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 $A, B, \ldots$ is a probability distribution that gives the probability that each of $A, B, \ldots$ 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:

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

1.4 Kullback-Leibler (KL) divergence

KL divergence is a non-symmetric 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.

2 Aggregation

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

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

3 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:

$$P_G(P_1, ..., P_n)(A \cap B) = P_G(P_1, ..., P_n)(A)\, P_G(P_1, ..., P_n)(B)$$

- **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:

$$\exists P_i | P_i(A) = 0 \text{ or } P_i(A) = 1 \Rightarrow P_G(A) = 0 \text{ or } P_G(A) = 1$$

4 Aggregation Methods

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

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

4.1 Conditional independence

Two event are say conditionally independent if

$$P(X_1, \ldots, X_n \mid A) = \prod_i P(X_i \mid A)$$

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

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

4.2 Full independence

If D_i are assumed to be independent, we get:

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

Which bring to the aggregation method:

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

4.3 Additive method

- **(Generalized) Linear Pooling.**

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

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

  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. No agreement between different source is stress.

- **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 ?)

4.4 Multiplicative method

- **(Generalized) Log-linear Pooling.**

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

  where w_i is the weight associated to each P_i and $S_w = \sum_i w_i$. Properties: verify Bayesianity, preserve unanimity and 0/1 forcing but do not preserve independence nor marginalization.

  If $w_i = 1 \forall i \neq 0$, this corresponds to the *conjunction of probabilities.*

  This is equivalent to a Bayesian notation of conditional propability:

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

  Therefore, the decomposition is exact if there is one weight w per combinaison $(A, D_i, \ldots D_n)$. 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 $S_w = 1$, P_0 is filtered out; if $S_w > 1$, P_G will be closer to P_i than P_0

- **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.

- Tau model
- Bordley formula
- Nu-model
- ...