigma_{iXX}}^{-1} \cdot (x - m_{ix}) \\
&\Sigma_{Y|X\,i} = \Sigma_{iYY} - \Sigma_{iYX} \cdot {\Sigma_{iXX}}^{-1} \cdot \Sigma_{iXY} \\
&\pi_{Y|X\,i} = \frac{\pi_i \mathcal{N}(x|m_{X\,i}, \Sigma_{X\,i})}{\sum_{j=1}^{K_X} \pi_j \mathcal{N}(x|m_{X\,j}, \Sigma_{X\,j})}
\end{aligned}
\tag{42}
$$

Using the Definition 5.1 and the Proposition A.6 the **Bayesian fusion of GMMs** can be easily derived. If the first GMM has K_a components and the second GMM has K_b components, the result is a GMM with $K_a \cdot K_b$ components, which is the result of pairing all the components of the two original GMMs.

Proposition B.3 (Bayesian Fusion for GMMs)**.**

$$
\begin{aligned}
\mathcal{G}_X(x|g_a) &= \sum_{i=1}^{K_a} \pi_{ai} \mathcal{N}_X(x|m_{ai}, \Sigma_{ai}) \\
\mathcal{G}_X(x|g_b) &= \sum_{j=1}^{K_b} \pi_{bj} \mathcal{N}_X(x|m_{bj}, \Sigma_{bj})
\end{aligned}
\tag{43}
$$

then The following holds:

$$
\begin{aligned}
&\mathcal{G}_X(x|g_a) \cdot \mathcal{G}_X(x|g_b) = \sum_{i=1}^{K_a} \sum_{j=1}^{K_b} c_{a_i b_j} \pi_{ai} \pi_{bj} \mathcal{N}_X(x|m_{a_i b_j}, \Sigma_{a_i b_j}) \\
&\int_\Omega \mathcal{G}_X(x|g_a) \cdot \mathcal{G}_X(x|g_b) \mathrm{d}x = \sum_{i=1}^{K_a} \sum_{j=1}^{K_b} c_{a_i b_j} \pi_{ai} \pi_{bj}
\end{aligned}
\tag{44}
$$

where $c_{a_i b_j}$, $m_{a_i b_j}$ and $\Sigma_{a_i b_j}$ are as in Proposition A.6. This gives a closed form for the Bayesian fusion of two independent information contribution.

B.1 Source mixture

If we have J independent random generators (sources) for one and the same quantity X and if each source is active with probability w^j, with $w^j \geq 0, \sum_{j=1}^{J} w^j = 1$; pooling all these sources

leads to a common PDF $p_X(x)$ that is the convex linear combination of the PDFs $p_X^j(x)$ of the sources; in summary:

$$p_X(x) = \sum_{j=1}^{J} w^j p_X^j(x). \tag{45}$$

In the case that the $p_X^j(x)$s are GMMs, the common PDF $p_X(x)$ becomes a mixture of GMMs. It can be easily seen that $p_X(x)$ is then also a GMM:

$$p_X(x) = \sum_{j=1}^{J} w^j \mathcal{G}_X(x|g_X^j) = \sum_{j=1}^{J} w^j \sum_{i=1}^{K_j} \pi_i^j \mathcal{N}_X(x|m_i^j, \Sigma_i^j) = \sum_{j=1}^{J} \sum_{i=1}^{K_j} w^j \pi_i^j \mathcal{N}_X(x|m_i^j, \Sigma_i^j) \tag{46}$$

thus, $p_X(x)$ is a GMM with $K = \sum_{j=1}^{J} K_j$ Gaussian components with weights $w^j \pi_i^j$. Note that the expectations m_i^j and covariance matrices Σ_i^j remain unmodified. This will lead to the following proposition.

Proposition B.4 (Source Mixture)**.** *Let $X^j \sim \mathcal{G}_{X^j}(x|g_X^j)$. The convex linear combination is also a GMM and:*

$$X \sim \mathcal{G}_X(x) = \sum_{j=1}^{J} w^j \mathcal{G}_X(x|g_X^j) = \sum_{j=1}^{J} w^j \sum_{i=1}^{K_j} \pi_i^j \mathcal{N}_X(x|m_i^j, \Sigma_i^j) = \sum_{j=1}^{J} \sum_{i=1}^{K_j} w^j \pi_i^j \mathcal{N}_X(x|m_i^j, \Sigma_i^j) \tag{47}$$

Having an explicit formula might be a bit overwhelming, but it would enable the reader for a rather easy and clean computer implementation. The common PDF can be written as:

$$p_X(x) = \mathcal{G}_X(x|g_X) = \sum_{k=1}^{K} w_k \mathcal{N}_X(x|m_k, \Sigma_k). \tag{48}$$

The connection between the indices (i, j) and k can be established with the following bijection:

$$\begin{aligned} (i,j) \mapsto k \text{ with } k(i,j) &= \begin{cases} \sum_{v=1}^{i \le K_1} 1 \text{ for } j = 1 \\ \sum_{u=1}^{j-1} K_u + \sum_{v=1}^{i \le K_j} 1 \text{ for } j \ge 2 \end{cases} \\ i(j) &= 1, \ldots, K_j \\ j &= 1, \ldots, J. \end{aligned} \tag{49}$$

The weights w_k are thus:

$$w_k = w_{k(i,j)} = w^j \pi_i^j. \tag{50}$$

Thus the implementation consists of concatenating all Gaussian components of the sources and multiplying their mixture weights by the corresponding source weights.

Proposition B.5 (Convolution of GMMs)**.** *If $X \sim \mathcal{G}_X(x|g_X)$ and $Y \sim \mathcal{G}_Y(x|g_Y)$ are stochastically independent, then $Z = X + Y \sim \mathcal{G}_Z(z|g_Z) = \mathcal{G}_X(x|g_X) * \mathcal{G}_Y(y|g_Y)$ where $*$ denotes the convolution operator and:*

$$\begin{aligned}
g_X &= (K_X, \pi_{Xi}, m_{Xi}, \Sigma_{Xi})_{i=1,\dots,K_X} \\
g_Y &= (K_Y, \pi_{Yj}, m_{Yj}, \Sigma_{Yj})_{j=1,\dots,K_Y} \\
g_Z &= (K_Z, \pi_{Zij}, m_{Zij}, \Sigma_{Zij})_{\substack{i=1,\dots,K_X \\ j=1,\dots,K_Y}} \\
K_Z &= K_X \cdot K_y \\
\pi_{Zij} &= \pi_{Xi} \cdot \pi_{Yj} \\
m_{Zij} &= m_{Xi} + m_{Yj} \\
\Sigma_{Zij} &= \Sigma_{Xi} + \Sigma_{Yj}.
\end{aligned} \tag{51}$$

Bayesian fusion and convolution are two fundamental operations for combining information in the context of GMMs. They result in an increase in the number of components, which can lead to computational challenges. If one can find a closed form integral based metric for the difference between two GMMs as in Equation (21), it would be possible to use optimization techniques to find a GMM with a smaller number of components that approximates the result of the Bayesian fusion or convolution, thus mitigating the computational challenges.

Proposition B.6 (GMM Sampling)**.** *To sample from a GMM, one first selects a component according to the mixture weights, treating them as probabilities of a categorical distribution. Once a component is chosen, a sample is drawn from the corresponding Gaussian distribution defined by that component's mean and covariance. The makes sampling from a GMM an exact process up to the exactness of the uniform samples that are used.*

References

[1] David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[2] Thomas Bayes. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, (53):370–418, 12 1763.

[3] Jürgen Beyerer. *Verfahren zur quantitativen statistischen Bewertung von Zusatzwissen in der Meßtechnik*. VDI Verlag, 1999.

[4] Jürgen Beyerer, Raphael Hagmanns, and Daniel Stadler. *Pattern Recognition: Introduction, Features, Classifiers and Principles*. De Gruyter, 2024.

[5] Christopher M. Bishop. *Pattern recognition and machine learning*. Information Science and Statistics. Springer-Verlag New York, 2006.

[6] G. Breit and E. Wigner. Capture of slow neutrons. *Phys. Rev.*, 49:519–531, Apr 1936.

[7] Jiahua Chen. *Statistical Inference Under Mixture Models*. Springer Nature Singapore, 2023.

[8] Francis Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*. Springer London, 2013.

[9] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons Ltd, second edition, 2006.

[10] Anirban DasGupta. *Fundamentals of Probability: A First Course*. Springer New York, 2010.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 39(1):1–22, 1977.

[12] Bruno Finetti. *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons, Ltd, 2017.

[13] Linus Fischer. *Reduktionsmethoden für multivariate Gaußsche Mischungen*. KIT, IES Chair, Master Thesis, 2026.

[14] Alan Genz and Frank Bretz. *Computation of Multivariate Normal and t-Probabilities*. Springer Berlin Heidelberg, 2009.

[15] Gerald Gerlach and Klaus-Dieter Sommer. *Messunsicherheit: Kurz und praktisch - für Ingenieure und Naturwissenschafler*. De Gruyter, 2024.

[16] Maurice Herlihy and Nir Shavit. *The Art of Multiprocessor Programming*. Morgan Kaufmann, 2012.

[17] Matthew D. Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014.

[18] Marco Huber. *Nonlinear Gaussian Filtering : Theory, Algorithms, and Applications*. KIT Scientific Publishing, Karlsruhe, 2015.

[19] J.D. Jackson. *Classical Electrodynamics, International Adaptation*. John Wiley & Sons Ltd, 2021.

[20] Joint Committee for Guides in Metrology. Evaluation of measurement data — guide to the expression of uncertainty in measurement. Technical report, Joint Committee for Guides in Metrology, 2008.

[21] Joint Committee for Guides in Metrology. Uncertainty of measurement — guide to the expression of uncertainty in measurement, propagation of distributions using a monte carlo method. Technical report, Joint Committee for Guides in Metrology, 2008.

[22] Eric Jondeau, Ser-Huang Poon, and Michael Rockinger. *Financial Modeling Under Non-Gaussian Distributions*. Springer London, 2007.

[23] Achim Klenke. *Probability Theory: A Comprehensive Course*. Springer Cham, 2020.

[24] A. Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer Berlin, 1933.

[25] V.A. Kotelnikov. *The Theory of Optimum Noise Immunity*. McGraw-Hill, 1959.

[26] Dennis V. Lindley. Scoring rules and the inevitability of probability. *International Statistical Review / Revue Internationale de Statistique*, 50(1):1–11, 1982.

[27] Dennis V. Lindley. The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Statistical Science*, 2(1):17–24, 1987.

[28] Yulong Lu, Andrew Stuart, and Hendrik Weber. Gaussian approximations for probability measures on $\mathbb{R}^d$. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1136–1165, 2017.

[29] Pierre Simon marquis de Laplace. *Théorie analytique des probabilités*. Courcier, 1820.

[30] Vladimir Gilelevic Mazja. *Approximate Approximations*. Number v.141 in Mathematical Surveys and Monographs. American Mathematical Society, 2014.

[31] Alvin C. Rencher and William F. Christensen. *Methods of Multivariate Analysis*. John Wiley & Sons Ltd, 2012.

[32] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer New York, 2004.

[33] Franz Schwabl. *Quantum Mechanics*. Springer Berlin Heidelberg, fourth edition, 2007.

[34] Terence Tao. *An introduction to measure theory*. Number 126 in Graduate Studies in Mathematics. American Mathematical Society, 2011. Description based upon print version of record.

[35] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons Ltd, 1985.

[36] Y. L. Tong. *The Multivariate Normal Distribution*. Springer New York, 1990.

[37] John Venn. *The Logic of Chance*. Dover Publications, 2013.

[38] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

[39] Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019.

[40] Gottfried Winkler. *Stochastische Systeme*. Akademische Verlagsgesellschaft, 1977.