Probability Aggregation

1 Basic Concept and Definition

Be an event A associated with a certain probability $P(A)$ to occurs.

1.1 Conditional Probability

The conditional probability of A given B (i.e.: knowing that B occurred) is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

1.2 Bayes Theorem

Bayes Theorem can be derive from the conditional probability

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

1.3 Joint Probability

The joint probability distribution for $A, B, \dots$ is a probability distribution that gives the probability that each of $A, B, \dots$ falls in any particular range of values specified for that variable.

$$P(A, B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

The sequential simulation algorithm is based on this equation:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \dots X_{i-1})$$

1.4 Independence

Two event are say independent if

$$P(A, B) = P(A)P(B)$$

and conditionally independent when:

$$P(D_1, \dots, D_n \mid A) = \prod_{i=1}^{n} P(D_i \mid A)$$

1.5 Odd ratios

$$O_i = \frac{P_i}{1 - P_i}, \quad 0 < P(A) < 1$$

Bordly (1982), Journel (2002) and Krishnan (2008) are using this notation. Although Allard (2012) said that back-transforming non-binary odds ratio of into probabilities using $P(\cdot) = O(\cdot)/(1 + O(\cdot))$

1.6 Kullback-Leibler (KL) divergence

KL divergence is a non-symmetric (therefore can't be qualify of distance) measure of the information lost when P_G is used to approximate P:

$$D_{\mathrm{KL}}(P \| P_G) = \int_{-\infty}^{\infty} P(A) \log \frac{P(A)}{P_G(A)} \, \mathrm{d}A = \sum_A P(A) \log \frac{P(A)}{P_G(A)}$$

It can be view as the expectation of the logarithmic difference between the probabilities.

There is a strong connection between entropy and KL divergence.

2 Aggregation of Probabilities

Aggregation targets to combine difference knowledge D_i related to a single event A :

$$\boxed{\begin{aligned} P(A \mid D_1, \ldots, D_n) &= P_G\left(P(A \mid D_1), \ldots, P(A \mid D_n)\right) \\ &= P_G\left(P_1, \ldots, P_n\right) \end{aligned}}$$

Combining the definition of the joint probability $P(A, D_1, \ldots, D_n)$ and $P(D_1, \ldots, D_n)$ we arrived to:

$$\begin{aligned} P(A \mid D_1, \ldots, D_n) &= \frac{P(A, D_1, \ldots, D_n)}{P(D_1, \ldots, D_n)} \\ &= \frac{P(A) \prod_{i=1}^{n} P(D_i \mid A_1, D_1, \ldots D_{i-1})}{P(D_1, \ldots, D_n)} \end{aligned}$$

Assuming independence of D_i conditionally to A suppress the dependence of the author $D_1, \ldots D_{i-1}$:

$$P(A \mid D_1, \ldots, D_n) = \frac{P(A) \prod_{i=1}^{n} P(D_i \mid A)}{P(D_1, \ldots, D_n)}$$

Which bring to the aggregation method:

$$P(A \mid D_1, \ldots, D_n) = P(A) \prod_{i=1}^{n} \frac{P(D_i \mid A)}{P(D_i)}$$

2.1 Properties

- **Dictatorship.** If the method has a probability P_i which overtake the others :
$$P_G(P_1, ..., P_i, ..., P_n)(A) = P_i(A)$$
- **Convexity.** If the method verify:
$$P_G \in [\min\{P_1, ..., P_n\}, \max\{P_1, ..., P_n\}]$$
- **Unanimity.** If the method verify:
$$\forall P_i = p \Rightarrow P_G = p$$
Convex method always preserve unanimity.
- **Independence Preservation.** If the method verify:ee
$$P_G(P_1, ..., P_n)(A \cap B) = P_G(P_1, ..., P_n)(A)\, P_G(P_1, ..., P_n)(B)$$
- **Marginalization.** If the method verify:
$$P_G\{M_k(P_1), ..., M_k(P_n)\} = M_k\{P_G(P_1, ..., P_n)\}$$
where,
$$M_k\{P(A)\} = P(A_k)$$
- **External Bayesianity** If the method verify:
$$P_G(P_1^L, ..., P_n^L)(A) = P_G^L(P_1, ..., P_n)(A)$$
where,
$$P_i^L(A) = \frac{L(A)P_i(A)}{\sum_A L(A)P_i(A)}$$
- **Certainty Effect** (or 0/1 forcing property). If the method verify:
$$\exists P_i | P_i(A) = 0 \text{ or } P_i(A) = 1 \Rightarrow P_G(A) = 0 \text{ or } P_G(A) = 1$$

3 Aggregation Methods

Allard et al. present table and formula to summarized all method explain below (Table 2 and 3, equation 33 in the paper).

3.1 Additive Methods (OR method)

3.1.1 (Generalized) Linear Pooling

$$P_G(A) = \sum_{i=0}^{n} w_i P_i(A), \qquad \sum_i w_i = 1$$

Properties: sub-optimal (not the best), do not preserve independence, 0/1 forcing nor Bayesianity. convex method (preserve unanimity) and marginalization (only possible aggregation method).

P_G result in a multi-modal distribution, each P_i represent a different population. This is equivalent to first sample a population P_i with weight w_i and then sample inside this distribution the event A (marginalization propertie, arithmetic average, disjunction of probabilities). No agreement (intersection) between different source is stress.

3.1.2 Beta-Transform Linear Pooling

$$P_G(A) = H_{\alpha,\beta}\left(\sum_{i=0}^{n} w_i P_i(A)\right)$$

Again, $\sum_i w_i = 1$, and $H_{\alpha,\beta}$ is the cumulative density function of a beta distribution:

$$H_{\alpha,\beta}(x) = B(\alpha, \beta)^{-1} \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$$

where $x \in [0, 1]$, $\alpha > 0$, $\beta > 0$ and $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$

The special case $\beta = \alpha = 1$ simplify to the linear pooling.

Properties: loose marginalization, non-convex aggr. prob., can be show to be better than LP (why ?)

3.2 Multiplicative Methods (AND method)

3.2.1 (Generalized) Log-linear Pooling

$$P_G(A) \propto P_0(A)^{1-sumw_i} \prod_{i=1}^{n} P_i(A)^{w_i}$$

Properties: verify external Bayesian (only possible aggregation method), preserve unanimity and 0/1 forcing but do not preserve independence nor marginalization.

If we don't include the prior and put each weight equal to 1 ($w_i = 1$), this corresponds to the conjunction of probabilities. If the sum of weight is not equal to one, all properties are lost.

This can be re-written with the conditional prob.:

$$\begin{aligned} P(A \mid D_1, \ldots, D_n) &\propto P_0(A) P(D_1, \ldots, D_n \mid A) \\ &\propto P_0(A)^{1-sumw_i} \prod_{i=1}^{n} P(A \mid D_i)^{w_{a,D_1,\ldots D_n}} \end{aligned}$$

Therefore, the decomposition is exact if there is one weight w_i per decomposition $(A, D_i, \dots D_n)$. Log-linear pooling make the assumption that the weight is constant.

The sum of weight S_w plays an important role with regards to the prior influence: If $\sum w_i = 1$, P_0 is filtered out; if $S_w > 1$, P_G will be closer to P_i than P_0

3.2.2 (Generalized) Logarithmic Pooling

$$P_G(A) \propto H(A) \prod_{i=1}^{n} P(A \mid D_i)^{w_i}$$

GLP allows P_G to depend upon A where again $\sum_i w_i = 1$, and $H(A)$ is the an arbitrary bounded function playing the role of likelihood on the elements of A.

3.2.3 Maximum Entropy Approach

The Kullback-Leiber divergence can be used as an objectif function to maximized in order to find the best P_G similar to P. Allard et al. (2012) showed that its is equivalent to conditional independence and is the special case of log-linear pooling for weight equalt to 1.

3.3 Multiplication of Odds Methods

Allard et al. (2012) emphasize that most of the literature is concerned about binary case and that odds methods are equivalent to probabilities method.

3.3.1 Bordley formula

For binary case, Bordley (1982) publish an aggregation technique based on the product of odd ratios:

3.3.2 Tau model

$$\frac{O}{O_0} = \prod_{i=0}^{n} \left(\frac{O_i}{O_0} \right)^{\tau_i}$$

with $O = O(A \mid D_1, \dots, D_n)$, $O_0 = O(A)$ and $O_i = O(A \mid D_i)$. Journel (2002) interpreted O_i as distances to the unknown A. The model verify convexity and any value τ_i measure the redundancy of the informations D_i with the informations $D_{1,\dots i-1}$ already used.

3.3.3 Nu-model