# Nonexistence of universal conditional objects

**Arthur Ramer**
School of Computer Science and Engineering
UNSW, Sydney 2052, Australia
ramer@cse.unsw.edu.au

*Par suite seulement d'une malheureuse coïncidence, ce résultat ne fut pas quantitativement , suivant le langage des savants, celui qui avait été prévu. En fait, il fut assez différent.*

Pierre Boulle (1953). Le règne des sages.

## Abstract

Construction of putative conditional objects is linked to Dirichlet generating functions. Lewis's triviality result and van Fraasen construction are retrieved. Nonexistence of a universal extension of a given boolean algebra where all its true probabilistic conditional objects can reside is proven.

**Keywords:** Conditional object, triviality lemma, information generating function, logic of conditionals.

## 1 Preamble

Conditional objects have been actively discussed for over one hundred years in the literature [4] on foundations and philosophy of probability. They are often introduced as *logical* statements in some form of probabilistic logic. To reason *quantitatively* about and with them one needs to postulate some probabilistic setting, at least a space of probabilistic events. Once being modeled this way, a few questions arise immediately:

- can these objects interact with the ordinal objects (events)? with one another?
- can one build iterated conditional objects, ie. how to condition a conditional?
- can these objects be modeled in he original space? or in a fixed, though perhaps large extension of the original space.

Several constructions of such objects have been proposed. They model some aspects of these objects, typically permitting a one-step iteration of conditioning. Depending on semantics of such intended objects, they can be modeled statically or with recognition of some temporal logical structure. Further, depending on the intended applications one can be satisfied with limited interaction among the objects. Perhaps one would define $(A|B)$ and $P((A|B)|C)$, but not necessarily $((A|B)|C)$ and $P(((A|B)|C)|D)$.

Then again, one could be satisfied with computing just the probabilities of these simpler conditional objects, without constructing a complete probability distribution on the entire space where these objects might live.

## 2 Intended properties and results

Conditional objects in probabilistic contexts are usually intended as an algebraic or combinatorial constructs that can, functionally, carry the conditional probability. Assuming a probability space $X$ and events $A, B, \ldots \subseteq X$ we want to have a function $f(A, B) =: A|B$, where $(A|B)$ is to live either in the original space $X$ or in some extension $E(X)$ thereof. It is posited that there be a natural way to construct probability distribution $P^E$ on $E(X)$, given $P$ on $X$.

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 978–982
Torremolinos (Málaga), June 22–27, 2008

We want $P^E((A|B)) = P(A|B)$, the latter understood as the usual conditional probability. It is often tacitly assumed that the conditional objects, once constructed, are suitable for an arbitrary probability distribution. Thus this equality would have to hold for any $P$ on $X$. A weaker solution would only produce such objects specialised to a specific probability assignment.

Logical process of conditioning should, in a sense, 'subtract' the information. Lifted to the probabilistic setting one might expect that the hypothetical $(A|B)$ is independent of $B$. Ideally, the joint probability distribution on $(A|B) \otimes B$ should be isomorphic to that on $A \cap B$. A weaker property would have some joint distribution on $(A|B) \times B$ that is isomorphic to one on $A \cap B$ and which would marginalise to those given on $(A|B)$ and $B$. In this weaker form we do not require independence of $(A|B)$ and $B$, just that we can 'derive' the common part $A \cap B$ from the conditional object $(A|B)$ and the conditioning object $B$. It is the main result of this paper that such a reconstruction is, in principle, not possible.

We approach the putative construction of such objects through the analysis of Dirichlet generating functions associated with probability distributions. We will be computing such functions for the hypothetical objects. The motivation for our method was the analysis of the entropy values of various conditionals. Accordingly, these Dirichlet sums can be termed *entropy* or *information generating functions*.

Lewis provided the *triviality* lemma [4], stating that a probability assignment to a boolean algebra with more than four values cannot admit conditional objects *within* the algebra itself. It is to mean that there is no boolean operation $o_P : \langle A, B \rangle \to (A|B)$ which would be congruent with probabilistic conditioning under $P$. We can give it a one-line proof using such generating function.

We then carry out a form of 'reverse engineering' known formally as *decategorification* to produce a combinatorial object to which a Dirichlet sum, approximating that of a 'true' conditional might be associated. This leads directly to Goodman-Nguyen-van Fraasen algebra. Compromises made in this construction process make apparent that these conditional are imperfect - the joint probability distribution constructed from the conditioning event $B$ and the conditional $(A|B)$ must be different than that of the combined event $A \cap B$, except for some trivial conditioning.

Finally we prove that is a necessary situation and there can be no algebra of conditional events, however complex in its structure, that would make the combined event match precisely the product of the conditional and the conditioning events.

## 3 Notation

We deal, almost exclusively, with discrete probability distributions. Given two independent distributions $P$ and $Q$, with domains $X$ and $Y$ respectively we denote $P \otimes Y$ their cartesian product - a joint distribution with $X \times Y$ as its domain. If an infinite series of such independent distributions $P_1, P_2, \ldots$ needs to be 'multiplied' we write $\prod_{j=1}^{\infty} P_j$ for the joint distribution on the cartesian product of $X_j$. Although this construction is, technically, not a discrete distribution, we always need only a 'weak product'

$$\bigcup_{k=1}^{\infty} \prod_{j=1}^{k} X_j = X_1 \cup X_1 \times X_2 \cup X_1 \times X_2 \times X_3 \cup \ldots$$

To each probability distribution we shall assign a certain analytic function $f(u)$ defined as a finite or infinite sum

$$\sum \frac{a_i}{z_i^u}.$$

Such sums are known as general Dirichlet series [5]. They are analytic in a half-plane, thus have an abscissa of convergence. It is the intercept on the real axis of the vertical line in the complex plane, such that there is convergence to the right of it, and divergence to the left. We need only the most basic fact

- that the arithmetic operations on such series or differentiation lead to other analytic functions that also can be expanded into such series.

## 4 Information generating functions

Golomb [3] defined information generating function as

$$f_P(x) = \sum p_j^x.$$

Its key property is $H(P) = -f'_P(1)$. Moreover, it has the multiplicative property $f_{P\otimes Q}(u) = f_P(u)f_Q(u)$, from which one can derive the additivity of entropies

$$\begin{aligned} f'_{P\otimes Q}(1) &= (f_P f_Q)'(1) \\ &= (f'_P f_Q + f'_Q f_P)(1) = f'_P(1) + f'_Q(1) \end{aligned}$$

as $f_P(1) = f_Q(1) = 1$.

The present author, independently though 20 odd years later, introduced an entropy (or uncertainty) generating function[1]

$$f_P(x) = \sum p_j^{1-x} = \sum \frac{p_j}{p_j^x}.$$

It is again a Dirichlet generating function, now making explicit which are the intended coefficients and the exponentials. This form has certain advantages over the Golomb's one. As the coefficients (written in numerator) are explicit, they can be taken from a different probability assignment than the exponentials (written in denominator). This leads to generating information distance - $I$-divergence. And as a minor nicety, now $H(P) = f'(0)$.

The author also noted that, as $f_P(u)$ is an analytic function, its logarithm is analytic as well. Writing $g_P(u) = \log f_P(u)$ gives

$$g'(u) = \frac{f'(u)}{f_P(u)}; \quad g_{P\otimes Q}(u) = g_P(u) + g_Q(u)$$

which for the entropies means that

$$\begin{aligned} H(P) &= g'(0) \\ H(P\otimes Q) &= (g_P + g_Q)'(0) = H(P) + H(Q). \end{aligned}$$

[1]Here and in the rest of the paper we avoid proliferation of function symbols by reusing $f(u)$ and $g(u)$ for various related generating functions.

Thus $g_P(u)$ is additive wrt formation of cartesian products.

Another advantage was pointed out by Csiszar (unrelated to any entropy computations). Namely, the latter function is a moment generating function of the logarithmic random variable $V : P(V = -\log p_i) = p_i$. This suggests the next step - using characteristic functions instead of moment-generating ones. Here, we would define

$$f_P(u) = \sum p_j^{1-ix}, i = \sqrt{-1}; g_P(u) = \log f(_P(u).$$

Properties of characteristic functions are better understood and this plays important role in the later nonexistence proof.

Along with these functions based on the probability assignment on the entire domain, we will use functions corresponding to the partial probability distribution on a subset of the domain $A \subseteq X$. We write[2]

$$f_{P;A} = \sum_A p_j^{1-x} \text{ or } f_{P;A} = \sum_A p_j^{1-ix}.$$

Differentiating and evaluating at 0 gives partial entropy

$$f'_{P;A}(0) = -\sum_A p_j \log p_j = H(P;A).$$

Though we refer to Shannon entropy, most of the reasoning about conditionals could be carried out using other entropies. For example, Renyi entropy of order $\alpha$ is $R^{(\alpha)}(P) = \frac{f_P(\alpha)-1}{\alpha}$. Now $1 = f_P(0)$ and $\lim_{\alpha\to 0} \frac{f_P(\alpha)-1}{\alpha} = f'_P(0)$. This gives the classical limit $\lim_{\alpha\to 0} R^{(\alpha)}(P) = H(P)$.

## 5 Constructing probabilistic conditionals

If a conditional object be constructed as an event ( a subset) in an a probability space then its function $F_P$ should satisfy

$$f_{P:(A|B)}(u) = \frac{f_{P;A\cap B}(u)}{f_{P;B}(u)}.$$

[2]We simplify the subscripts whenever possible without confusion; here these summations should read as $\sum_{j:x_j\in A}$. In other places we omit probability assignment or the domain if those can be inferred from the context.

This equation is a direct generalisation of the equality of total probability values

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

(One could use a more refined, symmetric formula $f_{P;(A|B)} \cdot f_{P;B} = f_{P;A\cap B} \cdot f_{P;X}$ corresponding to $P(A|B)P(B) = P(A \cap B)P(X)$, but it only leads to a more complicated construction.)

We discuss so far the generating functions but need the actual objects. To construct them we use the paradigm of decategorification [1], here a construction of a combinatorial species ('espèce de structure') [2]. Such a construction can be carried out faithfully for ordinary and exponential generating functions. Here we need to resort to a bit of 'cheating' to effect a useful construction.

As a minimum we are interested in probabilities of the objects that will represent $(A|B)$ for any pair of subsets $A, B$ of the domain $X$. We will be replacing the function $F_X$ with just the constant $f_X(0) = 1$, representing the probability of $X$. We can write

$$f_{A|B}(u) = \frac{f_{A\cap B}(u)}{f_B(u)} = \frac{f_{A\cap B}(u)}{f_X(u) - f_{\overline{B}}(u)}$$

After the replacement

$$\begin{aligned} f_{A|B}(u) &= \frac{f_{A\cap B}(u)}{1 - f_{\overline{B}}(u)} \\ &= f_{A\cap B}(u)(1 + f_{\overline{B}}(u) + f^2_{\overline{B}}(u) + \ldots) \\ &= f_{A\cap B} + f_{\overline{B}} f_{A\cap B} + f^2_{\overline{B}} f_{A\cap B} + \ldots \end{aligned}$$

The corresponding combinatorial object is not difficult to build. We take $E = X^\infty$ as the extension of our base space $X$. We associate to every product of functions $f_{A_1} \cdot f_{A_2} \cdots f_{A_k}$ the cartesian product $A_1 \times A_2 \times \ldots \times A_k$ and complete it to

$$A_1 \times A_2 \times \ldots \times A_k \times \prod_{i=k+1}^{\infty} X.$$

Then we take the union of such sets to correspond to the sum of the series. We find

$$\begin{aligned} (A|B) &= (A \cap B) \times X \times \cdots \\ &\cup\ \overline{B} \times (A \cap B) \times X \times \cdots \\ &\cup\ \overline{B} \times \overline{B} \times (A \cap B) \times X \times \cdots \\ &\cup\ \ldots\ldots \end{aligned}$$

where the products in different lines are trivially disjoint. Space $X^\infty$ has a natural probability structure, which defines the probabilities of all conditional objects immersed there. The resulting construction is known as Goodman-Nguyen-van Fraasen algebra [4]. Objects like $(A|B)$ can now be mixed among themselves using standard boolean operations.

Classical subsets of $A, B \subseteq X$ correspond to $A \times \prod_2^\infty X$. It is immediate that conditioning on the entire $X$ amounts to no action (as it should)

$$(A|X) = A = A \times \prod_2^\infty X$$

as the later terms of the set union are all empty. We can compute the values like $P(((A|B)|C)$ or $P((A|B)|(C|D))$. Though, these are only the numerical probability values; there are no corresponding conditional objects.

In our approach of linking conditionals with generating functions we can compute much more than the overall probabilities of conditional objects. For example, we can compute their entropies and other information parameters. However, the entire construction is imperfect in one very important aspect - we must pay the price for our light-hearted substitution of constant 1 for the entire function $f_X(u)$.

However, as long as we deal only with the probabilities of conditionals the model works very well. For example, $(A|B)$ and $B$ are stochastically independent, as we would expect of logically independent objects. It is easily checked that the intersection $(A|B) \cap B$

corresponds to $(A \cap B) \times \prod_{2}^{\infty} X$. We get

$$\begin{aligned} P((A|B) \cap B) &= P(A \cap B) \\ &= \frac{P(A \cap B)}{P(B)} P(B) \\ &= P(A|B)P(B) \end{aligned}$$

## 6 Nonexistence

Lewis stated that there can be no two-argument boolean operator $o : \langle A, B \rangle \mapsto (A|B)$, its values within the same boolean algebra. Using our approach we would ask whether

$$\frac{f_{A \cap B}(u)}{f_B(u)} = \frac{\sum_{A \cap B} p_j^{1-u}}{\sum_B p_j^{1-u}}$$

can be presented as $\sum_C p_j^{1-u}$ for some $C \subset X$.

Trivially, it is impossible in general, giving our promised one-line proof of triviality lemma.

The main result is that there can be no universal construction of $E(X)$ such that the complete conditional objects can live there. We proceed as before, but use now the characteristic function (of the logarithmic rv) $f_A(u) = \sum_A p_j^{1-iu}$. The question becomes whether $\frac{f_{A \cap B}(u)}{f_B(u)}$ can ever be a characteristic function. There are numerous conditions known to be required of a function for it to be a characteristic function of 'anything'. For one, it must go to zero as its argument goes to infinity. It is trivial to find examples when the ratio above does not satisfy this property.

## 7 Closing remarks

Our results do not affect most of the work on construction and analysis of conditional objects. On the contrary, we emphasise that many such constructions are possible and that, in light of our results, none can claim to be a completely perfect answer to the conditioning problem.

A similar discussion can be conducted in the possibilistic setting. It will be complicated somewhat by the fact that there are several methods of conditioning in that framework. However, we expect that most of them admit similar impossibility results.

We give only a few basic references; some of them contain excellent extensive overviews of the work on building conditional objects.

## References


[1] J Baez, J Dolan. *From finite sets to Feynman diagrams,* in B Engquist, W Schmid (eds.) Mathematics Unlimited - 2001 and Beyond. Springer Verlag 2001.

[2] F Bergeron, G Labelle, P Leroux. Combinatorial Species and Tree-like Structures. Cambridge University Press 1997.

[3] S Golomb,R Peile, R Scholtz. Basic Concepts in Information Theory and Coding. Plenum Press 1994.

[4] IR Goodman, HT Nguyen, EA Walker. Conditional Inference and Logic for Intelligent Systems: A Theory of Measure-Free Conditioning. North-Holland 1991.

[5] GH Hardy, M Riesz. The general theory of Dirichlet's series. Cambridge Tracts in Mathematics, No. 18. Cambridge University Press, 1915. *available on-line* http://dlxs2.library.cornell.edu/cgi/t/text/text-idx?c=math;idno=01480002

[6] B van Fraasen. *Probabilities of conditionals,* in W Harper, C Hooker (eds.) Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science. D Reidel 1976, 261–300.