This chapter addresses the problem that Bayesian inference cannot accommodate theory change, and proposes a framework for dealing with such changes. It first presents a Bayesian scheme for inferring predictions from observations by means of statistical hypotheses. An example shows how the hypotheses represent the theoretical structure underlying the scheme. This is followed by an example of a change of hypotheses. The chapter then presents a general framework for changing hypotheses, and proposes minimisation of the distance between hypotheses as a rationality criterion. Finally the chapter discusses the import of this for Bayesian statistical inference.

The present chapter can be read independently of the preceding chapters. There is considerable overlap with other chapters on the technical introduction of Bayesian schemes, but there is particular stress on some details that have not been given attention in the foregoing. Apart from that, chapter 1 will be helpful for situating this chapter within the general plan of this thesis. Chapter 2 may help to clarify the notion of a hypothesis employed in this chapter, in particular when it comes to the idea that changing the hypotheses amounts to a change of language. Finally, chapter 3 elaborates on the idea that partitions of hypotheses can be viewed as an expression of the theoretical structure underlying the inductive predictions.

8.1 Introduction

*Fixed theoretical structure.* In what follows I am concerned with Bayesian statistical inferences. These inferences are here considered in a scheme that generates predictions by means of hypotheses: Bayesian updating is used to adapt a probability over hypotheses to known observations, and this adapted probability is further used to generate predictions over unknown observations. The hypotheses in the scheme represent the theoretical structure that underlies the predictions. However, after we have chosen these hypotheses and a prior probability over them, updating fully determines the probabilities over the hypotheses at any later stage, and thus also the predictions resulting from that. There is no room
for any further amendments to the hypotheses or to the prior probability assign-

ment over them after they have been chosen. In Bayesian statistical inference, the theoretical structure is therefore fixed.

The fixity of the theoretical structure in the above schemes is a specific form of a more general problem for Bayesianism. Within the philosophy of science it has been formulated, among others by Earman (1992: 195–198), as the problem that Bayesianism fails to accommodate theory change. But the fact that Bayesian inference is in this sense dogmatic is at the origin of many other criticisms, including the criticism of Dawid (1982) that Bayesian inference is by definition calibrated. Furthermore, as hypotheses can be considered as specific terms in the observation language, changing the hypotheses in the scheme amounts to changing the language with which the predictions are made. The same problem can therefore be seen in light of the fact that Bayesianism fails to accommodate language change, as noted by Gillies (2000) and discussed elaborately by Williamson (2003).

This chapter addresses the above problems with Bayesianism. More in particular, it proposes a way of dealing with theory change within Bayesian statistical inference. The plan of the chapter is to introduce the Bayesian scheme for generating predictions from hypotheses, to present an example of such a scheme, then to show in the context of the example how hypotheses can be changed, and finally to give a general framework for such changes.

8.2 Hypotheses, conditioning and predictions

This section describes the Bayesian scheme for making predictions, as it has been presented in several of the preceding chapters. Observations and observational hypotheses are defined in terms of an observational algebra, and degrees of belief are represented by probability assignments over this algebra. The set-theoretical underpinning may seem unnecessary in the context of a short chapter. However, as will become apparent in sections 8.5 and 8.6, the underpinning is essential for a correct understanding of hypotheses change.

Observations and hypotheses. The predictions range over possible observations $K$, a set of consecutive natural numbers, say $\{0, 1\}$. At every time $t$ we observe one number $q_t \in K$. We can represent these observations in an observational algebra. Let $K^\omega$ be the space of all infinite observation sequences $e$:

$$e = q_1q_2q_3 \ldots$$  \hspace{1cm} (8.1)
The observational algebra \( Q \), a so-called cylindrical \( \sigma \)-algebra, consists of all possible subsets of the space \( K^\omega \). If we denote the \( t \)-th element in a series \( e \) with \( e(t) \), we can define an observation \( Q_t^q \) as an element of the algebra \( Q \) as follows:

\[
Q_t^q = \{ e \in K^\omega : e(t) = q \}.
\]

(8.2)

Note that there is a distinction between the observations \( Q_t^q \) and the values of observations \( q \). The values, represented with small letters, are natural numbers. The observations, denoted with large letters, are elements of the algebra \( Q \).

In the same way we can define an element in the algebra that refers to a finite sequence of observations. If we define the ordered sequence \( e_t = \langle q_1, q_2, \ldots, q_t \rangle \), we can write

\[
E_t = \{ e \in K^\omega : \forall t' \leq t : e(t') = q_{t'} \}.
\]

(8.3)

Again, it must be noted that the small letters \( e_t \) refer to a sequence of natural numbers, while the large letters \( E_t \) denote elements of the algebra, and carry a sequence of natural numbers as argument. The argument is sometimes omitted for sake of brevity. The observations and sequences of observations are related to each other in the natural way:

\[
Q_{t+1}^q \cap E_t = E_{t+1}.
\]

(8.4)

As in this equation, I normally refer to sequences of observations with the expression \( E_t \), suppressing the reference to the sequence \( e_t \).

Observational hypotheses can also be seen as elements of the observational algebra. If we say of an observational hypothesis \( h \) that its truth can be determined relative to an infinitely long sequence of observations \( e \), then we can define hypotheses as subsets of \( K^\omega \) in the following way:

\[
H = \{ e \in K^\omega : W_h(e) = 1 \}.
\]

(8.5)

Here \( W_h(e) = 1 \) if and only if the proposition \( h \) is true of \( e \), and \( W_h(e) = 0 \) otherwise. The hypotheses can thus be arguments of the same probability functions over the observational algebra. A partition of hypotheses is a collection \( \mathcal{H} = \{ H_0, H_1, \ldots, H_N \} \) defined by the following condition for the indicator functions \( W_{h_n} \):

\[
\forall e \in K^\omega : \sum_{n} W_{h_n}(e) = 1.
\]

(8.6)

This means that the hypotheses \( H_n \) are mutually exclusive and jointly exhaustive sets in \( K^\omega \).
The Bayesian scheme. Belief states are represented by probability functions over \( Q \). They take observations \( Q^n \), sequences \( E_t \), and hypotheses \( H_n \) as arguments. The functions are defined relative to a partition \( \mathcal{H} \) and a sequence of known observations \( e_t \): the function \( p[\mathcal{H}, e_t] \) represents the belief state upon observing \( E_t \) under the assumption of a partition \( \mathcal{H} \). It can be constructed by conditioning a prior probability function \( p[\mathcal{H}, e_0] \) on the observations \( E_t \):

\[
p[\mathcal{H}, e_t](\cdot) = p[\mathcal{H}, e_0](\cdot | E_t). \tag{8.7}
\]

Because of this, we have \( p[\mathcal{H}, e_t](E_t) = 1 \). Updating the probability by simple conditioning is known as Bayes’ rule. Both the probabilities assigned to observations and those assigned to hypotheses can be updated for new observations in this way. The probability before updating is called the prior probability, and the one after updating the posterior.

To calculate the predictions, we can employ a partition of hypotheses, and apply the law of total probability:

\[
p[\mathcal{H}, e_t](Q^{t+1}) = \sum_n p[\mathcal{H}, e_t](H_n) p[\mathcal{H}, e_t](Q^n | H_n). \tag{8.8}
\]

The terms \( p[\mathcal{H}, e_t](Q^n | H_n) \) are called the posterior likelihoods of the hypotheses \( H_n \) for \( Q^{t+1} \). The prediction is obtained by weighing these posterior likelihoods with the posterior probability over the hypotheses, \( p[\mathcal{H}, e_t](H_n) \).

Both posterior probabilities of equation (8.8) can be obtained from a Bayesian update of the prior probability \( p[\mathcal{H}, e_0] \) according to expression (8.7). In this chapter the likelihoods do not change upon conditioning. Such likelihoods are sometimes called non-inductive.

\[
p[\mathcal{H}, e_t](Q^{t+1} | H_n) = p[\mathcal{H}, e_0](Q^{t+1} | H_n). \tag{8.9}
\]

That is, the observations influence the predictions only via the probability over the hypotheses. Part of the input probabilities for generating the predictions \( p[\mathcal{H}, e_t](Q^{t+1}) \) are therefore the likelihoods \( p[\mathcal{H}, e_0](Q^n | H_n) \).

The predictions are further determined by the probability assignment over the hypotheses, \( p[\mathcal{H}, e_t](H_n) \). This probability can be determined by means of the relation

\[
p[\mathcal{H}, e_t](H_n) = p[\mathcal{H}, e_{t-1}](H_n) \frac{p[\mathcal{H}, e_{t-1}](Q^n | H_n)}{p[\mathcal{H}, e_{t-1}](Q^n)}, \tag{8.10}
\]

where \( q \) equals the last number in the sequence \( e_i \). Note that the denominator \( p[\mathcal{H}, e_{t-1}](Q^n) \) can be rewritten with equation (8.8), substituting \( t = i - 1 \). Recall further that the likelihoods \( p[\mathcal{H}, e_{t-1}](Q^n | H_n) \) are in this chapter equal
for all sequences \(e_{i-1}\), as expressed in equation (8.9). The posterior probability \(p_{\{\mathcal{H},e_0\}}(H_n)\) can therefore be determined recursively by the prior probability \(p_{\{\mathcal{H},e_0\}}(H_n)\) for all \(n\), and the likelihoods \(p_{\{\mathcal{H},e_0\}}(O_i|H_n)\) for all \(n\) and \(i \leq t\). These are the other input probabilities for generating the predictions.

In sum, predictions can be generated if we assume hypotheses, their likelihoods, and a prior probability assignment over them. The prior and the likelihoods are first used to determine the posterior probability assignment over the partition. The likelihoods are then used together with this probability over the partition for generating the prediction itself. The whole construction that uses hypotheses to generate predictions is called the Bayesian scheme.

8.3 Contaminated cows

This section gives an example of a Bayesian scheme. The reader must be warned that the case presented falls short of actual scientific cases in many respects. The focus here is on the conceptual issues rather than on actual applications.

The example case. Consider a veterinary surgeon investigating a herd of cows during an epidemic, classifying them into contaminated and uncontaminated. The farmer claims that the herd has been treated with a drug that reduces the risk of contamination. It is an accepted fact about the epidemic that the average incidence rate among untreated cows is 0.4, as more than half of the cows show a natural resistance against contamination from other cows. The incidence rate among treated cows is 0.2 on average, because the drug is not always effective. The aim of the investigation is to decide whether the cows have been treated with the drug, and further to predict the incidence rate of the contamination in the herd. To enhance the dramatic impact, it may be imagined that the effect of the epidemic only shows in a slightly diminished milk quality, but that the fate of the cows depends on the incidence rate being lower than 0.3. For higher incidence rates the milk production fails to meet the quality criteria. Furthermore, the farmer is liable to legal prosecution if he has not treated the cows.

Setting up the inductive inference. The observations of the veterinary surgeon consist in test results concerning a number of cows. The result of testing cow \(t\) can be that it is contaminated, \(q_t = 1\), or that it is not, \(q_t = 0\). The test results can then be framed in the observational algebra. The vet may set up a scheme using a partition \(\mathcal{D}\) of two hypotheses, which are associated with suppositions on treatment with the drug. The hypothesis \(D_1\) is associated with the supposition
that the cows are in fact treated, while $D_0$ means that they are not. It must be noted that the suppositions are thus not linked to observations directly, since the observations only concern contamination while the suppositions concern treatment. The relation that treatment bears to the observations is given by the incidence rates for treated and untreated cows, and this relation is laid down in the statistical hypotheses $D_0$ and $D_1$. For the observational content of the hypothesis on treatment $D_1$ we may take

$$W_{d_1}(e) = \begin{cases} 1 & \text{if } f(e) = 0.2, \\ 0 & \text{otherwise,} \end{cases}$$

(8.11)

where $f(e)$ is the relative frequency of results $q_t = 1$ in the infinite sequence $e$. The hypothesis $D_0$ may be defined in a similar way using $f(e) = 0.4$. A more precise definition is that the hypotheses comprise all so-called Von Mises Kollektivs for the given incidence rates, but for present purposes the loose definition suffices.

Being sets in the observational algebra, the hypotheses can also appear as arguments in the probability functions $p_{[D,e]}$. The fact that the veterinary surgeon is undecided on whether the farmer has treated his cows can be reflected in

$$p_{[D,e_0]}(D_0) = p_{[D,e_0]}(D_1) = 0.5.$$  

(8.12)

Hypotheses on other relative frequencies, which are strictly speaking part of the partition, are thus given a zero probability. The likelihoods, for cow $t$ being contaminated, of the hypotheses that it has or has not been treated are

$$p_{[D,e_0]}(Q_1^1|D_1) = 0.2,$$  

(8.13)

$$p_{[D,e_0]}(Q_1^1|D_0) = 0.4.$$  

(8.14)

I further assume that the estimated incidence rates are not affected by the running investigations, so that equation (8.9) holds.

**Conclusions from observations and theory.** With these values in place, the veterinary surgeon can start to predict the incidence rate in the herd, and decide over the treatment efforts by the farmer. Imagine that the first five test results are positive,

$$e_5 = 11111.$$  

(8.15)

Subsequent updating on these test results yields the following probabilities and predictions:
8.4 Careless vaccination

This section shows how the hypotheses employed in the above scheme can be changed. I describe this change, and illustrate that it allows us to derive different conclusions and predictions.

Extending the example. Imagine that the veterinary surgeon becomes suspicious of the test results. After all, more than half of the cows are normally
immune. The sequence of test results must therefore be a rather unusual stochastic fluctuation on the average relative frequency of 0.4. The vet therefore decides to reconsider the inductive assumptions that underly the scheme, and to run a number of additional tests with an adapted scheme. In particular, she investigates the drug that the farmer claims to have used, and finds that it is a vaccinate with rather strict instructions for application. In most cases it works very well, even reducing the risk of contamination to 0.025, but careless use turns the vaccinate into a substance that causes a portion of 0.9 cows to be, or at least to appear, contaminated after treatment. The hypotheses that the vet wants to add to the scheme are that the drug has been used either carefully or carelessly.

Refined partition. The additional hypotheses may be collected in a separate partition $C$, with $C_1$ associated with careful, and $C_0$ with careless treatment. Both hypotheses only apply to the case in which the cows have actually been treated, $D_1$. The combined partition is $\mathcal{B} = \{B_0, B_{10}, B_{11}\}$ in which $B_0 = D_0$, $B_{10} = D_1 \cdot C_0$, and $B_{11} = D_1 \cdot C_1$. Hypothesis $B_0$ is again defined with the relative frequency of 0.4, and the new hypotheses $B_{10}$ and $B_{11}$ can be defined with 0.9 and 0.025 respectively. These three relative frequencies define the new partition.

It is notable that the hypotheses $B_{10}$ and $B_{11}$ cannot be viewed as intersections $D_1 \cap C_0$ and $D_1 \cap C_1$: judged from the definition using relative frequencies, the original set $D_1$ and both sets $B_{10}$ and $B_{11}$ are disjoint. The relation between the old and the new hypotheses is a rather different one. We must imagine that within every infinite sequence $e \in D_1$, that is, within every possible world in which all cows are treated, we make a further selection of the observations $q_t$ into those concerning cows that have been vaccinated with care, and those concerning cows that have been vaccinated carelessly. So $B_{10}$ and $B_{11}$ can be distilled from the old one by breaking up every $e \in D_1$, for which $f(e) = 0.2$, into two subrows $e_0$ and $e_1$ by means of a place selection, taking care that the relative frequencies of the two subrows are 0.9 and 0.025 respectively, and by grouping these subrows into $B_{10}$ and $B_{11}$. Because $0.025 < 0.2 < 0.9$, such place selections can always be constructed.

The likelihoods of the hypotheses may again be equated to the relative frequencies that define the hypotheses:

$$p_{(B_{10})}(Q_1^1|B_{10}) = 0.9,$$  \hspace{1cm} (8.16)

$$p_{(B_{11})}(Q_1^1|B_{11}) = 0.025.$$  \hspace{1cm} (8.17)
In order to arrive at the overall incidence rate of 0.2 for treated cows, the veterinary surgeon may further assume that a portion of 0.2 of all farmers do not treat the vaccinate with the necessary care, as $0.2 \times 0.9 + (1 - 0.2) \times 0.025 = 0.2$. I come back to this choice in section 8.6. Finally, using the probability assignment after five tests, the combined probability of treatment with the drug and the lack of care is

$$p_{[B,c_5]}(B_{10}) = 0.03 \times 0.2 = 0.006$$  \hfill (8.18)

It must be noted that with the employment of $B$, the probability over the observational algebra really undergoes an external shock: instead of allocating 0.030 probability on the set $D_1$, we now allocate 0.006 on $B_{10}$ and 0.024 on $B_{11}$.

**Different conclusions.** With these new hypotheses and the associated inductive assumptions, the veterinary surgeon can run a number of additional tests. Let us say that the next ten test results are all positive too,

$$e_{15} = 111111111111111.$$  \hfill (8.19)

Subsequent updating on these test results yields the following probabilities and predictions:

<table>
<thead>
<tr>
<th>Number of tests $t$</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>11</th>
<th>13</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{[B,c_{15}]}(B_{10})$</td>
<td>0.01</td>
<td>0.03</td>
<td>0.14</td>
<td>0.49</td>
<td>0.80</td>
<td>0.95</td>
</tr>
<tr>
<td>$p_{[B,c_{15}]}(Q_{t+1}^1)$</td>
<td>0.39</td>
<td>0.42</td>
<td>0.47</td>
<td>0.62</td>
<td>0.80</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Now the probability for $B_{10}$ approaches 1, while the predictions for a cow in the herd to be contaminated tend to 0.9. Clearly these values differ from those that were to be expected on the basis of $D$.

The conclusions expressed in these values are that the farmer did treat his cows with the drug, but that he did not apply it with the necessary care. The further conclusion is that the incidence rate of the epidemic in his herd is 0.9. Again, these conclusions are drawn from the test results in combination with the inductive assumptions of partition $B$. It is only when compared to the other members of the partition that the hypothesis $B_{10}$, which prescribes an incidence rate of 0.9, fits the test results best. For present purposes, however, it is most notable that these conclusions differ dramatically from those derivable from $D$.

Note that this is again different if we further introduce the partition $I$ on whether the test material is itself infected, and stipulate that in the combined partition $A = \{ I_0 \cdot D_0, I_0 \cdot D_1 \cdot C_0, I_0 \cdot D_1 \cdot C_1, I_1 \}$ we have $p_{[A,c_4]}(Q_{t+1}^1|I_1) = 1$,.
while slightly adapting the values for the other likelihoods. Relative to the partition $\mathcal{A}$, the priors for $\mathcal{I}$ and some further observations, the conclusion may then be that the test material is infected. However, for the partition $\mathcal{B}$ and its associated inductive assumption, the conclusions must be as indicated above.

## 8.5 A Framework for Changing Partitions

The above illustrates how we can change a partition of hypotheses during an update procedure. This section gives a general framework for such changes, and draws attention to the need for new criteria of rationality to guide them.

**Capturing hypotheses change.** On the change of partition itself, as illustrated in figure 8.1, I can be relatively brief. Let us say that the old partition $\mathcal{H} = \{H_0, H_1, \ldots, H_N\}$ consists of hypotheses $H_n$ with likelihoods

$$p_{\mathcal{H},e_1}(Q_{t+1}^g|H_n) = \theta_n^g. \quad (8.20)$$

The addition of a partition $\mathcal{G} = \{F_0, F_1, \ldots, F_M\}$ to this partition generates a combined partition $\mathcal{G} = \mathcal{H} \times \mathcal{G}$, which consists of $N \times M$ hypotheses $G_{nm} = H_n \cdot F_m$. Each of these hypotheses may be associated with a relative frequency of the observation $q$, denoted $\gamma_{nm}^q$, so that

$$p_{\mathcal{G},e_1}(Q_{t+1}^g|G_{nm}) = \gamma_{nm}^q. \quad (8.21)$$

The details of the partition change may be such that for some of the $H_n$ we have that $\gamma_{nm}^q = \theta_n^q$ for all $q$ and $m$. We can then collect the hypotheses $G_{nm}$ under the single index number $n$, as for example $B_0$ above. More in general, if two hypotheses $G_{nm}$ and $G_{n'm'}$ are such that $\gamma_{nm}^q = \gamma_{n'm'}^q$ for all $q$, we can merge them into a single hypothesis. In the extreme case in which for all $q$ the $\gamma_{nm}^q$ vary only with $m$, the change of partition comes down to a replacement of $\mathcal{H}$ by $\mathcal{G}$.

**An external shock to the probability assignment.** With the introduction of new hypotheses, the probability over the observational algebra undergoes an external shock. First, the probability over the hypotheses themselves changes. But since the new hypotheses have different likelihoods, the probability over most other elements of the algebra changes as well. It is in this chapter assumed that at the time of change $\tau$, the new probability assignment over the hypotheses observes the following restriction:

$$p_{\mathcal{G},e_\tau}((\cup_m G_{nm}) = \sum_m p_{\mathcal{G},e_\tau}(G_{nm}) = p_{\mathcal{H},e_\tau}(H_n). \quad (8.22)$$
That is, the probability assignment arrived at by updating over $\mathcal{H}$ is taken over into the new partition $\mathcal{G}$. This restriction serves to link every collection $\bigcup_m G_{nm}$ to the original hypotheses $H_n$, but it can be dropped if further details of the partition change permit it. Finally, within the limits set by this restriction, the probabilities of the hypotheses $G_{nm}$ can vary freely.

It can be noted that the change in probability due to partition change is not one that can be represented as Bayesian conditioning. Conditioning determines how to adapt probability assignments if for some observation $Q^l_t$ or $E_t$ the probability is externally fixed to 1. It is quite different to set the probability of a number of hypotheses $H_n$ to zero, and to redistribute this probability over new hypotheses $G_{nm}$. A partition change is therefore an external shock to the probability assignment to which we cannot apply Bayesian updating. Now there are many arguments to the effect that Bayesian updating is the only rational way to adapt a probability assignment to new information, but these arguments do not apply in this case, since the new information can in this case not be represented, in any straightforward manner, as an element of the algebra. It seems that the possibility of partition change necessitates new criteria of rationality, and the definition of an associated update operation.
8.6 Distance between partitions

This section answers the need for a rationality criterion and an associated update operation. In particular, it elaborates on a distance function between the old and the new partition, and shows how to minimise this distance during the partition change.

Minimising cross-entropy. Partition change may be considered as an external shock to the probability assignment over the algebra. Williamson (2003) argues that changes in the probability assignment must be conservative, that is, as small as possible, and further that such conservatism can be explicated by a minimisation of the cross-entropy distance function between the old probability $p_0$ and the new probability $p$, under the restrictions imposed by the external shock. The distance function is defined by

$$\Delta(p, p_0) = \sum_U p(U) \log \frac{p(U)}{p_0(U)},$$

where the index $U$ runs over all sets in the finite algebra over which $p_0$ and $p$ are defined. As elaborated in Kullback (1959) and Paris (1994: 120–126), minimising this distance under the external restrictions effectively minimises the information change that is induced in the probability assignment by the external shock. Interestingly, the operation of minimising cross-entropy coincides with the operation of a Bayesian update in the case that some probability $p_{H, e, t}(Q_t^i)$ is restricted to 1. It therefore accords with Bayesian statistical inference to adopt the minimisation of cross-entropy as the update operation in cases of partition change.

We are not yet done with the update operation for partition change. For one thing, the above distance function blows up if the algebra contains an infinite number of elements, as is the case for the algebra $Q$. We need to select a finite collection of elements of the algebra, for which we may then minimise the distance between the old and the new probability assignment. Note that it is not desirable to minimise the difference between the old and the new predictions. The reason for the partition change is exactly that the old predictions do not match the pattern in the observations well. And note further that the probability assignment over the hypotheses is changed deliberately, so that we cannot apply the minimisation of the distance to the assignments over hypotheses either. In sum, we have to apply the minimisation of cross entropy to a collection of elements from the observational algebra that does not emphasise the predictions or the hypotheses themselves.
As already indicated in the example, it is rather intuitive to choose a minimisation of the distance between the likelihoods of the hypotheses $H_n$ and of the associated collections $\cup_m G_{nm}$. These likelihoods fully express the hypotheses, and the distance between the likelihoods is therefore an intuitive measure for the closeness of the two partitions. A further reason for choosing the collection $\cup_m G_{nm}$ can be found in the relation between the old and the new hypotheses. Recall that the likelihoods of $H_n$ for observations $Q^t_q$ are determined by the relative frequencies of the observations $q \in K$ within the infinite sequences of observations, or possible worlds, for which $H_n$ is true. With the change of hypotheses, we effectively make a further division of these possible worlds into the hypotheses $G_{nm}$. Specifically, each infinite sequence of observations $e \in H_n$, having a relative frequency $\theta_n^q$, must be split into $M$ infinite subsequences $e_m$, having relative frequencies $\gamma_{nm}^q$, and these subsequences can then be incorporated into separate hypotheses, $e_m \in G_{nm}$. Because the hypotheses $G_{nm}$ are derived from the original hypotheses $H_n$ in this way, we may expect the relative frequency associated with the aggregate $\cup_m G_{nm}$ to be the same as, or at least close to, the original relative frequency associated with $H_n$.

A note on Kollektivs. At this point we may recall the definition of hypotheses as sets of sequences with specific relative frequencies, which is developed in chapter 2. In the context of that chapter it seems more elegant to equate hypotheses with collections of Kollektivs. Section 2.3.2 argues that there are further reasons for the definition of hypotheses by means of relative frequencies only, and these reasons can become apparent if we consider the creation of hypotheses alluded to above. Let me first admit that this creation is not a neatly defined operation yet. However, I do think that such an operation can eventually be defined, and that it then mimics the kind of epistemic move involved in choosing a new partition. Indeed, the veterinary surgeon imagines that an infinite sequence of unobserved cows is broken up into finite segments, the herds, which are then marked as being treated carefully and carelessly. These finite segments are then concatenated to render two different infinite sequences. But if the infinite sequences $e \in H_n$ are taken to be Kollektivs, we simply cannot create these different sequences $e_m \in G_{nm}$ from the single hypothesis $H_n$ by means of a place selection. We must therefore maintain that the $e \in H_n$ are not Kollektivs.

Calculations. Any hypothesis prescribes the likelihoods for infinitely many observations $Q^t_{t+t}$, associated with different times $t \geq 0$. However, these likelihoods are in this chapter constant, and it seems natural to define the distance between the partitions as the distance between the likelihoods at a single time.
For \( p_0 \) we can use the old likelihoods \( p_{[H,e_\tau]}(Q_{\tau+t}^q|H_n) = \theta_q \). For \( p \) we use the aggregated likelihoods, given by

\[
\gamma_n^q = p_{[G,e_\tau]}(Q_{\tau+t}^q|\cup_m G_{nm}) = \sum_m \frac{p_{[G,e_\tau]}(G_{nm}) p_{[G,e_\tau]}(Q_{\tau+t}^q|G_{nm})}{\sum_m p_{[G,e_\tau]}(G_{nm}) p_{[G,e_\tau]}(Q_{\tau+t}^q|G_{nm})} = \sum_m \rho_{nm} \gamma_{nm}^q.
\]

(8.24)

(8.25)

Here the \( \rho_{nm} \) are defined by the fraction in equation (8.24), so that \( \sum_m \rho_{nm} = 1 \). The \( \gamma_n^q \) are a function of these \( \rho_{nm} \).

We can now use the distance function to find the aggregated likelihoods \( p_{[G,e_\tau]}(Q_{\tau+t}^q|\cup_m G_{nm}) \) that are closest to the likelihoods \( p_{[H,e_\tau]}(Q_{\tau+t}^q|H_n) \), for any time \( t \). These distances are defined for each hypothesis \( H_n \) separately:

\[
\Delta_n(\rho_{nm} \gamma_{nm}^q, \theta_q) = \sum_q \gamma_n^q \log \frac{\gamma_n^q}{\theta_q}.
\]

(8.26)

The distance for \( H_n \) is thus a function only of the fractions \( \rho_{nm} \), which determine how the probability of \( H_n \) is distributed over the \( G_{nm} \). The update operation after a hypotheses change is to find, for every \( H_n \) separately, the values of \( \rho_{nm} \) that minimise the distance function \( \Delta_n \).

This can be employed to provide a further underpinning for the choice of the probabilities \( p_{[B,e_\tau]}(B_{10}) \) and \( p_{[B,e_\tau]}(B_{11}) \) in the example. It was stated there that the veterinary surgeon chooses these probabilities in order to arrive at the overall incidence rate of 0.2. Note that the distance between the likelihoods of \( H \) and the aggregated likelihoods of \( G \) is zero and therefore minimal if we find values for \( \rho_{nm} \) so that \( \gamma_n^q = \sum_m \rho_{nm} \gamma_{nm} = \theta_q \). In the case of the partitions \( D \) and \( B \), the equation simply becomes \( 0.9 \times \rho_{10} + 0.025 \times (1 - \rho_{10}) = 0.2 \), for which \( \rho_{10} = 0.2 \) is the solution.

**Generalisations.** It must be stressed that the present exposition does not comprise the full story on partition change. There are many cases of partition change that are not covered by the above framework, but that can in principle be treated in a similar way. One such case deserves separate attention here. The example presents a probability assignment that is not open-minded: almost all hypotheses on relative frequencies are given a zero probability. This may cause the impression that the framework for partition change can only be applied if the old probability assignment is not open-minded. It may be hard to see what other hypotheses can be added if, for instance, the prior probability already
includes all possible hypotheses on relative frequencies. However, the above framework can also be used to change a partition of all hypotheses on relative frequencies into a partition of hypotheses that concern all Markov processes. The application of the framework for partition change is thus not limited to cases in which the prior is not open-minded.

8.7 Concluding remarks

In this chapter it has been shown how we can frame a partition change, and a procedure has been provided to render this change rational, employing a distance function between the partitions. I complete the chapter with a summary and some remarks on the proposed framework in the context of Bayesian statistical inference.

The proposed framework enables us to adapt the hypotheses that function in a scheme for making predictions. By writing down the predictions in terms of a Bayesian scheme, I locate the theoretical structure underlying the predictions inside the probability assignment. Theoretical developments can therefore be framed as external shocks to the probability assignment representing the current opinions, just as new observations. I then argue that the operation that updates the assignment for the external shock is a generalised version of Bayesian conditioning, namely cross-entropy minimisation. The framework is therefore a natural extension of Bayesian statistical inference. On the whole, the chapter proposes an answer to the problem that Bayesian statistical inference cannot accommodate theory change.

The chapter may also fulfil a role in an older discussion between inductivists and Popperians: it basically shows how we can incorporate a notion of conjecture within an inductivist setting. It is a typical feature of Carnapian inductive logic that there is no room for an explicit formulation of inductive assumptions, as such assumptions are part and parcel of the choice of language. Conjectures can therefore not be captured within a Carnapian logic. However, the above discussion associates the premisses with the hypotheses used in the Bayesian scheme, and further allows us to change them. It provides a truly nonmonotonic inductive Bayesian logic, in the sense that besides the set of available observations, also the inductive assumptions may be altered along the way. This chapter is thus a first step in generalising inductive Bayesian logic to incorporate changes in the projectability assumptions.