Noninformative Prior Weights for Dirichlet PDFs *

Audun Jøsang[1] and Jin-Hee Cho[2] and Feng Chen[3]

***Abstract*— The noninformative prior weight W of a Dirichlet PDF (Probability Density Function) determines the balance between the prior probability and the influence of new observations on the posterior probability distribution. In this work, we propose a method for dynamically converging the weight W in a way that satisfies two constraints. The first constraint is that the prior Dirichlet PDF (i.e. in the absence of evidence) must always be uniform, which dictates that $W = k$ where k is the cardinality of the domain. The second constraint is that the prior weight of large domains must not be so heavy that it prevents new observation evidence from having the expected influence over the shape of the Dirichlet PDF, which dictates that W quickly converges to a low constant C_W in the presence of observation evidence, where typically $C_W = 2$. In the case of a binary domain, the noninformative prior weight is normally set to $W = 2$, irrespective of the amount of evidence. In the case of a multidimensional domain with arbitrarily large cardinality k, the noninformative prior weight is initially equal to the domain cardinality k, but rapidly decreases to the constant convergence factor C_W as the amount of evidence increases.**

I. INTRODUCTION

Dirichlet PDFs (Probability Density Functions) are equivalent to subjective opinions which represent arguments in subjective logic (SL). A subjective opinion can explicitly represent epistemic uncertainty of probabilities through its belief mass distribution, its epistemic uncertainty mass, and its prior probability distribution [1]. Note that the concept of *prior probability* is normally called *base rate* in subjective logic. Subjective opinions implicitly also represent other uncertainty characteristics, such as vagueness, dissonance, and consonance [2], but these characteristics are not discussed in the present study.

In the present work, we propose a simple function to determine the noninformative prior weight W of Dirichlet PDFs (Probability Density Functions), which thereby also applies to multinomial subjective opinions. This prior weight is designed to provide a sound balance between the prior probability and the influence of new evidence on the posterior probability.

II. PRIOR PROBABILITY AND PRIOR WEIGHT

The concept of prior probabilities is central in the theory of probability. For example, prior probabilities are needed for default reasoning, for Bayes' theorem, for abductive reasoning and for Bayesian updating. This section describes the concept of prior probability distribution over random variables, and how the noninformative prior weight influences posterior probability distributions.

Given a domain $\mathbb{X}$ of cardinality k, the default prior probability of each singleton value in the domain is uniformely $1/k$. The default prior probability of a subset consisting of n singletons is n/k. This type of subset is called a composite state value, which has default prior probability equal to the number of singletons it contains, relative to the cardinality of the whole domain. The default prior probability of a subset is sometimes called 'relative atomicity' in the literature. In addition to defining default prior probability relative to the whole domain $\mathbb{X}$, we can also define *default relative prior probability* with respect to another fully or partially overlapping state value $x \subset \mathbb{X}$.

In contrast to default prior probabilities, it is possible and useful to apply realistic prior probabilities that reflect real background probabilities for real situations. Realistic prior probabilities are in general different from default prior probabilities. Considering for example the prior probability of a particular infectious medical condition in a given population, the domain can be defined as the binary set $\{$*'infected'*, *'not infected'*$\}$ with respect to that particular medical condition. Assuming that a random person enters a hospital, the physician would *a priori* not know whether that person is infected or not, because the physician does not have any evidence.

Applying default prior probabilities to the case of the infectious condition, would mean that the physician assumes a prior probability of 0.5 that the the person is infected, which would be totally inadequate in general. Typically, the background probability of an infectious condition is normally much lower than 0.5, and can typically be determined given relevant statistics from a given population.

Statistical data about medical conditions is collected from hospitals, clinics and other sources where people are treated for particular medical conditions. To determine infection rates for infectious conditions is precisely to determine prior probabilities for the same conditions. This can be determined through statistics, guidelines and expert opinions.

It is extremely useful to have available prior probabilities for medical conditions. This data can be used not only by policy makers, the prior probabilities can also be used with medical tests to provide a better indication of the likelihood that a patient has a specific medical condition [3].

It is also possible to dynamically update prior probabilities as a function of observed evidence. Assume e.g. an urn

* Published at the 2022 IEEE International Conference on Multisensor Fusion and Integration (MFI 2022), Cranfield, UK, September 2022. This work is partially supported by NSF under the Grant Number 2107449, 2107450, and 2107451.

[1]Audun Jøsang is with University of Oslo, Norway. josang@mn.uio.no

[2]Jin-Hee Cho is with Virginia Tech, USA. jicho@vt.edu

[3]Feng Chen is with University of Texas at Dallas, USA. Feng.Chen@UTDallas.edu

containing balls of red (x) and black ($\overline{x}$) colour of unknown proportions, then the initial prior probabilities of the two types of balls should be the default uniform prior probability $a(x) = a(\overline{x}) = 1/2$. Then, after picking (with return) and observing some balls, the prior probabilities can be adjusted closer to the relative proportions of observed balls.

The advantage of integrating prior probabilities with belief mass distribution and epistemic uncertainty in subjective opinions is to enables a simple intuitive interpretation of beliefs, and to provide a basis for conditional reasoning under uncertainty. When computing the projected (expected) probability distribution of a opinion, the contribution from the prior probability distribution is proportional to the uncertainty mass. In case of total uncertainty, the projected posterior probability is equal to the prior probability.

Prior probabilities are expressed in the form of a prior probability distribution, $\boldsymbol{a}_X$, so that $\boldsymbol{a}_X(x)$ represents the prior probability of the state value $x \in \mathbb{X}$. Prior probability distribution is formally defined below.

Definition 1 (Prior Probability Distribution): Let $\mathbb{X}$ be a domain, and let X be a random state variable in $\mathbb{X}$. The prior probability distribution $\boldsymbol{a}_X$ assigns a prior probability to possible values of $X \in \mathbb{X}$, and is an additive probability distribution, formally expressed as:

$$\begin{aligned} &\text{Prior probability distribution: } \boldsymbol{a}_X : \mathbb{X} \to [0,1], \\ &\text{with the additivity requirement: } \sum_{x \in \mathbb{X}} \boldsymbol{a}_X(x) = 1 \ . \end{aligned} \tag{1}$$

□

The prior probability distribution is normally assumed to be shared among analysts (i.e., not subjective) because it is based on general background information. Although different analysts can have different opinions about the same state variable, they normally share the same prior probability distribution. However, in case two observers or analysts do not share the same background information it is obvious that they can assign different prior probability distributions to the same state variable. Hence, prior probabilities can be shared or subjective.

This flexibility allows two different analysts to assign a different belief mass distribution and uncertainty mass, as well as a different prior probability distribution to the same state variable. In this way, the different views, analyses and interpretations of the same situation by different observers can be naturally expressed.

Events that can be repeated many times are typically frequentist in nature, meaning that prior probabilities for such events typically can be derived from statistical observations. For an event that happened once in the past, or that might happen once in the future, the analyst must typically determine prior probabilities from subjective intuition, or from analysing the nature of the phenomenon using scientific methods. However, in many cases this can lead to considerable vagueness about prior probabilities, and when nothing else is known, it is possible to use the default prior probability distribution for a random state variable. More specifically, when there are k singletons in the domain, the default prior probability of each singleton is $1/k$.

The difference between subjective and frequentist probabilities is that the former typically is defined as subjective betting odds – and the latter based on relative frequencies of empirically observed data. Note that subjective probability typicvally converges towards frequentist probability when empirical data is collected and becomes available [4]. The concepts of *subjective* and *empirical* prior probabilities can be interpreted in a similar fashion, where they also converge and merge into a single prior probability when empirical data about the population is collected and becomes available.

When no evidence other than the prior probability distribution is known, the projected probability distribution is equal to the prior probability distribution. As the amount of collected evidence increases, the projected probability distribution is increasingly determined by that evidence and decreasingly determined by the prior probability distribution. The concept of *prior weight*, here denoted by W, is the factor which determines how strongly the projected probability is determined by the prior probability distribution relative to the evidence. A heavy prior weight lets the prior probability distribution have more influence on the projected probability distribution, and a light prior weight lets the collected evidence have more influence.

The concept of *noninformative prior weight* expresses that prior probability based on background statistics or background beliefs should not be considered as observation evidence, hence the term 'noninformative'. However, the prior can not be totally weightless, it has to carry some weight. To determine a noninformative prior weight which gives sound and rational balance between background and observation is the main contribution of this paper.

III. BETA PDF AND BINOMIAL OPINIONS

The opinion type in subjective logic depends on the state variable it applies to, i.e., binomial opinions for binary variables, multinomial opinions for multidimensional state variables, and hyper-opinions for hypervariables. In this section, we provide a brief description of Beta PDFs (probability density functions), and how they correspond to binomial opinions [1].

A. Beta Distribution

Given the binary domain $\mathbb{X} = \{x, \bar{x}\}$ and the value $x \in \mathbb{X}$, $\text{Beta}(p_x)$ is the probability density function $\text{Beta}(p_x; \alpha, \beta) : [0,1] \to \mathbb{R}_{\geq 0}$ where $p_x + p_{\bar{x}} = 1$. The Beta PDF is given by:

$$\text{Beta}(p_x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (p_x)^{\alpha - 1} (1 - p_x)^{\beta - 1}, \tag{2}$$

where $\alpha > 0, \beta > 0$, $p(x) \neq 0$ if $\alpha < 1$ and $p(x) \neq 1$ if $\beta < 1$; and the additivity requirement should hold with $\int_0^1 \text{Beta}(p_x; \alpha, \beta)\, \mathrm{d}p_x = 1$. The α and β parameters can simply be represented by the prior probability a_x and the observation evidence (r_x, s_x) where r_x is the amount of positive evidence and s_x is the amount of negative evidence:

$$\alpha = r_x + a_x W, \qquad \beta = s_x + (1 - a_x)W \,. \tag{3}$$

W is the noninformative prior weight in the absence of positive evidence r_x or negative evidence s_x. The expected probability of the Beta PDF is given by Eq. (4):

$$\mathrm{E}(x) = \frac{\alpha}{\alpha + \beta} = \frac{r_x + a_x W}{r_x + s_x + W} \tag{4}$$

For the binomial Beta distribution, the prior weight $W = 2$ produces consistent and intuitive results for all types of evidence, from vacuous (no evidence) to dogmatic (an infinite amount of evidence) and anything in between.

The special case of $\alpha + \beta = 2$ leads to $a = \frac{\alpha}{2}$, and in particular when $\alpha = \beta = 1$ then $a = \frac{1}{2}$.

As an example of the mapping from opinions to Beta distributions, the opinion of Fig. 2 corresponds to the Beta PDF of Fig. 1.

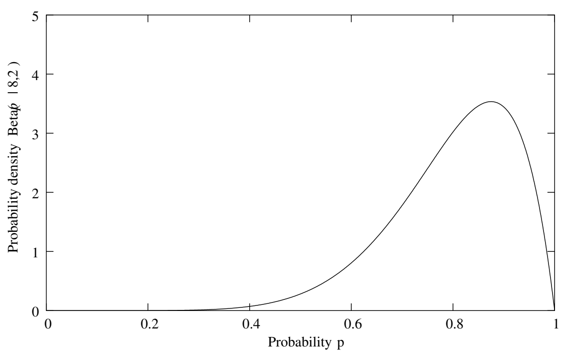

Fig. 1. Plot of function $\mathrm{Beta}(p \mid 8, 2)$ PDF

B. Binomial Opinion Representation

A *binomial opinion* is equivalent to a Beta PDF through the bijective mapping of Eq. (6). A binomial opinion on a given proposition x is represented by $\omega_x = (b_x, d_x, u_x, a_x)$ where the opinion applies to the value x in the binary domain $\mathbb{X} = \{x, \bar{x}\}$ with the additivity requirement $b_x + d_x + u_x = 1$. To be specific, each parameter indicates,

- b_x: *belief mass* supporting x being *TRUE*;
- d_x: *disbelief mass* supporting x being *FALSE*, i.e., supporting $\bar{x}$ being TRUE;
- u_x: *epistemic uncertainty*; and
- a_x: *prior probability* of x being TRUE.

The projected probability of x is computed as:

$$\mathrm{P}(x) = b_x + a_x u_x \tag{5}$$

A *vacuous opinion* is an opinion for which $u_x = 1$, meaning that $b_x = d_x = 0$. A binomial opinion is typically used for judging *true* vs. *false* or *agree* vs. *disagree*.

Fig. ˜2 illustrates the opinion $\omega_x = (0.7,\ 0.1,\ 0.2,\ 0.5)$ indicated by a black dot in the triangle.

The *projector* line through the opinion point, parallel to the line from the uncertainty top vertex to the prior probability point a_x, determines the projected probability $\mathrm{P}(x)$.

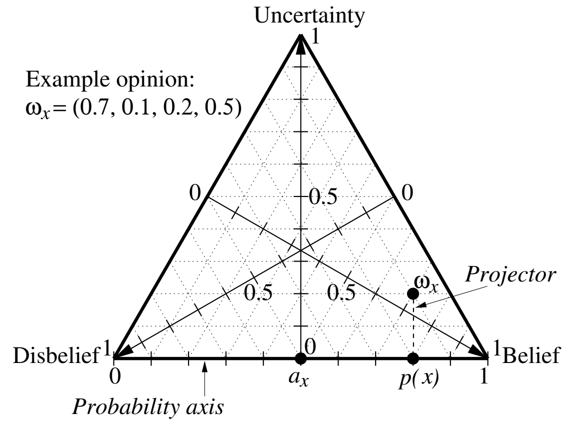

Fig. 2. Opinion triangle

Equivalence netween a binomial opinion and a Beta PDF can be achieved through the following mapping rule:

$$\begin{cases} b_x = \frac{r_x}{r_x + s_x + W}, \\ d_x = \frac{s_x}{r_x + s_x + W}, \\ u_x = \frac{W}{r_x + s_x + W} \ . \end{cases} \tag{6}$$

The mapping of Eq. (6) leads to $\mathrm{P}(x) = \mathrm{E}(x)$, i.e. to equality between the projected probability of the binomial opinion and the expected probability of the Beta PDF.

Note that the noninformative prior weight W is a factor in the mapping of Eq. (6). It can be seen that a large W produces a relatively larger epistemic uncertainty mass u_x in the presence of evidence r_x and s_x. For the mapping from a Beta PDF to a binomial opinion the noninformative prior weight is normally set to $W = 2$.

The uniform Beta distribution is $\mathrm{Beta}(1, 1)$ which corresponds to the opinion $w_x = (0,\ 0,\ 1,\ \frac{1}{2})$. This particular opinion thus represents the case when nothing is known about the probability of x, i.e. its distribution is uniform, as illustrated in Fig. 3.

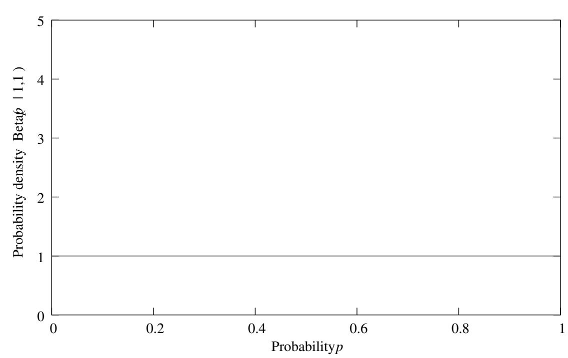

Fig. 3. Uniform $\mathrm{Beta}(p \mid 1, 1)$

IV. DIRICHLET PDF AND MULTINOMIAL OPINIONS

A. Dirichlet Distribution

Multinomial probability density over a domain of cardinality k is represented by the k-dimensional Dirichlet PDF, where the special case with $k = 2$ is the Beta PDF as discussed above. Let $\mathbb{X}$ denote a domain of k mutually disjoint state values, let α_X denote the strength vector over the values of $\mathbb{X}$ and let $\mathbf{p}_X$ denote the probability distribution over $\mathbb{X}$. Dirichlet PDF with $\mathbf{p}_X$ as k-dimensional probability variable is defined by:

$$\mathrm{Dir}(\boldsymbol{p}_X; \alpha_X) = \frac{\Gamma\Big(\sum_{x\in\mathbb{X}} \alpha_X(x)\Big)}{\prod_{x\in\mathbb{X}}(\alpha_X(x))} \prod_{x\in\mathbb{X}} \boldsymbol{p}_X(x)^{(\alpha_X(x)-1)}, \quad (7)$$

where $\alpha_X(x) \geq 0$ and $\mathbf{p}_X(x) \neq 0$ if $\alpha_X(x) < 1$. The total strength $\alpha_X(x)$ for each belief value $x \in \mathbb{X}$ can be given by:

$$\alpha_X(x) = \boldsymbol{r}_X(x) + \boldsymbol{a}_X(x)W, \text{where } \boldsymbol{r}_X(x) \geq 0. \ \forall x \in \mathbb{X}, \quad (8)$$

Here W is noninformative weight representing the amount of influence that the prior probability distribution $\boldsymbol{a}_X$ shall have on the expected probability distribution $\mathbf{E}_X$. Given the Dirichlet PDF, the expected probability distribution over $\mathbb{X}$ can be obtained by:

$$\mathbf{E}_X(x) = \frac{\alpha_X(x)}{\sum_{x_i\in\mathbb{X}} \alpha_X(x_i)} = \frac{\boldsymbol{r}_X(x) + \boldsymbol{a}_X(x)W}{W + \sum_{x_i\in\mathbb{X}} \boldsymbol{r}_X(x_i)} \quad (9)$$

B. Multinomial Opinion Representation

A *multinomial opinion* is equivalent to a Dirichlet PDF through the bijective mapping of Eq. (11). A multinomial opinion in a given proposition x is represented by $\omega_X = (\boldsymbol{b}_X, u_X, \boldsymbol{a}_X)$ over a domain $\mathbb{X}$ with random state variable $X \in \mathbb{X}$. We assume domain cardinality $k = |\mathbb{X}| > 2$ and the additivity requirement $\sum_{x\in\mathbb{X}} \boldsymbol{b}_X(x) + u_X = 1$. To be specific, each parameter indicates,

- $\boldsymbol{b}_X$: *belief mass distribution* over $\mathbb{X}$;
- u_X: *epistemic uncertainty mass*;
- $\boldsymbol{a}_X$: *prior probability distribution* over $\mathbb{X}$.

The projected probability distribution of multinomial opinions is given by:

$$\mathbf{P}_X(x) = \boldsymbol{b}_X(x) + \boldsymbol{a}_X(x)u_X, \quad \forall x \in \mathbb{X} \quad (10)$$

The only type of multinomial opinions that can be easily visualised is trinomial, which can be represented as a point inside a tetrahedron (3D simplex), which in fact is a barycentric coordinate system of four axes, as shown in Fig. 4.

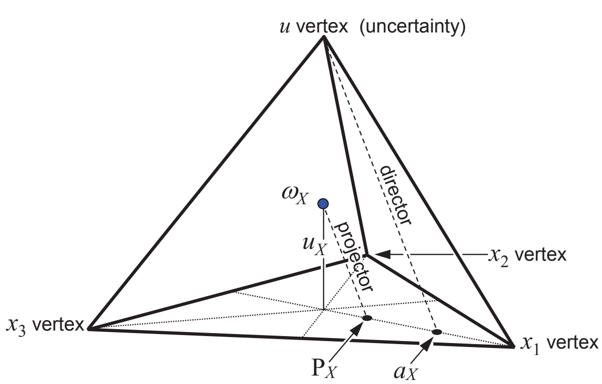

Fig. 4. Visualisation of trinomial opinion as barycentric tetrahedron

In Fig. 4, the vertical elevation of the opinion point inside the tetrahedron represents the uncertainty mass. The distances from each of the three triangular side planes to the opinion point represent the respective belief masses. The prior probability distribution $\boldsymbol{a}_X$ is indicated as a point on the base triangular plane. The line that joins the tetrahedron summit and the prior probability distribution point represents the director. The projected probability distribution point is geometrically determined by tracing a projection from the opinion point, parallel to the director, onto the base plane.

Assume a ternary domain $\mathbb{X} = \{x_1, x_2, x_3\}$, and a corresponding random state variable X. Fig. 4 shows the tetrahedron with the example multinomial opinion ω_X that has belief mass distribution $\boldsymbol{b}_X = \{0.20, 0.20, 0.20\}$, uncertainty mass $u_X = 0.40$, and prior probability distribution $\boldsymbol{a}_X = \{0.750, 0.125, 0.125\}$. Only the uncertainty axis is shown in Fig. 4. The belief axes for x_1, x_2 and x_3 are not shown due to the difficulty of 3D visualisation on the 2D plane of Fig. 4.

The triangle and tetrahedron are part of the *simplex* family of geometrical shapes. A multinomial opinion on a domain of cardinality k can be represented as a point in a simplex of dimension k. For example, a binomial opinion can be represented inside a triangle which is a 2D simplex, and a trinomial opinion can be represented inside a tetrahedron which is a 3D simplex.

By applying Eq. (10) to the example of Fig. 4, the projected probability distribution is $\mathbf{P}_X = \{0.50, 0.25, 0.25\}$.

The observed evidence in the Dirichlet PDF can be mapped to the multinomial opinions as:

$$\begin{cases} \boldsymbol{b}_X(x) = \dfrac{\boldsymbol{r}(x)}{\sum_{x_i\in\mathbb{X}} \boldsymbol{r}(x_i) + W} \\ \\ u_X = \dfrac{W}{\sum_{x_i\in\mathbb{X}} \boldsymbol{r}(x_i) + W} \end{cases} \quad (11)$$

The mapping of Eq. (11) leads to $\mathbf{P}_X = \mathrm{E}_X$, i.e. to equality between the projected probability distribution of the nultinomial opinion and the expected probability distribution of the Dirichlet PDF.

Note that a binomial or multinomial opinion assigns belief mass to singleton state values only. However, in real life, we sometimes have difficulty in distinguishing between possible states due to cognitive limitations or environmental noise. In such situations we might want to assign belief mass to composite values. This kind of opinions are called hyper-opinions, which correspond to Dirichlet hyper PDFs. However, hyper opinions are not discussed here.

V. PRIOR WEIGHT OF THE DIRICHLET PDF

For Dirichlet PDFs, which are equivalent to multinomial opinions, the problem with setting the prior weight as $W = 2$ is that a totally uncertain opinion doe not correspond to a uniform Dirichlet PDF, as illustrated in Fig. 5.

It is normally assumed that the prior probability density in case of a binary domain $\mathbb{X} = \{x, \overline{x}\}$ is uniform. This requires that $\alpha_X(x) = \alpha_X(\overline{x}) = 1$, which means that

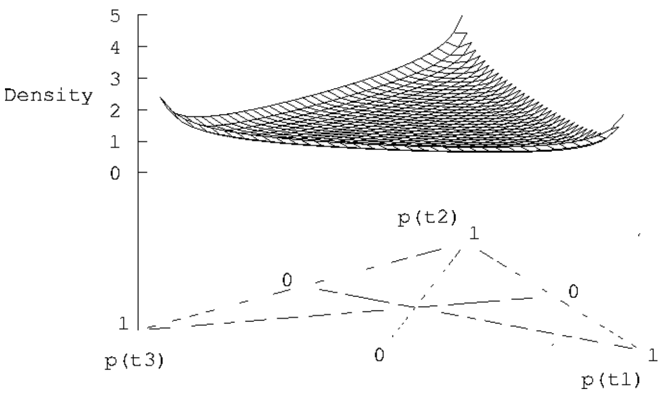

Fig. 5. Non-uniform vacuous trinomial Dirichlet PDF $\text{Dir}(\boldsymbol{p}_X \mid \{0.67, 0.67, 0.67\})$ with $W = 2$.

$W = \alpha_X(x) + \alpha_X(\overline{x}) = 2$. Similarly, a uniform prior probability density over a domain larger than binary would require an noninformative prior weight $W = k$, meaning that W must be equal to the cardinality k of the domain for which a uniform prior probability density is assumed.

However, selecting $W > 2$ would result in new observation evidence having relatively less influence over the Dirichlet PDF, and over the posterior probability distribution. In fact it would be unnatural to set $W = k$ in general for arbitrarily large domains, because a large W would make the PDF relatively insensitive to new observation evidence.

As an example, consider a domain of cardinality 100. To have a uniform prior PDF when no evidence has been received would require $W = 100$. Assume now that a specific event (state value) of interest has been observed 100 times, and no other event has been observed, then the projected probability of the event would only be about $^1/_2$, which would be highly counter-intuitive and inadequate. In contrast, when assuming a uniform PDF in the binary case, meaning that $W = 2$, and assuming the positive outcome has been observed 100 times, and the negative outcome has not been observed, then the projected probability of the positive outcome is close to 1, as intuition would dictate.

As a solution to making any prior Dirichlet PDF uniform for any domain cardinality, and still allowing new observations to have a normal influence on the PDF, we define the *convergent noninformative prior weight* expressed by Eq. (12) below.

Definition 2 (Convergent Noninformative Prior Weight): Let $\mathbb{X}$ be a domain with cardinality k and for which the vector $\boldsymbol{r}_X$ represents collected evidence. Assuming the convergence factor C_W, then the convergent noninformative prior weight for the Dirichlet PDF over $\mathbb{X}$ is defined by:

$$W = \frac{k + C_W k \sum_k \boldsymbol{r}_X(x)}{1 + k \sum_k \boldsymbol{r}_X(x)} \tag{12}$$

□

The convergence factor C_W determines the sensitivity of the Dirichlet PDF to new observations. The larger C_W, the less sensitive the Dirichlet PDF becomes to new observation evidence. If we assume that the sensitivity should always be the same as for the binomial case, then it is natural to set $C_W = 2$, which reflects the noninformative prior weight of the uniform Beta PDF over a binary domain.

The property of W according to Eq. (12) is, for example, that $W = 2$ when $k = 2$, that $W = k$ when $\sum_k \boldsymbol{r}_X(x) = 0$, and that $W = 2$ when $\sum_k \boldsymbol{r}_X(x) \to \infty$.

Fig. 6 shows how the noninformative prior weight W changes as a function of k and $\sum_k \boldsymbol{r}_X(x)$ with $C_W = 2$. It can be seen that W converges to C_W very rapidly.

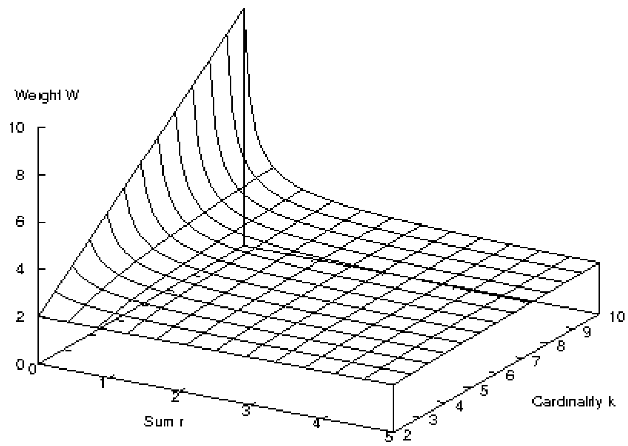

Fig. 6. W computed as a function of k and $\sum_k \boldsymbol{r}_X(x)$

Note that Fig. 6 shows the cardinality dimension k as a continuous parameter, but this assumption does not reflect reality. The domain cardinality k is of course an integer with minimum value $k = 2$. Hence, the plot of Fig. 6 is a simplification using a continuous k to show the effect it has on the noninformative prior weight W.

When determining the noninformative prior weight according to Eq. (12), the vacuous multinomial opinion corresponds to a uniform Dirichlet PDF, as illustrated in Fig. 7.

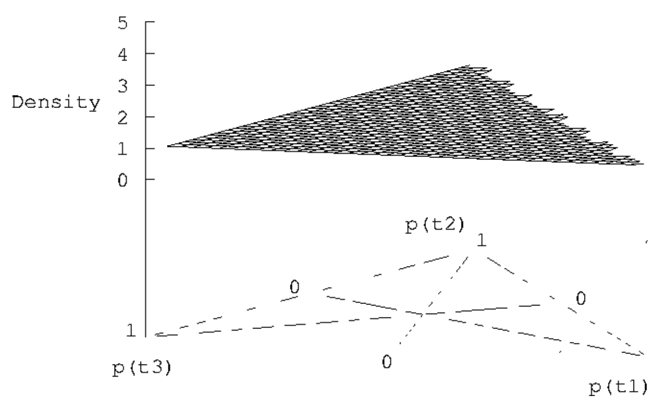

Fig. 7. Uniform vacuous trinomial Dirichlet PDF $\text{Dir}(\boldsymbol{p}_X \mid \{1, 1, 1\})$ with $W = 3$

Thanks to the dynamic nature of W according to Eq. (12), the influence of new evidence on the Dirichlet PDF is sound and intuitive. For example, if each state value has been observed once, then we have,

$$\boldsymbol{r}_X(x_1) = \boldsymbol{r}_X(x_2) = \boldsymbol{r}_X(x_3) = 1. \tag{13}$$

The computation of the prior weight according to Eq. (12) gives $W = {}^{21}/_{10} = 2.1$, and the strength parameters of the Dirichlet PDF are computed according to Eq. (8) as:

$$\alpha_X(x_1) = \alpha_X(x_2) = \alpha_X(x_3) = {}^{17}/_{10} = 1.7. \tag{14}$$

The corresponding Dirichlet PDF is shown in Fig. 8

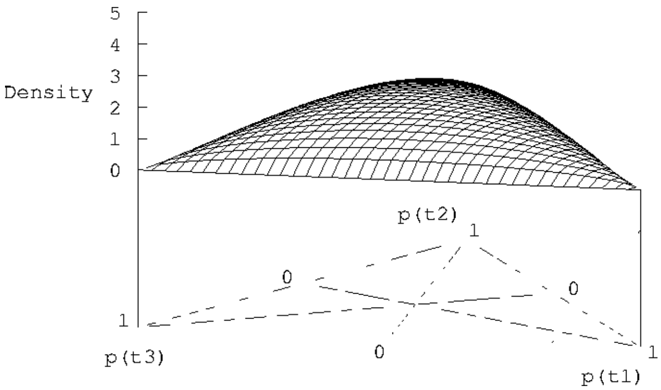

Fig. 8. Trinomial Dirichlet PDF $\mathrm{Dir}(\boldsymbol{p}_X \mid \{1.7, 1.7, 1.7\})$ with $W = 2.1$

VI. DISCUSSION

The challenge of defining a sound noninformative prior has been a topic for debate in the statistics community for many decades. In this regard, Jeffreys prior is often put forward as a candidate. Without going into detail, Jeffreys noninformative prior weight is $W = 1$ for the binomial Beta PDF [5], which gives the concave prior PDF illustrated in Fig. 9. The dashed uniform illustrates what the uniform Beta would have been with $W = 2$.

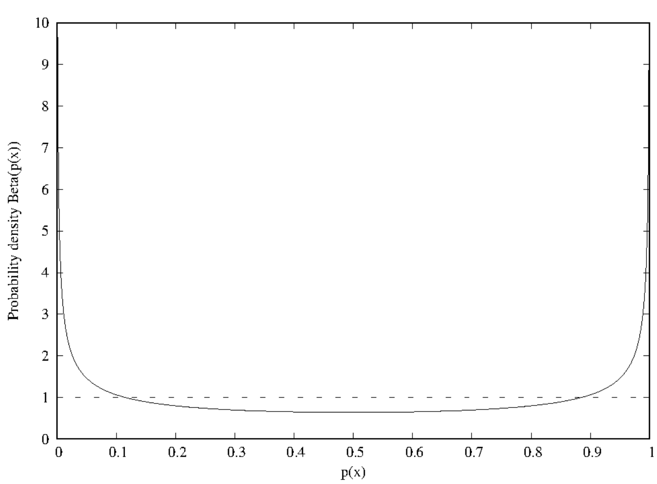

Fig. 9. Beta PDF for Jeffreys prior

As a generalisation of Jeffreys prior for the Beta PDF, Jeffreys noninforative prior weight for Dirichlet PDFs is $W = k/2$ in general, with k as domain cardinality [6], [7].

It can be noted that the Jeffreys noninformative prior does not satisfy the two criteria we proposed, i.e. 1) that the prior PDF shall be uniform, and 2) that the prior weight must not be so heavy that the Dirichlet is not properly influenced by new observations. Assume for example a domain with cardinality $k = 2000$. In that case, Jeffreys prior would dictate the noninformative prior weight to be $W = 1000$, meaning that it would take around 500 observations to have any significant influence on the Dirichlet, which obviously is a totally inadequate model. This observation indicates that Jeffreys prior does not represent a sound prior in general.

The concept of uncertainty is complex due to its multidimensional nature and causes. In case of relatively little evidence, which corresponds to large epistemic uncertainty, the prior probability has a correspondingly large influence on the expected probability. In case of total epistemic uncertainty, the corresponding Beta or Dirichlet PDFs should arguably be uniform. The cardinality of the domain should not change this property. At the same time, the influence of evidence should not depend on the domain cardinality. In order to satisfy both of these requirements, the noninformative prior weight cannot be a general constant. In fact, it must be dynamically determined as a function of the domain cardinality, an appropriate convergence factor, and the amount of collected evidence. The function expressed in Eq. (12) has precisely the required property.

VII. CONCLUSION

In this work, we propose the convergent noninformative prior weight for Dirichlet PDFs. This method of computing the noninformative prior guarantees that the vacuous Beta as well as the vacuous Dirichlet PDF for any domain cardinality is uniform, as would be expected. In addition, for situations where observation evidence is included, the function produces a dynamic noninformative prior which lets the included evidence influence the Dirichlet PDF in a sound and intuitive fashion.

REFERENCES

[1] A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer, Heidelberg, 2016.

[2] A. Jøsang, J.-H. Cho, and F. Chen, "Uncertainty Characteristics of Subjective Opinions," in *Proc. of the 21st International Conference on Information Fusion (FUSION 2018)*, 2018.

[3] U. Hoffrage, S. Lindsey, R. Hertwig, and G. Gigerenzer, "Communicating statistical information," *Science*, vol. 290, no. 5500, pp. 2261–2262, December 2000.

[4] B. de Finetti, "The true subjective probability problem," in *The Concept of Probability in Psychological Experiments*, C.-A. Staël von Holstein, Ed. D. Reidel, Dordrecht, 1974, pp. 15–23.

[5] C. L. Atwood, "Constrained noninformative priors in risk assessment," *Reliability Engineering and System Safety*, vol. 53, pp. 37–46, 1996.

[6] D. Kelly and C. Atwood, "Finding a Minimally Informative Dirichlet Prior Using Least Squares," in *ANS PSA 2011 International Topical Meeting on Probabilistic Safety Assesment and Analysis*, American Nuclear Society, March 2011, Idaho National Laboratory, Technical Report INL/CON-11-20971.

[7] N. O. Siu and D. L. Kelly, "Bayesian parameter estimation in probabilistic risk assessment," *Reliability Engineering and System Safety*, pp. 89–116, 1998.