Quantum-Like Sampling

Andreas Wichert

Department of Informatics, INESC-ID/IST-University of Lisboa, 1000-029 Lisboa, Portugal;
andreas.wichert@tecnico.ulisboa.pt

Abstract: Probability theory is built around Kolmogorov's axioms. To each event, a numerical degree of belief between 0 and 1 is assigned, which provides a way of summarizing the uncertainty. Kolmogorov's probabilities of events are added, the sum of all possible events is one. The numerical degrees of belief can be estimated from a sample by its true fraction. The frequency of an event in a sample is counted and normalized resulting in a linear relation. We introduce quantum-like sampling. The resulting Kolmogorov's probabilities are in a sigmoid relation. The sigmoid relation offers a better importability since it induces the bell-shaped distribution, it leads also to less uncertainty when computing the Shannon's entropy. Additionally, we conducted 100 empirical experiments by quantum-like sampling 100 times a random training sets and validation sets out of the Titanic data set using the Naïve Bayes classifier. In the mean the accuracy increased from 78.84% to 79.46%.

Keywords: quantum probabilities; sampling; quantum cognition; naïve bayes

1. Introduction

Quantum algorithms are based on different principles than classical algorithm. Here we investigate a simple quantum-like algorithm that is motivated by quantum physics. The incapability between Kolmogorov's probabilities and quantum probabilities results from the different norms that are used. In quantum probabilities the length of the vector in l_2 norm representing the amplitudes of all events is one. Usually, Kolmogorov's probabilities are converted in Quantum-like probabilities by the squared root operation, it is difficult to attach any meaning to the squared root operation. Motivated by the lack of interpretability we define quantum-like sampling, the l_2 sampling. The resulting Kolmogorov's probabilities are not any more linear but related to the sigmoid function. In the following we will introduce the traditional l_1 sampling Then we will introduce the quantum-like sampling, the l_2 sampling. We will indicate the relation between the l_2 sampling, the sigmoid function and the normal distribution. The quantum-like sampling leads to less uncertainty. We fortify this hypothesis by empirical experiments with a simple Naïve Bayes classifier on the Titanic dataset.

2. Kolmogorovs Probabilities

Probability theory is built around Kolmogorov's axioms (first published in 1933 [1]). All probabilities are between 0 and 1. For any proposition x,

$$0 \leq p(x) \leq 1$$

and

$$p(true) = 1, \ p(false) = 0.$$

To each sentence, a numerical degree of belief between 0 and 1 is assigned, which provides a way of summarizing the uncertainty. The last axiom expresses the probability of disjunction and is given by

$$p(x \vee y) = p(x) + p(y) - p(x \wedge y)$$

where do these numerical degrees of belief come from?

- Humans can believe in a subjective viewpoint, which can be determined by some empirical psychological experiments. This approach is a very subjective way to determine the numerical degree of belief.
- For a finite sample we can estimate the true fraction. We count the frequency of an event in a sample. We do not know the true value because we cannot access the whole population of events. This approach is called frequentist.
- It appears that the true values can be determined from the true nature of the universe, for example, for a fair coin, the probability of heads is 0.5. This approach is related to the Platonic world of ideas. However, we can never verify whether a fair coin exists.

Frequentist Approach and l_1 Sampling

Relying on the frequentist approach, one can determine the probability of an event x by counting. If Ω is the set of all possible events, $p(\Omega) = 1$ and the cardinality $\text{card}(\Omega)$ determines the number of elements of a set Ω and $\text{card}(x)$ is the number of elements of the set x and with $x \in \Omega$

$$p(x) = \frac{\text{card}(x)}{\text{card}(\Omega)}. \tag{1}$$

This kind of sampling is the l_1 sampling. With n events $x_1, x_2, ..., x_n$ that describe all possible events of the set Ω,

$$0 \leq \frac{\text{card}(x_i)}{\text{card}(\Omega)}, \quad \sum_{i=1}^{n} \frac{\text{card}(x_i)}{\text{card}(\Omega)} = 1 \Leftrightarrow \sum_{i=1}^{n} \text{card}(x_i) = \text{card}(\Omega)$$

We can interpret $\frac{\text{card}(\omega_i)}{\text{card}(\Omega)}$ as dimension of an n dimensional vector with the vector being

$$\begin{pmatrix} \frac{\text{card}(x_1)}{\text{card}(\Omega)} \\ \frac{\text{card}(x_2)}{\text{card}(\Omega)} \\ \cdots \\ \frac{\text{card}(x_n)}{\text{card}(\Omega)} \end{pmatrix}. \tag{2}$$

with the l_1 norm

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n| \tag{3}$$

it is as a unit-length vector in the norm l_1

$$\left| \frac{\text{card}(x_1)}{\text{card}(\Omega)} \right| + \left| \frac{\text{card}(x_2)}{\text{card}(\Omega)} \right| + \cdots + \left| \frac{\text{card}(x_n)}{\text{card}(\Omega)} \right| = 1. \tag{4}$$

3. Quantum Probabilities

Quantum physics evaluates a probability $p(x)$ of a state x as the squared magnitude of a probability amplitude $A(x)$, which is represented by a complex number

$$p(x) = |A(x)|^2 = A(x)^* \cdot A(x). \tag{5}$$

This is because the product of complex number with its conjugate is always a real number. With

$$A(x) = \alpha + \beta \cdot i \tag{6}$$

$$A(x)^* \cdot A(x) = (\alpha - \beta \cdot i) \cdot (\alpha + \beta \cdot i) \tag{7}$$

$$A(x)^* \cdot A(x) = \alpha^2 + \beta^2 = |A(x)|^2. \tag{8}$$

Quantum physics by itself does not offer any justification or explanation beside the statement that it just works fine, see [2]. This quantum probabilities are as well called von Neumann probabilities. Converting two amplitudes into probabilities leads to an interference term $2 \cdot \Re(A(x) \cdot A^*(y))$,

$$
\begin{aligned}
|A(x) + A(y)|^2 &= |A(x)|^2 + |A(y)|^2 + 2 \cdot \Re(A(x) \cdot A^*(y)) \\
|A(x) + A(y)|^2 &= p(x) + p(y) + 2 \cdot \Re(A(x) \cdot A^*(y)),
\end{aligned}
\tag{9}
$$

making both approaches, in general, incompatible

$$
|A(x) + A(y)|^2 \neq p(x) + p(y).
\tag{10}
$$

In other words, the summation rule of classical probability theory is violated, resulting in one of the most fundamental laws of quantum mechanics, see [2]. In quantum physics we interpret $A(x_i)$ as dimension of an n dimensional vector with the vector being

$$
\begin{pmatrix}
A(x_1) \\
A(x_2) \\
\cdots \\
A(x_n)
\end{pmatrix}.
\tag{11}
$$

In quantum physics the unit-length vector is computed in the in the norm l_2

$$
\|\mathbf{x}\|_2 = \sqrt{(|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2)}
\tag{12}
$$

instead of the norm l_1 with a unit-length vector in the norm l_2

$$
|A(x_1)|^2 + |A(x_2)|^2 + \cdots + |A(x_n)|^2 = 1.
\tag{13}
$$

By replacing the l_1 norm by the Euclidean l_2 norm in the classical probability theory (Kolmogorov probabilities), we obtain quantum mechanics [3] with all the corresponding laws. The incapability between the Kolmogorov's probabilities and quantum probabilities results from the simple fact

$$
\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1.
\tag{14}
$$

This incapability is as well the basis for quantum cognition. Quantum cognition is motivated by clues from psychology indicate that human cognition is based on quantum probability rather than the traditional probability theory as explained by Kolmogorov's axioms, see [4–7]. Empirical findings show that, under uncertainty, humans tend to violate the expected utility theory and consequently the laws of classical probability theory (e.g., the law of total probability [6–8]), In [4,9–14] leading to what is known as the "disjunction effect" which, in turn, leads to violation of the Sure Thing Principle. The violation results from an additional interference that influences the classical probabilities.

Conversion

The amplitude is the root of the belief multiplied with the corresponding phase [11–13] $\theta \in [0, 2 \cdot \pi)$

$$
A(x) = a(x, \theta) = \sqrt{p(x)} \cdot e^{i \cdot \theta}.
\tag{15}
$$

With l_1 sampling it is difficult to attach any meaning to $\sqrt{\mathrm{card}(x)}$,

$$
a(x, \theta) = \sqrt{\frac{\mathrm{card}(x)}{\mathrm{card}(\Omega)}} \cdot e^{i \cdot \theta}.
\tag{16}
$$

Motivated by the lack of interpretability we define the quantum-like sampling, the l_2 sampling with n events $x_1, x_2, ..., x_n$ that describe all possible events

$$\Phi = |\text{card}(x_1)|^2 + |\text{card}(x_2)|^2 + \cdots + |\text{card}(x_n)|^2 \tag{17}$$

$$p(x_i) = \frac{|\text{card}(x_i)|^2}{\Phi}, \quad \sum_{i=1}^{n} \frac{|\text{card}(\omega_i)|^2}{\Phi} = 1 \tag{18}$$

and

$$a(x, \theta) = \frac{\text{card}(x)}{\sqrt{\text{card}(\Phi)}} \cdot e^{i \cdot \theta}. \tag{19}$$

with

$$\Omega \leq \Phi \leq \Omega^2. \tag{20}$$

The l_2 sampling leads to an interpretation of the amplitudes as a normalized frequency of occurrence of an event multiplied by the phase. Φ is dependent on the distribution of all values $\text{card}(x_i)$. When developing empirical experiments that are explained by quantum cognition models l_2 sampling should be used rather than l_1 sampling. What is the interpretation of $|\text{card}(x_i)|^2$ outside quantum interpretation?

4. Quantum-Like Sampling and the Sigmoid Function

To understand the difference between l_1 sampling and l_2 sampling (quantum-like sampling) we will analyze a simple binary event x

$$p(x) + p(\neg x) = 1.$$

For l_1 sampling $p(x)$ and $p(\neg x)$ is defined as

$$p(x) = \frac{\text{card}(x)}{\text{card}(\Omega)}, \quad p(\neg x) = \frac{\Omega - \text{card}(x)}{\text{card}(\Omega)}$$

and for l_2 sampling

$$p(x) = \frac{\text{card}(x)^2}{\text{card}(x)^2 + (\Omega - \text{card}(x))^2}, \quad p(\neg x) = \frac{(\Omega - \text{card}(x))^2}{\text{card}(x)^2 + (\Omega - \text{card}(x))^2}.$$
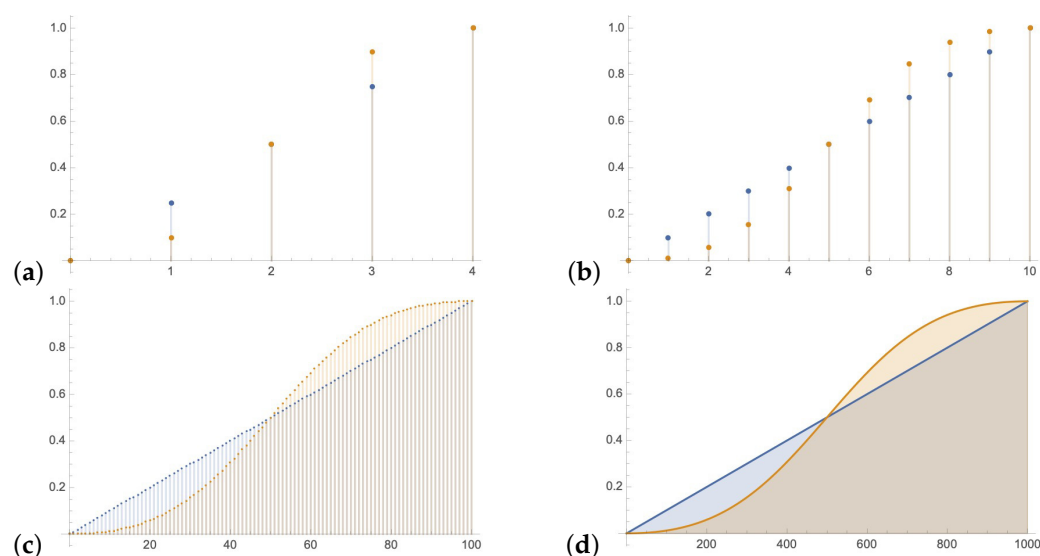
with

$$\omega = \text{card}(x)$$

we can define the functions for binary l_1 and l_2 sampling as

$$f(\omega) = \frac{\omega}{\Omega} \tag{21}$$

and

$$g(\omega) = \frac{\omega^2}{\omega^2 + (\Omega - \omega)^2}, \tag{22}$$

In Figure 1 $f(\omega) = \frac{\omega}{\Omega}$ is compared to $g(\omega) = \frac{\omega^2}{\omega^2 + (\Omega - \omega)^2}$ for Ω, (a) $\Omega = 4$, (b) $\Omega = 10$, (c) $\Omega = 100$ and (d) $\Omega = 1000$. With growing size of Ω the function $g(\omega)$ converge to a continuous sigmoid function. In Figure 2 the sigmoid function $g(x)$ with $\Omega = 10$ is compared to the well known scaled to the logistic function.

Figure 1. $f(\omega) = \frac{\omega}{\Omega}$ compared to $g(\omega) = \frac{\omega^2}{\omega^2 + (\Omega - \omega)^2}$ for Ω, (**a**) $\Omega = 4$, (**b**) $\Omega = 10$, (**c**) $\Omega = 100$ and (**d**) $\Omega = 1000$. With growing size of Ω the function $g(x)$ converge to a sigmoid function.
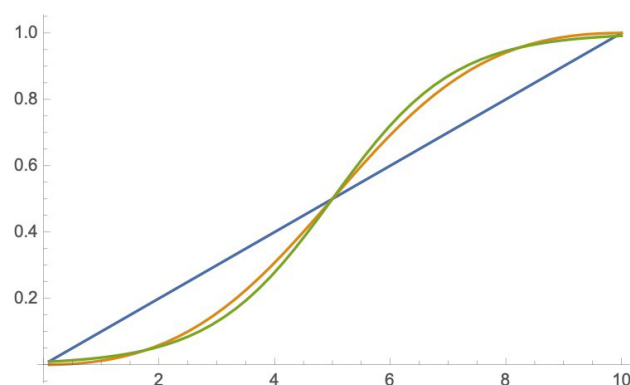
Figure 2. Sigmoid functions: $g(x) = \frac{x^2}{x^2 + (10-x)^2}$ versus logistic function $\sigma(x) = \frac{1}{1 + e^{-0.9523 \cdot (x-5)}}$.

The derivative of the sigmoid function $g(x)$ is a bell-shaped function. For the continuous function $g(x)$, the derivative is similar to the Gaussian distribution. The central limit theorem states that under certain (fairly common) conditions, the sum of many random variables will have an approximately Gausssian distribution. In Figure 3 the derivative of $g(x) = \frac{x^2}{x^2 + (10-x)^2}$ with $\Omega = 10$, $g'(x) = -\frac{5(-10+x)x}{(50-10x+x^2)^2}$ is indicated and compared to Gaussian distribution $\mathcal{N}$ with $\mu = 5$ and $\sigma^2 = 1.9^2$.
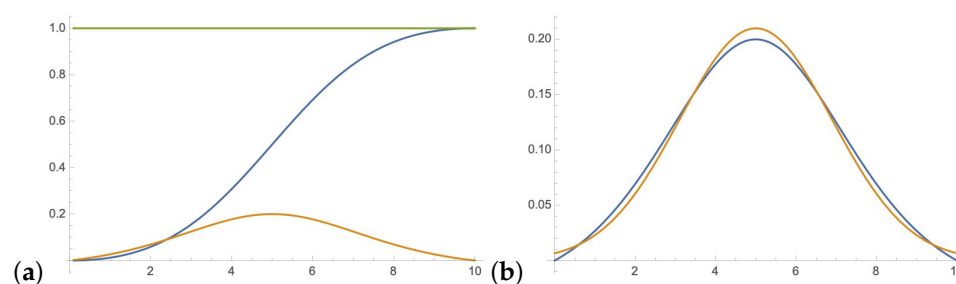

Figure 3. (**a**) $g(x) = \frac{x^2}{x^2 + (10-x)^2}$ and $g'(x) = -\frac{5(-10+x)x}{(50-10x+x^2)^2}$; (**b**) $g'(x) = -\frac{5(-10+x)x}{(50-10x+x^2)^2}$ versus $\mathcal{N}(x|5, 1.9^2) = \frac{1}{1.9 * \sqrt{2*\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-5}{1.9}\right)^2}$.

The derivative of the sigmoid function $g(\omega)$ is less similar to the probability mass function of the the binomial distribution.

The l_2 sampling leads to natural sigmoid representation of probabilities that reflects the nature of Gaussian/Normal distribution. This leads to less uncertainty using l_2 compared to l_1 sampling represented by Shannon's entropy.

$$H = -p(x) \cdot \log_2 p(x) - p(\neg x) \cdot \log_2 p(\neg x) \tag{23}$$

as can be seen in Figure 4 for $\Omega = 100$.
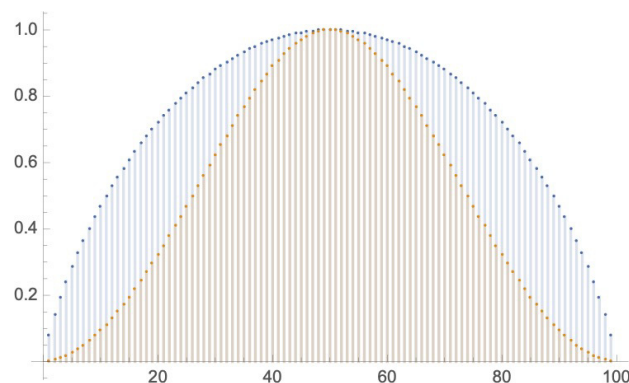
Figure 4. Shannon's entropy H with $\Omega = 100$, blue discrete plot for l_1 sampling and yellow discrete plot l_2 sampling.

Combination

Depending on l_1 or l_2 sampling we get different results when we combine probabilities. Multiplying two independent sampled events x and y results in the joint distribution

$$p_{l1}(x, y) = \frac{x}{10} \cdot \frac{y}{10} \tag{24}$$

when l_1 sampled with $\Omega = 10$ and

$$p_{l2}(x, y) = \frac{x^2}{x^2 + (10 - x)^2} \cdot \frac{y^2}{y^2 + (10 - y)^2} \tag{25}$$

when l_2 sampled with $\Omega = 10$ as indicated in the Figure 5. The same applies for weighted sum of information, corresponding to the Shannon's entropy as used in the *ID*3 algorithm [15–17] for symbolical machine learning to preform a greedy search for small decision trees. Computing

$$H_{l1} = -\frac{x}{10} \cdot \log_2 \frac{x}{10} - \frac{10 - x}{10} \cdot \log_2 \frac{10 - x}{10} - \frac{y}{10} \cdot \log_2 \frac{y}{10} - \frac{10 - y}{10} \cdot \log_2 \frac{10 - y}{10} \tag{26}$$

when l_1 sampled with $\Omega = 10$ and

$$H_{l2} = -\frac{x^2}{x^2 + (10 - x)^2} \cdot \log_2 \frac{x^2}{x^2 + (10 - x)^2} -$$

$$\left(1 - \frac{x^2}{x^2 + (10 - x)^2}\right) \cdot \log_2 \left(1 - \frac{x^2}{x^2 + (10 - x)^2}\right) -$$

$$\frac{x^2}{x^2 + (10 - x)^2} \cdot \log_2 \frac{x^2}{x^2 + (10 - x)^2} -$$

$$\left(1 - \frac{x^2}{x^2 + (10 - x)^2}\right) \cdot \log_2 \left(1 - \frac{x^2}{x^2 + (10 - x)^2}\right) \tag{27}$$

when l_2 sampled with $\Omega = 10$ leads to different H_1 and H_2 values as indicated in the Figure 6.

(a)

(b)

(c)

(d)
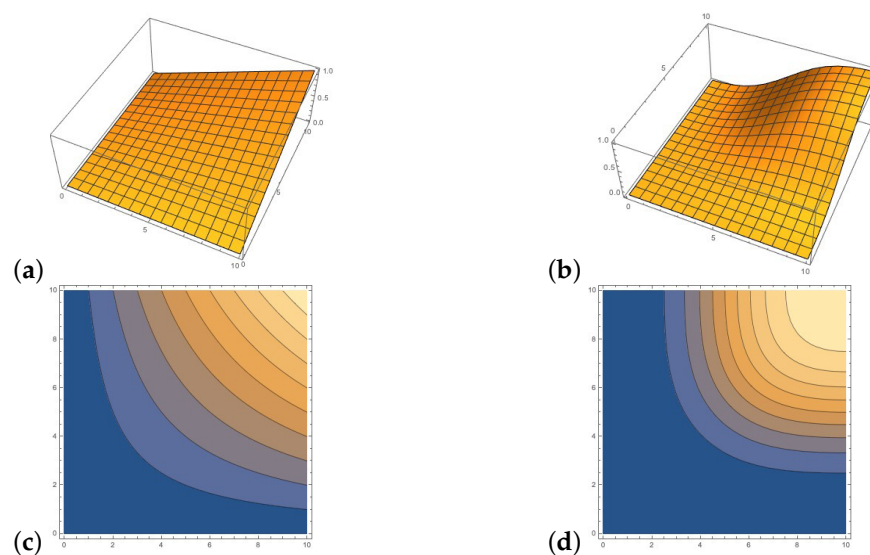
Figure 5. Multiplying two independent sampled events x and y with $\Omega = 10$ leads to different results: (**a**) $p_{l1}(x,y)$, (**b**) $p_{l2}(x,y)$, (**c**) counterplot of $p_{l1}(x,y)$, (**d**) counterplot of $p_{l2}(x,y)$.

(a)

(b)

(c)

(d)
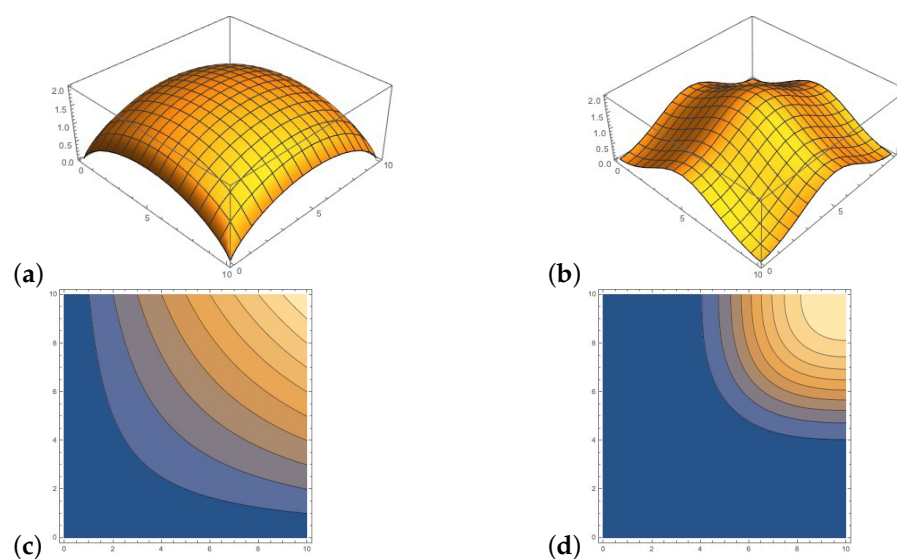
Figure 6. Different H_1 and H_2 values for sampled events x and y with $\Omega = 10$: (**a**) $H_1(x,y)$, (**b**) $H_2(x,y)$, (**c**) counterplot of $H_1(x,y)$, (**d**) counterplot of $H_2(x,y)$.

Less uncertainty would lead to better results in machine learning. We preform empirical experiments with a simple Naïve Bayes classifier to fortify this hypothesis.

5. Naïve Bayes Classifier

For a target function $f : X \rightarrow h$, where each instance x described by attributes $a_1, a_2 \cdots a_n$ Most probable value of $f(x)$ is:

$$h_{map} = \arg\max_{h_i} p(h_i | a_1, a_2, a_3, .., a_n) = \frac{p(a_1, a_2, a_3, ..., a_n | h_i) \cdot P(h_i)}{P(a_1, a_2, a_3, ..., a_n)}$$

The likelihood of the conditional probability $p(a_1, a_2, a_3, ..., a_n | h_i)$ is described by 2^n possible combinations. For n possible variables, the exponential growth of combinations

being true or false becomes an intractable problem for large n since all $2^n - 1$ possible combinations must be known. The decomposition of large probabilistic domains into weakly connected subsets via conditional independence,

$$p(a_1, a_2, a_3, ..., a_n | h_i) = \prod_{j=1}^{n} p(a_j | h_i).$$

is known as the Naïve Bayes assumption and is one of the most important developments in the recent history of Artificial Intelligence [16]. It assumes that a single cause directly influences a number of events, all of which are conditionally independent. The Naïve Bayes classifier is defined as

$$h_{map} = \arg\max_{h_i} \prod_{j=1}^{n} P(a_j | h_i) \cdot P(h_i).$$

Titanic Dataset

When the Titanic sank it killed 1502 out of 2224 passengers and crew. We are using a file titanic.cvs processed file from https://gist.github.com/michhar/ accessed on 1 July 2021 that corresponds to the file train.cvs from kaggle https://www.kaggle.com/c/titanic/data?select=train.csv accessed on 1 July 2021 that contains data for 891 of the real Titanic passengers. Each row represents one person. The columns describe different attributes about the person including their *PassengerId*, whether they *Survived*, heir passenger-class *Pclass*, their *Name*, their *Sex*, their *Age*, and other six attributes, see Figure 7. In our experiment we will only use the attributes *Survived*, *Pclass*, *Sex* and *Age*.

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C
11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7	G6	S
12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S
13	0	3	Saundercock, Mr. William Henry	male	20	0	0	A/5. 2161	8.05		S
14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.275		S
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542		S
16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16		S
17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125		Q
18	1	2	Williams, Mr. Charles Eugene	male		0	0	244373	13		S
19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31	1	0	345763	18		S
20	1	3	Masselmani, Mrs. Fatima	female		0	0	2649	7.225		C
21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26		S

Figure 7. The Titanic data set is represented in an Excel table that contains data for 891 of the real Titanic passengers, some entries are not defined.

The binary attribute *Survived* (=s) will be our target h resulting in the prior $p(s)$ and $p(\neg s)$. Out of the 891 passengers 342 survived and 549 did not survive. The attribute *Pclass* has three values 1 for upper class, 2 for middle class and 3 for lower class. It will be binarized into the attribute c indicating if the person had a lower class cabin by

$$c = 1 \; if \; Pclass == 3 \; else \; c = 0,$$

in the case *Pclass* is not defined the default value is 0. The attribute *Sex* has two binary values represented by the binary attribute m

$$m = 1 \; if \; Sex == male \; else \; m = 0,$$

in the case *Sex* is not defined the default value is 0. The attribute *Age* will be binarized into the attribute g indicating if the person is a child or grown up by

$$g = 0 \ \ if \ \ Age < 18 \ \ else \ \ g = 1,$$

in the case *Age* is not defined the default value is 1.

Using the l_1 sampling over the whole data set we get the values for the priors

$$p(s) = 0.3838, \ \ p(\neg s) = 0.6162,$$

and for the likelihoods,

$$p(c|s) = 0.3480, \ \ p(\neg c|s) = 0.6520, \ \ p(c|\neg s) = 0.6776, \ \ p(\neg c|\neg s) = 0.3224,$$

$$p(m|s) = 0.3187, \ p(\neg m|s) = 0.6813, \ p(m|\neg s) = 0.8525, \ p(\neg m|\neg s) = 0.14755,$$

$$p(g|s) = 0.8216, \ p(\neg g|s) = 0.178363, \ p(g|\neg s) = 0.9053, \ p(\neg g|\neg s) = 0.0947.$$

Using the l_2 sampling over the whole data set we get the values for the priors

$$p(s) = 0.27957406, \ \ p(\neg s) = 0.72042594$$

and for the likelihoods,

$$p(c|s) = 0.2217, \ p(\neg c|s) = 0.7783, \ p(c|\neg s) = 0.8154, \ p(\neg c|\neg s) = 0.1846,$$

$$p(m|s) = 0.1796, \ p(\neg m|s) = 0.8524, \ p(m|\neg s) = 0.9709, \ p(\neg m|\neg s) = 0.0291,$$

$$p(g|s) = 0.95499, \ p(\neg g|s) = 0.04501, \ p(g|\neg s) = 0.9892, \ p(\neg g|\neg s) = 0.0108.$$

The sampled probability values are quite different.

We measure the accuracy of our algorithm by dividing the number of entries it correctly classified by the total number of entries. For l_1 sampling there 190 are entries wrong classified resulting in the accuracy of 78.66%, for the l_2 sampling 179 entries are wrong classified resulting in the accuracy of 79.91%.

In the next step we separate our data set in a training set and a validation set resulting in 295 elements in the validation set We use sklearn.model_selection accessed on 1 July 2021 import train_test_split accessed on 1 July 2021 with the parameters train_test_split(XH,yy, test_size=0.33,random_state=42) accessed on 1 July 2021. The later is used to validate how well our algorithm is doing. For l_1 sampling 60 entries are wrong classified resulting in the accuracy of 79.66%, for the l_2 sampling 55 entries are wrong classified resulting in the accuracy of 81.36%.

In the next we sample the 100 times a random training set and validation set We use sklearn.model_selection accessed on 1 July 2021 import train_test_split accessed on 1 July 2021 with the parameters train_test_split(XH,yy,test_size=0.33) accessed on 1 July 2021. For l_1 sampling there were in the mean 62.42 entries wrong classified resulting in the accuracy of 78.84%, for the l_2 sampling there were 60.59 entries wrong classified resulting in the accuracy of 79.46%.

The trend from this simple evaluation indicates that l_2 sampling leads to better results compared to l_1 sampling in the empirical experiments.

So far we looked at the performance of classification models in terms of accuracy. Specifically, we measured error based on the fraction of mistakes. However, in some tasks, there are some types of mistakes that are worse than others. However in our simple task the correctly classified results when l_1 sampled are also correctly classified when l_2 sampled.

6. Conclusions

We introduced quantum-like sampling also called the l_2 sampling. The l_2 sampling leads to an interpretation of the amplitudes as a normalized frequency of occurrence of an event multiplied by the phase

$$a(x, \theta) = \frac{\mathrm{card}(x)}{\sqrt{\mathrm{card}(\Phi)}} \cdot e^{i \cdot \theta},$$

Φ is dependent on the distribution of all values $\mathrm{card}(x_i)$. When developing empirical experiments that are explained by quantum cognition models l_2 sampling should be used rather than l_1 sampling.

The quantum inspired l_2 sampling maps the probability values to a natural continuous sigmoid function, its derivative is is a bell-shaped function that is similar to the Gaussian distribution of events. The l_2 sampling improves the classification accuracy in machine learning models that are based on sampling as indicated by empirical experiments with Naïve Bayes classifier.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this paper are provided within the main body of the manuscript.

Conflicts of Interest: Compliance with Ethical Standards: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The author declare no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Kolmogorov, A. *Grundbegriffe der Wahrscheinlichkeitsrechnung*; Springer: Berlin, Germany, 1933.
2. Binney, J.; Skinner, D. *The Physics of Quantum Mechanics*; Oxford University Press: Oxford, UK, 2014
3. Aaronson, S. Is quantum mechanics an island in theoryspace? In *Proceedings of the Växjö Conference Quantum Theory: Reconsideration of Foundations*; Khrennikov, A., Ed.; University Press: Vaxjo, Sweden, 2004; Volume 10, pp. 15–28.
4. Busemeyer, J.; Wang, Z.; Trueblood, J. Hierarchical bayesian estimation of quantum decision model parameters. In Proceedings of the 6th International Symposium on Quantum Interactions, Paris, France, 27–29 June 2012; pp. 80–89.
5. Busemeyer, J.R.; Trueblood, J. Comparison of quantum and bayesian inference models. In *Quantum Interaction*; Lecture Notes in Computer Science; Bruza, P., Sofge, D., Lawless, W., van Rijsbergen, K., Klusch, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5494, pp. 29–43.
6. Busemeyer, J.R.; Wang, Z.; Lambert-Mogiliansky, A. Empirical comparison of markov and quantum models of decision making. *J. Math. Psychol.* **2009**, *53*, 423–433. [CrossRef]
7. Busemeyer, J.R.; Wang, Z.; Townsend, J.T. Quantum dynamics of human decision-making. *J. Math. Psychol.* **2006**, *50*, 220–241. [CrossRef]
8. Khrennikov, A. Quantum-like model of cognitive decision making and information processing. *J. Biosyst.* **2009**, *95*, 179–187. [CrossRef] [PubMed]
9. Busemeyer, J.; Pothos, E.; Franco, R.; Trueblood, J. A quantum theoretical explanation for probability judgment errors. *J. Psychol. Rev.* **2011**, *118*, 193–218. [CrossRef] [PubMed]
10. Busemeyer, J.; Wang, Z. Quantum cognition: Key issues and discussion. *Top. Cogn. Sci.* **2014**, *6*, 43–46. [CrossRef] [PubMed]
11. Wichert, A. *Principles of Quantum Artificial Intelligence: Quantum Problem Solving and Machine Learning*, 2nd ed.; World Scientific: Singapore, 2020.
12. Wichert, A.; Moreira, C. Balanced quantum-like model for decision making. In Proceedings of the 11th International Conference on Quantum Interaction, Nice, France, 3–5 September 2018; pp. 79–90.
13. Wichert, A.; Moreira, C.; Bruza, P. Quantum-like bayesian networks. *Entropy* **2020**, *22*, 170. [CrossRef] [PubMed]
14. Yukalov, V.; Sornette, D. Decision theory with prospect interference and entanglement. *Theory Decis.* **2011**, *70*, 283–328. [CrossRef]

15. Luger, G.F.; Stubblefield, W.A. *Artificial Intelligence, Structures and Strategies for Complex Problem Solving*, 3rd ed.; Addison-Wesley: Boston, MA, USA, 1998.
16. Mitchell, T. *Machine Learning*; McGraw-Hill: New York, NY, USA, 1997.
17. Winston, P.H. *Artificial Intelligence*, 3rd ed.; Addison-Wesley: Boston, MA, USA, 1992.