

3

Hypothetical Expectations

The notion of an expectation presented in the previous two chapters does not tap all of my ability to quantify my beliefs about random quantities. Often I will find myself thinking about ‘hypothetical’ expectations of the form ‘my expectation of X supposing that Q is true’, where Q is a random proposition. If I am to exploit this type of belief then I need to understand it better (i.e. to give it meaning), and also to integrate it into my inference, in order to better constrain other expectations that I cannot specify directly. This requires me to be more precise about ‘supposing’.

This chapter will strike many readers as rather ‘philosophical’. But it is no more philosophical than Chapter 1. That chapter discussed the meaning of an expectation, and this one discusses the meaning of a hypothetical expectation (also termed a conditional expectation, see Sec. 3.1). My approach in this chapter is more idiosyncratic than in Chapter 1, and it has taken me longer to come to the viewpoint that I now describe. Rest assured, though, that all of the usual results still hold; it is only the interpretation that differs. These usual results can be found in Sec. 3.3 and Sec. 3.5. Sec. 3.1, Sec. 3.2, and Sec. 3.4 are mainly concerned with meaning. You may or may not agree with my construction of meaning, but I hope at least you appreciate the importance of the struggle.

3.1 What do we mean by ‘supposing’?

I start with a trap.

Consider my expectation of sea-level rise in 2100 supposing that the Greenland ice-sheet melts before then. A simplistic treatment would be for me to add about 6 m to my expectation supposing the ice-sheet didn’t melt, this being my belief about the amount of sea-level rise contained in the Greenland ice. But maybe that’s the wrong supposition. There is a basic difference between supposing that the Greenland ice-sheet melts, and supposing that the only thing that happens is that the Greenland ice-sheet melts; the +6 m is the latter. But if the Greenland ice-sheet melts, it would be part of a much larger scenario, in which those factors which caused the ice-sheet to melt (changes in air and ocean temperature, changes in precipitation patterns) also affected other ice-sheets, which in turn

From *Lecture Notes on Statistical Inference*, Jonathan Rougier, Copyright © University of Bristol 2015.

affected sea-level rise. So both are valid suppositions, but they are not the same.

Therefore, when ‘supposing’ I must be wary of not confusing ‘interventions’ with ‘scenarios’. An intervention would be where I imagine doing something to melt the Greenland ice-sheet. A scenario would be a situation in which the Greenland ice-sheet melted. Pearl (2000) contains a detailed assessment of these different ways of supposing. We will simply accept the caution that Q should, where possible, be at or near the root of a causal tree.

I will denote the value I give to my belief about X supposing Q to be true as ‘ $E(X | Q)$ ’. Generally, I will refer to this as a *hypothetical expectation*. In textbooks this might be termed a ‘conditional expectation’, but this is an oversimplification. In the proper theory of expectation, a conditional expectation is a random quantity.¹ But $E(X | Q)$ is *not* a random quantity: it is a value. This issue could be finessed with a clever notation. Hence, $E(X | Y)$ is a conditional expectation (random quantity) because the symbol to the right of the bar is a random quantity, while $E(X | Q)$ is a value because the symbol to the right of the bar is a random proposition. But I prefer to sidestep this issue, and I will not use $E(X | Y)$ at all. Just to be clear, though, I will refer to $E(X | Q)$ as a ‘hypothetical expectation’ rather than a ‘conditional expectation’. This also serves to emphasise, for me, that when I specify a value for $E(X | Q)$ I am ‘supposing’.

Are there any intuitions about ‘supposing’? I submit that there is one crucial property that any rational notion of supposing ought to satisfy, if it is to be consistent with how we would like to use supposing in practice.

Definition 3.1 (Recursive property). $E(\cdot | Q)$ should satisfy the axioms of expectation, and the same relation between $E(\cdot)$ and $E(\cdot | Q)$ should hold as between $E(\cdot | R)$ and $E(\cdot | Q \wedge R)$.

Effectively, hypothetical expectations ought to have a ‘narrative’ contiguity: they put me into a world which is like my world, except for the truth of Q . I write ‘narrative’ here, to suggest that my beliefs are telling a story about my world, and that the truth of Q is something I incorporate into this story, with the smallest possible perturbation. So there is still a theory of expectation in this hypothetical world. And there is the opportunity for me to go deeper, and consider a hypothetical expectation within my hypothetical world, and so on. This seems unexceptionable to me—the idea that many features of my world carry over into my hypothetical world, and that coherence of expectations is one of them, and that this property is recursive.

How will this property be used? In the next section I will propose a relation between ‘original’ and hypothetical expectation, based on an analogy. If this relation implies that hypothetical expectations have the recursive property then we can accept it, tentatively, as a defining relation. We can go on to investigate other properties of hypothetical expectation that follow from this definition. If, on

¹ See Grimmett and Stirzaker (2001, sec. 3.7, sec. 7.7), or the books cited at the start of Chapter 1.

reflection, we find that these additional properties also concur with our intuition about supposing, then that is very encouraging. But if we find a property which seems quite contrary to our intuition about supposing then we might have to think again.

3.2 Definition through analogy

Recall the betting interpretation of $\Pr(Q)$, given in Sec. 1.4. $\Pr(Q)$ is my 'fair price' for a bet that pays 0 if Q is false, and 1 if Q is true. Hence it satisfies the equality

$$E\{\mathbb{1}_Q - \Pr(Q)\} = 0. \quad (3.1)$$

Now turn this interpretation around to provide a working definition of hypothetical probability:

The hypothetical probability $\Pr(A | Q)$ is my fair price for a bet on A which only goes ahead if Q is true.

This definition implies the following result. (This result also follows directly from Def. 3.2, below.)

Theorem 3.1 (Called Off Bet theorem).

Let A and Q be random propositions. Then, under the definition immediately above,

$$\Pr(A, Q) = \Pr(A | Q) \Pr(Q). \quad (3.2)$$

Proof. By the same reason that lead to (3.1),

$$E\left\{\mathbb{1}_{-Q} \cdot 0 + \mathbb{1}_Q(\mathbb{1}_A - \Pr(A | Q))\right\} = 0$$

and the result follows immediately. \square

Note that Thm 3.1 defines a unique value of $\Pr(A | Q)$ when $\Pr(Q) > 0$; but an arbitrary value when $\Pr(Q) = 0$, which implies that $\Pr(A, Q) = 0$, and that the equality has the form $0 = \Pr(A | Q) \cdot 0$.

Now we take the next step. The recursive property requires that a hypothetical probability is just the hypothetical expectation of the indicator function of a random proposition. If this were not true, my hypothetical world would not be contiguous with my actual world, since a basic property of expectations would have changed. And this strongly suggests that $E(X\mathbb{1}_Q) = E(X | Q) \Pr(Q)$ ought to be the right generalisation of Thm 3.1. So I will take this as the definition of conditional expectation, subject to checking that it satisfies the recursive property (which it does, as shown below).

Definition 3.2 (Hypothetical expectation and probability).

Let X be any scalar random quantity and Q be any random proposition. $E(X | Q)$ is the hypothetical expectation of X given Q exactly when it satisfies

$$E(X\mathbb{1}_Q) = E(X | Q) \Pr(Q). \quad (3.3)$$

Hypothetical probability is defined as $\Pr(A | Q) := E(\mathbb{1}_A | Q)$.

Note that Def. 3.2 defines a unique value of $E(X | Q)$ when $\Pr(Q) > 0$, but an arbitrary value when $\Pr(Q) = 0$, because in this case Schwarz's inequality (Thm 1.3) implies that

$$E(X\mathbb{1}_Q)^2 \leq E(X^2)E(\mathbb{1}_Q^2) = E(X^2)\Pr(Q) = 0$$

which implies that $E(X\mathbb{1}_Q)^2 = 0$ and the definition has the form $0 = E(X | Q) \cdot 0$. Def. 3.2 implies that $\Pr(A | Q)$ can indeed be interpreted as a called off bet, since it implies Thm 3.1.

Now we have to check Def. 3.2 satisfies the recursive property.

First, we must check to see whether $E(\cdot | Q)$ satisfies the axioms of expectation. The FTP (Thm 1.6) is useful here. If we can show that there is a hypothetical version of the FTP, then we know, using the same argument as in the FTP proof, that $E(\cdot | Q)$ satisfies the axioms, and all of the results we have proved for expectation and probability also hold for hypothetical expectation and probability. So the following result is exactly what we need.

Theorem 3.2 (Hypothetical FTP, HFTP).

If $\Pr(Q) > 0$, then

$$E\{g(\mathbf{X}) | Q\} = \sum_{\mathbf{x}} g(\mathbf{x}) \cdot \Pr(\mathbf{X} \doteq \mathbf{x} | Q)$$

for any g .

Note the restriction to $\Pr(Q) > 0$. This restriction will occur frequently in equalities involving hypothetical expectations or probabilities. These equalities cannot hold if $E(X | Q)$ or $\Pr(X | Q)$ are arbitrary, and hence $\Pr(Q) = 0$ must always be excluded.

Before proving this result, the following result is of independent interest.

Theorem 3.3 (Muddy Table theorem).

Let $Q := q(\mathbf{X})$ where $q(\mathbf{x})$ is a first-order sentence. If $\Pr(Q) > 0$ then

$$\Pr(\mathbf{X} \doteq \mathbf{x} | Q) = \frac{\mathbb{1}_{q(\mathbf{x})} \mathbf{p}(\mathbf{x})}{\Pr(Q)}.$$

This result uses the 'p' notation from Sec. 1.5.1; note also that this is an example of a functional relation which holds for all $\mathbf{x} \in \mathfrak{X}$. I have taken the name from a description of a similar result in van Fraassen (1989, ch. 7).

Proof. As $\Pr(Q) > 0$,

$$\Pr(\mathbf{X} \doteq \mathbf{x} | Q) = \frac{\Pr\{(\mathbf{X} \doteq \mathbf{x}) \wedge q(\mathbf{X})\}}{\Pr(Q)} = \frac{E(\mathbb{1}_{\mathbf{X} \doteq \mathbf{x}} \cdot \mathbb{1}_{q(\mathbf{X})})}{\Pr(Q)}$$

from Thm 3.1. Applying the FTP to the numerator gives

$$E(\mathbb{1}_{\mathbf{X} \doteq \mathbf{x}} \cdot \mathbb{1}_{q(\mathbf{X})}) = \sum_{\mathbf{x}'} \mathbb{1}_{\mathbf{x}' \doteq \mathbf{x}} \cdot \mathbb{1}_{q(\mathbf{x}')} \cdot \mathbf{p}(\mathbf{x}') = \mathbb{1}_{q(\mathbf{x})} \mathbf{p}(\mathbf{x}),$$

as required. □

Proof of Thm 3.2, the HFTP.

$$\begin{aligned}
 E\{g(\mathbf{X})\mathbb{1}_{q(\mathbf{X})}\} &= \sum_{\mathbf{x}} g(\mathbf{x})\mathbb{1}_{q(\mathbf{x})} \cdot p(\mathbf{x}) && \text{FTP} \\
 &= \sum_{\mathbf{x}} g(\mathbf{x}) \cdot \frac{\mathbb{1}_{q(\mathbf{x})} P(\mathbf{x})}{\Pr(Q)} \cdot \Pr(Q) && \text{as } \Pr(Q) > 0 \\
 &= \sum_{\mathbf{x}} g(\mathbf{x}) \cdot \Pr(\mathbf{X} = \mathbf{x} \mid Q) \cdot \Pr(Q) && \text{Muddy Table theorem}
 \end{aligned}$$

and the result follows on dividing through by $\Pr(Q)$, according to (3.3). \square

So the first part of the recursive property is satisfied: a hypothetical expectation as defined in (3.2) does indeed satisfy the axioms of expectation, provided that $\Pr(Q) > 0$. Of course this restriction on $\Pr(Q)$ also has a very intuitive justification: I doubt there is any narrative contiguity between my world and a hypothetical world in which I suppose that something I believe is impossible is actually true.²

Now for the second part. We need to establish that $E(\cdot)$ and $E(\cdot \mid Q)$ stand in the same relation to each other as $E(\cdot \mid R)$ and $E(\cdot \mid Q \wedge R)$. In other words, does everything still work if I am already in hypothetical world R ? The definition states that

$$E(X\mathbb{1}_Q) = E(X \mid Q) \Pr(Q).$$

Now if the recursive property holds then we should just be able to drop a ' R ' into each of the three expressions. So we need to show that

$$E(X\mathbb{1}_Q \mid R) = E(X \mid Q, R) \Pr(Q \mid R).$$

And, indeed, this follows if $\Pr(R) > 0$:

$$\begin{aligned}
 E(X\mathbb{1}_Q \mid R) &= \frac{E(\mathbb{1}_R \cdot X\mathbb{1}_Q)}{\Pr(R)} && \text{by (3.3) with } \Pr(R) > 0 \\
 &= \frac{E(\mathbb{1}_Q \mathbb{1}_R \cdot X)}{\Pr(R)} \\
 &= \frac{E(\mathbb{1}_{Q \wedge R} \cdot X)}{\Pr(R)} \\
 &= \frac{E(X \mid Q, R) \Pr(Q, R)}{\Pr(R)} && \text{(3.3) again} \\
 &= E(X \mid Q, R) \frac{\Pr(Q, R)}{\Pr(R)} \\
 &= E(X \mid Q, R) \Pr(Q \mid R) && \text{and again}
 \end{aligned}$$

as required.

3.3 Further properties of hypothetical expectation

The previous section established that hypothetical expectations as defined in Def. 3.2 satisfy the recursive property given in Def. 3.1. This section demonstrates some additional properties that hypothetical expectations have. In all cases, these properties seem intuitive

² Lewis (1973) provides a well-known analysis of such worlds: they are fascinating but not germane.

for ‘supposing’, giving us confidence that we can incorporate the beliefs we specify when supposing Q to be true with other expectations and probabilities. This incorporation will be discussed further in Sec. 3.4.

One of the key results for conditional expectations is the ‘tower property’. This provides another link between my original expectation and a set of hypothetical expectations. It depends on the notion of a belief partition.

Definition 3.3 (Belief partition).

Let $\mathbb{Q} := \{Q_1, \dots, Q_k\}$ be a finite set of random propositions. Then \mathbb{Q} is a belief partition exactly when

$$\sum_{j=1}^k \mathbb{1}_{Q_j} = 1.$$

Asserting that \mathbb{Q} is a belief partition is asserting an equality which holds for random quantities. It must be interpreted as a statement of my beliefs, which implies that $\sum_j \Pr(Q_j) = 1$ and that $\Pr(Q_i, Q_j) = 0$ for $i \neq j$. A stronger definition of ‘partition’ insists that each $Q_j := q_j(\mathbf{X})$ for which

$$\sum_{j=1}^k \mathbb{1}_{q_j(x)} = 1 \quad \text{for all } x \in \mathcal{X}. \quad (\dagger)$$

In other words, the q ’s partition the realm of \mathbf{X} into a set of mutually exclusive tiles. This definition of a partition has nothing to do with beliefs. But only the weaker property in Def. 3.3 is required below, in which the q ’s are not necessarily mutually exclusive, but where I do not believe that Q_i and Q_j can both be true.³

Here is a very intuitive and useful result. It implies the ‘obvious’ property that if $\Pr(Q) \leftarrow 1$, then $E\{g(\mathbf{X}) \mid Q\} = E\{g(\mathbf{X})\}$.

Theorem 3.4 (Conglomerability).

Let $\mathbb{Q} := \{Q_1, \dots, Q_k\}$ be a belief partition. Then

$$E\{g(\mathbf{X})\} = \sum_{j=1}^k E\{g(\mathbf{X}) \mid Q_j\} \Pr(Q_j)$$

for any g .

Proof. Because \mathbb{Q} satisfies (\dagger) ,

$$\begin{aligned} E\{g(\mathbf{X})\} &= E\{g(\mathbf{X}) \cdot 1\} \\ &= E\left\{g(\mathbf{X}) \cdot \sum_j \mathbb{1}_{Q_j}\right\} \\ &= \sum_j E\{g(\mathbf{X}) \mathbb{1}_{Q_j}\} \\ &= \sum_j E\{g(\mathbf{X}) \mid Q_j\} \Pr(Q_j) \quad \text{by (3.3)} \end{aligned}$$

as required. \square

³ I doubt this makes much difference in practice, but it is always interesting to note where weaker conditions will suffice.

Bruno de Finetti (1974) used a slightly different definition of conglomerability, which is that

$$\min_j E\{g(X) \mid Q_j\} \leq E\{g(X)\} \leq \max_j E\{g(X) \mid Q_j\},$$

i.e. the expectation of any random quantity is bounded below and above by the smallest and largest values that the conditional expectation can take on a belief partition. This is implied by Thm 3.4. Practically speaking, this bounding property is very powerful, being not obvious, *a priori*, but acceptable *a posteriori*. It is also very useful. It justifies, for supposing, the same kind of tactic we often use for other reasoning, which is to imagine the worst and the best possible outcomes.

A special case of Q gives the celebrated Law of Iterated Expectation, also known as the ‘tower property’ of expectation.

Theorem 3.5 (Law of Iterated Expectation, LIE).

$$E\{g(X)\} = \sum_y E\{g(X) \mid Y \doteq y\} p(y).$$

Proof. Follows immediately from Thm 3.4 on setting

$$Q \leftarrow \bigcup_y \{Y \doteq y\}.$$

□

Conglomerability can be extended in a useful way, into a general procedure I will call *drilling down*. Let Q be a *belief partition for* Q exactly when

$$\sum_{j=1}^k \mathbb{1}_{Q_j} = \mathbb{1}_Q.$$

Note that Q_j implies Q , so that $\Pr(Q_j, Q) = \Pr(Q_j)$. By the same reasoning as above, $\sum_j \Pr(Q_j \mid Q) = 1$ and $\Pr(Q_i, Q_j) = 0$ for $i \neq j$.

Theorem 3.6 (Drilling down).

Let Q be a *belief partition for* Q . If $\Pr(Q) > 0$, then

$$E\{g(X) \mid Q\} = \sum_{j=1}^k E\{g(X) \mid Q_j\} \Pr(Q_j \mid Q).$$

Proof. This almost follows directly from the recursive property, which suggests we can drop a ‘ $|Q$ ’ into Thm 3.4. But it is safer to

prove it directly:

$$\begin{aligned}
E\{g(\mathbf{X}) \mid Q\} &= \frac{E\{g(\mathbf{X})\mathbb{1}_Q\}}{\Pr(Q)} && \text{by (3.3) as } \Pr(Q) > 0 \\
&= \frac{E\{g(\mathbf{X}) \sum_j \mathbb{1}_{Q_j}\}}{\Pr(Q)} \\
&= \frac{\sum_j E\{g(\mathbf{X})\mathbb{1}_{Q_j}\}}{\Pr(Q)} \\
&= \frac{\sum_j E\{g(\mathbf{X}) \mid Q_j\} \Pr(Q_j)}{\Pr(Q)} && \text{(3.3) again} \\
&= \sum_j E\{g(\mathbf{X}) \mid Q_j\} \frac{\Pr(Q_j, Q)}{\Pr(Q)} && \text{as } Q_j \text{ implies } Q \\
&= \sum_j E\{g(\mathbf{X}) \mid Q_j\} \Pr(Q_j \mid Q) && \text{(3.3) again}
\end{aligned}$$

as required. \square

As an illustration, suppose that X was some aspect of tomorrow's weather, perhaps the amount of rain. I can simplify the task of specifying my beliefs about X by dividing the weather into weather types, for example 'anti-cyclonic' (high pressure in the UK), 'neutral', or 'cyclonic' (low pressure).⁴ Then Q_j would be the proposition that the weather was of type j , and $Q \leftarrow \{Q_1, Q_2, Q_3\}$. If I was in a hurry I could assess $E(X)$ by using my beliefs $E(X \mid Q_j)$ and my probabilities $\Pr(Q_j)$. My beliefs $E(X \mid Q_j)$ would be fairly standard, but $\Pr(Q_j)$ could change from day to day. If I had more time, then I could take the most probable type, say Q_1 , and then partition it further, perhaps by the direction of the wind in my neighbourhood, to give me a belief partition for Q_1 , say $Q_j \leftarrow \{Q_{11}, \dots, Q_{18}\}$, corresponding to the cardinal and ordinal directions. Then I could refine my assessment of $E(X \mid Q_1)$ by using my beliefs $E(X \mid Q_{1k})$ and my probabilities $\Pr(Q_{1k} \mid Q_1)$. In other words, I could 'drill down' into Q_1 .

Weather forecasting is an application where the hypothetical expectations $E(X \mid Q_j)$ and $E(X \mid Q_{jk})$ are fairly standard, but the probabilities $\Pr(Q_j)$ and $\Pr(Q_{jk} \mid Q_j)$ are changing from day to day (in the UK).

* * *

Here is another useful result although it does not have a name, as far as I know. I have borrowed a nearby one from Williams (1991, sec. 9.7).

Theorem 3.7 (Taking out what is known).

Let $\mathbf{X} := (\mathbf{Y}, \mathbf{Z})$ and suppose the truth of Q implies that $\mathbf{Y} = \mathbf{y}$. If $\Pr(Q) > 0$ then

$$E\{h(\mathbf{Y}, \mathbf{Z}) \mid Q\} = E\{h(\mathbf{y}, \mathbf{Z}) \mid Q\}.$$

⁴ See Jones *et al.* (2013) and <http://www.cru.uea.ac.uk/cru/data/lwt/>.

Proof. The crucial step below is

$$\begin{aligned} \mathbb{E}\{h(\mathbf{Y}, \mathbf{Z})\mathbb{1}_Q\} &= \sum_{\mathbf{x}'} h(\mathbf{y}', \mathbf{z}') \mathbb{1}_{q(\mathbf{x}')} \cdot p(\mathbf{x}') \\ &= \sum_{\mathbf{x}'} h(\mathbf{y}, \mathbf{z}') \mathbb{1}_{q(\mathbf{x}')} \cdot p(\mathbf{x}') \\ &= \mathbb{E}\{h(\mathbf{y}, \mathbf{Z})\mathbb{1}_Q\} \end{aligned}$$

by the FTP, the second equality following because $\mathbf{y}' \neq \mathbf{y}$ implies that $\mathbb{1}_{q(\mathbf{x}')} = 0$. Then

$$\begin{aligned} \mathbb{E}\{h(\mathbf{X}) \mid Q\} &= \mathbb{E}\{h(\mathbf{Y}, \mathbf{Z}) \mid Q\} \\ &= \frac{\mathbb{E}\{h(\mathbf{Y}, \mathbf{Z})\mathbb{1}_Q\}}{\Pr(Q)} && \text{by (3.3) with } \Pr(Q) > 0 \\ &= \frac{\mathbb{E}\{h(\mathbf{y}, \mathbf{Z})\mathbb{1}_Q\}}{\Pr(Q)} && \text{see above} \\ &= \mathbb{E}\{h(\mathbf{y}, \mathbf{Z}) \mid Q\} && \text{(3.3) again} \end{aligned}$$

as required. \square

One immediate corollary is that

$$\mathbb{E}\{g(\mathbf{Y})h(\mathbf{Z}) \mid Q\} = g(\mathbf{y}) \mathbb{E}\{h(\mathbf{Z}) \mid Q\}$$

under the same conditions as Thm 3.7. This is how Thm 3.7 gets its name.

3.4 Hypothetical expectations in inference

First, I deal with the practical issue of incorporating hypothetical expectations into my beliefs and my inferences. A quick review of Sec. 2.1 might be helpful; I will use the same notation here.

Imagine I have assigned the value w to my hypothetical expectation $\mathbb{E}\{g(\mathbf{X}) \mid Q\}$, where $Q := q(\mathbf{X})$ for some first-order sentence $q(\mathbf{x})$. How does this become a line in the matrix G in (2.2)? Use the FTP to write out the expectation on the left of (3.3) and the probability on the right, to give

$$\sum_j g(\mathbf{x}^{(j)}) \mathbb{1}_{q(\mathbf{x}^{(j)})} \cdot p_j = w \sum_j \mathbb{1}_{q(\mathbf{x}^{(j)})} \cdot p_j.$$

Then rearrange to give

$$\sum_j (g(\mathbf{x}^{(j)}) - w) \mathbb{1}_{q(\mathbf{x}^{(j)})} \cdot p_j = 0$$

which is a row of $[G, v]$ with

$$G_{ij} \leftarrow (g(\mathbf{x}^{(j)}) - w) \mathbb{1}_{q(\mathbf{x}^{(j)})} \quad \text{and} \quad v_i \leftarrow 0.$$

This is the key thing to appreciate: hypothetical expectations (and hypothetical probabilities of course) allows me to make another type of belief assessment, which can be used to constrain my expectations of other random quantities, exactly as expressed in Chapter 2.

It is for this reason that a hypothetical expectation has to be meaningful to me directly. It has to bring something new to my inference. In many textbooks, the hypothetical probability is simply defined as

$$\Pr(A | Q) = \frac{\Pr(A, Q)}{\Pr(Q)} \quad \Pr(Q) > 0$$

and undefined otherwise. This does not create a new type of belief, it simply relabels a function of my ordinary beliefs. It does not enable me to specify, say, $\Pr(A | Q)$ directly, and then infer my value for $\Pr(A, Q)$ on the basis of my $\Pr(Q)$. To specify $\Pr(A | Q)$ directly it would have to be meaningful to me. If it was simply some function of $\Pr(A, Q)$ and $\Pr(Q)$ then there is no way that I could infer $\Pr(A, Q)$ from $\Pr(Q)$, because of circularity. To claim otherwise would be mystical.

That is why I have tried to motivate $E(X | Q)$ and $\Pr(A | Q)$ in terms of a practice—‘supposing’—that we recognise and find meaningful. Without this step, I have no grounds for thinking I can incorporate new beliefs to my inference. Then it turns out that the definition in Def. 3.2 is consistent with the basic recursive property I ascribe to ‘supposing’, and all of the usual results follow.

3.5 Hypothetical probabilities

There is nothing new to say here! Hypothetical probabilities are just hypothetical expectations. This section presents some of the standard results from the definition of hypothetical probability in Def. 3.2 and its implication in (3.2).

The following two results may be generalised in the obvious way to any finite number of random propositions.

Theorem 3.8 (Factorisation theorem).

Let P , Q , and R be random propositions. Then

$$\Pr(P, Q, R) = \Pr(P | Q, R) \Pr(Q | R) \Pr(R).$$

Proof. Follows immediately from two applications of (3.2):

$$\begin{aligned} \Pr(P, Q, R) &= \Pr(P | Q, R) \Pr(Q, R) \\ &= \Pr(P | Q, R) \Pr(Q | R) \Pr(R), \end{aligned} \quad (\dagger)$$

because $\mathbb{1}_{P, Q, R} = \mathbb{1}_P \mathbb{1}_{Q, R}$ and $\mathbb{1}_{Q, R} = \mathbb{1}_Q \mathbb{1}_R$. □

This result leads immediately to the following.

Theorem 3.9 (Sequential conditioning).

Let P , Q , and R be random propositions. If $\Pr(R) > 0$ then

$$\Pr(P, Q | R) = \Pr(P | Q, R) \Pr(Q | R).$$

Proof. Divide (†) through by $\Pr(R)$ and use (3.2) on the lefthand side. □

Then there is the very useful *Law of Total Probability (LTP)*, also known as the *Partition Theorem*. This is the probability version of conglomerability and the LIE (Thm 3.4).

Theorem 3.10 (Law of Total Probability, LTP).

Let P be a random proposition and $\mathbb{Q} := \{Q_1, \dots, Q_k\}$ be a belief partition. Then

$$\Pr(P) = \sum_{i=1}^k \Pr(P | Q_i) \Pr(Q_i).$$

Proof. Just use $g(\mathbf{X}) \leftarrow \mathbb{1}_P$ in Thm 3.4. \square

Finally, there is the celebrated *Bayes's theorem*.

Theorem 3.11 (Bayes's theorem). If $\Pr(Q) > 0$ then

$$\Pr(P | Q) = \frac{\Pr(Q | P) \Pr(P)}{\Pr(Q)}.$$

Proof. Follows immediately from (3.2),

$$\Pr(P, Q) = \Pr(P | Q) \Pr(Q) = \Pr(Q | P) \Pr(P),$$

and then rearranging the second equality. \square

There are several other versions of Bayes's theorem. For example, there is a sequential Bayes's theorem:

$$\Pr(P | Q_2, Q_1) = \frac{\Pr(Q_2 | P, Q_1)}{\Pr(Q_2 | Q_1)} \Pr(P | Q_1)$$

if $\Pr(Q_2, Q_1) > 0$. And there is Bayes's theorem for a belief partition, $\mathbb{P} := \{P_1, \dots, P_k\}$:

$$\Pr(P_i | Q) = \frac{\Pr(Q | P_i) \Pr(P_i)}{\sum_j \Pr(Q | P_j) \Pr(P_j)} \quad i = 1, \dots, k$$

if $\Pr(Q) > 0$, which uses the LTP in the denominator. And there is a Bayes's theorem in odds form,

$$\frac{\Pr(P_i | Q)}{\Pr(P_j | Q)} = \frac{\Pr(Q | P_i) \Pr(P_i)}{\Pr(Q | P_j) \Pr(P_j)} \quad i, j = 1, \dots, k$$

if $\Pr(P_j, Q) > 0$.

3.5.1 Probability Mass Functions

I extend the 'p' notation of Sec. 1.5.1 to include hypothetical probabilities, writing

$$p(\mathbf{x} | \mathbf{y}) := \Pr(\mathbf{X} \doteq \mathbf{x} | \mathbf{Y} \doteq \mathbf{y}),$$

remembering that $p(\mathbf{x} | \mathbf{y})$ is arbitrary if $p(\mathbf{y}) = 0$. As is standard, I will continue to refer to $p(\cdot | \cdot)$ as a probability mass function (PMF).

All of the standard results from Sec. 3.5 can now be expressed in terms of PMFs, remembering the conventions from Sec. 1.5.1

that functional equalities are taken to hold for all points in the product space of the free arguments. This requires some restrictions when conditioning. These restrictions are expressed in terms of the 'support' of the conditioning random quantities,

$$\text{supp } \mathbf{X} := \{x : p(x) > 0\}.$$

Note that if $\mathbf{X} := (X_1, \dots, X_m)$ then

$$\text{supp } \mathbf{X} \subset \prod_{i=1}^m \text{supp } X_i.$$

Here are the standard results in terms of PMFs:

- Factorisation theorem:

$$p(x, y, z) = p(x | y, z) p(y | z) p(z).$$

- Sequential conditioning:

$$p(x, y | z) = p(x | y, z) p(y | z) \quad z \in \text{supp } Z.$$

- Law of total probability:

$$p(x) = \sum_y p(x | y) p(y).$$

- Bayes's theorem:

$$p(x | y) = \frac{p(y | x) p(x)}{p(y)} \quad y \in \text{supp } Y.$$

All of these results can be extended to vectors of random quantities in the obvious way.