A unified representation of conditioning rules for convex capacities

Mayumi Horie

Part-time Lecturer, Osaka Univesity of Economics

Abstract

This paper proposes a unified representation, called the G-updating rule, which includes three conditioning rules as special cases, the naïve Bayes rule, the Dempster-Shafer rule (Shafer(1976)), and the generalized Bayes' updating rule introduced by Dempster(1967) or Fagin and Halpern(1991). It is shown that the G-updating rule constitutes a three-step conditioning, where one of the three rules is applied in each step.

I wish to thank Atsushi Kajii for invaluable comments and discussions to bring a good deal of enlightenment. Remaining errors or shortcomings in this paper are, of course, my own responsibility.

Citation: Horie, Mayumi, (2006) "A unified representation of conditioning rules for convex capacities." *Economics Bulletin,* Vol. 4, No. 19 pp. 1-6

Submitted: April 15, 2006. **Accepted:** May 23, 2006.

URL: http://economicsbulletin.vanderbilt.edu/2006/volume4/EB-06D80005A.pdf

1 Introduction

Various update rules for non-additive probabilities have been proposed. In the economic literature, one of the most prevalent updating rules for non-additive measures is the Dempster-Shafer rule(Shafer(1976)) (the DS rule). This rule for convex capacities is examined in Denneberg(1994), which pointed out that the naïve Bayes rule (the NB rule) and the DS rule are two sides of the same coin. These rules are also examined in Gilboa and Schmeidler(1993). They showed that the DS rule and the naïve Bayes' rule (the NB rule) are characterized by way of an elegant, systematic method, the *f-Bayesian update rule*, and the DS rule is equivalent to the maximum likelihood estimation.

Fagin and Halpern(1991) presented an update rule (the FH rule) for inner/outer measures or belief/plausibility functions, which was already suggested by Dempster(1967). As shown in Wasserman and Kadane(1990), the FH rule is also applicable to convex capacities. The FH rule, which is also called the generalized Bayes' rule and *does indeed generalize Bayes' rule of conditioning* (Walley(1991)), inherits the nature of the traditional Bayes' rule for probability measures.

Our goal is to represent aforementioned three conditioning rules discussed independently in their own context, in a unified, systematic framework, called the *G-updating rule*. This rule includes the three rules as special cases. An explanatory strength of the G-updating rule is that it enables us to deal with apparently different conditioning rules as a single rule with each different parameter G, which takes the form of an ordered triplet of the global states. It is shown that our G-update rule constitutes a three-step conditioning, where one of the foregoing three rules is applied in each step. In a behavioral sense, this G can be considered as a decision maker's *a priori* trichotomy of plausibility to the occurrence of every state.

2 The Model and Main Results

Let Ω be a finite set of *states* and let Σ be an algebra, $\Sigma = 2^{\Omega}$. A non-empty set in Σ is called an *event*. A *capacity* on Ω is a set function $\mu : \Sigma \to [0,1]$ satisfying (i) $\mu(\varnothing) = 0$, $\mu(\Omega) = 1$ and (ii) monotonicity: for every A and B in Σ such that $A \subset B$, we have $\mu(A) \leq \mu(B)$. A capacity μ is *convex* if for every A and B in Σ, $\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B)$. Given an event E in Σ, a *conditional*, or *updated capacity* μ_E is a capacity on E, i.e. for every $A \in \Sigma$ such that $A \cap E = E$, $\mu_E(A) = 1$. For any E in Σ, μ_E has domain Σ. When $E = \Omega$, μ_Ω is interpreted as the *unconditional capacity* and we simply write it μ.

Suppose the set of states Ω is partitioned into three disjoint sets G_i, $i = 1,2,3$, where some G_i is possibly empty. Let us denote an ordered triplet of G_i, $i = 1,2,3$ by $G = \langle G_1, G_2, G_3 \rangle$ and let $\mathcal{G}$ consist of all such ordered triplets of Ω. For a $G \in \mathcal{G}$ and an event $E \in \Sigma$, define $T_i^{G,E} \equiv E^c \cap G_i$, $i = 1,2,3$.[1] Although every $T_i^{G,E}$ depends on G and E, we denote it by T_i, $i = 1,2,3$ instead of $T_i^{G,E}$ for brevity's sake.

Given a $G \in \mathcal{G}$ and an $E \in \Sigma$, we define the *G-updating rule for a capacity μ given E*

[1] Throughout this paper, the complement of any set is done *with respect to* Ω.

through, for every $A \in \Sigma$

$$\mu_E^G(A) = \frac{\mu((A\cap E)\cup T_2) - \mu(T_2)}{[\mu((A\cap E)\cup T_2) - \mu(T_2)] + [\mu(E\cup T_2\cup T_3) - \mu((A\cap E)\cup T_2\cup T_3)]}. \tag{1}$$

The updated μ_E^G has domain Σ.

Although the G-updating rule (1) is applicable to any capacity, we focus on the case where μ is a convex capacity throughout this paper. When μ is a convex capacity, μ_E^G is well-defined if E is such that $\mu(E\cup T_2) - \mu(T_2) > 0$. This is verified in the following way. By the convexity of μ, we have

$$\mu(E\cup T_2\cup T_3) - \mu((A\cap E)\cup T_2\cup T_3) \geq \mu(E\cup T_2) - \mu((A\cap E)\cup T_2) \tag{2}$$

Adding $\mu((A\cap E)\cup T_2) - \mu(T_2)$ (≥ 0) to both sides of (2), we have

$$\begin{aligned} & [\mu((A\cap E)\cup T_2) - \mu(T_2)] + [\mu(E\cup T_2\cup T_3) - \mu((A\cap E)\cup T_2\cup T_3)] \\ \geq\ & \mu(E\cup T_2) - \mu(T_2) \end{aligned} \tag{3}$$

It follows that, if $\mu(E\cup T_2) - \mu(T_2) > 0$, then the denominator in (1) is also strictly positive.

At a first glance, (1) appears very complicated. But a crucial advantage of the G-updating rule is that it provides a general form including three update rules, the naïve Bayes rule (the NB rule), the Dempster-Safer rule (the DS rule), and the Fagin-Halpern rule (the FH rule) as special cases. To see this, first recall that the updated capacity by the NB rule, μ_E^{NB}, is defined as, for any event $E \in \Sigma$ such that $\mu(E) > 0$,

$$\mu_E^{NB}(A) = \frac{\mu(A\cap E)}{\mu(E)} \text{ for every } A \in \Sigma. \tag{4}$$

When $G = \langle \Omega, \varnothing, \varnothing \rangle$, thus $T_1 = E^c$ and $T_2 = T_3 = \varnothing$, we see that formulation (1) is equal to (4).

Next, the DS conditional capacity μ_E^{DS} is defined through, for any $E \in \Sigma$ with $1 - \mu(E^c) > 0$,

$$\mu_E^{DS}(A) = \frac{\mu((A\cap E)\cup E^c) - \mu(E^c)}{1 - \mu(E^c)} \text{ for every } A \in \Sigma. \tag{5}$$

It is also verified that, when $G = \langle \varnothing, \Omega, \varnothing \rangle$ (i.e. $T_2 = E^c$ and $T_1 = T_3 = \varnothing$), (1) is simplified into the DS conditional capacity (5).

Furthermore, the conditional capacity updated by the FH rule is given as

$$\mu_E^{FH}(A) = \frac{\mu(A\cap E)}{\mu(A\cap E) + 1 - \mu(A\cup E^c)} \text{ for every } A \in \Sigma. \tag{6}$$

When $G = \langle \varnothing, \varnothing, \Omega \rangle$ (i.e. $T_3 = E^c$ and $T_1 = T_2 = \varnothing$), formula (1) is reduced to (6). As in Denneberg(1994), it can happen that $\mu_E^{FH}(A)$ is not defined for some pair A and E. However, as shown above, μ_E^{FH} is well-defined for any $A \in \Sigma$ if E satisfies $\mu(E) > 0$, which is also pointed out in Fagin and Halpern(1991) for a belief function case.

Of course, (1) yields the conditional probability through the traditional Bayes' rule if μ

is additive.

In general, a G-updating rule is not reduced to any of the rules above, but it can be interpreted as a three-step conditioning after an event E has occurred, where in each step one of the rules above is applied. We shall elaborate this in the followings.

Suppose first that an event E with $\mu(E \cup T_2) - \mu(T_2) > 0$ was observed. Let S^1, S^2 be sets of states such that $S^1 = E \cup T_2 \cup T_3$ and $S^2 = E \cup T_3$ respectively.

The first step A decision maker conforms an initial capacity μ to be adjusted as if μ is updated by the NB rule (4) given event S^1. Consider the following revised capacity $\mu^1 : \Sigma \to [0,1]$

$$\mu^1(A) = \frac{\mu(A \cap S^1)}{\mu(S^1)} \text{ for every } A \in \Sigma. \tag{7}$$

It is well-defined since $\mu(S^1) \geq \mu(E \cup T_2) - \mu(T_2) > 0$. To interpret, imagine that G_1 consists of those states which are conceivably possible *a priori*, and so $T_1 = G_1 \cap E^c$ is the set of *impossible* states given E. Thus, a decision maker is sufficiently confident to remove T_1 from Ω. Therefore the support of an initial capacity μ is condensed from Ω into S^1 in this step.

The second step In this stage, the adjusted capacity μ^1 is revised again as follows:

$$\mu^2(A) = \frac{\mu^1((A \cap S^2) \cup T_2) - \mu^1(T_2)}{1 - \mu^1(T_2)} \text{ for every } A \in \Sigma. \tag{8}$$

The expression above is equivalent to the DS updating (5) for μ^1 given event S^2 and it is well-defined for every $A \in \Sigma$, since $1 - \mu^1(T_2) \geq \mu(S_1) - \mu(T_2) \geq \mu(E \cup T_2) - \mu(T_2) > 0$. The construction made by this formula is that G_2 is a set of plausible states *a priori*, thus the states in $T_2 = G_2 \cap E^c$ are *not plausible* given E. A decision maker is quite confident that those states were never concurrent with E, however T_3 might be. Hence, revisions in μ^1 are made as if S^2 is observed.

The final step The final step is conducted through the FH rule (6) for μ^2 given E:

$$\mu^3(A) = \frac{\mu^2(A \cap E)}{\mu^2(A \cap E) + 1 - \mu^2(S^2 \backslash (E \backslash A))} \text{ for every } A \in \Sigma. \tag{9}$$

The revised capacity μ^2 in the second stage is finally refined to remove T_3 from E, in the manner of the FH conditioning. It is also well-defined since $\mu^2(E) \geq \mu^1(E \cup T_2) - \mu^1(T_2) \geq \mu(E \cup T_2) - \mu(T_2) > 0$. In our context, G_3 can be viewed as observational states *a priori*: in other words, revisions in G_3 are carried out according to observations. After all, $T_3 = G_3 \cap E^c$ is constituted of states that were *not observed*, so as to be eliminated necessary to this end.

To interpret, it helps to regard that each T_i, $i = 1, 2, 3$ represents information revised in the ith step. In fact, $T_i = \varnothing$ implies that there is no further information revised in that step. Hence we have $\mu^1 = \mu$ if $T_1 = \varnothing$, $\mu^2 = \mu^1$ if $T_2 = \varnothing$, and so forth. By our earlier discussion, we know that $G = \langle \Omega, \varnothing, \varnothing \rangle$ induces μ_E^{NB}. In the light of the three-step updating above,

$T_1 = E^c$ (i.e. $T_2 = T_3 = \varnothing$) implies that $\mu^1 = \mu_E^{NB}$, $\mu^2 = \mu^1$, and $\mu^3 = \mu^2$, hence we have $\mu^3 = \mu_E^{NB}$.

The following lemma proves that the three-step conditioning above does indeed work as claimed.

Lemma 2.1 *Let μ be a convex capacity on Σ. Then, for every G in $\mathcal{G}$ and E in Σ with $\mu(E \cup T_2) - \mu(T_2) > 0$, we have $\mu_E^G = \mu^3$.*

Proof. Choose arbitrary $G \in \mathcal{G}$ and $E \in \Sigma$ with $\mu(E \cup T_2) - \mu(T_2) > 0$. We have to show that for every $A \in \Sigma$, $\mu_E^G(A) = \mu^3(A)$. Since E, T_2 and T_3 are disjoint, for any $A \in \Sigma$ we have

$$\begin{aligned} A \cap E &= (A \cap E) \cap (E \cup T_3), \\ (A \cap E) \cup T_2 \cup T_3 &= ((A \cap E) \cup T_3) \cup T_2 \\ &= ((A \cup T_3) \cap (E \cup T_3)) \cup T_2, \text{ and} \\ A \cup T_3 &= (E \cup T_3) \setminus (E \backslash A). \end{aligned}$$

Then

$$\begin{aligned} \mu_E^G(A) &= \frac{\mu((A \cap E) \cup T_2) - \mu(T_2)}{[\mu((A \cap E) \cup T_2) - \mu(T_2)] + [\mu(E \cup T_2 \cup T_3) - \mu((A \cap E) \cup T_2 \cup T_3)]} \\ &= \frac{\frac{\mu\left(((A\cap E)\cup T_2)\cap S^1\right)}{\mu(S^1)} - \frac{\mu\left(T_2 \cap S^1\right)}{\mu(S^1)}}{\left[\frac{\mu(((A\cap E)\cup T_2)\cap S^1)}{\mu(S^1)} - \frac{\mu(T_2\cap S^1)}{\mu(S^1)}\right] + \left[1 - \frac{\mu(((A\cap E)\cup T_2 \cup T_3)\cap S^1)}{\mu(S^1)}\right]} \\ &= \frac{\mu^1((A\cap E)\cup T_2) - \mu^1(T_2)}{[\mu^1((A\cap E)\cup T_2) - \mu^1(T_2)] + [1 - \mu^1(T_2) - \mu^1((A\cap E)\cup T_2\cup T_3) + \mu^1(T_2)]} && \text{(by (7))} \\ &= \frac{\frac{\mu^1\left(\left((A\cap E)\cap S^2\right)\cup T_2\right) - \mu^1(T_2)}{1-\mu^1(T_2)}}{\frac{\mu^1(((A\cap E)\cap S^2)\cup T_2) - \mu^1(T_2)}{1-\mu^1(T_2)} + \left[1 - \frac{\mu^1(((A\cup T_3)\cap S^2)\cup T_2) - \mu^1(T_2)}{1-\mu^1(T_2)}\right]} \\ &= \frac{\mu^2(A \cap E)}{\mu^2(A \cap E) + 1 - \mu^2(A \cup T_3)} && \text{(by (8))} \\ &= \frac{\mu^2(A \cap E)}{\mu^2(A \cap E) + 1 - \mu^2(S^2 \backslash (E \backslash A))} \\ &= \mu^3(A). && \text{(by (9))} \end{aligned}$$

■

The following theorem shows that the G-updating preserves convexity.

Theorem 2.1 *Let μ be a convex capacity on Σ. Then, for every $G \in \mathcal{G}$ and $E \in \Sigma$ with $\mu(E \cup T_2) - \mu(T_2) > 0$, μ_E^G defined in (1) is a convex capacity on Σ.*

Proof. By Lemma 2.1 above, it is suffice to show that μ^3 is a convex capacity on Σ. By the assumption that μ is a convex capacity, it is straightforward to see that μ^1 is a convex capacity, therefore μ^2 is also a convex capacity since both rules give affine transformations of μ. Furthermore, the convexity is preserved under the FH updating, which is the well-known result proved in Walley(1991), Sundberg and Wagner(1992), Chateauneuf and Jaffray(1991), or Denneberg(1994). Hence μ^3 is a convex capacity on Σ. ■

From Theorem 2.1, it enables us to represent a specific form of the conditional capacity μ_E^G in terms of a parameter G. To explore implications of this update rule, it is natural to see how μ_E^G varies with G. The following proposition shows that μ_E^G dominates μ_E^{FH}.

Proposition 2.2 *Let μ be a convex capacity. Then, for every G in $\mathcal{G}$ and E in Σ with $\mu(E \cup T_2) - \mu(T_2) > 0$, $\mu_E^{FH} \leq \mu_E^G$.*

Proof. We have to show that for any $A \in \Sigma$, $\mu_E^{FH}(A) \leq \mu_E^G(A)$. From the convexity of μ, we have for any $A \in \Sigma$ and $T_2, T_3 \subset E^c$

$$\begin{aligned} \mu(A \cap E) &\leq \mu((A \cap E) \cup T_2) - \mu(T_2) \text{ and} \\ 1 - \mu(A \cup E^c) &\geq \mu(E \cup T_2 \cup T_3) - \mu((A \cap E) \cup T_2 \cup T_3). \end{aligned}$$

Then

$$\begin{aligned} &\frac{\mu(A \cap E)}{\mu(A \cap E) + 1 - \mu(A \cup E^c)} \\ \leq\ &\frac{\mu((A \cap E) \cup T_2) - \mu(T_2)}{[\mu((A \cap E) \cup T_2) - \mu(T_2)] + [\mu(E \cup T_2 \cup T_3) - \mu((A \cap E) \cup T_2 \cup T_3)]}, \end{aligned}$$

since $\frac{a}{a+b}$ is increasing in a and decreasing in b. ■

3 Concluding Remarks

We introduced and investigated the G-updating rule. This G is a decision maker's *a priori* prescription for revising information. As we pointed out, Ω is divided into three sets according to the degree of intensity of plausibility, and we interpret the three-step conditioning to be implemented in an ascending order of intensity, i.e. $T_1 \to T_2 \to T_3$. In fact, we still obtain the same updated capacity μ_E^G even if the first step and second step are interchanged. That is, the first and second step are mutually commutative as long as step 3 is conducted by the FH rule. It is quite natural because the NB and the DS rule are commutative update rules as seen in Gilboa and Schmeidler(1993). In our context, G_3 consists of observational states, so it stands to reason that those states are revised in the closing step. We may, of course, replace step 1 or 2 by step 3, and obtain a profoundly different conditioning rule, whose property is the matter for future investigation.

References

Dempster, A. P. (1967) "Upper and Lower Probabilities Induced by a Multivalued Mapping" *Annals of Mathematical Statistics* **38**, 205–247.

Denneberg D. (1994) "Conditioning (Updating) Non-Additive Measures" *Annals of Operations Research* **52**, 21–42.

Fagin, R., and J. Y. Halpern (1991) "A New Approach to Updating Beliefs" in *Uncertainty in Artificial Intelligence 6* by P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, Eds., North-Holland: Amsterdam; Tokyo, 347–374.

Gilboa, I., and D. Schmeidler (1993) "Updating Ambiguous Beliefs" *Journal of Economic Theory* **59**, 33–49.

Jaffray, J. Y. (1992) "Bayesian Updating and Belief Functions" *IEEE Transactions on Systems, Man and Cybernetics* **22**, 1144–1152.

Shafer, G. (1976) *A Mathematical Theory of Evidence*, Princeton University Press: Princeton.

Sundberg, C., and C. Wagner (1992) "Generalized Finite Differences and Bayesian Conditioning of Choquet Capacities" *Advances in Applied Mathematics* **13**, 262-272.

Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall: London.

Wasserman, L. A., and J. B. Kadane. (1990) "Bayes' Theorem for Choquet Capacities" *The Annals of Statistics* **18**, 1328–1339.