JEFFREY CONDITIONING AND EXTERNAL BAYESIANITY

CARL WAGNER
MATHEMATICS DEPT.
UNIVERSITY OF TENNESSEE

FEW 2008 – MADISON

1. Combining Probability Distributions

Ω = countable set of possible states of the world

$p: \Omega \rightarrow [0,1]$ is a *probability mass function*

(henceforth, pmf) iff $\sum_{\omega \in \Omega} p(\omega) = 1$

Each pmf p on Ω induces a *probability measure* (also denoted by p) on 2^{Ω} by

$$p(E) := \sum_{\omega \in E} p(\omega)$$

$\Delta :=$ set of all pmfs on Ω ; $\Delta^n :=$ the n-fold Cartesian product of Δ.

$$\Delta^{n+} := \{ (p_1, \ldots, p_n) \in \Delta^n : \cap_i \mathrm{Supp}(p_i) \neq \emptyset \},$$

where $\mathrm{Supp}(p_i) := \{\omega \in \Omega : p_i(\omega) > 0\}$

A *pooling operator* is any map $T: \Delta^{n+} \rightarrow \Delta$. In what follows, we think of $T(p_1,\ldots, p_n)$ as a compromise between n individuals who have assessed pmfs $p_1,\ldots, p_n$ over Ω. Several other interpretations are possible.

2. Pooling and (strict) conditioning

Given n individuals with priors $p_1,\ldots, p_n$, suppose each individual discovers that the true state of the world belongs to the subset E of Ω, but learns nothing that would change the prior odds between any two states in E. Their goal is a compromise pmf that takes this discovery into account .

Should they

1. update their individual priors p_i to posteriors q_i by conditioning on E, then pool the q_i using pooling operator T ; or

2. pool their priors to a compromise prior using T, and then update the result by conditioning on E ?

If either procedure always results in the same final distribution, we say T *commutes with conditioning* (CC). Formally expressed,

CC: For all subsets E of Ω and all $(p_1,\ldots, p_n) \in \Delta^{n+}$ such that $p_i(E) > 0$ for all i, and $(p_1(.|E),\ldots, p_n(.|E)) \in \Delta^{n+}$, it is the case that

(2.1) $T(p_1,\ldots, p_n)(E) > 0$, and

(2.2) $T(p_1(.|E),\ldots, p_n(.|E)) = T(p_1,\ldots, p_n)(.|E)$

"Update-then-pool = pool-then-update"

Remark: Dictatorial pooling (for some fixed i, and all $(p_1,\ldots, p_n) \in \Delta^{n+}$, $T(p_1,\ldots, p_n) = p_i$) satisfies CC, but weighted arithmetic means do not in general furnish pooling operators that commute with conditioning.

3. External Bayesianity

● A function $\lambda : \Omega \to \mathbb{R}^+$ is a *likelihood for* a given $(p_1,\ldots, p_n) \in \Delta^{n+}$ iff, for all i,

$$(3.1) \quad 0 < \sum_{\omega \in \Omega} \lambda(\omega)p_i(\omega) < \text{infinity}$$

and $(q_1,\ldots, q_n) \in \Delta^{n+}$, where

$$(3.2) \quad q_i(\omega) := \lambda(\omega)p_i(\omega) / \sum_{\omega \in \Omega} \lambda(\omega)p_i(\omega)$$

Example: If $p_i(E) > 0$ for all i, then $\lambda(\omega) = [\omega \in E]$, the indicator function of E, is a likelihood for $(p_1,\ldots, p_n)$, with $q_i(\omega) = p_i(\omega|E)$, since $\sum_{\omega \in \Omega} \lambda(\omega)p_i(\omega) = p_i(E)$.

- $T: \Delta^{n+} \rightarrow \Delta$ is *externally Bayesian*

(Madansky 1964, Genest, McConway, Schervish 1986) iff the following condition holds:

EB: If $(p_1,\ldots, p_n) \in \Delta^{n+}$ and λ is a likelihood for $(p_1,\ldots, p_n)$, then

(3.3) $0 < \sum_{\omega \in \Omega} \lambda(\omega)T(p_1,\ldots, p_n)(\omega) <$ infinity

and the following commutativity property holds:

(3.4) $T(\lambda\, p_1 / \sum_{\omega \in \Omega} \lambda(\omega)\, p_1(\omega),\ldots,$

$\lambda\, p_n / \sum_{\omega \in \Omega} \lambda(\omega)\, p_n(\omega))$

$= \lambda\, T(p_1,\ldots, p_n) / \sum_{\omega \in \Omega} \lambda(\omega)T(p_1,\ldots, p_n)(\omega)$

"Update-then-pool = pool-then-update"

Theorem 1. (Hammond) Let $w(1),\ldots,w(n)$ be a sequence of nonnegative real numbers summing to 1. Define $T: \Delta^{n+} \rightarrow \Delta$ by

(3.5) $T(p_1,\ldots, p_n)(\omega) :=$

$$\Pi_{1\leq i \leq n}\, p_i(\omega)^{w(i)} / \sum_{\omega \in \Omega} \Pi_{1\leq i \leq n}\, p_i(\omega)^{w(i)},$$

where $0^0 := 1$. Then T is EB.

A complete characterization of EB pooling operators is given in Genest, et al (1986).

Theorem 2. EB implies CC.

Beweis: klar (let $\lambda(\omega) = [\omega \in E]$).

Question: Do EB pooling operators commute with probability revision by Jeffrey conditioning ? It depends….

4. Jeffrey Conditioning (JC)

- p, q : pmfs on countable set Ω

- $\mathbf{E} = \{ E_k \}$: a family of nonempty, pairwise disjoint subsets of Ω, with $p(E_k) > 0$ for all k

- q comes from p by Jeffrey conditioning on $\mathbf{E}$ iff there exists a sequence (e_k) of positive real numbers summing to 1 such that, for every $\omega \in \Omega$,

$$q(\omega) = \sum_k e_k \, p(\omega|E_k) \tag{4.1}$$

$$= \sum_k e_k \, p(\omega)[\omega \in E_k] / p(E_k).$$

- q is the appropriate revision of p in the light of new evidence iff ……

(1) based on the total evidence, old as well as new, you judge that, for each k, the posterior probability $q(E_k)$ should take the value e_k ; and

(2) for each k, you judge that nothing in the new evidence should alter the odds between any two states of the world in E_k .

Remark. When $\mathbf{E} = \{E\}$, JC reduces to ordinary conditioning.

Remark. If p and q are *any* pmfs on the countable set Ω such that
$q(\omega) > 0 \Rightarrow p(\omega) > 0$, then q comes from p by JC on $\mathbf{E} = \{ \{\omega\} : q(\omega) > 0 \}$, i.e., JC includes the case of total reassessment of a discrete distribution, as long as "zeros are not raised."

5. Alternative Parameterizations of JC

● Recall: If q is a revision of the probability measure p, and A and B are events, the *relevance quotient* $\rho_{q,p}(A)$ is the ratio

(5.1) $\rho_{q,p}(A) := q(A) / p(A)$

of new to old probabilities, and the *Bayes factor* $\beta_{q,p}(A:B)$ is the ratio

(5.2) $\beta_{q,p}(A:B) := q(A)/q(B) / p(A)/p(B)$

of new to old odds.

Remark: $\beta_{q,p}(A:B) = \rho_{q,p}(A) / \rho_{q,p}(B)$

Remark: When $q = p(.|E)$, then

$\beta_{q,p}(A:B) = p(E|A)/p(E|B)$ (likelihood ratio).

Theorem 3. Suppose that q comes from p by JC on **E**. Then

$$(5.3)\quad q(\omega) = \textstyle\sum_k\, e_k\, p(\omega|E_k) \qquad (\, e_k = q(E_k)\,)$$

$$(5.4)\qquad = \textstyle\sum_k\, r_k\, p(\omega)[\,\omega \in E_k]$$

$$(\, r_k = \rho_{q,p}(E_k)\,)$$

$$(5.5)\qquad = \textstyle\sum_k\, b_k\, p(\omega)[\,\omega \in E_k] \,/\, \sum_k\, b_k\, p(E_k)$$

$$(\, b_k = \beta_{q,p}\,(E_k\!: E_1)\,).$$

Question: Which parameterization is the most promising candidate to make JC commute with externally Bayesian pooling?

Strategy: Suppose $P \neq p$ is another pmf for which $P(E_k) > 0$ for all k. How should P be revised in light of new learning identical to that which prompted the revision of p to q?

Answer: Revise P to, call it Q, by the analogue of (5.5), i.e., let

$$Q(\omega) = \sum_k b_k P(\omega)[\omega \in E_k] / \sum_k b_k P(E_k),$$

with $b_k = \beta_{q,p}(E_k : E_1)$

Justification: Only the Bayes factors b_k capture what is learned from new evidence alone. The quantities e_k and r_k fail to efface all traces of the prior.

- A pooling operator T commutes with Jeffrey conditioning (CJC) iff the following condition holds:

CJC: For all families **E** = $\{ E_k \}$ of nonempty, pairwise disjoint subsets of Ω, all $(p_1,\ldots, p_n) \in \Delta^{n+}$ such that $p_i(E_k) > 0$ for all i and all k, and all sequences (b_k) of positive real numbers such that $b_1 = 1$ and

(5.6) $\sum_k b_k\, p_i(E_k) <$ infinity, $i = 1,..,n$

and such that $(q_1,\ldots, q_n) \in \Delta^{n+}$, where

(5.7) $q_i(\omega) =$

$$\sum_k b_k\, p_i(\omega)[\omega \in E_k] \,/\, \sum_k b_k\, p_i(E_k),$$

it is the case that

(5.8) $0 < \sum_k b_k\, T(p_1,\ldots, p_n)(E_k) <$ infinity

and

(5.9) $T(q_1,\ldots, q_n)(\omega) =$

$$\frac{\sum_k b_k T(p_1,\ldots, p_n) [\omega \in E_k]}{\sum_k b_k T(p_1,\ldots, p_n)(E_k)}$$

"Update-then-pool = pool-then-update"

Theorem 4. A pooling operator $T\colon \Delta^{n+} \to \Delta$ is externally Bayesian if and only if it commutes with Jeffrey conditioning in the sense of CJC.

Proof. Necessity. Take

$$\lambda(\omega) = \sum_k b_k [\omega \in E_k] .$$

Sufficiency. Let $(\omega_1,\omega_2,\ldots)$ be a list of all those $\omega \in \Omega$ with $\lambda(\omega) > 0$, let $\mathbf{E} = \{ \{\omega_1\},\{\omega_2\}, \ldots\}$ and let $b_k = \lambda(\omega_k) / \lambda(\omega_1)$

Conclusion: Theorem 4 provides

(1) a salient reformulation of external Bayesianity in terms of Jeffrey conditionalization, a probability revision method familiar to philosophers ; and

(2) further support for the thesis that identical new learning should be reflected in identical Bayes factors.

(see also “Probability Kinematics and Commutativity,” *Philosophy of Science* **69**(2002), 266-278.