Exact Likelihood-based Evidential Prediction of an Ordinal Variable

Sébastien Ramel, Frédéric Pichon

Univ. Artois, UR 3926,
Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A),
F-62400 Béthune, France
{sebastien.ramel,frederic.pichon}@univ-artois.fr

Abstract

Given past observations of an ordinal variable, we want to predict a future observation. This paper provides the solution, according to the likelihood-based evidential method for statistical inference and prediction, of this problem, in an algebraic form. This result is obtained after establishing that the prediction of an ordinal variable can be computed, under some conditions on the possibility distribution representing the estimation uncertainty in this method, by integrating the marginals of this distribution.

Keywords— Prediction, Belief function, Ordinal variable, Likelihood.

1 Introduction

Consider observed counts of K ordered categories and that we wish to predict a future observation, as illustrated by Example 1.

Example 1. *The data in Table 1 are January precipitations, in inches, recorded during the period 1895-2004 in Arizona and categorized, as in [5, Example 7], into $K = 6$ ordered categories. The problem is to predict the precipitation category of the following January (2005).*

Table 1: Categorized Arizona January precipitation data, with observed count for each category.

precipitation	category	count
$[0, 0.75)$	'1'	48
$[0.75, 1.25)$	'2'	17
$[1.25, 1.75)$	'3'	19
$[1.75, 2.25)$	'4'	11
$[2.25, 2.75)$	'5'	6
≥ 2.75	'6'	9

One of the main results of this paper is that we provide the solution, according to the method for statistical inference and prediction proposed in [12, 13], of this basic, yet important, problem, in an algebraic form.

This so-called likelihood-based method is framed in the Dempster-Shafer theory of belief functions [3, 20] and satisfies the important requirement of being compatible with Bayesian reasoning. It assumes a parametric model, with parameter $\boldsymbol{\theta}$, for the random variable to be predicted. It is based on three steps. First, estimation uncertainty on $\boldsymbol{\theta}$ is quantified by a possibility distribution; in the example above, this possibility distribution is nothing but the relative likelihood of $\boldsymbol{\theta}$ given the observed counts. Next, the random variable to be predicted is expressed as a function, called a φ-equation, of a pivotal random variable and $\boldsymbol{\theta}$. Finally, the possibility distribution is combined with the pivotal distribution to yield a predictive belief function (PBF) that quantifies the uncertainty about the future observation.

There exist a few other methods to statistical inference and prediction, based on Dempster-Shafer theory. In particular, Dempster's original approach [4], which relies as well on a φ-equation, satisfies also the property of being compatible with Bayesian inference. However, its application poses severe technical difficulties in general [9, 10]. The other main methods, such as Martin and Liu's Inferential Model approach [15], which is an adaptation of Dempster's approach and uses also a φ-equation, or Denœux's confidence level-based approach [5, 10], satisfy frequentist-oriented requirements and are incompatible with Bayesian reasoning, as highlighted in the works [6, 9, 10] to which we refer for in-depth comparisons, as well as connections, between the main Dempster-Shafer theory-based methods to statistical inference and prediction.

The likelihood-based approach has been used for the prediction of quantitative variables (see, *e.g.*, [12, 13]) as well as qualitative variables. The simplest qualitative case, *i.e.*, the binary case, which involves only a parameter $\boldsymbol{\theta} \in [0, 1]$, has been addressed for possibility distributions on $\boldsymbol{\theta}$ obtained in various contexts: observation of a binomial variable [13], calibration of binary classifiers [22, 17], and binary classification through logistic and choquistic regressions [19]. Of particular interest with the binary case is that the PBF can be computed exactly, simply by integrating the possibility distribution on $\boldsymbol{\theta}$, under the condition that this distribution is unimodal and continuous. The PBF even admits an algebraic expression when the possibility distribution is the relative likelihood of $\boldsymbol{\theta}$ given observed data having a binomial distribution with proportion $\boldsymbol{\theta}$ [22]; we have thus an elegant and convenient solution to a basic, yet fundamental, problem, which can readily be used to address all applied problems that involve predicting a binary variable from its past observations, such as the binning-based approach to calibration as done in [22].

The nominal case, which involves a "structure of the second kind" [3], was considered recently in [9] for a possibility distribution on $\boldsymbol{\theta}$ obtained in the context of multinomial logistic regression. The last qualitative case, *i.e.*, the ordinal case, which relies on a "structure of the first kind" [3], is at play in [21] with a possibility distribution on $\boldsymbol{\theta}$ obtained in a context of calibrating multi-class classifiers. Notably, in both the nominal and ordinal cases, contrarily to the binary case, no condition on the possibility distribution on $\boldsymbol{\theta}$ simplifying the expression of the PBF (and even less leading to an algebraic form for it) has been identified and thus for these cases the computation of the PBF always has to be carried out only approximately, through Monte Carlo simulation, as was the case in [9, 21].

In this paper[1], we fill this gap, for the ordinal case. First (Section 3), we bring to light that the PBF can be computed by integrating the possibility distribution on $\boldsymbol{\theta}$, under some conditions on this distribution, which basically extend those of the binary case. Second (Section 4), we show that the PBF even admits an algebraic form, generalizing that of the binary case, when the possibility distribution is the relative likelihood of $\boldsymbol{\theta}$ given observed data having non empty categories and following a multinomial distribution whose underlying categorical distribution has parameter $\boldsymbol{\theta}$. The paper starts (Section 2) by recalling the prediction approach introduced in [12, 13] and, in particular, the results related to the exact computation of the PBF in the binary case. We will assume that the reader has some basic knowledge of the Dempster-Shafer theory of belief functions (a recent reminder can be found in [8]).

2 Likelihood-based evidential prediction

Let Z be a random variable, with probability function $g_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta} \in \Theta$ is the unknown parameter. Consider the problem of predicting a future observation $z \in \mathcal{Z}$ of Z, having observed a realisation $\mathbf{y} \in \mathcal{Y}$ of a random vector $\mathbf{Y}$ with probability function $f_{\boldsymbol{\theta}}$. This section summarizes first (Section 2.1) the necessary elements of the method introduced in [12, 13] for this general problem and then (Section 2.2) recalls results with respect to the specific case of predicting a binary variable Z, using this approach.

2.1 Method

Estimation uncertainty about $\boldsymbol{\theta}$ given the observation $\mathbf{y}$ is quantified by a possibility distribution pl^{Θ} on Θ, interpreted as the contour function of a consonant belief function Bel^{Θ} and defined as the relative likelihood of any value θ of $\boldsymbol{\theta}$ after observing $\mathbf{Y} = \mathbf{y}$, *i.e.*,

$$pl^{\Theta}(\theta) = \mathcal{L}(\theta; \mathbf{y}) / \mathcal{L}(\hat{\theta}; \mathbf{y}),$$

for all $\theta \in \Theta$, where $\mathcal{L}(\theta; \mathbf{y}) = c f_{\boldsymbol{\theta}}(\mathbf{y})$ with $c > 0$ an arbitrary constant and where $\hat{\theta}$ is a maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$.

We recall that Bel^{Θ}, being consonant, is characterized by pl^{Θ} and its focal sets are the sets

$$\Gamma(w) = \{\theta \in \Theta \,|\, pl^{\Theta}(\theta) \geq w\},$$

for all $w \in [0, 1]$. Moreover, let W be a random variable with distribution λ, where λ is the uniform probability measure on $[0, 1]$. Then Bel^{Θ} is induced by the random set $\Gamma(W)$, meaning that we have

$$Bel^{\Theta}(A) = \lambda(w \in [0, 1] \,|\, \Gamma(w) \subseteq A),$$

for all $A \subseteq \Theta$.

Now, Z can always be expressed as a function, called a φ-equation, of θ and a pivotal variable V whose distribution μ does not depend on θ:

$$Z = \varphi(\theta, V).$$

Such φ-equation can be constructed canonically by inverting the cumulative distribution function $F_Z(\cdot; \theta)$ of Z, i.e., we have $\varphi(\theta, V) = F_Z^{-1}(V; \theta)$ with $V \sim \mathcal{U}([0, 1])$ [6, 12]. Hereafter, we assume only φ-equations such that $V \sim \mathcal{U}([0, 1])$.

[1] This paper is an extended and revised version of [18].

Combining the estimation uncertainty about $\boldsymbol{\theta}$ with the φ-equation yields prediction uncertainty about a future realisation of Z, quantified by a PBF noted $Bel^{\mathcal{Z}}$ and induced by the random set $\varphi(\Gamma(W), V)$, *i.e.*, for all $A \subseteq \mathcal{Z}$:

$$Bel^{\mathcal{Z}}(A) = \lambda \otimes \mu(\{(w,v) \in [0,1]^2 \mid \varphi(\Gamma(w), v) \subseteq A\}), \tag{1}$$

where $\lambda \otimes \mu$ is the uniform probability measure on $[0,1]^2$. The focal sets of $Bel^{\mathcal{Z}}$ are the sets $\varphi(\Gamma(w), v)$ for all $(w,v) \in [0,1]^2$.

2.2 Prediction of a binary variable

Let $Z \in \mathcal{Z} = \{1, 2\}$ be a binary random variable with (unknown) parameter $P_1 \in \mathcal{P}_1 = [0,1]$, where $P_1 := \mathbb{P}(Z = 1)$. Z can be expressed as follows:

$$Z = \varphi(P_1, V) = \begin{cases} 1 & \text{if } V \leq P_1, \\ 2 & \text{otherwise.} \end{cases} \tag{2}$$

Assume estimation uncertainty about P_1 quantified by a possibility distribution $pl^{\mathcal{P}_1}$. Let $Bel^{\mathcal{P}_1}$ be the consonant belief function with contour function $pl^{\mathcal{P}_1}$. If $pl^{\mathcal{P}_1}$ is unimodal, continuous and such that $pl^{\mathcal{P}_1}(0) = pl^{\mathcal{P}_1}(1) = 0$ (thus with mode $0 < \hat{P}_1 < 1$), the focal sets $\Gamma(w) = \{P_1 \in \mathcal{P}_1 \mid pl^{\mathcal{P}_1}(P_1) \geq w\}$ of $Bel^{\mathcal{P}_1}$ form closed intervals $\Gamma(w) = [L_1(w), U_1(w)]$, for all $w \in [0,1]$, where $L_1(w)$ and $U_1(w)$ are the two roots of the equation $pl^{\mathcal{P}_1}(P_1) = w$. Hence, $Bel^{\mathcal{P}_1}$ is induced by the random interval $[L_1(W), U_1(W)]$ with $W \sim \mathcal{U}([0,1])$. In this case, the PBF $Bel^{\mathcal{Z}}$ about a future realisation of Z is given by (1) with φ defined by (2). It satisfies[2] [13]

$$\begin{aligned}
Bel^{\mathcal{Z}}(\{1\}) &= \lambda \otimes \mu(\{(w,v) \in [0,1]^2 \mid \varphi(\Gamma(w), v) \subseteq \{1\}) \\
&= \lambda \otimes \mu(\{(w,v) \in [0,1]^2 | L_1(w) \geq v)) \\
&= \int_0^1 \int_0^{L_1(w)} 1 dv dw \\
&= \int_0^1 L_1(w) dw \\
&= \int_{pl^{\mathcal{P}_1}(0)}^{pl^{\mathcal{P}_1}(\hat{P}_1)} L_1(w) dw \\
&= \int_0^{\hat{P}_1} L_1(pl^{\mathcal{P}_1}(t)) pl'^{\mathcal{P}_1}(t) dt \\
&= \int_0^{\hat{P}_1} t\, pl'^{\mathcal{P}_1}(t) dt \\
&= [t\, pl^{\mathcal{P}_1}(t)]_0^{\hat{P}_1} - \int_0^{\hat{P}_1} pl^{\mathcal{P}_1}(t) dt \\
&= \hat{P}_1 - \int_0^{\hat{P}_1} pl^{\mathcal{P}_1}(P_1) dP_1,
\end{aligned} \tag{3}$$

[2]The derivation followed here is slightly different from the original one in [13] and is useful for further results of this paper. Let us also mention that, as shown in [9], Eqs. (3) and (4) can also be derived using the recent refinement of the prediction method recalled in Section 2.1, based on the theory of epistemic random fuzzy sets [7, 8], which views the relative likelihood as a possibility distribution and the PBF as induced by a random fuzzy set. This refinement is more satisfactory conceptually with regard to handling independent samples. However, since we are concerned in this paper only with the case of a single sample, the original version of the method, which involves only the simpler notion of a random set, is sufficient and is therefore followed.

and, similarly, we can obtain

$$Bel^{\mathcal{Z}}(\{2\}) = 1 - \hat{P}_1 - \int_{\hat{P}_1}^{1} pl^{\mathcal{P}_1}(P_1)dP_1. \tag{4}$$

A particular situation where the possibility distribution $pl^{\mathcal{P}_1}$ has the aforementioned properties (*i.e.*, being continuous, unimodal and zero at the bounds of the unit interval), is when it is defined as the relative likelihood of P_1 given an observation y_1, $0 < y_1 < n$, of a random variable Y_1 following a binomial distribution with parameters n known and P_1 unknown, and with y_1 the number of successes out of the n experiments. We have then

$$pl^{\mathcal{P}_1}(P_1) = \frac{P_1^{y_1}(1-P_1)^{n-y_1}}{\hat{P}_1^{y_1}(1-\hat{P}_1)^{n-y_1}}, \quad \forall P_1 \in [0,1], \tag{5}$$

with $\hat{P}_1 = y_1/n$ the MLE of P_1. In this case, it has even been shown [22] that Eqs. (3) and (4) admit algebraic expressions

$$Bel^{\mathcal{Z}}(\{1\}) = \hat{P}_1 - \frac{\underline{B}(\hat{P}_1; y_1+1, n-y_1+1)}{\hat{P}_1^{y_1}(1-\hat{P}_1)^{n-y_1}}, \tag{6}$$

$$Bel^{\mathcal{Z}}(\{2\}) = 1 - \hat{P}_1 - \frac{\overline{B}(\hat{P}_1; y_1+1, n-y_1+1)}{\hat{P}_1^{y_1}(1-\hat{P}_1)^{n-y_1}}, \tag{7}$$

with $\underline{\mathrm{B}}$ and $\overline{B}$, respectively, the lower and upper incomplete beta functions defined as

$$\underline{B}(z; a, b) = \int_0^z t^{a-1}(1-t)^{b-1}dt, \tag{8}$$

$$\overline{B}(z; a, b) = \int_z^1 t^{a-1}(1-t)^{b-1}dt = \underline{B}(1-z; b, a), \tag{9}$$

that admit algebraic forms for integer values of a and b as

$$\underline{B}(z; a, b) = \sum_{j=a}^{a+b-1} \frac{(a-1)!(b-1)!}{j!(a+b-1-j)!} z^j (1-z)^{a+b-1-j}.$$

Example 2. *Suppose $y_1 = 5$ successes out of $n = 15$ experiments. Estimation uncertainty about P_1 is therefore represented by the belief function $Bel^{\mathcal{P}_1}$ with contour function of the form* (5) *illustrated by Figure 1. The predictive uncertainty about a future realisation of Z is quantified by the PBF $Bel^{\mathcal{Z}}$ defined by Eqs.* (6) *and* (7)*. The mass function $m^{\mathcal{Z}}$ associated to $Bel^{\mathcal{Z}}$ verifies $m^{\mathcal{Z}}(\{i\}) = Bel^{\mathcal{Z}}(\{i\})$ for $i = 1, 2$, and $m^{\mathcal{Z}}(\mathcal{Z}) = 1 - \sum_{i=1}^{2} m^{\mathcal{Z}}(\{i\})$. We find $m^{\mathcal{Z}}(\{1\}) \approx .20, m^{\mathcal{Z}}(\{2\}) \approx .51$ and $m^{\mathcal{Z}}(\mathcal{Z}) \approx .29$. The areas associated with these masses are shown on Figure 1.*

3 Prediction of an ordinal variable

Let Z be an ordinal random variable taking its value in $\mathcal{Z} = \{1, \ldots, K\}$, where the K categories are, without lack of generality, denoted by the integers from 1 to K and ordered according to the natural order between integers. The probability measure of Z is characterized by the vector $\mathbf{P} = (P_1, \ldots, P_{K-1})$ of (unknown) parameters $P_j := \sum_{i=1}^{j} \mathbb{P}(Z = i)$, $1 \leq j < K$. We have $\mathbf{P} \in \mathcal{P}$, where

$$\mathcal{P} = \{(P_1, \ldots, P_{K-1}) \in [0,1]^{K-1} \mid P_1 \leq \cdots \leq P_{K-1}\}.$$

Following [3], Z can be expressed as, with $P_K := 1$:

$$Z = \varphi(\mathbf{P}, V) = \begin{cases} j\,, & \text{if } P_{j-1} < V \leq P_j \text{ for some } j \text{ such that } 1 < j \leq K, \\ 1\,, & \text{otherwise } (i.e., V \leq P_1). \end{cases} \tag{10}$$

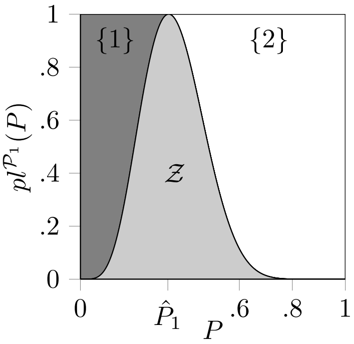


Figure 1: Contour function $pl^{\mathcal{P}_1}$ for $y_1 = 5$ successes out of $n = 15$ experiments (—), induced predictive mass function $m^{\mathcal{Z}}$ (■ $m^{\mathcal{Z}}(\{1\})$, □ $m^{\mathcal{Z}}(\{2\})$, ■ $m^{\mathcal{Z}}(\mathcal{Z})$).

It can easily be verified that $\mathbb{P}(\varphi(\mathbf{P}, V) = j) = \mathbb{P}(Z = j)$, for all $1 \leq j \leq K$.

Consider estimation uncertainty about $\mathbf{P}$ quantified by a possibility distribution $pl^{\mathcal{P}}$. Let $Bel^{\mathcal{P}}$ be the consonant belief function with contour function $pl^{\mathcal{P}}$. Its focal sets are $\Gamma(w) = \{\mathbf{P} \in \mathcal{P} \,|\, pl^{\mathcal{P}}(\mathbf{P}) \geq w\}$, with $w \in [0, 1]$. Under such estimation uncertainty about $\mathbf{P}$, the PBF $Bel^{\mathcal{Z}}$ about a future realisation of Z is given by (1) with φ defined by (10).

In this section, we show that under four conditions on $Bel^{\mathcal{P}}$, $Bel^{\mathcal{Z}}$ admits a simple expression. More precisely, first, we provide in Section 3.1 a simple expression for $Bel^{\mathcal{Z}}(\llbracket i, j \rrbracket)$ for all $1 \leq i \leq j \leq K$, which holds under three of these conditions. Then, in Section 3.2, we add a fourth condition to $Bel^{\mathcal{P}}$, which implies that $Bel^{\mathcal{Z}}$ is characterized by $Bel^{\mathcal{Z}}(\llbracket i, j \rrbracket)$, for all $1 \leq i \leq j \leq K$.

3.1 Expression of $Bel^{\mathcal{Z}}(\llbracket i, j \rrbracket)$ based on the marginals of $pl^{\mathcal{P}}$

In the binary case (Section 2.2), the simple expression given by Eqs. (3) and (4) of the PBF was obtained under the assumptions that estimation uncertainty about P_1 is quantified by a possibility distribution $pl^{\mathcal{P}_1}$, which is continuous, unimodal and zero at the bounds. These assumptions yield the roots $L_1(w)$ and $U_1(w)$, $w \in [0, 1]$, such that $0 < L_1(1) = U_1(1) = \hat{P}_1 < 1$, which play a central role in the derivation of the PBF, in particular through the property that we have, for all $(w, v) \in [0, 1]^2$ and for any $A \subseteq \mathcal{Z} = \{1, 2\}$,

$$
\begin{aligned}
& \varphi(\Gamma(w), v) \quad \subseteq \quad A \\
\Leftrightarrow \min_{z \in A} z \leq \min_{P_1 \in \Gamma(w)} \varphi(P_1, v) \quad & \text{and} \quad \max_{P_1 \in \Gamma(w)} \varphi(P_1, v) \leq \max_{z \in A} z \qquad (11) \\
\Leftrightarrow \min_{z \in A} z \leq \varphi(U_1(w), v) \quad & \text{and} \quad \varphi(L_1(w), v) \leq \max_{z \in A} z \qquad (12)
\end{aligned}
$$

with φ defined by (2). This property is at play when moving from the first line to the second line of Eq. (3), where Eq. (12) reduces to $\varphi(L_1(w), v) \leq 1$, which is in turn equivalent to $L_1(w) \geq v$.

The simple expression of $Bel^{\mathcal{Z}}(\llbracket i, j \rrbracket)$, for all $1 \leq i \leq j \leq K$, that will be presented in this section, holds under extensions of these assumptions of the binary case. They allow us to obtain an extension of the roots $L_1(w)$ and $U_1(w)$ that respects a similar property as above (thanks to extending the equivalence between (11) and (12)) with respect to φ defined by (10), which is instrumental in the derivation of $Bel^{\mathcal{Z}}(\llbracket i, j \rrbracket)$.

We start by extending the assumption of the binary case that $pl^{\mathcal{P}_1}$ is zero at the bounds. This assumption means that the "extreme" distributions $\mathbb{P}_1$ and $\mathbb{P}_2$, such that $\mathbb{P}_1(Z = 1) = 0$ and $\mathbb{P}_2(Z = 2) = 0$, are considered impossible. We recall that, as we have seen in Section 2.2, this assumption is satisfied when a binomial variable has been observed with $0 < y_1 < n$ successes, i.e., at least one success (category 1) and at least one failure (category 2) have been observed.

Let us introduce a condition about $Bel^{\mathcal{P}}$, which extends this assumption to the ordinal case: any $\mathbf{P} \in \mathcal{P}$ for which there exists $1 \le i \le K$ such that $\mathbb{P}(Z = i) = 0$, is impossible (equivalently, plausible $\mathbf{P}$'s are those such that for all $1 \le i \le K$, $\mathbb{P}(Z = i) > 0$). This condition is formally stated by Assumption 1 and, as will be shown in Section 4, is satisfied when $Bel^{\mathcal{P}}$ represents estimation uncertainty about $\mathbf{P}$ induced by past observations of Z, where each category has been observed at least once.

Assumption 1. *$pl^{\mathcal{P}}(\mathbf{P}) > 0$ if and only if $\mathbf{P} \in \mathcal{P}^* := \{(P_1, \ldots, P_{K-1}) \in (0,1)^{K-1} \mid P_1 < \cdots < P_{K-1}\}$.*

The second condition that we need to introduce about $Bel^{\mathcal{P}}$ is an extension of the assumptions of continuity and unimodality of $pl^{\mathcal{P}_1}$ in the binary case. It relies on the notion of marginal contour function, which we recall below.

Let $\mathcal{P}_j = [0,1]$, for all $1 \le j < K$, and $pl^{\mathcal{P}_j}$ be the marginal contour function of $pl^{\mathcal{P}}$ for the j-th component of $\mathbf{P}$, defined by [12, Eq. (76)]:

$$pl^{\mathcal{P}_j}(P_j) = \sup_{\mathbf{P}_{-j}} pl^{\mathcal{P}}(\mathbf{P}), \quad \forall P_j \in [0,1], \tag{13}$$

where $\mathbf{P}_{-j}$ is the subvector of $\mathbf{P}$ with component j removed.

We can now state the second condition about $Bel^{\mathcal{P}}$, which bears on its marginal contour functions:

Assumption 2. *$pl^{\mathcal{P}_j}$, $1 \le j < K$, is continuous and unimodal (with mode denoted $\hat{P}_j$).*

Much as the assumptions of continuity, unimodality and being zero at the bounds in the binary case imply that the mode of $pl^{\mathcal{P}_1}$ lies in the open unit interval, the following Lemma 1 shows that their extensions to the ordinal case, i.e., Assumptions 1 and 2, imply that the modes of the marginals of $Bel^{\mathcal{P}}$ lie in the open unit interval and, in addition, that the mode of the i-th marginal is lower than that of the j-th marginal, for all $1 \le i < j < K$.

Lemma 1. *Assumptions 1 and 2 imply that $0 < \hat{P}_i < 1$ for all $1 \le i < K$, and that $\hat{P}_i < \hat{P}_j$, for all $1 \le i < j < K$.*

Proof. From Assumption 2, mode $\hat{P}_i$ is the unique value $P_i \in [0,1]$ such that $pl^{\mathcal{P}_i}(P_i) = 1$, for all $1 \le i < K$. Furthermore, for all $1 \le i < K$, we have $pl^{\mathcal{P}_i}(0) = 0$ since any $\mathbf{P} \in \mathcal{P}$ whose i-th component is equal to 0 does not belong to $\mathcal{P}^*$ and therefore, from Assumption 1, its possibility $pl^{\mathcal{P}}(\mathbf{P})$ is equal to 0. We can show similarly that, for all $1 \le i < K$, we have $pl^{\mathcal{P}_i}(1) = 0$. Therefore, the mode $\hat{P}_i$ must be in $(0,1)$.

Now, consider all $\mathbf{P} \in \mathcal{P}$ that have $\hat{P}_i$ as i-th component. $pl^{\mathcal{P}_i}(\hat{P}_i) = 1$ implies that there is at least one of those $\mathbf{P}$, which is such that $pl^{\mathcal{P}}(\mathbf{P}) = 1$, hence it belongs to $\mathcal{P}^*$, and thus it has value P_j for its j-th component, $1 \le i < j < K$, such that $P_j > \hat{P}_i$. Moreover, for this $\mathbf{P}$, since $pl^{\mathcal{P}}(\mathbf{P}) = 1$, then $pl^{\mathcal{P}_j}(P_j) = 1$ and thus, from Assumption 2, we have $\hat{P}_j = P_j$. Overall, we have $\hat{P}_j = P_j > \hat{P}_i$. ☐

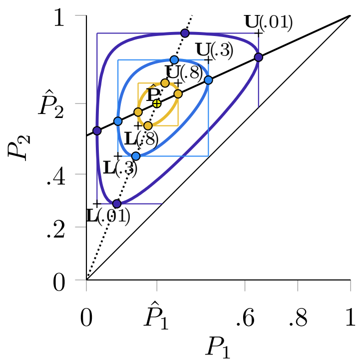
Now that extensions of the assumptions of the binary case have been introduced, let us unveil an extension of the roots $L_1(w)$ and $U_1(w)$ that they allow us to obtain. These "roots" are two particular elements, denoted by $\mathbf{L}(w)$ and $\mathbf{U}(w)$, of $\mathcal{P}$, obtained from the marginal contour functions of $Bel^{\mathcal{P}}$. Their definition rely on the following simple lemma about the marginal contour functions of $Bel^{\mathcal{P}}$, which extends the fact that in the binary case $pl^{\mathcal{P}_1}(P_1) = w$ has roots $L_1(w)$ and $U_1(w)$:

Lemma 2. *Assumptions 1 and 2 imply that, for all $1 \le j < K$, the equation $pl^{\mathcal{P}_j}(P_j) = w$ has only two roots for all $w \in [0,1)$ and only one root ($0<\hat{P}_j<1$) for $w = 1$.*

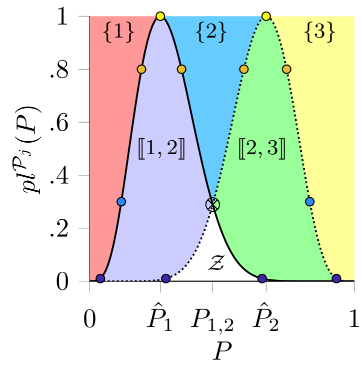
Proof. From Lemma 1, we know that Assumptions 1 and 2 imply that $0 < \hat{P}_j < 1$ for all $1 \le j < K$. The lemma follows then directly from Assumption 2. ☐

For all $1 \le j < K$ and $w \in [0,1)$, we denote the two roots of the equation $pl^{\mathcal{P}_j}(P_j) = w$ by $L_j(w)$ and $U_j(w)$, where $L_j(w) < U_j(w)$. Furthermore, let $L_j(1) := \hat{P}_j$ and $U_j(1) := \hat{P}_j$ and let us introduce, for all $w \in [0,1]$, $\mathbf{L}(w) := (L_1(w), \ldots, L_{K-1}(w))$ and $\mathbf{U}(w) := (U_1(w), \ldots, U_{K-1}(w))$, i.e., $\mathbf{L}(w)$ and $\mathbf{U}(w)$ are the vectors composed of the "lower" and "upper" roots, respectively, of all the marginals.

Example 3 below illustrates the main notions covered so far in this section, i.e., the extensions (Assumptions 1 and 2) to the ordinal case of the assumptions of the binary case, the extensions (Lemmas 1 and 2) of what they imply and the extensions $\mathbf{L}(w)$ and $\mathbf{U}(w)$ of the roots $L_1(w)$ and $U_1(w)$.



(a) Contour plot of $pl^{\mathcal{P}}$. Focal sets $\Gamma(.01)$(—), $\Gamma(.3)$(—) and $\Gamma(.8)$(—). Suprema (○) associated to marginals $pl^{\mathcal{P}_1}$(-○-) and $pl^{\mathcal{P}_2}$(···○···). Roots (+) $\mathbf{L}(w)$ and $\mathbf{U}(w)$ for $w \in \{.01, .3, .8, 1\}$.



(b) Marginals $pl^{\mathcal{P}_1}$(-○-) and $pl^{\mathcal{P}_2}$(···○···). Modes $\hat{P}_1$ and $\hat{P}_2$. Intersection (⊗) at value $P_{1,2}$. Mass function $m^{\mathcal{Z}}$ associated to $Bel^{\mathcal{Z}}$ (see Corollary 1): ■ $m^{\mathcal{Z}}(\{1\})$, ■ $m^{\mathcal{Z}}(\{2\})$, ■ $m^{\mathcal{Z}}(\{3\})$, ■ $m^{\mathcal{Z}}([\![1,2]\!])$, ■ $m^{\mathcal{Z}}([\![2,3]\!])$, □ $m^{\mathcal{Z}}(\mathcal{Z})$.

Figure 2: $Bel^{\mathcal{P}}$ respecting Assumptions 1-4 (focal sets in Fig. 2a, marginals in Fig. 2b), associated $m^{\mathcal{Z}}$ (Fig. 2b).

Example 3. *Consider the case where $Bel^{\mathcal{P}}$ represents estimation uncertainty about $\mathbf{P}$ given n past observations of Z; we denote by y_i the number of times the i-th category has been observed and we assume that $y_i > 0$, for all $1 \le i \le K$. In this case, using the approach recalled in Section 2.1, $Bel^{\mathcal{P}}$ is characterized by the relative likelihood $pl^{\mathcal{P}}$, which has the following expression and satisfies Assumptions 1 and 2*[3]*:*

$$pl^{\mathcal{P}}(\mathbf{P}) = \left(\frac{P_1}{\hat{P}_1}\right)^{y_1} \left(\frac{1-P_{K-1}}{1-\hat{P}_{K-1}}\right)^{y_K} \prod_{j=2}^{K-1} \left(\frac{P_j - P_{j-1}}{\hat{P}_j - \hat{P}_{j-1}}\right)^{y_j}, \quad \forall \mathbf{P} \in \mathcal{P}. \tag{14}$$

An example of such $Bel^{\mathcal{P}}$ for a case where $K = 3$, $n = 15$, $y_1 = 4$, $y_2 = 6$ and $y_3 = 5$, is provided in Figure 2: Figure 2a illustrates Assumption 1 by showing, for $w \in \{.01, .3, .8\}$, the focal sets $\Gamma(w)$ of $Bel^{\mathcal{P}}$; Figure 2b shows the marginals of its contour function $pl^{\mathcal{P}}$ and their modes at play in Assumption 2, Lemma 1 and Lemma 2. $\mathbf{U}(w)$ and $\mathbf{L}(w)$ are illustrated by Figure 2a for $w \in \{.01, .3, .8, 1\}$.

Contrarily to the binary case, in the ordinal case, there are several (marginal) contour functions involved. In order to obtain our simple expression for $Bel^{\mathcal{Z}}([\![i,j]\!])$, for all $1 \le i \le j \le K$, we need that these marginal contour functions satisfy a particular condition with respect to one another, specifically each pair of marginals should intersect only once in the open unit interval. This condition is formally stated by Assumption 3 and illustrated[4] in Figure 2b where the two marginals of Example 3 intersect only at the value $P_{1,2}$.

Assumption 3. *For all $1 \le i < j < K$, there is a value $P_{i,j} \in (0,1)$ such that $pl^{\mathcal{P}_i}(P_{i,j}) = pl^{\mathcal{P}_j}(P_{i,j})$ and, for all $P \in (0,1), P \neq P_{i,j}$, $pl^{\mathcal{P}_i}(P) \neq pl^{\mathcal{P}_j}(P)$.*

We are now ready to provide our first main result, which is that under the preceding conditions, $Bel^{\mathcal{Z}}([\![i,j]\!])$, for all $1 \le i \le j \le K$, can be computed by integrating the marginal contour functions (which is a generalization of Eqs. (3) and (4)):

[3] See Section 4.1 for the complete derivation of this expression and Section 4.2 for the fact that it satisfies these assumptions.

[4] It is shown in Section 4.2 that the marginals of $pl^{\mathcal{P}}$ defined by (14) satisfy Assumption 3.

Theorem 1. *Under Assumptions 1-3, $Bel^{\mathcal{Z}}([\![i,j]\!])$ is equal to*

$$\begin{cases} 1\,, & \text{if } i=1, j=K,\\ \hat{P}_j - \int_0^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP\,, & \text{if } i=1, j<K,\\ 1-\hat{P}_{i-1} - \int_{\hat{P}_{i-1}}^{1} pl^{\mathcal{P}_{i-1}}(P)dP\,, & \text{if } 1<i, j=K,\\ \hat{P}_j - \hat{P}_{i-1} - \int_{\hat{P}_{i-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{P_{i-1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP\,, & \text{if } 1<i, j<K. \end{cases} \tag{15}$$

Proof. From Eq. (1), we obtain, for all $1 \le i \le j \le K$,

$$\begin{aligned} Bel^{\mathcal{Z}}([\![i,j]\!]) &= \lambda \otimes \mu(\{(w,v) \in [0,1]^2 \,|\, \varphi(\Gamma(w),v) \subseteq [\![i,j]\!]\}) \\ &= \lambda \otimes \mu(\{(w,v) \in [0,1]^2 | i \le \ell(w,v), u(w,v) \le j\}), \end{aligned} \tag{16}$$

with, for all $(w,v) \in [0,1]^2$,

$$\ell(w,v) := \min_{\mathbf{P}\in\Gamma(w)} \varphi(\mathbf{P},v), \tag{17}$$

$$u(w,v) := \max_{\mathbf{P}\in\Gamma(w)} \varphi(\mathbf{P},v). \tag{18}$$

The first part of this proof consists in showing that, for all $(w,v) \in [0,1]^2$, the "bounds", i.e., the minimum $\ell(w,v)$ and the maximum $u(w,v)$, of the focal set $\varphi(\Gamma(w),v)$ of the PBF $Bel^{\mathcal{Z}}$, are attained for $\mathbf{U}(w)$ and $\mathbf{L}(w)$, i.e., we have

$$\ell(w,v) = \varphi(\mathbf{U}(w),v), \tag{19}$$

$$u(w,v) = \varphi(\mathbf{L}(w),v). \tag{20}$$

Equalities (19) and (20) are shown in Appendix A. We then have the equivalence

$$\begin{aligned} & i \le \ell(w,v) \quad \text{and} \quad u(w,v) \le j \\ \Leftrightarrow\; & i \le \varphi(\mathbf{U}(w),v) \quad \text{and} \quad \varphi(\mathbf{L}(w),v) \le j, \end{aligned} \tag{21}$$

which is an extension of the equivalence between (11) and (12) in the binary case. Hence, Eq. (16) can be rewritten, for all $1 \le i \le j \le K$,

$$Bel^{\mathcal{Z}}([\![i,j]\!]) = \lambda \otimes \mu(\{(w,v) \in [0,1]^2 | i \le \varphi(\mathbf{U}(w),v), \varphi(\mathbf{L}(w),v) \le j\}). \tag{22}$$

As in the binary case, where Eq. (12) reduces for all $A \subseteq \mathcal{Z} = \{1,2\}$ to simple expressions based on the roots $L_1(w)$ and $U_1(w)$, e.g., for $A=\{1\}$ it reduces to $L_1(w) \ge v$, which in turn leads to being able to compute the PBF by integrating the contour function $pl^{\mathcal{P}_1}$, it happens that Eq. (21) reduces for all $1 \le i \le j \le K$ to simple expressions based on the roots $L_j(w)$ and $U_{i-1}(w)$, which in turn leads to being able to compute $Bel^{\mathcal{Z}}([\![i,j]\!])$ (defined by (22)) by integrating the marginal contour functions $pl^{\mathcal{P}_j}$ and $pl^{\mathcal{P}_{i-1}}$ according to Eq. (15); this is shown in Appendix B. □

We remark that thanks to Theorem 1, we have simple expressions for the cumulative belief and plausibility functions about Z, defined respectively as $cbel^{\mathcal{Z}}(j) := Bel^{\mathcal{Z}}(([\![1,j]\!])$ and $cpl^{\mathcal{Z}}(j) := Pl^{\mathcal{Z}}(([\![1,j]\!]) = 1 - Bel^{\mathcal{Z}}([\![j+1,K]\!])$, for all $1 \le j \le K$, which are quantities that are often used when the variable of interest is ordered; for instance, in Example 1 (which as will be seen in Section 4 respects Assumptions 1-3), one may be interested in the degrees of belief and of plausibility that the precipitation will be, *e.g.*, below 1.25 inches, which amounts to the quantities $cbel^{\mathcal{Z}}(2) = Bel^{\mathcal{Z}}(([\![1,2]\!])$ and $cpl^{\mathcal{Z}}(2) = Pl^{\mathcal{Z}}(([\![1,2]\!])$, respectively.

3.2 Simple expression of $Bel^{\mathcal{Z}}$

In the binary case, the assumptions on $pl^{\mathcal{P}_1}$ (continuous, unimodal, zero at the bounds) imply the focal sets of $Bel^{\mathcal{P}_1}$ to be, for all $w \in [0,1]$, closed intervals $[L_1(w), U_1(w)]$ such that $\hat{P}_1 \in [L_1(w), U_1(w)]$ with $0 < \hat{P}_1 < 1$. The considered extensions (Assumptions 1 and 2) to the ordinal case of these assumptions lead clearly to the superlevel sets of the marginal contour functions of $Bel^{\mathcal{P}}$ to exhibit a similar property; they are closed intervals $[L_j(w), U_j(w)]$ such that $\hat{P}_j \in [L_j(w), U_j(w)]$ with $0 < \hat{P}_j < 1$. Moreover, they also imply that the focal sets $\Gamma(w)$ of $Bel^{\mathcal{P}}$ include, for all $w \in [0,1]$, the element $\hat{\mathbf{P}} = (\hat{P}_1, \ldots, \hat{P}_{K-1})$ of $\mathcal{P}^*$, as shown by Lemma 3 below, hence extending the property in the binary case that $0 < \hat{P}_1 < 1$ belongs to all focal sets of $Bel^{\mathcal{P}_1}$.

Lemma 3. *Assumptions 1 and 2 imply that* $\hat{\mathbf{P}} \in \mathcal{P}^*$ *and, for all* $w \in [0,1]$, $\hat{\mathbf{P}} \in \Gamma(w)$ *with* $\hat{\mathbf{P}} = (\hat{P}_1, \ldots, \hat{P}_{K-1})$.

Proof. From Lemma 1, we obtain $\hat{\mathbf{P}} \in \mathcal{P}^*$. From Assumption 2, $\hat{P}_i$ is the unique value such that $pl^{\mathcal{P}_i}(\hat{P}_i) = 1$, for all $1 \leq i < K$. Moreover, $pl^{\mathcal{P}_1}(\hat{P}_1) = 1$ implies that there is at least one $\mathbf{P} \in \mathcal{P}$ that has $\hat{P}_1$ as first component and such that $pl^{\mathcal{P}}(\mathbf{P}) = 1$, hence it belongs to $\mathcal{P}^*$ from Assumption 1, and thus it has value P_2 for its second component, such that $P_2 > \hat{P}_1$. Moreover, for this $\mathbf{P}$, since $pl^{\mathcal{P}}(\mathbf{P}) = 1$, then $pl^{\mathcal{P}_2}(P_2) = 1$ and thus, from Assumption 2, we have $P_2 = \hat{P}_2$. More generally, by induction, we obtain that this $\mathbf{P}$, which is such that $pl^{\mathcal{P}}(\mathbf{P}) = 1$, satisfies $\mathbf{P} = \hat{\mathbf{P}}$. Hence, for all $w \in [0,1]$, $\hat{\mathbf{P}} \in \Gamma(w)$. □

However, it does not seem that Assumptions 1 and 2 (even complemented with Assumption 3) are sufficient to imply that the focal sets $\Gamma(w)$ of $Bel^{\mathcal{P}}$ satisfy a simple corresponding property to that, in the binary case, of the focal sets of $Bel^{\mathcal{P}_1}$ being intervals. This leads us to consider, in this section, an additional condition about $Bel^{\mathcal{P}}$.

An usual extension to $\mathbb{R}^d$, $d > 1$, of the notion of an interval of $\mathbb{R}$ is that of a convex set. Another, less common but more general[5], extension is that of a star-convex set [2]: a set $S \subseteq \mathbb{R}^d$ is said star-convex if there exists an $x_0 \in S$ (called star center) such that the line segment from x_0 to any point x in S is contained in S. The fourth and last condition that we introduce about $Bel^{\mathcal{P}}$ is that its focal sets be star-convex, as stated by Assumption 4.

Assumption 4. *For all* $w \in (0,1]$, $\Gamma(w)$ *is star convex.*

Example 4. $Bel^{\mathcal{P}}$ *in Example 3 satisfies Assumption 4, as its focal sets respect the stronger condition of being convex*[6].

Assumption 4, together with Assumption 1, imply that the focal sets of $Bel^{\mathcal{Z}}$ are intervals, as shown by the following lemma.

Lemma 4. *Under Assumptions 1 and 4, we have, for all* $(w,v) \in [0,1]^2$,

$$\varphi(\Gamma(w), v) = [\![\ell(w,v), u(w,v)]\!], \tag{23}$$

with $\ell(w,v)$ *and* $u(w,v)$ *defined by* (17) *and* (18)*, respectively.*

Proof. See Appendix C. □

As shown in [5], when the focal sets of some belief function $Bel_0^{\mathcal{Z}}$ on $\mathcal{Z}$ are intervals, then this belief function is characterized by $Bel_0^{\mathcal{Z}}([\![i,j]\!])$, for all $1 \leq i \leq j \leq K$. This latter fact, together with Theorem 1 and Lemma 4, yield directly our second main result (Theorem 2), which is that under the four preceding Assumptions 1-4, we have a simple expression for $Bel^{\mathcal{Z}}$.

Theorem 2. *Under Assumptions 1-4,* $Bel^{\mathcal{Z}}$ *is characterized by* $Bel^{\mathcal{Z}}([\![i,j]\!])$ *for all* $1 \leq i \leq j \leq K$*, with* $Bel^{\mathcal{Z}}([\![i,j]\!])$ *defined by* (15).

[5] If a set is convex, then it is star-convex, but the converse is not true.

[6] This is shown in Section 4.2.

For completeness, Corollary 1 provides the expression of the mass function $m^{\mathcal{Z}}$ associated to $Bel^{\mathcal{Z}}$ of Theorem 2; this corollary shows that $m^{\mathcal{Z}}$ can also be expressed in terms of integration of the marginal contour functions. This expression for $m^{\mathcal{Z}}$ is illustrated graphically by Figure 2b: similarly as in the binary case (see Figure 1 of Example 2), we can see that the masses allocated to the intervals $[\![i, j]\!]$, $1 \leq i \leq j \leq K$, are in one-to-one correspondence with the areas of the elements of the partition of $[0, 1]^2$ resulting from the marginal contour function overlay.

Corollary 1. *The mass function $m^{\mathcal{Z}}$ associated to $Bel^{\mathcal{Z}}$ is characterized by $m^{\mathcal{Z}}([\![i, j]\!])$ for all $1 \leq i \leq j \leq K$, with $m^{\mathcal{Z}}([\![i, j]\!])$ equal to*

$$\begin{cases} A_{1,K}\,, & \text{if } i = 1, j = K, \\ A_{1,j} - \int_{P_{0,j}}^{P_{1,j}} pl^{\mathcal{P}_j}(P)dP & \text{if } i = 1, j < K, \\ A_{i,K} - \int_{P_{i-1,K-1}}^{P_{i-1,K}} pl^{\mathcal{P}_{i-1}}(P)dP, & \text{if } 1 < i, j = K, \\ A_{i,j} - \int_{P_{i-1,j-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{P_{i-1,j}}^{P_{i,j}} pl^{\mathcal{P}_j}(P)dP & \text{if } 1 < i, j < K, \end{cases} \tag{24}$$

with

$$A_{i,j} := \begin{cases} P_{j,j} - P_{j-1,j-1} & \text{if } i = j, \\ \int_{P_{i-1,j-1}}^{P_{i,j-1}} pl^{\mathcal{P}_{j-1}}(P)dP + \int_{P_{i,j-1}}^{P_{i,j}} pl^{\mathcal{P}_i}(P)dP & \text{otherwise}, \end{cases} \tag{25}$$

$P_{j,j} := \hat{P}_j, P_{0,j} := 0$, $P_{j,K} := 1$, for all $1 \leq j < K$ and $P_{0,0} := 0$, $P_{K,K} := 1$.

Proof. See Appendix D. □

4 Prediction given past observations

Consider an ordinal variable Z with K categories and with probability measure characterized by some (unknown) $\mathbf{P} \in \mathcal{P}$. In this section, we consider the problem where we want to predict a future observation of this variable and where estimation uncertainty about $\mathbf{P}$ comes from n past observations of the variable, with y_i denoting the number of times the i-th category has been observed and y_i assumed to be such that $y_i > 0$, $1 \leq i \leq K$. An illustration of this problem is provided by Example 1. In Section 4.1, the definition of the relative likelihood function in this case is recalled, and some useful notations are also introduced. In Section 4.2, it is shown that this function satisfies the assumptions considered in Section 3 and therefore that Theorem 2 can be evoked to compute the predictive uncertainty about the future observation. Then, in Section 4.3, it is shown that this predictive uncertainty can even be computed algebraically. Finally, in Section 4.4, this solution to the problem of predicting a future observation of an ordinal variable given past observations of the variable, is compared to two Dempster-Shafer theory-based alternative solutions.

4.1 Relative likelihood for multinomial data

Formally, the past observations amounts to having observed a realisation $\mathbf{y} = (y_1, \ldots, y_K)$, with non empty[7] categories (*i.e.*, $y_i > 0$, $1 \leq i \leq K$), of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_K)$, where $Y_j = \sum_{i=1}^{n} I(X_i = j)$, with $I(\cdot)$ the indicator function and $X_1, \ldots, X_n$ an iid sample of parent random

[7]The case of an empty category may be addressed by fusing it with a neighbouring category, *i.e.*, by considering a coarsening of $\mathcal{Z}$, or, similarly to [11], by removing it, *i.e.*, by converting the problem to a problem with $K-1$ categories. Alternatively, one may employ a kind of Laplace correction [14], where an imaginary observation is added to each category, *i.e.*, by replacing $\mathbf{y}$ by $\mathbf{y}' = (y'_1, \ldots, y'_K)$ with $y'_i = y_i + 1$, for all $1 \leq i \leq K$. A detailed study of this situation is beyond the scope of this paper and is left for further research.

variable X following a categorical distribution with K categories such that $\mathbb{P}(X \leq j) = P_j$, $1 \leq j < K$, with $(P_1, \ldots, P_{K-1}) \in \mathcal{P}$ unknown and n known. $\mathbf{Y}$ follows a multinomial distribution with parameters n and $\mathbf{p} \in \Pi$ where

$$\Pi := \{(p_1, \ldots, p_K) \in [0,1]^{K-1} \mid \sum_{i=1}^{K} p_i = 1\}$$

with

$$p_j := P_j - P_{j-1}, \quad \forall 1 \leq j \leq K, \tag{26}$$

where $P_0 := 0$ and $P_K := 1$.

The relative likelihood of any $\mathbf{p} \in \Pi$ given $\mathbf{y}$ is $pl^{\Pi}(\mathbf{p})$ with pl^{Π} the possibility distribution such that

$$pl^{\Pi}(\mathbf{p}) = \frac{\mathcal{L}(\mathbf{p}; \mathbf{y})}{\mathcal{L}(\hat{\mathbf{p}}; \mathbf{y})} = \prod_{i=1}^{K} \left(\frac{p_i}{\hat{p}_i}\right)^{y_i}, \tag{27}$$

with $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_K)$ the MLE of $\mathbf{p}$ where $\hat{p}_i = y_i/n$, $1 \leq i \leq K$. Since $y_i > 0$, $1 \leq i \leq K$, we clearly have $pl^{\Pi}(\mathbf{p}) > 0$ if and only if $\mathbf{p} \in \Pi^*$ where

$$\Pi^* := \{(p_1, \ldots, p_K) \in (0,1)^{K-1} \mid \sum_{i=1}^{K} p_i = 1\}.$$

It will be convenient to express the one-to-one correspondence between elements of $\mathcal{P}$ and of Π ($\mathbf{p} \in \Pi$ is obtained from $\mathbf{P} \in \mathcal{P}$ using (26), and $\mathbf{P}$ is obtained from $\mathbf{p}$ using $P_j = \sum_{i=1}^{j} p_i$, $1 \leq j < K$) in a matrix form. Specifically, let $T := (\Delta, \mathbf{0})$ be the $(K-1 \times K)$-matrix formed by the concatenation of the $(K-1 \times K-1)$ lower triangular matrix of ones Δ and the (column) vector $\mathbf{0}$ of $K-1$ zeros. Then, the element $\mathbf{P} \in \mathcal{P}$ in correspondence with $\mathbf{p} \in \Pi$ is $\mathbf{P} = T\mathbf{p}$.

Given this one-to-one correspondence, it is clear that the relative likelihood $pl^{\mathcal{P}}(\mathbf{P})$ of any $\mathbf{P} \in \mathcal{P}$ given $\mathbf{y}$ is equal to the relative likelihood $pl^{\Pi}(\mathbf{p})$ of $\mathbf{p}$ given $\mathbf{y}$, for $\mathbf{p}$ the unique vector in Π such that $\mathbf{P} = T\mathbf{p}$, i.e., we have

$$pl^{\mathcal{P}}(\mathbf{P}) = \frac{\mathcal{L}(\mathbf{p}; \mathbf{y})}{\mathcal{L}(\hat{\mathbf{p}}; \mathbf{y})} \tag{28}$$

for $\mathbf{P} = T\mathbf{p}$. It is also clear that the MLE $\hat{\mathbf{P}} = (\hat{P}_1, \ldots, \hat{P}_{K-1})$ of $\mathbf{P}$ satisfies $\hat{\mathbf{P}} = T\hat{\mathbf{p}}$ and we have thus that the MLE $\hat{P}_j$ of P_j is

$$\hat{P}_j = \sum_{i=1}^{j} \hat{p}_i = \sum_{i=1}^{j} \frac{y_i}{n}. \tag{29}$$

Using Eqs. (26), (27), (28) and (29), we obtain that $pl^{\mathcal{P}}$ is defined by Eq. (14).

4.2 Satisfied assumptions

We consider in this section the relative likelihood function $pl^{\mathcal{P}}$ defined by Eq. (14) when $y_i > 0$, $1 \leq i \leq K$. We show that it satisfies the four assumptions introduced in Section 3, starting with Assumption 1 (Lemma 5), proceeding then to Assumption 4 (Lemma 6, which shows that its superlevel sets are convex thereby satisfying Assumption 4) and Assumption 2 (Lemma 7), and concluding with Assumption 3 (Lemma 8).

Lemma 5. *$pl^{\mathcal{P}}$ satisfies Assumption 1.*

Proof. As remarked in Section 4.1, given $y_i > 0$, $1 \leq i \leq K$, we have $pl^{\Pi}(\mathbf{p}) > 0$ if and only if $\mathbf{p} \in \Pi^*$, i.e., $\mathbf{p} \in \Pi^* \Leftrightarrow pl^{\Pi}(\mathbf{p}) > 0$.

We have, for $\mathbf{P} = T\mathbf{p}$, $pl^{\Pi}(\mathbf{p}) > 0 \Leftrightarrow pl^{\mathcal{P}}(\mathbf{P}) > 0$. Hence, $\mathbf{p} \in \Pi^* \Leftrightarrow pl^{\mathcal{P}}(\mathbf{P}) > 0$, for $\mathbf{P} = T\mathbf{p}$.

Furthermore, it is easy to show that $\mathbf{p} \in \Pi^* \Leftrightarrow \mathbf{P} \in \mathcal{P}^*$, for $\mathbf{P} = T\mathbf{p}$. Hence, we have $pl^{\mathcal{P}}(\mathbf{P}) > 0 \Leftrightarrow \mathbf{P} \in \mathcal{P}^*$. □

Lemma 6. $\Gamma(w)$*, for all* $w \in (0, 1]$*, is convex.*

Proof. See Appendix E. $\square$

Lemma 7. *The marginal contour functions* $pl^{\mathcal{P}_j}$*,* $1 \leq j < K$*, of* $pl^{\mathcal{P}}$ *satisfy Assumption 2.*

Proof. See Appendix F. $\square$

Lemma 8. *The marginal contour functions* $pl^{\mathcal{P}_j}$*,* $1 \leq j < K$*, of* $pl^{\mathcal{P}}$ *satisfy Assumption 3.*

Proof. See Appendix G. $\square$

In sum, in this section, we have shown the following proposition:

Proposition 1. $pl^{\mathcal{P}}$ *defined by* (14) *satisfies Assumptions 1-4.*

From Proposition 1 and Theorem 2, we have that, given past observations $\mathbf{y} = (y_1, \dots, y_K)$ of an ordinal variable, our predictive uncertainty about a future observation of this variable is quantified by the belief function $Bel^{\mathcal{Z}}$ characterized by $Bel^{\mathcal{Z}}([\![i, j]\!])$ for all $1 \leq i \leq j \leq K$, with $Bel^{\mathcal{Z}}([\![i, j]\!])$ defined by (15). Next section will show that, actually, in this particular setting, $Bel^{\mathcal{Z}}$ even has a closed-form.

4.3 Algebraic formula for $Bel^{\mathcal{Z}}$

The last main result of this paper is that, when estimation uncertainty is characterized by (14), then Eq. (15) admits an algebraic expression:

Theorem 3. *Under estimation uncertainty given by Eq.* (14)*,* $Bel^{\mathcal{Z}}$ *is characterized by* $Bel^{\mathcal{Z}}([\![i, j]\!])$ *for all* $1 \leq i \leq j \leq K$*, with* $Bel^{\mathcal{Z}}([\![i, j]\!]) =$

$$\begin{cases} 1\,, & \text{if } i = 1, j = K, \\ \hat{P}_j - \frac{\underline{B}(\hat{P}_j; n_j+1, n-n_j+1)}{c_j}\,, & \text{if } i = 1, j < K, \\ 1 - \hat{P}_{i-1} - \frac{\overline{B}(\hat{P}_{i-1}; n_{i-1}+1, n-n_{i-1}+1)}{c_{i-1}}\,, & \text{if } 1 < i, j = K, \\ \hat{P}_j - \hat{P}_{i-1} - \frac{\underline{B}(P_{i-1,j}; n_{i-1}+1, n-n_{i-1}+1) - \underline{B}(\hat{P}_{i-1}; n_{i-1}+1, n-n_{i-1}+1)}{c_{i-1}} & \text{if } 1 < i, j < K, \\ \quad - \frac{\underline{B}(\hat{P}_j; n_j+1, n-n_j+1) - \underline{B}(P_{i-1,j}; n_j+1, n-n_j+1)}{c_j}\,, & \end{cases} \tag{30}$$

where $n_j := \sum_{i=1}^{j} y_i$*,* $c_j := \hat{P}_j^{n_j}(1 - \hat{P}_j)^{n-n_j}$*, for all* $1 \leq j < K$*, and*

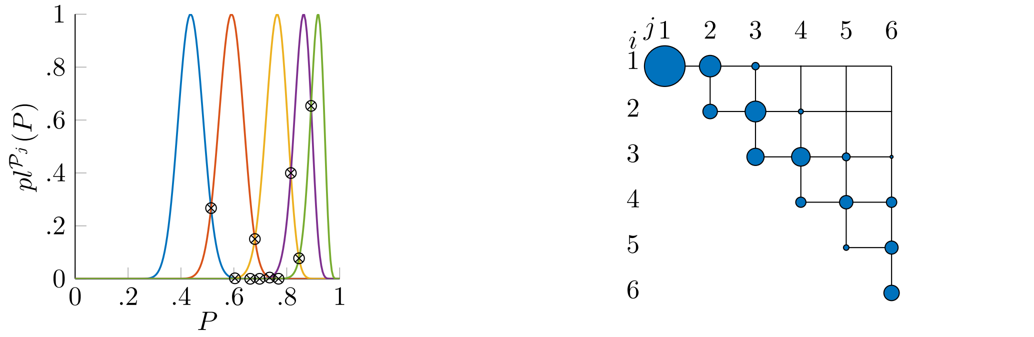
$$P_{i,j} = \left(\left(\left(\frac{\hat{P}_j}{1 - \hat{P}_j} \right)^{\hat{P}_j} \left(\frac{1 - \hat{P}_i}{\hat{P}_i} \right)^{\hat{P}_i} \left(\frac{1 - \hat{P}_j}{1 - \hat{P}_i} \right) \right)^{-1/(\hat{P}_j - \hat{P}_i)} + 1 \right)^{-1}, \quad \forall 1 \leq i < j < K. \tag{31}$$

Proof. See Appendix H. $\square$

Theorem 3 is illustrated by Examples 5 and 6.

Example 5. *In Example 3, we considered estimation uncertainty about* $\mathbf{P}$ *for a case* $K = 3$*, induced by the observed realisation* $\mathbf{y} = (4, 6, 5)$ *of an ordinal variable, and therefore represented by the belief function* $Bel^{\mathcal{P}}$ *with contour function of the form* (14)*. According to Theorem 3, the predictive uncertainty about a future observation of this variable is quantified by the PBF* $Bel^{\mathcal{Z}}$ *defined by* (30)*. The mass function* $m^{\mathcal{Z}}$*, associated to* $Bel^{\mathcal{Z}}$ *and which, we recall, is illustrated graphically by Fig. 2b, is such that* $m^{\mathcal{Z}}(\{1\}) \approx .15$*,* $m^{\mathcal{Z}}(\{2\}) \approx .12$*,* $m^{\mathcal{Z}}(\{3\}) \approx .20$*,* $m^{\mathcal{Z}}([\![1, 2]\!]) \approx .24$*,* $m^{\mathcal{Z}}([\![2, 3]\!]) \approx .25$*,* $m^{\mathcal{Z}}(\mathcal{Z}) \approx .04$*.*

Suppose now that estimation uncertainty is induced by the observed realisation $\mathbf{y} = (266, 400, 334)$*, which amounts to* $n = 1000$ *observations (the empirical proportions associated to this realisation are approximately the same as that of* $\mathbf{y} = (4, 6, 5)$ *where* $n = 15$*). In this case, the mass function* $m^{\mathcal{Z}}$ *is such that* $m^{\mathcal{Z}}(\{1\}) \approx .25$*,* $m^{\mathcal{Z}}(\{2\}) \approx .36$*,* $m^{\mathcal{Z}}(\{3\}) \approx .31$*,* $m^{\mathcal{Z}}([\![1, 2]\!]) \approx .04$*,* $m^{\mathcal{Z}}([\![2, 3]\!] \approx .04$*,* $m^{\mathcal{Z}}(\mathcal{Z}) = 0$*. As can be seen, with more observations, the mass function obtained is more specific (less uncertain).*



(a) Maginal contour functions : $pl^{\mathcal{P}_1}$(—), $pl^{\mathcal{P}_2}$(—), $pl^{\mathcal{P}_3}$(—), $pl^{\mathcal{P}_4}$(—), $pl^{\mathcal{P}_5}$(—). Intersections (⊗) between marginals.

(b) Predictive mass values $m^{\mathcal{Z}}(\llbracket i,j \rrbracket)$ for all $1 \leq i \leq j \leq K$, are proportional to areas of corresponding circles.

Figure 3: Marginal contour functions (Fig. 3a) and predictive mass function (Fig. 3b) for the Arizona January precipitation data.

Example 6. *Example 1 corresponds to $n = 110$ past observations of an ordinal variable with $K = 6$ categories, grouped in the vector of counts $\mathbf{y} = (48, 17, 19, 11, 6, 9)$. The marginal contour functions $pl^{\mathcal{P}_j}$, $1 \leq j < 6$, are illustrated by Fig 3a; we recall (see the proof of Lemma 7 in Appendix F) that $pl^{\mathcal{P}_j}$ is nothing but the relative likelihood of a binomial variable for which we have observed $n_j = \sum_{i=1}^{j} y_i$ successes out of n experiments. According to Lemma 8, each pair of marginals $pl^{\mathcal{P}_i}$ and $pl^{\mathcal{P}_j}$ intersects exactly once within the open unit interval. This intersection happens at the value $P_{i,j}$ given by Eq.* (31)*. These marginals, together with their intersections, create a partition of the unit square $[0,1]^2$. This partition contains $6(6+1)/2 = 21$ elements, which are assigned according to Corollary 1 to the (interval) focal sets of the mass function $m^{\mathcal{Z}}$ associated to the PBF $Bel^{\mathcal{Z}}$ and whose areas are equal to the masses allocated to these focal sets. Mass function $m^{\mathcal{Z}}$ is illustrated in Fig. 3b: the node at the intersection of the i-th row and the j-th column corresponds to the interval focal set $\llbracket i,j \rrbracket$ and the area of the circle at that node is proportional to the mass value allocated to this interval. The degrees of belief and of plausibility that the precipitation will be,* e.g.*, below* 1.25 *inches, amounts to computing the cumulative belief $cbel^{\mathcal{Z}}(2)$ and cumulative plausibility $cpl^{\mathcal{Z}}(2)$, respectively. We find, using Theorem 3, $cbel^{\mathcal{Z}}(2) \approx 0.53$ and $cpl^{\mathcal{Z}}(2) \approx 0.65$.*

4.4 Predictive belief functions at confidence level $1 - \alpha$

As mentioned in the introduction, there exist a few other Dempster-Shafer theory-based approaches to statistical inference and prediction, besides the likelihood-based one. Among them, to our knowledge, only the work [5] has handled the problem of the prediction of an ordinal variable given past observations of the variable. Instead of satisfying the requirement of being compatible with Bayesian reasoning, the solution proposed in [5] satisfies the requirement of being frequency-calibrated in the following sense: the PBF $Bel^{\mathcal{Z}}(\cdot; \mathbf{y})$ should be less committed than the true probability distribution $P_Z(\cdot; \mathbf{P})$, for at least a proportion $1 - \alpha \in (0, 1)$ of the samples $\mathbf{y}$, under repeated sampling [10, 5]. Formally, $Bel^{\mathcal{Z}}$ should satisfy, for any $\mathbf{P} \in \mathcal{P}$:

$$\mathbb{P}_{\mathbf{Y}}\left(Bel^{\mathcal{Z}}(\cdot; \mathbf{Y}) \leq \mathbb{P}_Z(\cdot\,; \mathbf{P})\,; \mathbf{P}\right) \geq 1 - \alpha. \tag{32}$$

A belief function satisfying (32), at least asymptotically (as the sample size n tends to infinity), will be called hereafter a PBF at confidence level $1 - \alpha$ ($1 - \alpha$-level PBF for short). It should be underlined that, as remarked in [10], a major drawback of $1 - \alpha$-level PBFs is that they require the user to specify the value α, e.g., 0.05, this value being arbitrary.

The $1 - \alpha$-level PBF obtained in [5] will be denoted $Bel^{\mathcal{Z}}_G$ due to the fact that it relies on Goodman confidence intervals (we refer the reader to [5] for details). Its focal sets are intervals and thus $Bel^{\mathcal{Z}}_G$ is characterized by $Bel_G^{\mathcal{Z}}(\llbracket i,j \rrbracket)$, $1 \le i \le j \le K$. For a vector of counts $\mathbf{y}$, it is defined as:

$$Bel_G^{\mathcal{Z}}(\llbracket i,j \rrbracket) \quad = \quad \max\left(\sum\nolimits_{k\in\llbracket i,j \rrbracket} p_k^-,\, 1-\sum\nolimits_{k\notin\llbracket i,j \rrbracket} p_k^+\right), \tag{33}$$

with, for all $1 \le k \le K$,

$$p_k^- = \frac{b+2y_k-\sqrt{\Delta_k}}{2(n+b)}, \qquad \text{and} \qquad p_k^+ = \frac{b+2y_k+\sqrt{\Delta_k}}{2(n+b)}, \tag{34}$$

with $b = \chi^2_{1;1-\alpha/K}$ and $\Delta_k = b\,(b+4y_k\,(n-y_k)\,/n)$.

Interestingly, thanks to some technical results obtained earlier in this paper as well as [10, Setion 2.2], which provides an adaptation of the prediction method of Section 2.1 allowing one to compute a PBF at a given confidence level, we can derive another $1-\alpha$-level PBF that admits a simple expression. From [10, Setion 2.2], it is straightforward to obtain that the belief function denoted[8] $Bel_W^{\mathcal{Z}}$ and induced by the random set $\varphi(\Gamma(w_\alpha), V)$, with $V \sim \mathcal{U}([0,1])$ and $\Gamma(w_\alpha)$ the w_α-superlevel set of $pl^{\mathcal{P}}$ defined by (14) where $w_\alpha = \exp(-0.5\chi^2_{K-1;1-\alpha})$, *i.e.*,

$$Bel_W^{\mathcal{Z}}(A) = \mu(\{v \in [0,1] \,|\, \varphi(\Gamma(w_\alpha), v) \subseteq A\}), \quad \forall A \subseteq \mathcal{Z}, \tag{35}$$

with μ the uniform probability measure on $[0,1]$, is a $1-\alpha$-level PBF. The simple expression for this PBF is provided by Theorem 4.

Theorem 4. *$Bel_W^{\mathcal{Z}}$ defined by (35) is characterized by $Bel_W^{\mathcal{Z}}(\llbracket i,j \rrbracket)$ for all $1 \le i \le j \le K$, with*

$$Bel_W^{\mathcal{Z}}(\llbracket i,j \rrbracket) = \begin{cases} 1, & \text{if } i=1, j=K,\\ L_j(w_\alpha), & \text{if } i=1, j<K,\\ 1-U_{i-1}(w_\alpha), & \text{if } 1<i, j=K,\\ \max(L_j(w_\alpha)-U_{i-1}(w_\alpha),0) & \text{if } 1<i, j<K, \end{cases} \tag{36}$$

with, for all $1 \le j < K$, $L_j(w_\alpha)$ and $U_j(w_\alpha)$ the two roots of the equation $pl^{\mathcal{P}_j}(P_j) = w_\alpha$, such that $L_j(w_\alpha) \le U_j(w_\alpha)$, and where $pl^{\mathcal{P}_j}$ is the marginal of $pl^{\mathcal{P}}$ for its j-th component.

Proof. See Appendix I. □

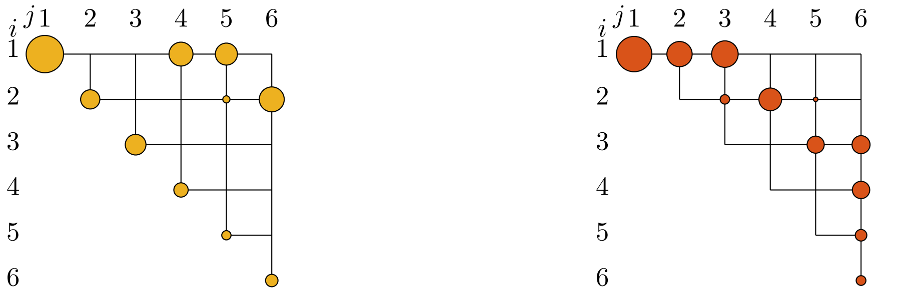
Example 7 illustrates PBFs $Bel_G^{\mathcal{Z}}$ and $Bel_W^{\mathcal{Z}}$, and compares them to the likelihood-based PBF $Bel^{\mathcal{Z}}$ defined by (30).

Example 7. *Figure 4 shows the PBFs $Bel_G^{\mathcal{Z}}$ and $Bel_W^{\mathcal{Z}}$, at confidence level 0.95, for the Arizona January precipitation data of Example 1. Comparing visually Fig. 4b and Fig. 3b, we see that $Bel_W^{\mathcal{Z}}$, which relies on the $w_{0.05}$-cut[9] of the relative likelihood $pl^{\mathcal{P}}$ defined by (14), allocates masses to less specific focal sets than the likelihood-based PBF $Bel^{\mathcal{Z}}$, which relies on the entire relative likelihood. Numerically, we find that $Bel_W^{\mathcal{Z}}(A) \le Bel^{\mathcal{Z}}(A)$ for all $A \subseteq \mathcal{Z}$, i.e., $Bel_W^{\mathcal{Z}}$ is less committed than $Bel^{\mathcal{Z}}$, which is an expected behaviour [6, Section 4.2]. Even though $Bel^{\mathcal{Z}}$ is not more committed than $Bel_G^{\mathcal{Z}}$, i.e., we do not have $Bel_G^{\mathcal{Z}}(A) \le Bel^{\mathcal{Z}}(A)$ for all $A \subseteq \mathcal{Z}$, we note that $Bel_G^{\mathcal{Z}}$ also tends to allocate masses to less specific focal sets than $Bel^{\mathcal{Z}}$, as can be seen by comparing Fig. 4a and Fig. 3b.*

To further compare the three PBFs $Bel^{\mathcal{Z}}$, $Bel_G^{\mathcal{Z}}$ and $Bel_W^{\mathcal{Z}}$, in the remainder of this section we check experimentally, similarly as done in [7], the frequentist property (32) of the $1-\alpha$-level PBFs $Bel_G^{\mathcal{Z}}$ and $Bel_W^{\mathcal{Z}}$, and we also examine this property with respect to the likelihood-based PBF $Bel^{\mathcal{Z}}$, which, we recall, was not derived to satisfy this requirement.

In our experiment, we considered an ordinal variable with $K = 6$ categories and assumed that $\mathbf{Y}$ follows a multinomial distribution with parameters $n = 110$ and $\mathbf{p}_0 \in \Pi$ such that $\mathbf{p}_0 = (p_{01}, \ldots, p_{06}) = (0.45, 0.15, 0.17, 0.10, 0.05, 0.08)$ (this arbitrary multinomial distribution was chosen to be close to the

[8]This notation is due to the fact that $Bel_W^{\mathcal{Z}}$ relies on Wilks' theorem, see [10, Setion 2.2] for details.
[9]We have $w_{0.05} = 3.9\text{e}{-03}$.



(a) $m_G^{\mathcal{Z}}$. (b) $m_W^{\mathcal{Z}}$.

Figure 4: Predictive mass values $m^{\mathcal{Z}}(\llbracket i,j \rrbracket)$ for all $1 \leq i \leq j \leq K$, for the Arizona January precipitation data, according to the 0.95-level PBFs $Bel_G^{\mathcal{Z}}$ (Fig. 4a) and $Bel_W^{\mathcal{Z}}$ (Fig. 4b).

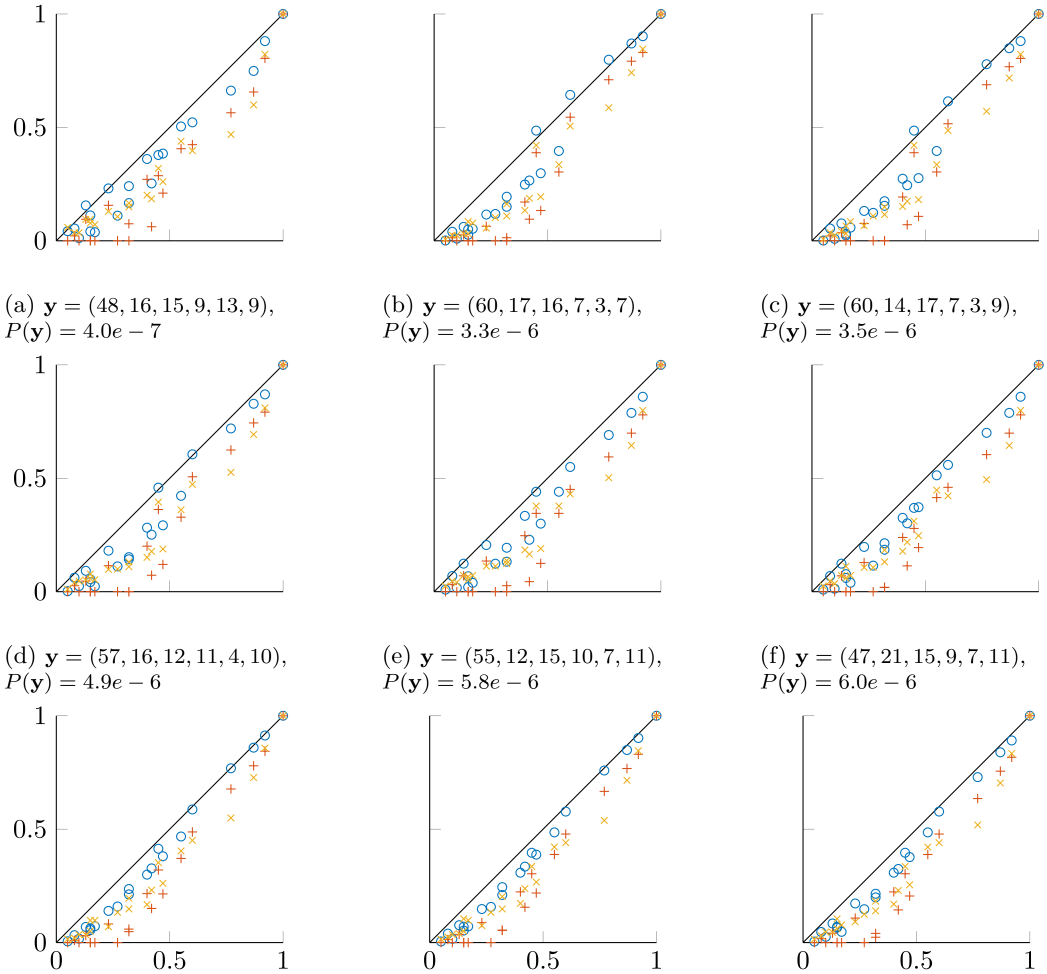
empirical proportions of the data of Example 1). Equivalently, this multinomial distribution can be defined with parameters $\mathbf{P}_0 = (P_{01}, \ldots, P_{0K-1})$, such that $P_{0k} = \sum_{j=1}^{k} p_{0j}$, for all $1 \leq j < K$, which yields $\mathbf{P}_0 = (0.45, 0.60, 0.77, 0.87, 0.92)$. We drew nine realisations $\mathbf{y}$ of $\mathbf{Y}$. For each realisation drawn, we computed the degrees of belief allocated to the $6(6+1)/2 = 21$ intervals $\llbracket i,j \rrbracket \subseteq \mathcal{Z}$, by the three PBFs $Bel^{\mathcal{Z}}$, $Bel_G^{\mathcal{Z}}$ and $Bel_W^{\mathcal{Z}}$ (using $\alpha = 0.05$ for both $Bel_G^{\mathcal{Z}}$ and $Bel_W^{\mathcal{Z}}$). We also computed the true probabilities of these intervals, which are given by $\mathbb{P}_Z(i \leq Z \leq j; \mathbf{P}_0) = \mathbb{P}_Z(Z \leq j; \mathbf{P}_0) - \mathbb{P}_Z(Z < i; \mathbf{P}_0) = P_{0j} - P_{0i-1}$.

Figure 5 shows the results of this experiment. We can check that, for all nine realisations, we have $Bel_G^{\mathcal{Z}}(\llbracket i,j \rrbracket) \leq \mathbb{P}_Z(\llbracket i,j \rrbracket)$ and $Bel_W^{\mathcal{Z}}(\llbracket i,j \rrbracket) \leq \mathbb{P}_Z(\llbracket i,j \rrbracket)$ for all $1 \leq i \leq j \leq K$. In addition, we can observe that $Bel_W^{\mathcal{Z}}$ is competitive with $Bel_G^{\mathcal{Z}}$: for low probability intervals, it is typically not as good, i.e., more cautious, but for high probability intervals, it is usually better, i.e., closer to the true probabilities. We can also notice that, for each of the three PBFs, the most probable the realisation $\mathbf{y}$ is, the less scattered the degrees of belief seem to be. Importantly, for the five most probable realisations (Figs. 5e-5i), we can observe that the likelihood-based PBF $Bel^{\mathcal{Z}}$ is better calibrated than the $1-\alpha$-level PBFs: it respects as well $Bel^{\mathcal{Z}}(\llbracket i,j \rrbracket) \leq \mathbb{P}_Z(\llbracket i,j \rrbracket)$, for all $1 \leq i \leq j \leq K$, whilst being closer to the diagonal. For the four least probable realisations (Figs. 5a-5d), when $Bel^{\mathcal{Z}}$ does go over the diagonal for some events, then it does not go over by much. Overall, this experiment suggests that while the likelihood-based PBF $Bel^{\mathcal{Z}}$ is not designed to have the frequentist property (32), it might still exhibit some useful calibration behaviour in terms of being close to the true probabilities. Moreover, if property (32) needs to be enforced, then $Bel_W^{\mathcal{Z}}$, which has a simple expression and may be regarded as the frequency-calibrated adaptation of $Bel^{\mathcal{Z}}$, is an interesting alternative to the $1-\alpha$-level PBF $Bel_G^{\mathcal{Z}}$ proposed in [5].

5 Conclusion

In this paper, we considered the problem of the prediction of an ordinal variable, according to the likelihood-based evidential method for statistical inference and prediction. First, we established that this prediction can be computed, under some conditions on the possibility distribution representing the estimation uncertainty in this method, by integrating the marginals of this distribution. Then, we showed that the prediction even admits an algebraic expression, when the estimation uncertainty is obtained from past observations of the variable.

In the case of a binary variable, a simple expression, based on the integration of the estimation uncertainty, was obtained for the exact prediction under the condition that the estimation uncertainty is unimodal and continuous. An algebraic expression was even obtained in the particular case where the estimation uncertainty comes from past observations of the binary variable to be predicted. These two results were used in several works in the binary classification domain, which have introduced evi-



(a) $\mathbf{y} = (48, 16, 15, 9, 13, 9)$, $P(\mathbf{y}) = 4.0e-7$

(b) $\mathbf{y} = (60, 17, 16, 7, 3, 7)$, $P(\mathbf{y}) = 3.3e-6$

(c) $\mathbf{y} = (60, 14, 17, 7, 3, 9)$, $P(\mathbf{y}) = 3.5e-6$

(d) $\mathbf{y} = (57, 16, 12, 11, 4, 10)$, $P(\mathbf{y}) = 4.9e-6$

(e) $\mathbf{y} = (55, 12, 15, 10, 7, 11)$, $P(\mathbf{y}) = 5.8e-6$

(f) $\mathbf{y} = (47, 21, 15, 9, 7, 11)$, $P(\mathbf{y}) = 6.0e-6$

(g) $\mathbf{y} = (52, 19, 19, 9, 5, 6)$, $P(\mathbf{y}) = 1.7e-5$

(h) $\mathbf{y} = (50, 20, 19, 9, 5, 7)$, $P(\mathbf{y}) = 1.9e-5$

(i) $\mathbf{y} = (50, 20, 16, 11, 5, 8)$, $P(\mathbf{y}) = 2.1e-5$

Figure 5: Degrees of belief (vertical axes) vs. true probabilities (horizontal axes) of the 21 interval subsets of $\mathcal{Z} = [\![1, 6]\!]$, for 9 draws $\mathbf{y}$ (ordered in ascending probability), for the different methods: likelihood-based PBF $Bel^{\mathcal{Z}}$ (○), 0.95-level PBF $Bel^{\mathcal{Z}}_{G}$ (×), 0.95-level PBF $Bel^{\mathcal{Z}}_{W}$ (+).

dential extensions of binary classifier calibration techniques [22, 17] and of binary logistic and choquistic regressions [19], allowing a finer uncertainty quantification than their probabilistic counterparts. Since our results generalize these two results of the binary case, a natural perspective is to try and use them in the more general, ordinal, classification domain. In particular, preliminary experiments suggest that the four conditions considered in Section 3 might be respected when deriving an evidential extension of ordinal logistic regression [16] in the same way that the evidential extension of binary logistic regression was derived. Another important line of future work is to try and bridge the gap, in terms of conditions for the exact prediction, also for the nominal case.

Acknowledgements

The authors thank Thierry Denœux and Jonathan Williams for their constructive comments that helped to improve this manuscript.

References

[1] S. Boyd, L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[2] S. Braga de Freitas, J. Orrillo, W. Sosa. From Arrow–Debreu condition to star shape preferences. *Optimization*, 69(11):2405-2419, 2019.

[3] A. P. Dempster. New methods for reasoning towards posterior distributions based on sample data. *The Annals of Mathematical Statistics*, 355-374, 1966.

[4] A. P. Dempster. A generalization of Bayesian inference (with discussion). *J. R. Statistical Society B*, 30:205-247, 1968.

[5] T. Denœux. Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, 42(3):228-252, 2006.

[6] T. Denœux. Quantifying Predictive Uncertainty Using Belief Functions: Different Approaches and Practical Construction. In V. Kreinovich, S. Sriboonchitta and N. Chakpitak, editors, *Predictive Econometrics and Big Data*, Springer-Verlag, SCI 753, pages 157-176, 2018.

[7] T. Denœux. Belief functions induced by random fuzzy sets: A general framework for representing uncertain and fuzzy evidence. *Fuzzy Sets and Systems*, 424:63-91, 2021.

[8] T. Denœux. Reasoning with fuzzy and uncertain evidence using epistemic random fuzzy sets: general framework and practical models. *Fuzzy Sets and Systems*, 453:1-36, 2023.

[9] T. Denœux. Uncertainty quantification in logistic regression using random fuzzy sets and belief functions. *International Journal of Approximate Reasoning*, 168:109159, 2024.

[10] T. Denœux, S. Li. Frequency-calibrated belief functions: review and new insights. *International Journal of Approximate Reasoning*, 92:232-254, 2018.

[11] P. E. Jacob, R. Gong, P. T. Edlefsen, A. P. Dempster. A Gibbs sampler for a class of random convex polytopes. *Journal of the American Statistical Association*, 116(535):1181-1192, 2021.

[12] O. Kanjanatarakul, T. Denœux, S. Sriboonchitta. Prediction of future observations using belief functions: a likelihood-based approach. *International Journal of Approximate Reasoning*, 72:71-94, 2016.

[13] O. Kanjanatarakul, S. Sriboonchitta, T. Denœux. Forecasting using belief functions: an application to marketing econometrics. *International Journal of Approximate Reasoning*, 55(5):1113-1128, 2014.

[14] D. Koller, N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT Press, 2009.

[15] R. Martin, C. Liu. *Inferential Models: Reasoning with Uncertainty*, CRC Press, Boca Raton, 2016.

[16] P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society*. 42(2), 1980.

[17] P. Minary, F. Pichon, D. Mercier, E. Lefevre, B. Droit. Evidential joint calibration of binary SVM classifiers. *Soft Computing*, 23:4655-4671, 2019.

[18] F. Pichon, S. Ramel. Algebraic expression for the relative likelihood-based evidential prediction of an ordinal variable. In Y. Bi, A.-L. Jousselme, and T. Denœux, editors, *Belief Functions: Theory and Applications, Proc. of the 8th International Conference, BELIEF 2024*, Belfast, UK, September 2-4, 2024, volume 14909 of Lecture Notes in Computer Science, pages 150–158, Springer, 2024.

[19] S. Ramel, F. Pichon, F. Delmotte. A reliable version of choquistic regression based on evidence theory. *Knowledge-Based Systems*, 205:106-252, 2020.

[20] G. Shafer. *A mathematical theory of evidence.* Princeton university press, 1976.

[21] P. Xu, F. Davoine, T. Denœux. Evidential multinomial logistic regression for multiclass classifier calibration. *International Conference on Information Fusion*, pp. 1106-1112, 2015.

[22] P. Xu, F. Davoine, H. Zha, T. Denœux. Evidential calibration of binary SVM classifiers. *International Journal of Approximate Reasoning*, 72:55-70, 2016.

A Proof of the equalities (19) and (20)

First, given the definition of φ (Eq. (10)), we need to make sure that, as can be observed on Fig. 2a, $\mathbf{L}(w)$ and $\mathbf{U}(w)$ belong to $\mathcal{P}$, which is formally proven by Lemma 9:

Lemma 9. *For all* $w \in [0,1]$, $\mathbf{U}(w) \in \mathcal{P}$ *and* $\mathbf{L}(w) \in \mathcal{P}$. *Specifically, we have* $\mathbf{L}(0) \in \mathcal{P}\backslash\mathcal{P}^*$, $\mathbf{U}(0) \in \mathcal{P}\backslash\mathcal{P}^*$, *and, for all* $w \in (0,1]$, $\mathbf{U}(w) \in \mathcal{P}^*$ *and* $\mathbf{L}(w) \in \mathcal{P}^*$.

Proof.

- Case $w = 0$.

  For all $1 \leq j < K$, given Assumption 1, it is clear that $pl^{\mathcal{P}_j}(P_j) = 0$ if and only if $P_j = 0$ or $P_j = 1$, *i.e.*, we have $L_j(0) = 0$ and $U_j(0) = 1$. We have thus $\mathbf{L}(0) \in \mathcal{P}\backslash\mathcal{P}^*$ and $\mathbf{U}(0) \in \mathcal{P}\backslash\mathcal{P}^*$.

- Case $w \in (0,1]$.

  $\mathbf{L}(w) \in \mathcal{P}^*$ is equivalent to $L_{j-1}(w) < L_j(w)$, for all $1 < j < K$, with $L_j(w) \in (0,1)$ for all $1 \leq j < K$.

  Assume $L_j(w) \in \{0,1\}$. Given Assumption 1, $pl^{\mathcal{P}_j}(P_j) = 0$ if and only if $P_j = 0$ or $P_j = 1$. Hence, $pl^{\mathcal{P}_j}(L_j(w)) = 0$, which contradicts the definition of $L_j(w)$ (it is a P_j such that $pl^{\mathcal{P}_j}(P_j) = w > 0$).

  We now show $L_{j-1}(w) < L_j(w)$, for all $1 < j < K$.

  First, notice that necessarily $\hat{P}_i < P_{i,j} < \hat{P}_j$ for all $1 \leq i < j < K$, since the marginals being unimodal (Assumption 2) and given $\hat{P}_i < \hat{P}_j$ (Lemma 1), on $[\hat{P}_i, \hat{P}_j]$, $pl^{\mathcal{P}_i}$ is strictly decreasing from $pl^{\mathcal{P}_i}(\hat{P}_i) = 1$ to $pl^{\mathcal{P}_i}(\hat{P}_j) < 1$ while $pl^{\mathcal{P}_j}$ is strictly increasing from $pl^{\mathcal{P}_j}(\hat{P}_i) < 1$ to $pl^{\mathcal{P}_j}(\hat{P}_j) = 1$ and therefore it exists $P \in [\hat{P}_i, \hat{P}_j]$, such that $pl^{\mathcal{P}_i}(P) = pl^{\mathcal{P}_j}(P)$, which must be $P = P_{i,j}$ given Assumption 3.

  Next, we show that, for any $1 \leq i < j < K$,

  $$pl^{\mathcal{P}_i}(P) > pl^{\mathcal{P}_j}(P), \quad \forall P \in (0, P_{i,j}). \tag{37}$$

  As there is no $P \in (0, \hat{P}_i]$ such that $pl^{\mathcal{P}_i}(P) = pl^{\mathcal{P}_j}(P)$ (from Assumption 3 and, necessarily, $\hat{P}_i < P_{i,j} < \hat{P}_j$), this implies that, for any $P \in (0, \hat{P}_i]$, either $pl^{\mathcal{P}_i}(P) > pl^{\mathcal{P}_j}(P)$ or $pl^{\mathcal{P}_i}(P) < pl^{\mathcal{P}_j}(P)$. However, $pl^{\mathcal{P}_i}(P) = 1 > pl^{\mathcal{P}_j}(P)$ for $P = \hat{P}_i$ and consequently, for all $P \in (0, \hat{P}_i]$, we have $pl^{\mathcal{P}_i}(P) > pl^{\mathcal{P}_j}(P)$. Moreover, on $[\hat{P}_i, P_{i,j}]$, $pl^{\mathcal{P}_i}$ is decreasing from $pl^{\mathcal{P}_i}(\hat{P}_i) = 1$ to $w_{i,j} := pl^{\mathcal{P}_i}(P_{i,j})$, and $pl^{\mathcal{P}_j}$ is increasing from $pl^{\mathcal{P}_j}(\hat{P}_i)$ to $pl^{\mathcal{P}_j}(P_{i,j}) = w_{i,j}$, with necessarily $pl^{\mathcal{P}_j}(\hat{P}_i) \leq w_{i,j}$, so that we have $pl^{\mathcal{P}_i}(P) > pl^{\mathcal{P}_j}(P)$, for all $P \in [\hat{P}_i, P_{i,j})$. We have thus shown (37).

  We can then show $\mathbf{L}(w) \in \mathcal{P}^*$ as follows. For all $w \in (0,1]$, we have, for all $1 < j < K$, $L_{j-1}(w) \leq \hat{P}_{j-1}$. Since, for all $P \in (0, \hat{P}_{j-1}]$, we have, by (37), $pl^{\mathcal{P}_{j-1}}(P) > pl^{\mathcal{P}_j}(P)$, we deduce that

  $$pl^{\mathcal{P}_j}(L_{j-1}(w)) < pl^{\mathcal{P}_{j-1}}(L_{j-1}(w)) = w. \tag{38}$$

  In addition, for all $w \in (0,1]$, we have $L_j(w) \leq \hat{P}_j$ and $pl^{\mathcal{P}_j}(L_j(w)) = w$. From that and Eq. (38), we obtain $pl^{\mathcal{P}_j}(L_{j-1}(w)) < pl^{\mathcal{P}_j}(L_j(w))$. $pl^{\mathcal{P}_j}$ is strictly increasing on $(0, \hat{P}_j]$ and $L_j(w) \leq \hat{P}_j$, consequently $L_{j-1}(w) < L_j(w)$.

  It can be shown similarly that $\mathbf{U}(w) \in \mathcal{P}^*$.

□

Let us consider Eqs. (19) and (20) for $w = 0$ and for all $v \in [0,1]$. When $w = 0$, we have $\Gamma(w) = \mathcal{P}$. If $v = 0$ then $\varphi(\Gamma(w), v) = 1$ (since $\varphi(\mathbf{P}, 0) = 1$, for all $\mathbf{P} \in \mathcal{P}$) and therefore $\ell(0,0) = u(0,0) = 1$. If $v \in (0,1]$ then $\varphi(\Gamma(w), v) = \mathcal{Z}$ (since, for all $k \in \mathcal{Z}$, there exists $\mathbf{P} \in \mathcal{P}$ such that $P_{k-1} < v \leq P_k$) and therefore $\ell(0, v) = 1$ and $u(0, v) = K$. Now, from the proof of Lemma 9 ("Case $w = 0$"), we have $\mathbf{L}(0) = (L_1(0), \ldots, L_{K-1}(0)) = (0, \ldots, 0)$ and $\mathbf{U}(0) = (U_1(0), \ldots, U_{K-1}(0)) = (1, \ldots, 1)$. It is clear therefore that $\varphi(\mathbf{L}(0), 0) = \varphi(\mathbf{U}(0), 0) = 1$ and, for $v \in (0,1]$, $\varphi(\mathbf{L}(0), v) = K$ and $\varphi(\mathbf{U}(0), v) = 1$. Hence, Eqs. (19) and (20) hold when $w = 0$ for all $v \in [0,1]$.

Let us now turn our attention to Eqs. (19) and (20) for the case $w \in (0,1]$ (and $v \in [0,1]$). Consider, for $w \in (0,1]$, the set

$$\underline{\Gamma}(w) := \left\{ \mathbf{P} \in \mathcal{P}^* \;\middle|\; \mathbf{P} \in \prod_{j=1}^{K-1} [L_j(w), U_j(w)] \right\}.$$

Let $\mathbf{P} \in \Gamma(w)$, hence $\mathbf{P} \in \mathcal{P}^*$ from Assumption 1. Let P_j be the j-th component of $\mathbf{P}$, $1 \leq j < K$. We have, by (13), $pl^{\mathcal{P}_j}(P_j) \geq pl^{\mathcal{P}}(\mathbf{P}) \geq w$. From Assumption 2, we have $P_j \in [L_j(w), U_j(w)]$. Hence, $\mathbf{P} \in \prod_{j=1}^{K-1} [L_j(w), U_j(w)]$ and thus $\mathbf{P} \in \underline{\Gamma}(w)$. We have thus shown that we have

$$\underline{\Gamma}(w) \supseteq \Gamma(w). \tag{39}$$

Let $\tilde{\ell}(w,v) := \min_{\mathbf{P} \in \underline{\Gamma}(w)} \varphi(\mathbf{P}, v)$ and $\tilde{u}(w,v) = \max_{\mathbf{P} \in \underline{\Gamma}(w)} \varphi(\mathbf{P}, v)$. From (39), we have:

$$\tilde{\ell}(w,v) \leq \ell(w,v) \leq u(w,v) \leq \tilde{u}(w,v). \tag{40}$$

Given (40), Eqs. (19) and (20) are shown to hold when $w \in (0,1]$ for all $v \in [0,1]$, by showing first (Lemma 10) that

$$\begin{aligned} \tilde{\ell}(w,v) &= \varphi(\mathbf{U}(w), v), \\ \tilde{u}(w,v) &= \varphi(\mathbf{L}(w), v), \end{aligned}$$

and then (Lemma 11) that

$$\begin{aligned} \tilde{\ell}(w,v) &\geq \ell(w,v), \\ u(w,v) &\geq \tilde{u}(w,v). \end{aligned}$$

Lemma 10. *For all $w \in (0,1]$ and all $v \in [0,1]$,*

$$\begin{aligned} \tilde{\ell}(w,v) &= \varphi(\mathbf{U}(w), v), \\ \tilde{u}(w,v) &= \varphi(\mathbf{L}(w), v). \end{aligned}$$

Proof. Assume in this proof some $w \in (0,1]$.

From Lemma 9, $\mathbf{L}(w) \in \mathcal{P}^*$ and $\mathbf{U}(w) \in \mathcal{P}^*$ thus $\mathbf{L}(w) \in \underline{\Gamma}(w)$ and $\mathbf{U}(w) \in \underline{\Gamma}(w)$, hence we can show that $\tilde{u}(w,v) = \varphi(\mathbf{L}(w), v)$ and $\tilde{\ell}(w,v) = \varphi(\mathbf{U}(w), v)$ merely by showing that $\varphi(\mathbf{L}(w), v) \geq \varphi(\mathbf{P}, v)$ and $\varphi(\mathbf{U}(w), v) \leq \varphi(\mathbf{P}, v)$ for all $\mathbf{P} \in \underline{\Gamma}(w)$ and all $v \in [0,1]$. This is done as follows.

For all $v \in [0,1]$ and for all $\mathbf{P} \in \underline{\Gamma}(w)$, assume $\varphi(\mathbf{P}, v) = i$ for some $1 \leq i \leq K$. This means there exists $1 < i \leq K$ such that $P_{i-1} < v \leq P_i$, otherwise $v \leq P_1$. Assume first the former, hence $P_{i-1} < v$. In this case, since $L_{i-1}(w) \leq P_{i-1}$, then $L_{i-1}(w) < v$ and, from $\mathbf{L}(w) \in \mathcal{P}^*$, we have $L_{i-1}(w) < L_i(w)$, thus $\varphi(\mathbf{L}(w), v) \geq i$. Assume now the latter, *i.e.*, $v \leq P_1$. Since $L_1(w) \leq P_1$, either $v \leq L_1(w)$ in which case $\varphi(\mathbf{L}(w), v) = \varphi(\mathbf{P}, v)$, or $v > L_1(w)$, thus $\varphi(\mathbf{L}(w), v) > \varphi(\mathbf{P}, v)$. Overall, we have thus $\varphi(\mathbf{L}(w), v) \geq \varphi(\mathbf{P}, v)$ for all $v \in [0,1]$ and for all $\mathbf{P} \in \underline{\Gamma}(w)$

We show $\varphi(\mathbf{U}(w), v) \leq \varphi(\mathbf{P}, v)$ similarly as follows.

For all $v \in [0,1]$ and for all $\mathbf{P} \in \underline{\Gamma}(w)$, assume $\varphi(\mathbf{P}, v) = i$ for some $1 \leq i \leq K$. This means there exists $1 < i \leq K$ such that $P_{i-1} < v \leq P_i$, otherwise $v \leq P_1$. Assume first the former, hence $v \leq P_i$. In this case, since $U_i(w) \geq P_i$, then $U_i(w) \geq v$. Furthermore, $U_{i-1}(w) \geq P_{i-1}$, hence either $U_{i-1}(w) < v \leq U_i(w)$ (using $\mathbf{U}(w) \in \mathcal{P}^*$), in which case $\varphi(\mathbf{U}(w), v) = \varphi(\mathbf{P}, v)$, or $U_{i-1}(w) \geq v$, in which case $\varphi(\mathbf{U}(w), v) < \varphi(\mathbf{P}, v)$. Assume now the latter, *i.e.*, $v \leq P_1$. Since $U_1(w) \geq P_1$, then $v \leq U_1(w)$, thus $\varphi(\mathbf{U}(w), v) = \varphi(\mathbf{P}, v)$. Overall, we have thus $\varphi(\mathbf{U}(w), v) \leq \varphi(\mathbf{P}, v)$ for all $v \in [0,1]$ and for all $\mathbf{P} \in \underline{\Gamma}(w)$.

□

Lemma 11. *For all $w \in (0,1]$ and all $v \in [0,1]$,*

$$\begin{aligned} \tilde{\ell}(w,v) &\geq \ell(w,v), \\ u(w,v) &\geq \tilde{u}(w,v). \end{aligned} \tag{41}$$

Proof. Assume in this proof some $w \in (0,1]$.

If $\varphi(\mathbf{L}(w), v) = 1$, then, since necessarily $u(w,v) \geq 1$, we have $u(w,v) \geq \varphi(\mathbf{L}(w), v) = \tilde{u}(w,v)$.

Assume now $\varphi(\mathbf{L}(w), v) = \tilde{u}$ for some $1 < \tilde{u} \leq K$. This implies that we have $L_{\tilde{u}-1}(w) < v \leq L_{\tilde{u}}(w)$, with the convention $L_{\tilde{u}}(w) = 1$ for $\tilde{u} = K$.

Let $P_{\tilde{u}-1} := L_{\tilde{u}-1}(w)$. Let

$$\mathbf{P}_{\tilde{u}-1} := \arg \sup_{\mathbf{P}^{\downarrow \mathcal{P}_{\tilde{u}-1}} = P_{\tilde{u}-1}} pl^{\mathcal{P}}(\mathbf{P}),$$

where $\mathbf{P}^{\downarrow \mathcal{P}_{\tilde{u}-1}}$ denotes the value of component $\tilde{u}-1$ of $\mathbf{P}$. Since $pl^{\mathcal{P}_{\tilde{u}-1}}(P_{\tilde{u}-1}) = w$, then $pl^{\mathcal{P}}(\mathbf{P}_{\tilde{u}-1}) = w$, *i.e.*, $\mathbf{P}_{\tilde{u}-1} \in \Gamma(w)$. Furthermore, $\mathbf{P}_{\tilde{u}-1}^{\downarrow \mathcal{P}_{\tilde{u}-1}} = P_{\tilde{u}-1}$ hence, since $P_{\tilde{u}-1} < v$, we have $\varphi(\mathbf{P}_{\tilde{u}-1}, v) \geq \tilde{u}$ and thus $\varphi(\mathbf{P}_{\tilde{u}-1}, v) \geq \varphi(\mathbf{L}(w), v)$.

From $\mathbf{P}_{\tilde{u}-1} \in \Gamma(w)$ and $u(w,v)$ maximum of $\varphi(\mathbf{P}, v)$ under the constraint $\mathbf{P} \in \Gamma(w)$, we have $u(w,v) \geq \varphi(\mathbf{P}_{\tilde{u}-1}, v)$, hence $u(w,v) \geq \varphi(\mathbf{L}(w), v) = \tilde{u}(w,v)$.

Eq. (41) is shown (almost) similarly as follows.

If $\varphi(\mathbf{U}(w), v) = K$, then, since necessarily $\ell(w,v) \leq K$, we have $\ell(w,v) \leq \varphi(\mathbf{U}(w), v) = \tilde{\ell}(w,v)$.

Assume now $\varphi(\mathbf{U}(w), v) = \tilde{\ell}$ for some $1 < \tilde{\ell} < K$. This implies that we have $U_{\tilde{\ell}-1}(w) < v \leq U_{\tilde{\ell}}(w)$.

Let $P_{\tilde{\ell}} := U_{\tilde{\ell}}(w)$. Let

$$\mathbf{P}_{\tilde{\ell}} := \arg \sup_{\mathbf{P}^{\downarrow \mathcal{P}_{\tilde{\ell}}} = P_{\tilde{\ell}}} pl^{\mathcal{P}}(\mathbf{P}).$$

Since $pl^{\mathcal{P}_{\tilde{\ell}}}(P_{\tilde{\ell}}) = w$, then $pl^{\mathcal{P}}(\mathbf{P}_{\tilde{\ell}}) = w$, *i.e.*, $\mathbf{P}_{\tilde{\ell}} \in \Gamma(w)$. Furthermore, $\mathbf{P}_{\tilde{\ell}}^{\downarrow \mathcal{P}_{\tilde{\ell}}} = P_{\tilde{\ell}}$ hence, since $P_{\tilde{\ell}} \geq v$, we have $\varphi(\mathbf{P}_{\tilde{\ell}}, v) \leq \tilde{\ell}$ and thus $\varphi(\mathbf{P}_{\tilde{\ell}}, v) \leq \varphi(\mathbf{U}(w), v)$.

Assume now $\varphi(\mathbf{U}(w), v) = 1$. This implies that we have $v \leq U_1(w)$.

Let $P_1^U := U_1(w)$. Let

$$\mathbf{P}_1 := \arg \sup_{\mathbf{P}^{\downarrow \mathcal{P}_1} = P_1^U} pl^{\mathcal{P}}(\mathbf{P}).$$

Since $pl^{\mathcal{P}_1}(P_1^U) = w$, then $pl^{\mathcal{P}}(\mathbf{P}_1) = w$, *i.e.*, $\mathbf{P}_1 \in \Gamma(w)$. Furthermore, $\mathbf{P}_1^{\downarrow \mathcal{P}_1} = P_1^U$ hence, since $v \leq P_1^U$, we have $\varphi(\mathbf{P}_1, v) = 1$ and thus $\varphi(\mathbf{P}_1, v) = \varphi(\mathbf{U}(w), v)$.

Therefore, when $\varphi(\mathbf{U}(w), v) = \tilde{\ell}$ for some $1 \leq \tilde{\ell} < K$, we can find a $\mathbf{P} \in \Gamma(w)$ such that $\varphi(\mathbf{P}, v) \leq \varphi(\mathbf{U}(w), v)$. Since $\ell(w,v)$ is the minimum of $\varphi(\mathbf{P}, v)$ under the constraint $\mathbf{P} \in \Gamma(w)$, we have $\ell(w,v) \leq \varphi(\mathbf{U}(w), v) = \tilde{\ell}(w,v)$.

□

B Proof of the expression (15) of $Bel^{\mathcal{Z}}([\![i,j]\!])$

We start from expression (22) of $Bel^{\mathcal{Z}}([\![i,j]\!])$.

Let us first consider the case $i = 1$ and $j = K$. For $w = 0$ and $v = 0$, we have $\varphi(\mathbf{U}(w), v) = \varphi(\mathbf{L}(w), v) = 1$. For $w = 0$ and $v \in (0,1]$, we have $\varphi(\mathbf{U}(w), v) = 1$ and $\varphi(\mathbf{L}(w), v) = K$. For $w \in (0,1]$, $\mathbf{U}(w) \in \mathcal{P}^*$ and $\mathbf{L}(w) \in \mathcal{P}^*$, hence $\varphi(\mathbf{U}(w), v) \geq 1$ and $\varphi(\mathbf{L}(w), v) \leq K$, for all $v \in [0,1]$. In sum, for all $(w,v) \in [0,1]^2$, $1 \leq \varphi(\mathbf{U}(w), v)$ and $\varphi(\mathbf{L}(w), v) \leq K$ hold, hence $Bel^{\mathcal{Z}}([\![1,K]\!]) = \lambda \otimes \mu(\{(w,v) \in [0,1]^2\}) = 1$, which gives us the first case in Eq.(15).

For $i = 1$ and $j < K$, since we have established that for all $(w,v) \in [0,1]^2$, $1 \leq \varphi(\mathbf{U}(w), v)$, Eq. (22)

reduces to

$$
\begin{aligned}
& \lambda \otimes \mu(\{(w,v) \in [0,1]^2 | \varphi(\mathbf{L}(w), v) \leq j\}) \\
= \; & \lambda \otimes \mu(\{(w,v) \in [0,1]^2 | L_j(w) \geq v\}) \\
= \; & \int_0^1 \int_0^{L_j(w)} 1 dv dw \\
= \; & \int_0^1 L_j(w) dw \\
= \; & \int_{pl^{\mathcal{P}_j}(0)}^{pl^{\mathcal{P}_j}(\hat{P}_j)} L_j(w) dw \\
= \; & \int_0^{\hat{P}_j} L_j(pl^{\mathcal{P}_j}(t)) pl'^{\mathcal{P}_j}(t) dt
\end{aligned}
$$

which reduces to (given that $pl^{\mathcal{P}_j}$ is strictly increasing on $[0, \hat{P}_j]$, and therefore $L_j : [0,1] \to [0, \hat{P}_j]$ is its inverse function on this interval)

$$
\begin{aligned}
= \; & \int_0^{\hat{P}_j} t\, pl'^{\mathcal{P}_j}(t) dt \\
= \; & [t\, pl^{\mathcal{P}_j}(t)]_0^{\hat{P}_j} - \int_0^{\hat{P}_j} pl^{\mathcal{P}_j}(t) dt \\
= \; & \hat{P}_j - \int_0^{\hat{P}_j} pl^{\mathcal{P}_j}(t) dt,
\end{aligned}
$$

which gives us the second case in Eq. (15).

For $1 < i$ and $j = K$, since we have established that for all $(w,v) \in [0,1]^2$, $K \geq \varphi(\mathbf{L}(w), v)$, Eq. (22) reduces to

$$
\begin{aligned}
& \lambda \otimes \mu(\{(w,v) \in [0,1]^2 | i \leq \varphi(\mathbf{U}(w), v)\}) \\
= \; & \lambda \otimes \mu(\{(w,v) \in [0,1]^2 | U_{i-1}(w) < v)\}) \\
= \; & \int_0^1 \int_{U_{i-1}(w)}^1 1 dv dw \\
= \; & \int_0^1 1 - U_{i-1}(w) dw \\
= \; & 1 - \int_0^1 U_{i-1}(w) dw \\
= \; & 1 - \int_{pl^{\mathcal{P}_{i-1}}(1)}^{pl^{\mathcal{P}_{i-1}}(\hat{P}_{i-1})} U_{i-1}(w) dw \\
= \; & 1 - \int_1^{\hat{P}_{i-1}} U_{i-1}(pl^{\mathcal{P}_{i-1}}(t)) pl'^{\mathcal{P}_{i-1}}(t) dt \\
= \; & 1 + \int_{\hat{P}_{i-1}}^1 U_{i-1}(pl^{\mathcal{P}_{i-1}}(t)) pl'^{\mathcal{P}_{i-1}}(t) dt
\end{aligned}
$$

which reduces to (given that $pl^{\mathcal{P}_{i-1}}$ is strictly decreasing on $[\hat{P}_{i-1}, 1]$, and therefore $U_{i-1} : [0,1] \to$

$[\hat{P}_{i-1}, 1]$ is its inverse function on this interval)

$$\begin{aligned}
&= 1 + \int_{\hat{P}_{i-1}}^{1} t\, {pl'}^{\mathcal{P}_{i-1}}(t)dt \\
&= 1 + [t\, pl^{\mathcal{P}_{i-1}}(t)]_{\hat{P}_{i-1}}^{1} - \int_{\hat{P}_{i-1}}^{1} pl^{\mathcal{P}_{i-1}}(t)dt \\
&= 1 - \hat{P}_{i-1} - \int_{\hat{P}_{i-1}}^{1} pl^{\mathcal{P}_{i-1}}(t)dt,
\end{aligned}$$

which gives us the third case in Eq. (15).

For $1 < i$ and $j < K$, since for all $(w, v) \in [0, 1]^2$, we have $1 < i \le \varphi(\mathbf{U}(w), v) \Leftrightarrow U_{i-1}(w) < v$ and $\varphi(\mathbf{L}(w), v) \le j < K \Leftrightarrow v \le L_j(w)$, Eq. (22) reduces to

$$\begin{aligned}
& \lambda \otimes \mu(\{(w, v) \in [0, 1]^2 | U_{i-1}(w) < v, L_j(w) \ge v)\}) \\
= \; & \lambda \otimes \mu(\{(w, v) \in [0, 1]^2 | U_{i-1}(w) < v \le L_j(w)\}) \qquad (42)
\end{aligned}$$

Let $w_{i-1,j} := pl^{\mathcal{P}_{i-1}}(P_{i-1,j}) = pl^{\mathcal{P}_j}(P_{i-1,j})$. On $[\hat{P}_{i-1}, 1]$, $pl^{\mathcal{P}_{i-1}}$ is (strictly) decreasing from $pl^{\mathcal{P}_{i-1}}(\hat{P}_{i-1}) = 1$ to $pl^{\mathcal{P}_{i-1}}(1) = 0$, hence its inverse function (for this interval) U_{i-1} is decreasing on $[0, 1]$ (and we also have $0 < w_{i-1,j} < 1$ since $0 < \hat{P}_{i-1} < P_{i-1,j} < \hat{P}_j < 1$). On $[0, w_{i-1,j}]$, U_{i-1} decreases from $U_{i-1}(0) = 1$ to $U_{i-1}(w_{i-1,j}) = P_{i-1,j}$, and on $[w_{i-1,j}, 1]$ it decreases from $U_{i-1}(w_{i-1,j}) = P_{i-1,j}$ to $U_{i-1}(1) = \hat{P}_{i-1}$.

Similarly, we obtain that L_j increases on $[0, w_{i-1,j}]$ from $L_j(0) = 0$ to $L_j(w_{i-1,j}) = P_{i-1,j}$, and on $[w_{i-1,j}, 1]$ it increases from $L_j(w_{i-1,j}) = P_{i-1,j}$ to $L_j(1) = \hat{P}_j$. Consequently, for all $w \in [0, w_{i-1,j})$, we have $U_{i-1}(w) > L_j(w)$, and for all $w \in [w_{i-1,j}, 1]$, $U_{i-1}(w) \le L_j(w)$. Eq.(42) reduces thus to

$$\begin{aligned}
& \lambda \otimes \mu(\{v \in [0, 1], w_{i-1,j} \le w \le 1 | U_{i-1}(w) < v \le L_j(w)) \\
= \; & \int_{w_{i-1,j}}^{1} \int_{U_{i-1}(w)}^{L_j(w)} 1 dv dw \\
= \; & \int_{w_{i-1,j}}^{1} L_j(w) - U_{i-1}(w) dw \\
= \; & \int_{w_{i-1,j}}^{1} L_j(w) dw - \int_{w_{i-1,j}}^{1} U_{i-1}(w) dw. \qquad (43)
\end{aligned}$$

We have

$$\begin{aligned}
\int_{w_{i-1,j}}^{1} L_j(w) dw &= \int_{pl^{\mathcal{P}_j}(P_{i-1,j})}^{pl^{\mathcal{P}_j}(\hat{P}_j)} L_j(w) dw \\
&= \int_{P_{i-1,j}}^{\hat{P}_j} L_j(pl^{\mathcal{P}_j}(t)) {pl'}^{\mathcal{P}_j}(t) dt \\
&= \int_{P_{i-1,j}}^{\hat{P}_j} t\, {pl'}^{\mathcal{P}_j}(t) dt \\
&= [t\, pl^{\mathcal{P}_j}(t)]_{P_{i-1,j}}^{\hat{P}_j} - \int_{P_{i-1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(t) dt \\
&= \hat{P}_j - P_{i-1,j}\, w_{i-1,j} - \int_{P_{i-1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(t) dt,
\end{aligned}$$

and

$$\begin{aligned}
\int_{w_{i-1,j}}^{1} U_{i-1}(w)dw &= \int_{pl^{\mathcal{P}_{i-1}}(P_{i-1,j})}^{pl^{\mathcal{P}_{i-1}}(\hat{P}_{i-1})} U_{i-1}(w)dw \\
&= \int_{P_{i-1,j}}^{\hat{P}_{i-1}} U_{i-1}(pl^{\mathcal{P}_{i-1}}(t))pl'^{\mathcal{P}_{i-1}}(t)dt \\
&= -\int_{\hat{P}_{i-1}}^{P_{i-1,j}} U_{i-1}(pl^{\mathcal{P}_{i-1}}(t))pl'^{\mathcal{P}_{i-1}}(t)dt \\
&= -\int_{\hat{P}_{i-1}}^{P_{i-1,j}} t\,pl'^{\mathcal{P}_{i-1}}(t)dt \\
&= -([t\,pl^{\mathcal{P}_{i-1}}(t)]_{\hat{P}_{i-1}}^{P_{i-1,j}} - \int_{\hat{P}_{i-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(t)dt) \\
&= \hat{P}_{i-1} - P_{i-1,j}\,w_{i-1,j} + \int_{\hat{P}_{i-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(t)dt.
\end{aligned}$$

Eq. (43) simplifies then to

$$\begin{aligned}
&= \hat{P}_j - P_{i-1,j}\,w_{i-1,j} - \int_{P_{i-1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(t)dt - (\hat{P}_{i-1} - P_{i-1,j}\,w_{i-1,j} + \int_{\hat{P}_{i-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(t)dt) \\
&= \hat{P}_j - \hat{P}_{i-1} - \int_{P_{i-1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(t)dt - \int_{\hat{P}_{i-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(t)dt,
\end{aligned}$$

which gives us the last case in Eq. (15).

C Proof of Lemma 4

In the cases where $\ell(w,v) = u(w,v)$ or $\ell(w,v) = u(w,v) - 1$, Eq. (23) clearly holds.

Consider the case $w = 0$, in which case $\Gamma(w) = \mathcal{P}$. If $v = 0$ then $\varphi(\Gamma(w), v) = 1$ since $\varphi(\mathbf{P}, 0) = 1$, for all $\mathbf{P} \in \mathcal{P}$. If $v \in (0,1]$ then $\varphi(\Gamma(w), v) = \mathcal{Z}$ since, for all $k \in \mathcal{Z}$, there exists $\mathbf{P} \in \mathcal{P}$ such that $P_{k-1} < v \leq P_k$. In other words, Eq. (23) holds for $w = 0$.

It remains thus to consider the cases for which $\ell(w,v) < u(w,v) - 1$ and $w \in (0,1]$.

Let $w \in (0,1]$ and $v \in [0,1]$ such that $\ell(w,v) < u(w,v) - 1$. For short, we denote $\ell(w,v)$ by ℓ and $u(w,v)$ by u. Let $(\mathbf{P}^{(\ell)}, \mathbf{P}^{(u)}) \in \Gamma(w)^2$ such that $\varphi(\mathbf{P}^{(\ell)}, v) = \ell$ and $\varphi(\mathbf{P}^{(u)}, v) = u$. Furthermore, given Assumption 4, let $\mathbf{P}^* \in \Gamma(w)$ be the star center of $\Gamma(w)$. Let $z := \varphi(\mathbf{P}^*, v)$. Since $\mathbf{P}^* \in \Gamma(w)$, then necessarily $z \in [\![\ell, u]\!]$ and thus $[\![\ell, u]\!] = [\![\ell, z]\!] \cup [\![z, u]\!]$. Hence, for Eq. (23) to hold, it remains to show that for any k, $\ell < k < z$, there exists $\mathbf{P}^{(k)} \in \Gamma(w)$ such that $\varphi(\mathbf{P}^{(k)}, v) = k$ and that for any k', $z < k' < u$, there exists $\mathbf{P}^{(k')} \in \Gamma(w)$ such that $\varphi(\mathbf{P}^{(k')}, v) = k'$.

If $\ell = z$ or $\ell = z - 1$, then there is no k such that $\ell < k < z$, so we merely need to consider the cases where $\ell < z - 1$. Remark first that in such cases, and given Assumption 1, we have

$$\varphi(\mathbf{P}^*, v) = z \implies 0 < P^*_{z-1} < v$$

with $z > 2$. Similarly,

$$\varphi(\mathbf{P}^{(\ell)}, v) = \ell \implies v \leq P^{(\ell)}_\ell < 1$$

with $\ell < K - 1$.

Furthermore, consider some $k \in \mathcal{Z}$ such that $1 \leq \ell < k < z \leq K$. From Assumption 1 we have

$$\begin{aligned}
0 <& P^{(\ell)}_\ell < P^{(\ell)}_k < 1, \\
0 <& P^*_k \leq P^*_{z-1} < 1.
\end{aligned}$$

Therefore, overall, we have

$$0 < P^*_k \le P^*_{z-1} < v \le P^{(\ell)}_\ell < P^{(\ell)}_k < 1. \tag{44}$$

Let

$$t_k := \frac{v - P^*_k}{P^{(\ell)}_k - P^*_k}. \tag{45}$$

From (44), we have $v - P^*_k > 0$, $P^{(\ell)}_k - P^*_k > 0$ and $v - P^*_k < P^{(\ell)}_k - P^*_k$, thus $t_k \in (0, 1)$.

Since $\Gamma(w)$ is a star convex set (Assumption 4), $\mathbf{P}^{(k)} := t_k\mathbf{P}^{(\ell)} + (1 - t_k)\mathbf{P}^* \in \Gamma(w)$. Furthermore, given (45), $\mathbf{P}^{(k)}$ satisfies

$$\begin{aligned}
P^{(k)}_k &= \frac{v - P^*_k}{P^{(\ell)}_k - P^*_k} P^{(\ell)}_k + \left(1 - \frac{v - P^*_k}{P^{(\ell)}_k - P^*_k}\right) P^*_k \\
&= \frac{v - P^*_k}{P^{(\ell)}_k - P^*_k} P^{(\ell)}_k + \frac{P^{(\ell)}_k - v}{P^{(\ell)}_k - P^*_k} P^*_k \\
&= \frac{vP^{(\ell)}_k - P^*_k P^{(\ell)}_k + P^{(\ell)}_k P^*_k - vP^*_k}{P^{(\ell)}_k - P^*_k} \\
&= \frac{v(P^{(\ell)}_k - P^*_k)}{P^{(\ell)}_k - P^*_k} \\
&= v.
\end{aligned}$$

Finally, from Assumption 1 and $\mathbf{P}^{(k)} \in \Gamma(w)$, we have $P^{(k)}_{k-1} < P^{(k)}_k = v$, so that $\varphi(\mathbf{P}^{(k)}, v) = k$.

The fact that for any k', $z < k' < u$, there exists $\mathbf{P}^{(k')} \in \Gamma(w)$ such that $\varphi(\mathbf{P}^{(k')}, v) = k'$, is shown similarly as follows.

If $z = u$ or $z = u - 1$, then there is no k' such that $z < k' < u$, so we merely need to consider the cases where $z < u - 1$. In such cases, and given Assumption 1, we have

$$\varphi(\mathbf{P}^{(u)}, v) = u \implies 0 < P^{(u)}_{u-1} < v$$

with $u > 2$. Similarly,

$$\varphi(\mathbf{P}^*, v) = z \implies v \le P^*_z < 1$$

with $z < K - 1$.

Furthermore, consider some $k' \in \mathcal{Z}$ such that $1 \le z < k' < u \le K$. From Assumption 1 we have

$$\begin{aligned}
0 <& P^*_z < P^*_{k'} < 1 \\
0 <& P^{(u)}_{k'} \le P^{(u)}_{u-1} < 1
\end{aligned}$$

Therefore, overall, we have

$$0 < P^{(u)}_{k'} \le P^{(u)}_{u-1} < v \le P^*_z < P^*_{k'}, \tag{46}$$

Let

$$t_{k'} := \frac{v - P^{(u)}_{k'}}{P^*_{k'} - P^{(u)}_{k'}}. \tag{47}$$

From (46), we have $v - P^{(u)}_{k'} > 0$, $P^*_{k'} - P^{(u)}_{k'} > 0$ and $v - P^{(u)}_{k'} < P^*_{k'} - P^{(u)}_{k'}$, thus $t_{k'} \in (0, 1)$.

Since $\Gamma(w)$ is a star convex set (Assumption 4), $\mathbf{P}^{(k')} := t_{k'}\mathbf{P}^* + (1 - t_{k'})\mathbf{P}^{(u)} \in \Gamma(w)$. Furthermore, given (47), $\mathbf{P}^{(k')}$ satisfies

$$
\begin{aligned}
P_{k'}^{(k')} &= \frac{v - P_{k'}^{(u)}}{P_{k'}^* - P_{k'}^{(u)}} P_{k'}^* + \left(1 - \frac{v - P_{k'}^{(u)}}{P_{k'}^* - P_{k'}^{(u)}}\right) P_{k'}^{(u)} \\
&= \frac{v - P_{k'}^{(u)}}{P_{k'}^* - P_{k'}^{(u)}} P_{k'}^* + \frac{P_{k'}^* - v}{P_{k'}^* - P_{k'}^{(u)}} P_{k'}^{(u)} \\
&= \frac{v P_{k'}^* - P_{k'}^{(u)} P_{k'}^* + P_{k'}^* P_{k'}^{(u)} - v P_{k'}^{(u)}}{P_{k'}^* - P_{k'}^{(u)}} \\
&= \frac{v(P_{k'}^* - P_{k'}^{(u)})}{P_{k'}^* - P_{k'}^{(u)}} \\
&= v.
\end{aligned}
$$

Finally, from Assumption 1 and $\mathbf{P}^{(k')} \in \Gamma(w)$, we have $P_{k'-1}^{(k')} < P_{k'}^{(k')} = v$, so that $\varphi(\mathbf{P}^{(k')}, v) = k'$.

D Proof of Corollary 1

According to [5, Eq. (29)], we have

$$
m^{\mathcal{Z}}([\![i,j]\!]) = \tag{48}
$$
$$
\begin{cases}
Bel^{\mathcal{Z}}(\{j\}), & \text{if } j = i, \\
Bel^{\mathcal{Z}}([\![i,j]\!]) - Bel^{\mathcal{Z}}([\![i+1,j]\!]) - Bel^{\mathcal{Z}}([\![i,j-1]\!]), & \text{if } j = i+1, \\
Bel^{\mathcal{Z}}([\![i,j]\!]) - Bel^{\mathcal{Z}}([\![i+1,j]\!]) - Bel^{\mathcal{Z}}([\![i,j-1]\!]) + Bel^{\mathcal{Z}}([\![i+1,j-1]\!]), & \text{otherwise.}
\end{cases}
$$

Consider the first case of Eq. (24) where $i = 1$ and $j = K$. According to Eq. (48), we have

$$
\begin{aligned}
m^{\mathcal{Z}}([\![1,K]\!]) &= Bel^{\mathcal{Z}}([\![1,K]\!]) - Bel^{\mathcal{Z}}([\![2,K]\!]) - Bel^{\mathcal{Z}}([\![1,K-1]\!]) + Bel^{\mathcal{Z}}([\![2,K-1]\!]) \\
&= 1 - \left(1 - \hat{P}_1 - \int_{\hat{P}_1}^{1} pl^{\mathcal{P}_1}(P)dP\right) - \left(\hat{P}_{K-1} - \int_{0}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&\quad + \left(\hat{P}_{K-1} - \hat{P}_1 - \int_{\hat{P}_1}^{P_{1,K-1}} pl^{\mathcal{P}_1}(P)dP - \int_{P_{1,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&= \int_{0}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP - \int_{P_{1,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP \\
&\quad + \int_{\hat{P}_1}^{1} pl^{\mathcal{P}_1}(P)dP - \int_{\hat{P}_1}^{P_{1,K-1}} pl^{\mathcal{P}_1}(P)dP \\
&= \int_{0}^{P_{1,K-1}} pl^{\mathcal{P}_{K-1}}(P)dP + \int_{P_{1,K-1}}^{1} pl^{\mathcal{P}_1}(P)dP \\
&= \int_{P_{0,K-1}}^{P_{1,K-1}} pl^{\mathcal{P}_{K-1}}(P)dP + \int_{P_{1,K-1}}^{P_{1,K}} pl^{\mathcal{P}_1}(P)dP \\
&= A_{1,K}.
\end{aligned}
$$

Consider now the second case of Eq. (24) where $i = 1$ and $j < K$. For $j = i = 1$, we have

$$\begin{aligned} m^{\mathcal{Z}}(\{1\}) &= Bel^{\mathcal{Z}}(\{1\}) \\ &= \hat{P}_1 - \int_0^{\hat{P}_1} pl^{\mathcal{P}_1}(P)dP \\ &= (P_{1,1} - P_{0,0}) - \int_{P_{0,1}}^{P_{1,1}} pl^{\mathcal{P}_1}(P)dP \\ &= A_{1,1} - \int_{P_{0,1}}^{P_{1,1}} pl^{\mathcal{P}_1}(P)dP. \end{aligned}$$

For $j = i + 1 = 2$ we have

$$\begin{aligned} m^{\mathcal{Z}}([\![1,2]\!]) &= Bel^{\mathcal{Z}}([\![1,2]\!]) - Bel^{\mathcal{Z}}(\{2\}) - Bel^{\mathcal{Z}}(\{1\}) \\ &= \left(\hat{P}_2 - \int_0^{\hat{P}_2} pl^{\mathcal{P}_2}(P)dP\right) \\ &\quad - \left(\hat{P}_2 - \hat{P}_1 - \int_{\hat{P}_1}^{P_{1,2}} pl^{\mathcal{P}_1}(P)dP - \int_{P_{1,2}}^{\hat{P}_2} pl^{\mathcal{P}_2}(P)dP\right) \\ &\quad - \left(\hat{P}_1 - \int_0^{\hat{P}_1} pl^{\mathcal{P}_1}(P)dP\right) \\ &= \int_0^{\hat{P}_1} pl^{\mathcal{P}_1}(P)dP + \int_{\hat{P}_1}^{P_{1,2}} pl^{\mathcal{P}_1}(P)dP \\ &\quad + \int_{P_{1,2}}^{\hat{P}_2} pl^{\mathcal{P}_2}(P)dP - \int_0^{\hat{P}_2} pl^{\mathcal{P}_2}(P)dP \\ &= \left(\int_0^{\hat{P}_1} pl^{\mathcal{P}_1}(P)dP + \int_{\hat{P}_1}^{P_{1,2}} pl^{\mathcal{P}_1}(P)dP\right) - \int_0^{P_{1,2}} pl^{\mathcal{P}_2}(P)dP \\ &= \left(\int_{P_{0,1}}^{P_{1,1}} pl^{\mathcal{P}_1}(P)dP + \int_{P_{1,1}}^{P_{1,2}} pl^{\mathcal{P}_1}(P)dP\right) - \int_{P_{0,2}}^{P_{1,2}} pl^{\mathcal{P}_2}(P)dP \\ &= A_{1,2} - \int_{P_{0,2}}^{P_{1,2}} pl^{\mathcal{P}_2}(P)dP. \end{aligned}$$

For $2 < j < K$ we have

$$
\begin{aligned}
m^{\mathcal{Z}}([\![1,j]\!]) &= Bel^{\mathcal{Z}}([\![1,j]\!]) - Bel^{\mathcal{Z}}([\![2,j]\!]) - Bel^{\mathcal{Z}}([\![1,j-1]\!]) + Bel^{\mathcal{Z}}([\![2,j-1]\!]) \\
&= \left(\hat{P}_j - \int_0^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP\right) \\
&\quad - \left(\hat{P}_j - \hat{P}_1 - \int_{\hat{P}_1}^{P_{1,j}} pl^{\mathcal{P}_1}(P)dP - \int_{P_{1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP\right) \\
&\quad - \left(\hat{P}_{j-1} - \int_0^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP\right) \\
&\quad + \left(\hat{P}_{j-1} - \hat{P}_1 - \int_{\hat{P}_1}^{P_{1,j-1}} pl^{\mathcal{P}_1}(P)dP - \int_{P_{1,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP\right) \\
&= \int_0^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP - \int_{P_{1,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP \\
&\quad + \int_{\hat{P}_1}^{P_{1,j}} pl^{\mathcal{P}_1}(P)dP - \int_{\hat{P}_1}^{P_{1,j-1}} pl^{\mathcal{P}_1}(P)dP \\
&\quad + \int_{P_{1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP - \int_0^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP \\
&= \left(\int_0^{P_{1,j-1}} pl^{\mathcal{P}_{j-1}}(P)dP + \int_{P_{1,j-1}}^{P_{1,j}} pl^{\mathcal{P}_1}(P)dP\right) - \int_0^{P_{1,j}} pl^{\mathcal{P}_j}(P)dP \\
&= \left(\int_{P_{0,j-1}}^{P_{1,j-1}} pl^{\mathcal{P}_{j-1}}(P)dP + \int_{P_{1,j-1}}^{P_{1,j}} pl^{\mathcal{P}_1}(P)dP\right) - \int_{P_{0,j}}^{P_{1,j}} pl^{\mathcal{P}_j}(P)dP \\
&= A_{1,j} - \int_{P_{0,j}}^{P_{1,j}} pl^{\mathcal{P}_j}(P)dP.
\end{aligned}
$$

Consider now the third case of Eq. (24) where $1 < i$ and $j = K$. For $i = j = K$ we have

$$
\begin{aligned}
m^{\mathcal{Z}}(\{K\}) &= Bel^{\mathcal{Z}}(\{K\}) \\
&= 1 - \hat{P}_{K-1} - \int_{\hat{P}_{K-1}}^{1} pl^{\mathcal{P}_{K-1}}(P)dP \\
&= (P_{K,K} - P_{K-1,K-1}) - \int_{P_{K-1,K-1}}^{P_{K-1,K}} pl^{\mathcal{P}_{K-1}}(P)dP \\
&= A_{K,K} - \int_{P_{K-1,K-1}}^{P_{K-1,K}} pl^{\mathcal{P}_{K-1}}(P)dP.
\end{aligned}
$$

For $i = j - 1 = K - 1$ we have

$$
\begin{aligned}
m^{\mathcal{Z}}(\llbracket K-1, K \rrbracket) &= Bel^{\mathcal{Z}}(\llbracket K-1, K \rrbracket) - Bel^{\mathcal{Z}}(\{K\}) - Bel^{\mathcal{Z}}(\{K-1\}) \\
&= \left(1 - \hat{P}_{K-2} - \int_{\hat{P}_{K-2}}^{1} pl^{\mathcal{P}_{K-2}}(P)dP\right) \\
&\quad - \left(1 - \hat{P}_{K-1} - \int_{\hat{P}_{K-1}}^{1} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&\quad - \left(\hat{P}_{K-1} - \hat{P}_{K-2} - \int_{\hat{P}_{K-2}}^{P_{K-2,K-1}} pl^{\mathcal{P}_{K-2}}(P)dP - \int_{P_{K-2,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&= \int_{\hat{P}_{K-1}}^{1} pl^{\mathcal{P}_{K-1}}(P)dP + \int_{P_{K-2,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP \\
&\quad + \int_{\hat{P}_{K-2}}^{P_{K-2,K-1}} pl^{\mathcal{P}_{K-2}}(P)dP - \int_{\hat{P}_{K-2}}^{1} pl^{\mathcal{P}_{K-2}}(P)dP \\
&= \left(\int_{P_{K-2,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP + \int_{\hat{P}_{K-1}}^{1} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&\quad - \int_{P_{K-2,K-1}}^{1} pl^{\mathcal{P}_{K-2}}(P)dP \\
&= \left(\int_{P_{K-2,K-1}}^{P_{K-1,K-1}} pl^{\mathcal{P}_{K-1}}(P)dP + \int_{P_{K-1,K-1}}^{P_{K-1,K}} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&\quad - \int_{P_{K-2,K-1}}^{P_{K-2,K}} pl^{\mathcal{P}_{K-2}}(P)dP \\
&= A_{K-1,K} - \int_{P_{K-2,K-1}}^{P_{K-2,K}} pl^{\mathcal{P}_{K-2}}(P)dP.
\end{aligned}
$$

For $1 < i < K - 1$ we have

$$
\begin{aligned}
m^{\mathcal{Z}}(\llbracket i, K \rrbracket) &= Bel^{\mathcal{Z}}(\llbracket i, K \rrbracket) - Bel^{\mathcal{Z}}(\llbracket i+1, K \rrbracket) - Bel^{\mathcal{Z}}(\llbracket i, K-1 \rrbracket) + Bel^{\mathcal{Z}}(\llbracket i+1, K-1 \rrbracket) \\
&= \left(1 - \hat{P}_{i-1} - \int_{\hat{P}_{i-1}}^{1} pl^{\mathcal{P}_{i-1}}(P)dP\right) \\
&\quad - \left(1 - \hat{P}_{i} - \int_{\hat{P}_{i}}^{1} pl^{\mathcal{P}_{i}}(P)dP\right) \\
&\quad - \left(\hat{P}_{K-1} - \hat{P}_{i-1} - \int_{\hat{P}_{i-1}}^{P_{i-1,K-1}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{P_{i-1,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&\quad + \left(\hat{P}_{K-1} - \hat{P}_{i} - \int_{\hat{P}_{i}}^{P_{i,K-1}} pl^{\mathcal{P}_{i}}(P)dP - \int_{P_{i,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP\right) \\
&= \int_{\hat{P}_{i}}^{1} pl^{\mathcal{P}_{i}}(P)dP - \int_{\hat{P}_{i}}^{P_{i,K-1}} pl^{\mathcal{P}_{i}}(P)dP \\
&\quad + \int_{\hat{P}_{i-1}}^{P_{i-1,K-1}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{\hat{P}_{i-1}}^{1} pl^{\mathcal{P}_{i-1}}(P)dP
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{P_{i-1,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP-\int_{P_{i,K-1}}^{\hat{P}_{K-1}} pl^{\mathcal{P}_{K-1}}(P)dP \\
&=\left(\int_{P_{i-1,K-1}}^{P_{i,K-1}} pl^{\mathcal{P}_{K-1}}(P)dP+\int_{P_{i,K-1}}^{1} pl^{\mathcal{P}_{i}}(P)dP\right)-\int_{P_{i-1,K-1}}^{1} pl^{\mathcal{P}_{i-1}}(P)dP \\
&=\left(\int_{P_{i-1,K-1}}^{P_{i,K-1}} pl^{\mathcal{P}_{K-1}}(P)dP+\int_{P_{i,K-1}}^{P_{i,K}} pl^{\mathcal{P}_{i}}(P)dP\right)-\int_{P_{i-1,K-1}}^{P_{i-1,K}} pl^{\mathcal{P}_{i-1}}(P)dP \\
&=A_{i,K}-\int_{P_{i-1,K-1}}^{P_{i-1,K}} pl^{\mathcal{P}_{i-1}}(P)dP.
\end{aligned}
$$

Finally, consider the fourth case of Eq. (24) where $1<i$ and $j<K$. For $i=j$ we have

$$
\begin{aligned}
m^{\mathcal{Z}}(\{j\}) &= Bel^{\mathcal{Z}}(\{j\}) \\
&=\left(\hat{P}_j-\hat{P}_{j-1}\right)-\int_{\hat{P}_{j-1}}^{P_{j-1,j}} pl^{\mathcal{P}_{j-1}}(P)dP-\int_{P_{j-1,j}}^{\hat{P}_{j}} pl^{\mathcal{P}_{j}}(P)dP \\
&=(P_{j,j}-P_{j-1,j-1})-\int_{P_{j-1,j-1}}^{P_{j-1,j}} pl^{\mathcal{P}_{j-1}}(P)dP-\int_{P_{j-1,j}}^{P_{j,j}} pl^{\mathcal{P}_{j}}(P)dP \\
&=A_{j,j}-\int_{P_{j-1,j-1}}^{P_{j-1,j}} pl^{\mathcal{P}_{j-1}}(P)dP-\int_{P_{j-1,j}}^{P_{j,j}} pl^{\mathcal{P}_{j}}(P)dP.
\end{aligned}
$$

For $j=i+1$ we have

$$
\begin{aligned}
m^{\mathcal{Z}}([\![j-1,j]\!]) &= Bel^{\mathcal{Z}}([\![j-1,j]\!])-Bel^{\mathcal{Z}}(\{j\})-Bel^{\mathcal{Z}}(\{j-1\}) \\
&=\left(\hat{P}_j-\hat{P}_{j-2}-\int_{\hat{P}_{j-2}}^{P_{j-2,j}} pl^{\mathcal{P}_{j-2}}(P)dP-\int_{P_{j-2,j}}^{\hat{P}_{j}} pl^{\mathcal{P}_{j}}(P)dP\right) \\
&\quad-\left(\hat{P}_j-\hat{P}_{j-1}-\int_{\hat{P}_{j-1}}^{P_{j-1,j}} pl^{\mathcal{P}_{j-1}}(P)dP-\int_{P_{j-1,j}}^{\hat{P}_{j}} pl^{\mathcal{P}_{j}}(P)dP\right) \\
&\quad-\left(\hat{P}_{j-1}-\hat{P}_{j-2}-\int_{\hat{P}_{j-2}}^{P_{j-2,j-1}} pl^{\mathcal{P}_{j-2}}(P)dP-\int_{P_{j-2,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP\right) \\
&=\int_{P_{j-1,j}}^{\hat{P}_{j}} pl^{\mathcal{P}_{j}}(P)dP-\int_{P_{j-2,j}}^{\hat{P}_{j}} pl^{\mathcal{P}_{j}}(P)dP \\
&\quad+\int_{P_{j-2,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP+\int_{\hat{P}_{j-1}}^{P_{j-1,j}} pl^{\mathcal{P}_{j-1}}(P)dP \\
&\quad+\int_{\hat{P}_{j-2}}^{P_{j-2,j-1}} pl^{\mathcal{P}_{j-2}}(P)dP-\int_{\hat{P}_{j-2}}^{P_{j-2,j}} pl^{\mathcal{P}_{j-2}}(P)dP \\
&=\left(\int_{P_{j-2,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP+\int_{\hat{P}_{j-1}}^{P_{j-1,j}} pl^{\mathcal{P}_{j-1}}(P)dP\right) \\
&\quad-\int_{P_{j-2,j-1}}^{P_{j-2,j}} pl^{\mathcal{P}_{j-2}}(P)dP-\int_{P_{j-2,j}}^{P_{j-1,j}} pl^{\mathcal{P}_{j}}(P)dP \\
&=\left(\int_{P_{j-2,j-1}}^{P_{j-1,j-1}} pl^{\mathcal{P}_{j-1}}(P)dP+\int_{P_{j-1,j-1}}^{P_{j-1,j}} pl^{\mathcal{P}_{j-1}}(P)dP\right)
\end{aligned}
$$

$$
\begin{aligned}
& - \int_{P_{j-2,j-1}}^{P_{j-2,j}} pl^{\mathcal{P}_{j-2}}(P)dP - \int_{P_{j-2,j}}^{P_{j-1,j}} pl^{\mathcal{P}_j}(P)dP \\
&= A_{j-1,j} - \int_{P_{j-2,j-1}}^{P_{j-2,j}} pl^{\mathcal{P}_{j-2}}(P)dP - \int_{P_{j-2,j}}^{P_{j-1,j}} pl^{\mathcal{P}_j}(P)dP.
\end{aligned}
$$

For $j > i+1$ we have

$$
\begin{aligned}
m^{\mathcal{Z}}(\llbracket i,j \rrbracket) &= Bel^{\mathcal{Z}}(\llbracket i,j \rrbracket) - Bel^{\mathcal{Z}}(\llbracket i+1,j \rrbracket) - Bel^{\mathcal{Z}}(\llbracket i,j-1 \rrbracket) + Bel^{\mathcal{Z}}(\llbracket i+1,j-1 \rrbracket) \\
&= \left(\hat{P}_j - \hat{P}_{i-1} - \int_{\hat{P}_{i-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{P_{i-1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP \right) \\
&\quad - \left(\hat{P}_j - \hat{P}_i - \int_{\hat{P}_i}^{P_{i,j}} pl^{\mathcal{P}_i}(P)dP - \int_{P_{i,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP \right) \\
&\quad - \left(\hat{P}_{j-1} - \hat{P}_{i-1} - \int_{\hat{P}_{i-1}}^{P_{i-1,j-1}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{P_{i-1,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP \right) \\
&\quad + \left(\hat{P}_{j-1} - \hat{P}_i - \int_{\hat{P}_i}^{P_{i,j-1}} pl^{\mathcal{P}_i}(P)dP - \int_{P_{i,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP \right) \\
&= \int_{P_{i-1,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP - \int_{P_{i,j-1}}^{\hat{P}_{j-1}} pl^{\mathcal{P}_{j-1}}(P)dP \\
&\quad + \int_{\hat{P}_i}^{P_{i,j}} pl^{\mathcal{P}_i}(P)dP - \int_{\hat{P}_i}^{P_{i,j-1}} pl^{\mathcal{P}_i}(P)dP \\
&\quad + \int_{\hat{P}_{i-1}}^{P_{i-1,j-1}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{\hat{P}_{i-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(P)dP \\
&\quad + \int_{P_{i,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP - \int_{P_{i-1,j}}^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP \\
&= \left(\int_{P_{i-1,j-1}}^{P_{i,j-1}} pl^{\mathcal{P}_{j-1}}(P)dP + \int_{P_{i,j-1}}^{P_{i,j}} pl^{\mathcal{P}_i}(P)dP \right) \\
&\quad - \int_{P_{i-1,j-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{P_{i-1,j}}^{P_{i,j}} pl^{\mathcal{P}_j}(P)dP \\
&= A_{i,j} - \int_{P_{i-1,j-1}}^{P_{i-1,j}} pl^{\mathcal{P}_{i-1}}(P)dP - \int_{P_{i-1,j}}^{P_{i,j}} pl^{\mathcal{P}_j}(P)dP.
\end{aligned}
$$

E Proof of Lemma 6

For any $\mathbf{p} \in \Pi^*$, we have

$$
\begin{aligned}
\log pl^{\Pi}(\mathbf{p}) &= \sum_{i=1}^{K} y_i (\log(p_i) - \log(\hat{p}_i)), \\
\frac{\partial \log pl^{\Pi}(\mathbf{p})}{\partial p_i} &= \frac{y_i}{p_i}, \quad \forall 1 \le i \le K,
\end{aligned}
$$

and

$$
\frac{\partial^2 \log pl^{\Pi}(\mathbf{p})}{\partial p_i \partial p_j} = \begin{cases} -y_i/p_i^2 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{49}
$$

Consider the Hessian matrix H of $\log pl^{\Pi}(\mathbf{p})$, whose i-th row and j-th column entry is given by Eq. (49). H is negative definite since, for any vector $\mathbf{r} = (r_1, \ldots, r_K)^\intercal \in \mathbb{R}^K$, $\mathbf{r} \neq \mathbf{0}$, we have

$$\mathbf{r}^\intercal H \mathbf{r} = -\sum_{i=1}^{K} \frac{y_i r_i^2}{p_i^2} < 0.$$

Hence, pl^{Π} is strictly log-concave in Π^*, which implies that its w-superlevel sets $\Lambda(w) := \{\mathbf{p} \in \Pi^* | pl^{\Pi}(\mathbf{p}) \geq w\}$, for all $w \in (0,1]$, are convex [1].

Since $T : \mathbb{R}^K \to \mathbb{R}^{K-1}$ is a linear transformation, $T(\Lambda(w))$ is convex, for any $w \in (0,1]$. In addition, for any $w \in (0,1]$, we have $T(\Lambda(w)) = \{\mathbf{P} \in \mathcal{P}^* | pl^{\mathcal{P}}(\mathbf{P}) \geq w\}$, since for $\mathbf{P} := T\mathbf{p}$, $\mathbf{p} \in \Pi^* \Leftrightarrow \mathbf{P} \in \mathcal{P}^*$ and $pl^{\Pi}(\mathbf{p}) \geq w \Leftrightarrow pl^{\mathcal{P}}(\mathbf{P}) \geq w$. Since $pl^{\mathcal{P}}$ satisfies Assumption 1 (Lemma 5), any $\mathbf{P} \in \mathcal{P}$ such that $pl^{\mathcal{P}}(\mathbf{P}) \geq w$ for some $w \in (0,1]$ belongs to $\mathcal{P}^*$. Therefore, for $w \in (0,1]$, $\Gamma(w) = \{\mathbf{P} \in \mathcal{P} | pl^{\mathcal{P}}(\mathbf{P}) \geq w\} = \{\mathbf{P} \in \mathcal{P}^* | pl^{\mathcal{P}}(\mathbf{P}) \geq w\} = T(\Lambda(w))$ and is thus convex.

F Proof of Lemma 7

Given (28), $pl^{\mathcal{P}_j}(P) = \sup_{\mathbf{P}_{-j}} pl^{\mathcal{P}}(\mathbf{P})$ is obtained by finding a $\mathbf{p}^{\star j} = (p_1^{\star j}, \ldots, p_K^{\star j}) \in \Pi$, which maximises $\mathcal{L}(\mathbf{p}; \mathbf{y})$ under the constraint $\sum_{i=1}^{j} p_i = P$. Accordingly, $pl^{\mathcal{P}_j}$ can be written as

$$\begin{aligned} pl^{\mathcal{P}_j}(P) &= \sup_{\mathbf{p} \in \Pi : \sum_{i=1}^{j} p_i = P} \frac{\mathcal{L}(\mathbf{p}; \mathbf{y})}{\mathcal{L}(\hat{\mathbf{p}}; \mathbf{y})} \\ &= \frac{\mathcal{L}(\mathbf{p}^{\star j}; \mathbf{y})}{\mathcal{L}(\hat{\mathbf{p}}; \mathbf{y})}. \end{aligned} \tag{50}$$

Maximising $\mathcal{L}(\mathbf{p}; \mathbf{y})$ is equivalent to maximising $\log \mathcal{L}(\mathbf{p}; \mathbf{y})$. Such a $\mathbf{p}^{\star j}$ can thus be obtained by considering the Lagrangian function

$$\begin{aligned} L_j(\mathbf{p}, \boldsymbol{\lambda}) &= \log \mathcal{L}(\mathbf{p}; \mathbf{y}) + \lambda_1 (\sum_{i=1}^{j} p_i - P) + \lambda_2 (\sum_{i=1}^{K} p_i - 1) \\ &= \sum_{i=1}^{K} y_i \log p_i + \log n! - \sum_{i=1}^{K} \log y_i! + \lambda_1 (\sum_{i=1}^{j} p_i - P) + \lambda_2 (\sum_{i=1}^{K} p_i - 1) \end{aligned}$$

The partial derivatives of L_j with respect to p_i for $i = 1, \ldots, K$, as well as λ_1 and λ_2, are the following:

$$\frac{\partial L_j(\mathbf{p}, \boldsymbol{\lambda})}{\partial p_i} = \frac{y_i}{p_i} + \lambda_1 \cdot I(i \leq j) + \lambda_2, \quad i = 1, \ldots, K, \tag{51}$$

$$\frac{\partial L_j(\mathbf{p}, \boldsymbol{\lambda})}{\partial \lambda_1} = \sum_{i=1}^{j} p_i - P, \tag{52}$$

$$\frac{\partial L_j(\mathbf{p}, \boldsymbol{\lambda})}{\partial \lambda_2} = \sum_{i=1}^{K} p_i - 1. \tag{53}$$

$\mathbf{p}^{\star j} = (p_1^{\star j}, \ldots, p_K^{\star j})$ is obtained when the K partial derivatives in Eq. (51) are zeros, in which case we have

$$p_i^{\star j} = \frac{-y_i}{\lambda_1 \cdot I(i \leq j) + \lambda_2}, \quad i = 1, \ldots, K. \tag{54}$$

By substituting these expressions in Eq. (52) and by imposing that the latter cancels out, we obtain

$$\lambda_1 + \lambda_2 = \frac{-\sum_{i=1}^{j} y_i}{P}. \tag{55}$$

By substituting Eq. (54) in Eq. (53) and by imposing that the latter cancels out, we obtain

$$\sum_{i=1}^{j} \frac{-y_i}{\lambda_1 + \lambda_2} + \sum_{i=j+1}^{K} \frac{-y_i}{\lambda_2} \quad = \quad 1,$$

which, thanks to Eq. (55), becomes

$$\begin{aligned} P + \sum_{i=j+1}^{K} \frac{-y_i}{\lambda_2} &= 1 \\ \Leftrightarrow \lambda_2 &= \frac{\sum_{i=j+1}^{K} y_i}{P-1}. \end{aligned} \tag{56}$$

Let $n_l^k := \sum_{i=l}^{k} y_i$ for all $1 \leq l \leq k \leq K$. By substituting both Eqs. (55) and (56), in Eq. (54), we obtain

$$p_i^{\star j} = \begin{cases} (n_i^i / n_1^j) P & \text{if } i \leq j, \\ (n_i^i / n_{j+1}^K)(1-P) & \text{else.} \end{cases}$$

Let $\mathbf{P}^{\star j} = (P_1^{\star j}, \ldots, P_{K-1}^{\star j})$ be the vector such that $P_i^{\star j} = \sum_{l=1}^{i} p_l^{\star j}$. We then have

$$P_i^{\star j} = \begin{cases} (n_1^i / n_1^j) P & \text{if } i \leq j, \\ (n_{j+1}^i / n_{j+1}^K)\,(1-P) + P & \text{else.} \end{cases} \tag{57}$$

Consequently, given Eqs. (28) and (50), we obtain

$$pl^{\mathcal{P}_j}(P) = pl^{\mathcal{P}}(\mathbf{P}^{\star j}). \tag{58}$$

From Eqs. (14), (58) and (57), we have

$$\begin{aligned} pl^{\mathcal{P}_j}(P) &= \left(\frac{P_1^{\star j}}{\hat{P}_1}\right)^{y_1} \left(\prod_{i=2}^{K-1} \left(\frac{P_i^{\star j} - P_{i-1}^{\star j}}{\hat{P}_i - \hat{P}_{i-1}}\right)^{y_i}\right) \left(\frac{1 - P_{K-1}^{\star j}}{1 - \hat{P}_{K-1}}\right)^{y_K} \\ &= \left(\frac{P_1^{\star j}}{\hat{P}_1}\right)^{y_1} \left(\prod_{i=2}^{j} \left(\frac{P_i^{\star j} - P_{i-1}^{\star j}}{\hat{P}_i - \hat{P}_{i-1}}\right)^{y_i}\right) \left(\frac{P_{j+1}^{\star} - P_j^{\star}}{\hat{P}_{j+1} - \hat{P}_j}\right)^{y_{j+1}} \left(\prod_{i=j+2}^{K-1} \left(\frac{P_i^{\star j} - P_{i-1}^{\star j}}{\hat{P}_i - \hat{P}_{i-1}}\right)^{y_i}\right) \left(\frac{1 - P_{K-1}^{\star j}}{1 - \hat{P}_{K-1}}\right)^{y_K} \\ &= \left(\frac{n_1^K P}{n_1^j}\right)^{y_1} \left(\prod_{i=2}^{j} \left(\frac{n_1^K P}{n_1^j}\right)^{y_i}\right) \left(\frac{n_1^K (1-P)}{n_{j+1}^K}\right)^{y_{j+1}} \left(\prod_{i=j+2}^{K-1} \left(\frac{n_1^K (1-P)}{n_{j+1}^K}\right)^{y_i}\right) \left(\frac{n_1^K (1-P)}{n_{j+1}^K}\right)^{y_K} \\ &= \left(\prod_{i=1}^{j} \left(\frac{n_1^K P}{n_1^j}\right)^{y_i}\right) \left(\prod_{i=j+1}^{K} \left(\frac{n_1^K (1-P)}{n_{j+1}^K}\right)^{y_j}\right) \\ &= \left(\frac{n_1^K P}{n_1^j}\right)^{n_1^j} \left(\frac{n_1^K (1-P)}{n_{j+1}^K}\right)^{n_{j+1}^K}, \end{aligned}$$

which, with $n_j := \sum_{i=1}^{j} y_i$, $1 \leq j < K$, is equal to

$$\left(\frac{nP}{n_j}\right)^{n_j} \left(\frac{n(1-P)}{n-n_j}\right)^{n-n_j} \quad = \quad \left(\frac{P}{\hat{P}_j}\right)^{n_j} \left(\frac{1-P}{1-\hat{P}_j}\right)^{n-n_j}. \tag{59}$$

We can then remark that $pl^{\mathcal{P}_j}$ is nothing other than the relative likelihood function of a binomial variable Y_j with parameters n and P_j for which we have observed n_j successes out of n experiments, where $n > n_j > 0$. It is therefore unimodal, with mode $\hat{P}_j = \sum_{i=1}^{j} y_i / n$, and continuous, and thus it satisfies Assumption 2.

G Proof of Lemma 8

Let $h(P) := \log pl^{\mathcal{P}_i}(P) - \log pl^{\mathcal{P}_j}(P)$ for $P \in (0,1)$. From (59), we obtain for all $P \in (0,1)$

$$
\begin{aligned}
h(P) &= \log\left(\left(\frac{P}{\hat{P}_i}\right)^{n_i}\left(\frac{1-P}{1-\hat{P}_i}\right)^{n-n_i}\right) - \log\left(\left(\frac{P}{\hat{P}_j}\right)^{n_j}\left(\frac{1-P}{1-\hat{P}_j}\right)^{n-n_j}\right)\\
&= \log\left(\frac{P}{\hat{P}_i}\right)^{n_i} + \log\left(\frac{1-P}{1-\hat{P}_i}\right)^{n-n_i} - \log\left(\frac{P}{\hat{P}_j}\right)^{n_j} - \log\left(\frac{1-P}{1-\hat{P}_j}\right)^{n-n_j}\\
&= n_i \log\frac{P}{\hat{P}_i} + (n-n_i)\log\frac{1-P}{1-\hat{P}_i} - n_j\log\frac{P}{\hat{P}_j} - (n-n_j)\log\frac{1-P}{1-\hat{P}_j}\\
&= n_i\log P - n_i\log\hat{P}_i + (n-n_i)\log(1-P) - (n-n_i)\log(1-\hat{P}_i) - n_j\log P + n_j\log\hat{P}_j\\
&\quad -(n-n_j)\log(1-P) + (n-n_j)\log(1-\hat{P}_j)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial h(P)}{\partial P} &= \frac{n_i}{P} - \frac{n-n_i}{1-P} - \frac{n_j}{P} + \frac{n-n_j}{1-P}\\
&= (n_i - n_j)\left(\frac{1}{P} + \frac{1}{1-P}\right)\\
&< 0,
\end{aligned}
$$

hence $h(P)$ is decreasing on $(0,1)$. In addition, $\log pl^{\mathcal{P}_i}(\hat{P}_i) = 0$ and $\log pl^{\mathcal{P}_j}(\hat{P}_i) < 0$, so that $h(\hat{P}_i) > 0$ and, similarly, we obtain $h(\hat{P}_j) < 0$. Hence there exists an unique $P_{i,j} \in (0,1)$ (satisfying $P \in (\hat{P}_i, \hat{P}_j)$) such that $h(P_{i,j}) = 0 \Leftrightarrow pl^{\mathcal{P}_i}(P_{i,j}) = pl^{\mathcal{P}_j}(P_{i,j})$, *i.e.*, Assumption 3 holds.

H Proof of Theorem 3

The unique "intersection" value $P_{i,j}$ between $pl^{\mathcal{P}_i}$ and $pl^{\mathcal{P}_j}$, $\forall 1 \le i < j < K$, is such that

$$
\begin{aligned}
& pl^{\mathcal{P}_i}(P_{i,j}) = pl^{\mathcal{P}_j}(P_{i,j})\\
\iff & \left(\frac{P_{i,j}}{\hat{P}_i}\right)^{n_i}\left(\frac{1-P_{i,j}}{1-\hat{P}_i}\right)^{n-n_i} = \left(\frac{P_{i,j}}{\hat{P}_j}\right)^{n_j}\left(\frac{1-P_{i,j}}{1-\hat{P}_j}\right)^{n-n_j}\\
\iff & \left(\frac{(\hat{P}_j)^{n_j}(1-\hat{P}_j)^{n-n_j}}{(\hat{P}_i)^{n_i}(1-\hat{P}_i)^{n-n_i}}\right)\left(\frac{(P_{i,j})^{n_i}(1-P_{i,j})^{n-n_i}}{(P_{i,j})^{n_j}(1-P_{i,j})^{n-n_j}}\right) = 1.
\end{aligned}
$$

Setting $c_{i,j} := ((\hat{P}_j)^{n_j}(1-\hat{P}_j)^{n-n_j})/((\hat{P}_i)^{n_i}(1-\hat{P}_i)^{n-n_i})$, we have

$$
\begin{aligned}
& pl^{\mathcal{P}_i}(P_{i,j}) = pl^{\mathcal{P}_j}(P_{i,j})\\
\iff & c_{i,j}(P_{i,j})^{n_i-n_j}(1-P_{i,j})^{n_j-n_i} = 1\\
\iff & \log c_{i,j} + (n_i-n_j)\log P_{i,j} + (n_j-n_i)\log(1-P_{i,j}) = 0\\
\iff & \log c_{i,j} = (n_j-n_i)(\log P_{i,j} - \log(1-P_{i,j}))\\
\iff & \log\left(\frac{P_{i,j}}{1-P_{i,j}}\right) = \frac{\log c_{i,j}}{n_j-n_i}\\
\iff & P_{i,j} = \left(\exp\left(\frac{-\log c_{i,j}}{n_j-n_i}\right)+1\right)^{-1}
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
c_{i,j} &= \frac{(\hat{P}_j)^{n_j}(1-\hat{P}_j)^{n}(1-\hat{P}_i)^{n_i}}{(1-\hat{P}_j)^{n_j}(\hat{P}_i)^{n_i}(1-\hat{P}_i)^{n}}\\
&= \left(\frac{\hat{P}_j}{1-\hat{P}_j}\right)^{n_j}\left(\frac{1-\hat{P}_i}{\hat{P}_i}\right)^{n_i}\left(\frac{1-\hat{P}_j}{1-\hat{P}_i}\right)^{n}
\end{aligned}
$$

and

$$\begin{aligned}
\exp\left(\frac{-\log c_{i,j}}{n_j - n_i}\right) &= c_{i,j}^{-1/(n_j-n_i)} \\
&= \left(\left(\frac{\hat{P}_j}{1-\hat{P}_j}\right)^{n_j}\left(\frac{1-\hat{P}_i}{\hat{P}_i}\right)^{n_i}\left(\frac{1-\hat{P}_j}{1-\hat{P}_i}\right)^{n}\right)^{-1/(n_j-n_i)} \\
&= \left(\left(\frac{\hat{P}_j}{1-\hat{P}_j}\right)^{\hat{P}_j}\left(\frac{1-\hat{P}_i}{\hat{P}_i}\right)^{\hat{P}_i}\left(\frac{1-\hat{P}_j}{1-\hat{P}_i}\right)\right)^{-1/(\hat{P}_j-\hat{P}_i)}.
\end{aligned}$$

Consequently

$$P_{i,j} = \left(\left(\left(\frac{\hat{P}_j}{1-\hat{P}_j}\right)^{\hat{P}_j}\left(\frac{1-\hat{P}_i}{\hat{P}_i}\right)^{\hat{P}_i}\left(\frac{1-\hat{P}_j}{1-\hat{P}_i}\right)\right)^{-1/(\hat{P}_j-\hat{P}_i)} + 1\right)^{-1},$$

which establishes (31).

Now, Eq. (30) is obtained from Eq. (15) as follows.

The case where $i = 1, j = K$ is immediate.

The case where $i = 1, j < K$, is obtained as follows. From (59), we obtain

$$\int_0^{\hat{P}_j} pl^{\mathcal{P}_j}(P)dP = \int_0^{\hat{P}_j} \frac{P^{n_j}(1-P)^{n-n_j}}{\hat{P}_j^{n_j}(1-\hat{P}_j)^{n-n_j}},$$

which, using (8), equals

$$\frac{\underline{B}(\hat{P}_j; n_j+1, n-n_j+1)}{\hat{P}_j^{n_j}(1-\hat{P}_j)^{n-n_j}},$$

from which the equation for the case where $i = 1, j < K$ follows.

The case where $1 < i, j = K$, is obtained similarly by noticing that, using (59) and (9),

$$\int_{\hat{P}_{i-1}}^1 pl^{\mathcal{P}_{i-1}}(P)dP = \frac{\overline{B}(\hat{P}_{i-1}; n_{i-1}+1, n-n_{i-1}+1)}{\hat{P}_{i-1}^{n_{i-1}}(1-\hat{P}_{i-1})^{n-n_{i-1}}}.$$

Finally, the last case is obtained by noticing that for all $0 \le z_1 < z_2 \le 1$, we have, for all $1 \le j < K$,

$$\begin{aligned}
\int_{z_1}^{z_2} pl^{\mathcal{P}_j}(P)dP &= \int_0^{z_2} pl^{\mathcal{P}_j}(P)dP - \int_0^{z_1} pl^{\mathcal{P}_j}(P)dP \\
&= \frac{\underline{B}(z_2; n_j+1, n-n_j+1) - \underline{B}(z_1; n_j+1, n-n_j+1)}{\hat{P}_j^{n_j}(1-\hat{P}_j)^{n-n_j}}.
\end{aligned}$$

I Proof of Theorem 4

From Proposition 1, $pl^{\mathcal{P}}$ respects Assumptions 1 and 4, and thus from Lemma 4, we have, for all $v \in [0,1]$,

$$\varphi(\Gamma(w_\alpha)), v) = [\![\ell(w_\alpha, v), u(w_\alpha, v)]\!]. \tag{60}$$

Hence, the focal sets of $Bel_W^{\mathcal{Z}}$ are intervals and therefore $Bel_W^{\mathcal{Z}}$ is characterized by $Bel_W^{\mathcal{Z}}([\![i,j]\!])$, for all $1 \le i \le j \le K$, with

$$\begin{aligned}
Bel_C^{\mathcal{Z}}([\![i,j]\!]) &= \mu(\{v \in [0,1] \,|\, \varphi(\Gamma(w_\alpha), v) \subseteq [\![i,j]\!]\}) \\
&= \mu(\{v \in [0,1] \,|\, i \le \ell(w_\alpha, v), u(w_\alpha, v) \le j\}).
\end{aligned} \tag{61}$$

Furthermore, from Proposition 1, $pl^{\mathcal{P}}$ respects Assumptions 1-3. In this case, as shown in the proof of Theorem 1, the equalities (19) and (20) hold for all $(w, v) \in [0,1]^2$, from which we obtain

$$Bel_C^{\mathcal{Z}}([\![i,j]\!]) \quad = \quad \mu(\{v \in [0,1] \,|\, i \leq \varphi(\mathbf{U}(w_\alpha), v), \varphi(\mathbf{L}(w_\alpha), v) \leq j\}), \tag{62}$$

with $\mathbf{L}(w_\alpha) = (L_1(w_\alpha), \ldots, L_{K-1}(w_\alpha))$ and $\mathbf{U}(w_\alpha) = (U_1(w_\alpha), \ldots, U_{K-1}(w_\alpha))$.

Let us first consider the case $i = 1$ and $j = K$. From Lemma 9, $\mathbf{U}(w_\alpha) \in \mathcal{P}^*$ and $\mathbf{L}(w_\alpha) \in \mathcal{P}^*$, hence, for all $v \in [0,1]$, $\varphi(\mathbf{U}(w_\alpha), v) \geq 1$ and $\varphi(\mathbf{L}(w_\alpha), v) \leq K$, and thus $Bel_C^{\mathcal{Z}}([\![1,K]\!]) = \mu(\{v \in [0,1]\}) = 1$, which gives us the first case in Eq.(36).

For $i = 1$ and $j < K$, since we have established that for all $v \in [0,1]$, $1 \leq \varphi(\mathbf{U}(w_\alpha), v)$, Eq. (62) reduces to

$$\begin{aligned}
& \mu(\{v \in [0,1] \,|\, \varphi(\mathbf{L}(w_\alpha), v) \leq j\}) \\
= \;& \mu(\{v \in [0,1] \,|\, L_j(w_\alpha) \geq v\}) \\
= \;& \int_0^{L_j(w_\alpha)} 1 dv \\
= \;& L_j(w_\alpha),
\end{aligned} \tag{63}$$

which gives us the second case in Eq. (36).

For $1 < i$ and $j = K$, since we have established that for all $v \in [0,1]$, $K \geq \varphi(\mathbf{L}(w_\alpha), v)$, Eq. (62) reduces to

$$\begin{aligned}
& \mu(\{v \in [0,1] \,|\, i \leq \varphi(\mathbf{U}(w_\alpha), v)\}) \\
= \;& \mu(\{v \in [0,1] \,|\, U_{i-1}(w_\alpha) < v)\}) \\
= \;& \int_{U_{i-1}(w)}^{1} 1 dv \\
= \;& 1 - U_{i-1}(w_\alpha),
\end{aligned}$$

which gives us the third case in Eq. (36).

For $1 < i$ and $j < K$, since for all $v \in [0,1]$, we have $1 < i \leq \varphi(\mathbf{U}(w_\alpha), v) \Leftrightarrow U_{i-1}(w_\alpha) < v$ and $\varphi(\mathbf{L}(w_\alpha), v) \leq j < K \Leftrightarrow v \leq L_j(w_\alpha)$, Eq. (62) reduces to

$$\begin{aligned}
& \mu(\{v \in [0,1] \,|\, U_{i-1}(w_\alpha) < v, L_j(w_\alpha) \geq v\}) \\
= \;& \mu(\{v \in [0,1] \,|\, U_{i-1}(w_\alpha) < v \leq L_j(w_\alpha)\})
\end{aligned} \tag{64}$$

Let $w_{i-1,j} := pl^{\mathcal{P}_{i-1}}(P_{i-1,j}) = pl^{\mathcal{P}_j}(P_{i-1,j})$. As established in Appendix B, for all $w \in [0, w_{i-1,j})$, we have $U_{i-1}(w) > L_j(w)$. Hence, for $w_\alpha < w_{i-1,j}$, Eq. (64) reduces thus to

$$\mu(\{v \in [0,1] \,|\, U_{i-1}(w_\alpha) < v \leq L_j(w_\alpha)\}) = 0. \tag{65}$$

In addition, as established in Appendix B, for all $w \in [w_{i-1,j}, 1]$, we have $U_{i-1}(w) \leq L_j(w)$. Hence, for $w_\alpha \geq w_{i-1,j}$, Eq. (64) reduces thus to

$$\mu(\{v \in [0,1] \,|\, U_{i-1}(w_\alpha) < v \leq L_j(w_\alpha)\}) = L_j(w_\alpha) - U_{i-1}(w_\alpha) \geq 0. \tag{66}$$

From (65) and (66), Eq. (64) simplifies to

$$\mu(\{v \in [0,1] \,|\, U_{i-1}(w_\alpha) < v \leq L_j(w_\alpha)\}) = \max(L_j(w_\alpha) - U_{i-1}(w_\alpha), 0),$$

which gives us the last case in Eq. (36).