This chapter proposes a frequentist interpretation of statistical hypotheses. In this interpretation, statistical hypotheses are associated not with probability models over a whole algebra $\mathcal{Q}_0$, but rather with strict subsets of the observation so-called $\sigma$-algebra $\mathcal{Q}$, the extension of $\mathcal{Q}_0$. The Bayesian scheme can then be taken as a further specification of the Carnapian scheme, in which hypotheses appear as convenient extensions to the observation language.

The chapter first discusses statistical hypotheses in the logical picture, and indicates how a frequentist interpretation can elucidate their use. Then it defines a specific set of statistical hypotheses, for which such an interpretation can indeed be given. Under this interpretation the Bayesian scheme is seen to be formally, and not just extensionally, equivalent to the Carnapian scheme: both schemes take a completely specified probability $p$ over a single observational algebra as input. Some considerations on hypotheses and models complete the chapter.

This chapter presupposes chapter 1 as a whole. It is itself useful reading for chapters 3 and 8.

2.1 Statistical hypotheses

This section shows that statistical hypotheses are identical in terms of the observational algebra, and different only in the probability models associated with them. It is thus natural to view the probability over hypotheses as a second order probability, but such a probability seems at odds with the logical picture sketched in the preceding chapter. On the other hand, if this introduction of second order probability is avoided, more positive reasons for adopting the frequentist view remain.

Statistical hypotheses: algebraic or probabilistic? The Bayesian scheme of the preceding chapter employs the algebra $\mathcal{H} \times \mathcal{Q}_0$. Hypotheses are identified with sets $H_j = \{h_j\} \times \mathcal{Q}_0$, which each consist of the same observational algebra $\mathcal{Q}_0$, while their elements are labelled $h_j$ differently. In terms of algebraic structure, there is nothing in the hypotheses $H_j$ to tell them apart. That is, they all have
the same observational content. This reflects the fact that purely statistical hypotheses are consistent with any finite sequence of observations and therefore cannot be verified or falsified. But it may then seem rather strange that we are at the same time using observations to decide between statistical hypotheses. In technical terms, if within a sequence of observations $E_t$ the hypotheses coincide, these observations cannot be used to distinguish between the hypotheses. The use of conditioning to decide between hypotheses seems to lack intuitive basis if the hypotheses cannot somehow be told apart in the observational algebra.

Clearly the hypotheses $H_j = \{h_j\} \times Q_0$ are distinct in another aspect: the probability models defined over them are different. This difference in probability models is what connects the labels $h_j$ to the observations, and thus provides the hypotheses with distinct observational content. In other words, hypotheses indeed overlap in the algebra, but within specific sequences of observations $E_t$ the probabilities assigned to the hypotheses differ. However, this seems to land us in another puzzle. Recall that the Bayesian scheme offers a probability assignment $p[e_0]$ over the hypotheses. If the hypotheses are only distinct because of the probability models, it seems that we are in fact assigning probabilities to these probability models. This seems to turn the probability $p[e_0](H_j)$ into a kind of second order probability assignment, ranging over models $M_j$ and not just over the hypotheses $H_j$. And we may then wonder how this squares with the Kolmogorov definition of probability, in which probability is only assigned to elements of an algebra.

All this is portrayed as problematic a bit too eagerly. After all, in the Bayesian scheme the probabilities $p[e_t](H_j)$ are assigned to sets $H_j = \{h_j\} \times Q_0$, which are elements of the algebra $H \times Q_0$, just as the observations $Q_{t+1}$ and $E_t$. The sets $H_j$ may be identical in terms of the observational algebra, but nevertheless they are different sets if only for the mere fact of their different labelling. The fact that, apart from the labelling, these sets differ solely because the probability over the observations within them is different must not distract us too much. Furthermore, even if it is conceded that the probability assignments to hypotheses are essentially of second order, there is nothing inconsistent or flatly wrong in introducing such probability assignments. The use of second order probabilities has many proponents, as for example Sahlin (1983). Moreover, second order probabilities are at the heart of so-called expert systems, as discussed by Gaifman (1986), van Fraassen (1989) and others.

Motivating a frequentist semantics. The foregoing leads up to a number of reasons for developing an alternative interpretation of statistical hypotheses.
Consider the case in which the probability over hypotheses is taken as second order. Recall that in the logical picture, the premises of inductive arguments consist of observations together with a probability assignment \( p_{[e]} \). In this picture, the assignment may be read as a generalised truth valuation, which comprises a continuum of truth values. But the probability assignment \( p(h_j) \) is taken as some kind of second order probability over the models, \( p_{[e]}(p(h_j)) \), and not as a probability on the level of sets, \( p_{[e]}(H_j) \). This move of taking probability assignments as arguments of the probability assignment runs parallel to taking propositions on truth valuations in classical logic as propositions themselves. And it is well known that this opens the door for problems such as the liar paradox. As an example, the inconsistency of Bayesian updating as revealed in Maher (1993:105-29) crucially depends on the use of probability assignments within statements that are themselves assigned probability. In the logical picture sketched above, it seems much safer, as well as more in line with classical deductive logic, to determine premisses in terms of a single probability assignment.

Hypotheses were in the preceding chapter introduced in this way: they were presented as sets in the algebra, \( H_j = \{h_j\} \times \mathcal{Q}_0 \), and not as probabilistic models. However, also in this presentation, a number of reasons for an alternative interpretation may be advanced. Firstly, note that the logical picture of chapter 1 stays close to the empiricist roots of inductive logic. As suggested in that chapter, it is too much to strive for the derivation of a completely analytic probability assignment from the structure of the language. But I do feel that, where possible, we must attempt to associate probability assignments to observations, or more specifically, to elements in an observational algebra. It seems to me that the Bayesian use of hypotheses as separate observational algebras, \( H_j = \{h_j\} \times \mathcal{Q}_0 \), removes us unnecessarily far away from the empiricist roots of inductive logic.

Secondly, there is a rather natural way in which the statistical hypotheses can be given an interpretation as statements in an observational algebra after all. This interpretation is based on frequentism. The idea is to connect statistical hypotheses \( h_j \) to sets of infinite sequences \( e \) that have the probabilities \( p_{[e]}(h_j) \) as their limiting relative frequencies. As will become apparent, not all statistical hypotheses lend themselves for such an interpretation. But for those hypotheses that do allow for a frequentist interpretation, the theoretical interpretation of hypotheses appears as conceptual decadence: it employs a multitude of algebras \( \mathcal{Q}_0 \), where in fact we can do with just one extended algebra, or \( \sigma \)-algebra,
CHAPTER 2. A FREQUENTIST SEMANTICS OF HYPOTHESES

$Q = \sigma(Q_0)$. In my preferred terminology: the frequentist view of von Mises can be used as a razor to cut the beard that Kolmogorov is sporting.

The general idea of this chapter is that statistical hypotheses, or probability models, need to be given some kind of empirical content. Without such a content, they cannot be given a natural place in an empiricist inductive logic. The next two sections provide this observational content for a specific class of hypotheses. The concluding section of this chapter will return to the advantages of this alternative interpretation of statistical hypotheses.

2.2 A restricted class of hypotheses

This section gives a formal definition of a class of probabilistic hypotheses, associated with a specific collection of models. Only statistical hypotheses from this restricted class can be connected to elements of the observational algebra. The definition of the class is based on a frequentist interpretation of probability.

Relation with Von Mises. Von Mises (1928) introduced the so-called frequentist interpretation of probability as part of a systematic study into statistical phenomena. The interpretation is not an attempt to derive probability models from finite or even infinite sequences of observations. Rather it is an attempt to specify what probability means by defining this notion in terms of specific infinite sequences of observations $e$, the so-called Kollektivs. In the following I employ the frequentist interpretation in the exact opposite direction: I start with defining a certain class of statistical hypotheses, associated with certain probability models, and after that I define the hypotheses as sets of specific sequences $e$ by employing the frequentist interpretation of probability. I thereby leave aside many of the subtleties involved in the frequentist interpretation itself. It must be stressed that I do not attempt to justify the frequentist interpretation of probability, or to somehow prove its adequacy. Rather I am using the frequentist interpretation to provide an alternative semantics of the hypotheses $h_j$. This alternative semantics does not associate hypotheses with the complete observational algebra $H_j = \{h_j\} \times Q_0$, but rather with strict subsets $H_j \subseteq K^\omega$, and thus with elements in the extended algebra $Q$.

Defining statistical hypotheses. Before making this restricted class of statistical hypotheses precise, let me sketch the idea behind it. Every statistical hypothesis in it is associated with a probability model, and thus prescribes, for a range of possible circumstances or states, a probability for the observations, denoted $Q_{i+1}^t$. The states must at every position $t$ in the string be determined by the
observations within $e$ that are already given, $e_t = (e(1), e(2), \ldots, e(t))$, and they must further occur infinitely often in the infinitely long sequence of observations $e$. For every such $e$, the probability of $Q^q_{t+1}$ in some state is associated with a relative frequency of $q$’s occurring in this state. The subset of a hypothesis can then be identified with the set of infinitely long sequences $e$ for which all the relative frequencies associated with the states match the probability model.

The definition of this class of statistical hypotheses has two ingredients: a set of identity functions marking the states, and a set of probability vectors, each of them associated to one selection function annex state. The identity functions serve to characterise the states in which the corresponding probability vector applies.

**Definition** Let $w(e_t)$ be a function assigning a natural number $\{0, 1, \ldots, M\}$ to all sequences $e_t$. Further let $\theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$ be a set of fixed probability vectors $\theta_m$ of which the components $\theta_{qm} \in [0, 1]$ satisfy $\sum_{q \in K} \theta_{qm} = 1$ for each $0 < m \leq M$. Then the statistical hypothesis $h_{w\theta}$ determines a, possibly partial, probability model

\[ p_{h_{w\theta}}(Q^q_{t+1}|E^e_t) = \theta_{qw(e_t)}. \tag{2.1} \]

If $w(e_t) = 0$ the conditional probability remains undefined.

Hypotheses that can be written down in this way are called statistical. If all $\theta_{qm} > 0$, the hypothesis is called purely statistical. Note that these hypotheses do not yet specify the class of hypotheses that may be interpreted in a frequentist manner.

Some remarks may help to clarify the foregoing. First, the hypotheses $h_{w\theta}$ distinguish different states $w(e_t) = m$, depending on the sequence of observations $e_t$. Further, they associate with each of these states a probability vector $\theta_m$ ranging over possible next observations $q \in K$. For any sequence of observations $e_t$, not more than one such probability vector is chosen by the hypotheses. Note also that the probability model can still be partial, because the function $w$ need not assign a number $m > 0$ to all sequences $e_t$. Below I define two further restrictions on statistical hypotheses which clarify the point of this complication. Finally, recall that if every $e_t$ is assigned an $m > 0$, the probability model is defined completely. From this we can derive a complete prediction rule.

**Restrictions for frequentist hypotheses.** I now impose two further restrictions, which, together with the above, define the intended class of hypotheses $\mathcal{F}$. Recall from section 1.4 that with the direct probabilities we can recursively derive
values for all $p(E_t^e)$. A sequence $e_t$ is deemed possible by the probability model of $h_{w\theta}$ if and only if $p(h_{w\theta})(E_t^e) > 0$. The first restriction is that there may at most be finitely many sequences $e_t$ which are deemed possible by $h_{w\theta}$, but which nevertheless have $w(e_t) = 0$. If we collect the corresponding $E_t$ in a special set, denoted $E_{w=0}$, we can formulate this requirement in the following way:

$$\forall t > 0, \forall E_t \notin E_{w=0} : p(h_{w\theta})(E_t^e) > 0 \Rightarrow w(E_t^e) > 0.$$ \hspace{1cm} (2.2)

This means that $h_{w\theta}$ must assign a probability to the observation $Q_{q+1}$ in all those cases in which the observation set $E_t$ that preceded the observation $Q_{t+1}$ has nonzero probability and does not belong to $E_{w=0}$. Note that this presupposes that the probabilities for all observations $Q_t$ for which $E_t \subset Q_t$ were also nonzero. This complicates the requirement, but it does not present any real difficulty.

The second restriction is that in any possible infinite string $e$, all states $m$ are repeated infinitely often:

$$\forall m, \forall E_t' : \exists E_t^e \subset E_t' : p(h_{w\theta})(E_t) > 0 \land w(e_t') = m.$$ \hspace{1cm} (2.3)

This is to make sure that it makes sense, eventually, to talk of relative frequencies of observations in all the states. The class of frequentist statistical hypotheses comprises all statistical hypotheses, as defined with the above definition, that satisfy restrictions (2.2) and (2.3). This class of hypotheses covers a particular subset of possible statistical hypotheses. The identification of statistical hypotheses with elements $H_j \in \mathcal{Q}$ only works for this limited class.

An example hypothesis. To get to know the class of frequentist hypotheses, let me consider the example of chapter 1 again. In that example we have $K = \{0, 1, 2\}$, referring to observations of the empty pond, ducks and a tiger. Now recall the hypothesis $h$ on tigers hunting ducks. It may be defined in the terms of the function $w$, given in

$$w(e_t) = e_t(t) + 1$$

and in terms of probability vectors $\theta$, here consisting of nine components:

$$\theta_1 = \langle \frac{2}{3}, \frac{1}{3}, 0 \rangle,$$
$$\theta_2 = \langle \frac{1}{6}, \frac{1}{3}, \frac{1}{2} \rangle,$$
$$\theta_3 = \langle 1, 0, 0 \rangle.$$  

In words, this hypothesis states that there are three possible cases. If a tiger has not appeared in the last observation, $e_t(t) \neq 2$ and if there are no ducks in the
2.2. A RESTRICTED CLASS OF HYPOTHESSES

pond, $e_t(t) \neq 1$, they may appear with a chance of $\frac{1}{3}$ while the pond may stay empty with a chance of $\frac{2}{3}$. If a tiger has not appeared in the last observation and if there are ducks, they may stay with a chance of $\frac{1}{3}$, leave with a chance of $\frac{1}{6}$, and a tiger may appear with a chance of $\frac{1}{2}$. If, finally, a tiger has appeared in the last observation, the pond stays empty for one time unit with certainty.

It must be noted that not all sequences of observations $e_t$ are assigned a positive probability. Specifically, sequences such as $e_3 = \langle 0, 0, 2 \rangle$ or $e_5 = \langle 0, 1, 2, 1, 0 \rangle$ are assigned zero probability in the probability model of $h_1$. Thus frequentist hypotheses are not necessarily purely statistical. Note also that the probability model associated with $h_1$ is complete, because we have $m(e_t) > 0$ for all $e_t$. But this need not always be the case. For hypotheses in the frequentist class there can always be some finite number of sequences $e_t$ for which the hypothesis does not prescribe probabilities. The above hypothesis $h$ does not illustrate this possibility, but I return to it in chapter 3.

The reach of frequentist hypotheses. The class of frequentist hypotheses comprises many more hypotheses like $h$. There are hardly any restrictions on what kind of statements may be used as hypotheses. For example, they also include the formal equivalent of the statements that tigers operate alone, that ducks wander in packs, that when we see a duck a tiger is not far away, and any other such statement. The only restriction for the hypotheses is that their formal equivalent must be of the form presented above. As will be argued below, this comes down to the requirement that the probability models have an observational content, so that they can be identified with an element in the observation algebra.

On the other hand, many hypotheses cannot be deemed frequentist. As an example, consider the hypothesis that as a result of hungry tigers the number of ducks decreases:

$$\theta_{10} = 1 - \frac{1}{3} e^{-\rho t}, \quad (2.4)$$

$$\theta_{11} = \frac{1}{3} e^{-\rho t} \quad (2.5)$$

In a hypothesis with probabilities that change in such a way, there are no repeatable states with fixed probabilities. Another example is presented by the Carnapian $\lambda_\gamma$ rule, conceived as a probability model. The probability for some $Q_{t+1}$ given earlier observations $E_t$ are according to this rule determined by two statistics, to wit, the index $t$ and the fraction of the number of earlier occurrences of $q$ in $e_t$, denoted $t_q$. With every fraction $\frac{t_q}{t}$ and value of $t$ we can effectively associate another state $m$, but depending on $e$ there may be infinitely many of
such states, and moreover, some of these states are not repeated infinitely often. So the Carnapian $\lambda\gamma$ rule is not frequentist either.

While some hypotheses are not frequentist because their probabilistic models cannot be associated with limiting relative frequencies, other hypotheses are excluded because they are more specific than what limiting relative frequencies allow us to express. An example of the latter kind is presented by the so-called constituents in the $\alpha\lambda$-system of Hintikka (1966). To illustrate, consider the constituent $H_{\neg 2} = \{e : \forall i(e(i) \neq 2)\}$, which for obvious reasons may be called duck heaven. At first sight it may seem that this constituent is covered by the union of all statistical hypotheses that assign a zero probability to the observation of a tiger, $q = 2$. But the Hintikka-constituent $H_{\neg 2}$ is more specific, because it does not only mean that the limiting relative frequency for tigers in the $e$ included in $H_{\neg 2}$ must be zero, but also that in these sequences $e$ there are no tigers at all. The sequence $e = 012000\ldots$ has zero limiting relative frequency for 2, but it is not part of the Hintikka constituent. Because of this notorious measure-zero gap, the Hintikka-constituents are not frequentist, even while such constituents seem to be among the most basic patterns at hand.

In sum, it appears that the class of observational patterns is wider than the class covered by the notion of frequentist statistical hypothesis. Against this, one can also argue that, for example, Hintikka constituents are not observational patterns at all, or in any case much less observational than their frequentist variants.

2.3 Hypotheses as elements of $Q$

This section presents an interpretation of hypotheses as elements of the extended observation algebra $Q$. After that it briefly discusses the relation between these elements and the Kollektivs of Von Mises, and it elaborates on the notion of a partition.

2.3.1 Definition of the elements $H_{w\theta}$

Hypotheses as sets of infinite sequences. We are now in the position to define the set $H_{w\theta}$ that is associated with a frequentist hypothesis $h_{w\theta}$. Central to this definition is the identification of probabilities and relative frequencies. A
The relative frequency of some result is defined by the following:

\[ W_{qi}(e) = \begin{cases} 1 & \text{if } e(i) = q, \\ 0 & \text{otherwise}, \end{cases} \]  
(2.6)

\[ f_q(e) = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} W_{qi}(e). \]  
(2.7)

The frequentist interpretation is thus used to translate probabilities on observations, as prescribed by a hypothesis, into properties of infinite strings of observations. It is then possible to connect a probability assignment to a set of all those \( e \) for which the above relative frequencies exist, and for which they have the matching values.

To pin down the relative frequencies of observations occurring in some state determined by \( w(e_t) = m \), we may define for every \( e \) the subsequence \( e^m \) of all those observations \( q_{t+1} \) following the positions \( t \) at which \( w(e_t) = m \). First define \( I_m(e_i) = 1 \) if \( w(e_i) = m \), and \( I_m(e_i) = 0 \) otherwise. Then define

\[ s_m(e, t) = I_m(e_t) \sum_{i=1}^{t} I_m(e_i) \]  
(2.8)

with \( e_i = \langle e(1), e(2), \ldots, e(i) \rangle \) the first \( i \) entries of \( e \). Then \( s_m(e, t) \) is a sequence that has an increasing number on positions \( t \) where \( w(e_t) = m \), and 0 on all other positions \( t \). Then define

\[ e^m(s_m(e, t)) = e(t + 1). \]  
(2.9)

for all \( s_m(e, t) > 0 \), while \( e^m(0) \) remains undefined. The resulting sequence \( e^m \) contains exactly those observations made in the state \( m \).

The set which corresponds to a hypothesis \( h_{w\theta} \in \mathcal{F} \) can now be obtained by selecting all those \( e \in K^\omega \) for which all the subrows \( e^m \) have exactly \( \theta_{qm} \) as the relative frequencies of the observation results \( q \). Formally,

\[ H_{w\theta} = \{ e : \forall m, q \ [f_q(e^m) = \theta_{qm}] \} \]  
(2.10)

This is a subset \( H_{w\theta} \subseteq K^\omega \). Note also that only the statistical hypotheses \( h_{w\theta} \) that comply to restriction (2.3) can be identified with such a strict subset. This is because the definition of the relative frequencies \( f_q \) only works for subsequences \( e^m \) that have infinite length, and because these subsequences have infinite length only if (2.3) is fulfilled.
To illustrate the foregoing, consider the hypothesis \( h_0 \) of the hunting example, which is associated with the probability model

\[
p_{[h_0]}(Q^*_t \mid E_t) = \begin{cases} 
\frac{9}{10} & \text{if } q = 0, \\
\frac{9}{100} & \text{if } q = 1, \\
\frac{1}{100} & \text{if } q = 2.
\end{cases}
\]  

(2.11)

For this probability model there is no need for distinguishing different states \( w(e_t) \). The corresponding hypothesis may therefore be defined as an element \( H_0 \) quite easily:

\[
H_0 = \{ e : f_0(e) = \frac{9}{10} \land f_1(e) = \frac{9}{100} \}.
\]  

(2.12)

To this element we can now assign a probability \( p_{[e_0]}(H_0) \). Furthermore, within the set of sequences \( H_0 \) the probability of the observations is fixed by the model.

**Extending the observation algebra.** Statistical hypotheses are in the above presented as subsets of \( K^\omega \), but they are not made part of any observational algebra yet. I now discuss the extension of the observation algebra \( Q_0 \) that is needed for accommodating hypotheses as elements of it.

By defining frequentist hypotheses as strict subsets of the space \( K^\omega \), I am providing them with something of an observational content. But this content is not observational in the ordinary manner. Note that any finitely decidable observational hypothesis \( h_w \) can be associated with the element \( H_w \) of the finite observation algebra \( Q_0 \). As an example, the hypothesis that more than half of the first \( n \) observations have the result \( q \) can be decided within \( n \) observations. Hypotheses \( H_{w\theta} \) are not finitely decidable in this way. That is, the probability models prescribed by the frequentist hypotheses cannot be verified or falsified by any finite sequence of observations. Therefore frequentist hypotheses are not part of the finite observation algebra \( Q_0 \).

With the above definition in place, however, we can associate the hypotheses \( h_{w\theta} \) with an element of the \( \sigma \)-algebra \( Q \), the infinite extension of the observation algebra \( Q_0 \). A hypothesis \( H_{w\theta} \) is a so-called tail event in this algebra. It is an event in the observation algebra whose occurrence can only be verified or falsified at infinity. This corresponds to the fact that the hypothesis \( h_{w\theta} \) is not finitely decidable, but that it is, in the vocabulary of Kelly (1996), refutable in the limit. It may be noted that hypotheses that are higher up in Kelly’s hierarchy of decidability can still be associated with elements of an algebra \( Q \). Hypotheses may also be gradually refutable or verifiable, and they may have an even more complicated structure. Moreover, Bayesian updating can perfectly
well accommodate hypotheses that are undecidable to various degrees. However, the frequentist semantics proposed here restricts hypotheses to ones that are either gradually refutable or gradually verifiable.

With frequentist hypotheses as elements in the observational algebra, the hypotheses need not be treated as probability models over the observational algebra anymore. This means that we can assign probability to hypotheses just as we can assign probability to observations, which accords well with the empiricist roots of inductive logic. Moreover, the frequentist view on hypotheses leads to a picture in which the inductive inferences are all made from a single probability assignment over a single algebra $Q$, and in which there is no need for second-order probability. This is discussed further in section 2.5. Note also that the proposal to view frequentist hypotheses in terms of a partition of the $\sigma$-algebra shows similarities to the formal characterisation of Hintikka-constituents in terms of a partition of the algebra by Kuipers (1978). As made clear in the preceding section, there are differences between statistical hypotheses and these constituents, but the general idea may very well be the same.

Finally, and most importantly, note again that frequentist hypotheses are not part of the finite algebra $Q_0$ that is used in the Carnapian scheme. Statistical hypotheses can be used as the result of an enrichment of the algebra, or alternatively, observation language. This enrichment accounts for a larger expressive force that the frequentist hypotheses allow us. As will become clear in chapter 3, this enlarged expressive force is one of the driving forces behind the conceptual innovations that this thesis offers.

2.3.2 Collections of Kollektivs

I now deal with the relation between the frequentist interpretation of probability and the above definition of hypotheses as elements in $Q$. More specifically, I investigate the connection between the set of infinite sequences $H_{w\theta}$ and collections of so-called Kollektivs.

*Infinite sequences as Kollektivs.* Following the discussions in Von Plato (1994) and Van Lambalgen (1987), a Kollektiv is a specific infinite sequence of observation results $e$. Two properties define a Kollektiv: the limiting relative frequencies of the observations in $e$ must exist, and it must be impossible to select, with some fixed procedure, positions within the sequence $e$ such that the selected subsequence has different limiting relative frequencies. This latter property has become known as the so-called law of excluded gambling systems, which nicely expresses the idea behind the property: when selectively gambling
on results in a Kollektiv, we cannot find a gambling procedure that changes the probabilities for any of the results. Or in again other words, apart from the patterns fixed by the relative frequencies, there is no further weak pattern in the observations.

We can use this notion of Kollektiv to elaborate the above definition of frequentist hypotheses. Recall that the hypotheses themselves already present a specific selection procedure, namely \( w(e_t) \). But within the subsequences \( e^m \) created with this selection, the notion of Kollektiv becomes applicable. As a start, we can characterise the hypothesis \( H_{w\theta} \) as sets containing all those sequences \( e \) of which the subsequences \( e^m \) are Kollektivs associated with the probabilities \( \theta_{qm} \). However, this is not a suitable characterisation. The definition of the hypotheses does not preclude the existence of further selections within the subsequences \( e^m \), within which the relative frequencies are different from the probabilities \( \theta_{qm} \).

To see this, consider hypothesis \( h_0 \) above, which prescribes a single vector of probabilities for the whole sequence \( e \). Because this is the only criterion for membership of the corresponding element \( H_0 \), the hypothesis \( H_0 \) also includes sequences \( e' \) in which every even indexed observation is of an empty pond with certainty, so that the relative frequencies for even observations are simply \( \langle 1, 0, 0 \rangle \), while the relative frequencies of the results on the odd positions are \( \langle \frac{8}{100}, \frac{18}{100}, \frac{2}{100} \rangle \). The resulting relative frequencies in such sequences \( e' \) is then in accordance with the probabilities prescribed in the model for \( h_0 \), while the sequences \( e' \) are not Kollektivs for these probabilities.

It is perhaps appealing to sharpen the definition of hypotheses as subsets in \( K^\omega \), and to include only the sequences whose subsequences \( e^m \) are Kollektivs for the corresponding probabilities \( \theta_{qm} \). This involves formalising the law of excluded gambling systems. The notion of admitted selection procedure can be given a proper mathematical formulation by means of recursive functions, which is here employed implicitly as domain for the function \( w \). However, sharpening the definition of hypotheses to include only the Kollektivs involves a more detailed treatment of these recursive functions, which leads us too far away from the main line of this chapter. Moreover, there are independent reasons for preferring the looser definition given in the foregoing. The reason lies in the possibility of finding further structure in the observation results. They are made explicit in chapter 8.

In the following hypotheses \( H_{w\theta} \) are sets of sequences \( e \) whose subsequences \( e^m \), defined with a function \( w \), have limiting relative frequencies matching the probability model, \( \theta_{qm} \). Some of these subsequences are Kollektivs for these
probabilities, but others are Kollektivs of a more complicated probability structure. So frequentist hypotheses are composed of collections of Kollektivs.

Other intentions than Von Mises. As a last remark on hypotheses and frequentism, let me stress again that the use of frequentist notions here is opposite to von Mises original use of it. For von Mises the emergence of Kollektivs from sequences of actual observations was an empirical matter, which has to do with statistical phenomena. In this chapter Kollektivs and limiting relative frequencies are a purely formal tool. But more importantly, the aim of von Mises was to use these Kollektivs for understanding probabilities, that is, to interpret the notion of probability with these Kollektivs. In this chapter, by contrast, I give priority to the probability models. The use of the frequentism is only to provide an interpretation of these probability models in an observation algebra. Consequently, the interpretation is only given after the models have been specified. In view of this it is natural that, next to frequentist probabilities, we can also use the subjectively interpreted probabilities over the hypotheses.

2.3.3 Partitions

The remainder of this section discusses the notion of a partition. It may be recalled from chapter 1 that a Bayesian scheme employs collections of hypotheses, which were there called partitions. The foregoing only shows how, within a restricted class, we can construct strict subsets representing these hypotheses. But now we can also make clear in what way the collections of hypotheses form a partition. That is, the hypotheses themselves can be said to partition the observational algebra, meaning that they can form a collection of mutually exclusive and jointly exhaustive sets in $K^\omega$.

A patchwork of hypotheses. To see the general idea, consider a collection of hypotheses $H_{w0}$ based on some fixed selection function $w$, but with different probability vectors. For any such collection, we may define a complete partition of the observation algebra, by adding hypotheses with the same selection function $w$, but with probability vectors $\theta$ that are not covered in the collection yet. In other words, the frequentist interpretation allows to cover the whole algebra with a patchwork of hypotheses.

Let me elaborate this with an example. Consider again the hypothesis $h_0$, which has the selection function $w(e_t) = 1$ for any $e_t$ and a single probability vector $\theta = \langle \frac{2}{11}, \frac{9}{11}, \frac{1}{11} \rangle$. Consider the alternative hypothesis $H'_0$, which uses the same selection function but prescribes probabilities $\theta' = \langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle$. The two sets
CHAPTER 2. A FREQUENTIST SEMANTICS OF HYPOTHESES

\( \text{H}_0 \) and \( \text{H}'_0 \) are then mutually exclusive: sequences \( e \in \text{H}_0 \) have limiting relative frequencies that differ from the probabilities of \( \theta' \), and vice versa. However, the probability vectors \( \theta \) and \( \theta' \) are just two elements from the set of all possible vectors, \( C = \{ \theta : \sum_{q} \theta_q = 1 \} \). There are many more hypotheses with a uniform selection function, and the hypotheses \( \text{H}_0 \) and \( \text{H}'_0 \) are therefore not jointly exhaustive. A partition must minimally include all hypotheses \( \text{H}_{\omega \theta} \) with the selection function \( w(e_t) = 1 \) that have a probability vector from \( C \).

But we are not ready with defining the partition. Even when taken together, the hypotheses \( \text{H}_\theta \) with \( \theta \in C \) are not jointly exhaustive. Some sequences \( e \) do not have limiting relative frequencies for the observations at all. As an example, take the sequence \( e = 010011000000111111 \ldots \) in which the result 2 does not occur, the number of consecutive 0’s always equals the total number of observations preceding those 0’s, and the number of consecutive 1’s equals the number of consecutive 0’s that precede it. The observed relative frequencies of this sequence will keep fluctuating between \( \langle \frac{3}{4}, \frac{1}{4}, 0 \rangle \), which is reached after each package of consecutive 0’s, and \( \langle \frac{1}{2}, \frac{1}{2}, 0 \rangle \), which is reached after the consecutive 1’s. On the whole, there is no limiting relative frequency. Therefore, only if we also provide a hypothesis \( \text{H}_{\neg \theta} \) that contains all sequences \( e \) for which the limiting relative frequencies \( f_q(e) \) are not all defined, we can define a real partition of the space \( K^\omega \).

In sum, a Bayesian scheme with the hypotheses \( \text{H}_0 \) and \( \text{H}'_0 \) involves a partition \( C = \{ \text{H}_{\neg \theta}, \{ \text{H}_\theta \}_{\theta \in C} \} \). In this partition we may then assign zero prior probability to all hypotheses other than \( \text{H}_0 \) and \( \text{H}'_0 \) to return to the original collection.

Generalisations. Two further remarks on partitions conclude this section. First, note that partitions, as introduced above, can easily be generalised. Every admitted selection function \( w \) leads to a specific generalised partition. Such partitions generate predictions that are, as it is called, partially exchangeable. To give an intuitive idea, a partition based on a selection function \( w \) results in predictions that are invariant under permutations of any two observations \( q_{t+1} \) and \( q'_{t+1} \) as long as we have \( w(e_t) = w(e_{t'}) \) for the observations preceding these observations. The notion of exchangeability is more elaborately discussed in chapter 3. For partial exchangeability I refer to De Finetti (1972) and Diaconis and Freedman (1980). It leads us too far away from the aim of this chapter to discuss it here.

Second, we may also consider partitions in which more than one selection function is used. As an example, we may want to add the hypothesis \( \text{H}_1 \),
discussed at the start of subsection 1.3.1, in the partition of hypotheses $C$, which is based on the uniform selection function. To do this, we must first find the vector $\theta \in C$ that results from the probability model of $H_1$. We must refine this specific hypothesis $H_\theta$ in the partition $C$ into the hypothesis $H_\theta^* \cup H_1$. The hypothesis $H_\theta^*$ contains all those infinite sequences that have the limiting relative frequencies $\theta$, except for those sequences $e$ that also have the more complicated pattern described in $h_1$. If we employ the adapted hypothesis $H_\theta^*$, we may then simply add $H_1$ to the partition $C$. Clearly, many much more complicated combinations of hypotheses may be united in a single partition in the same way. I hope that the example here suffices to suggest the general idea.

2.4 Objective probability models

Traditionally the frequentist interpretation of probability serves to describe physical probability or, in one word, chance. In the context of inductive logic chances are used mainly to give an objective content to epistemic probability. Below it will be discussed how this use of the frequentist interpretation relates to hypotheses and probability models in inductive logic. It turns out that chances connect naturally to the definition of hypotheses as elements in $Q$.

2.4.1 Objective probability

*Probability models as models of the world.* Now that we have a frequentist interpretation of statistical hypotheses $H_{w\theta}$, it is natural to link the probability models $M_{w\theta} = (Q_0, p_{[h_{w\theta}]}$) to models of the world. The probabilities are then interpreted as physical probabilities or chances. Chance models may be taken as weakened versions of deterministic models, in which the probabilities are fixed to 0 or 1 and thus specify a single string $e$ in the observational algebra. An example of a deterministic model is that directly after the appearance of a pack of ducks a tiger appears, after the tiger the ducks hide for one time unit, after which the ducks appear again, and so on. The associated sequence is $e = 0120120120120\ldots$, assuming an empty pond at the onset.

Probability models can be viewed in very much the same way. In the hunting example, we can say that at any time the chances of there being an empty pond, a pack of ducks or a tiger are $\frac{2}{10}$, $\frac{9}{100}$ and $\frac{1}{100}$ respectively. As in the deterministic model, the numbers may here be interpreted as tendencies or propensities of events in the world, and not as expectations of observations. They pertain to something physical. However, if we subsequently take these probability models as input to an inductive argument, they cannot be interpreted as chances
anymore. In that case they are objective epistemic probabilities. They are epistemic because they serve as input to an inductive scheme, which concerns opinions and expectations. But they are also objective, because they refer to and are informed by physical probabilities or, in other words, chances in the world.

*Subjectivist criticisms.* Subjectivists such as Ramsey, De Finetti, Savage and Howson may be opposed to an interpretation of probability models as objective. And because of the strong link between subjectivism and Bayesianism, it is certainly not a standard practice in Bayesianism conceived more broadly to interpret the likelihoods of the hypotheses in terms of objective probability models. Many Bayesians explicitly deny that probability can have objective content. They adhere to a purely subjectivist view, which states that any probability assignment is an expression of belief, and in this role cannot concern some objective aspect of the world. But it goes too far to discuss the relation between strict subjectivism and Bayesianism here. I want to leave it at two remarks to put possible criticisms of the use of objective probability in perspective.

Firstly, it must be emphasised again that the schemes of this thesis already endorse an epistemic interpretation of the probability functions \( p[e_i] \). They are explicitly intended to represent beliefs. The fact that, on top of that, the probabilities are objective means that these beliefs refer to and are informed by chances in the world. It accords with generally objectivist statistical inference in science to interpret the probability models associated with hypotheses in this objective epistemic manner. Furthermore, as argued in the foregoing, this objective interpretation helps us to arrive at hypotheses as elements of the observational algebra, which leads to a more natural logical picture of inductive inference. Secondly, the use of a Bayesian scheme does not make the simultaneous use of objective and subjective epistemic probabilities impossible. Pluralist views on probability are as old as Poincaré (1952) and as current as Gillies (2000). Nothing in the Bayesian scheme itself precludes the use of objective probability.

2.4.2 Deriving direct probabilities

But let me return to the main line of this section, which is to connect the probability models to the hypotheses as elements of the observation algebra.

*Reformulating the Principal Principle.* It may be argued that there is at least the following connection between the objective chances \( p[h_{w^r}] \) in the model of some
hypothesis, and the beliefs associated with that hypothesis: if we conditionally accept some model as the model of the world, we must declare its objective chances on observations to be the correct probabilities of the observations. It is a small step to develop this connection into the following principle:

\[ p_{[e_0]}(E_t | p_{[h_{w\theta}]}(E_t) = f(w, \theta)) = f(w, \theta). \] (2.13)

In words, conditional on some probability model \( p_{[h_{w\theta}]} \) the belief we assign to observations must be the same as the chance that this model prescribes for the observations. This principle has first been formulated by Jeffreys (1939) as the principle of direct probabilities. Later it was given the name of Principal Principle by Lewis (1980), who linked it to a notion of admissible evidence and who took it not as a restriction on subjective probabilities, but rather as an implicit definition of chance. In one interpretation or another, the principle now enjoys wide popularity.

Note that the principle is formulated in terms of a second order epistemic probability \( p_{[e_0]} \) over both observations and a probability assignment \( p_{[h_{w\theta}]} \). With the above interpretation of hypotheses as elements in \( \mathcal{Q} \), we may present an alternative formulation. Now the hypothesis \( H_{w,\theta} \), as opposed to the probability model \( p_{[h_{w\theta}]} \), may be included in the condition:

\[ p_{[e_0]}(E_t | H_{w,\theta}) = f(w, \theta) \] (2.14)

This accords much better with the Kolmogorovian theory of probability, in which probability can only be assigned to elements in the algebra. Moreover, in this reformulation there is a rather natural argument for adopting the principle of direct probability, as I show below. This in itself presents yet another reason for adopting the interpretation of hypotheses as elements in \( \mathcal{Q} \).

**From hypothesis to probability model.** The natural question is whether the definition of a hypothesis, as an element \( H_{w,\theta} \in \mathcal{Q} \), determines the probability model associated with it, that is, whether it determines the likelihoods.

For one thing, the likelihoods are restricted by the definition of conditional probability in combination with the axioms of probability:

\[ p_{[e_0]}(E_t | H_{w,\theta}) = \frac{p_{[e_0]}(H_{w,\theta} \cap E_t)}{p_{[e_0]}(H_{w,\theta})}, \] (2.15)

where I am assuming that \( p_{[e_0]}(H_{w,\theta}) > 0 \). Whenever a hypothesis \( h_{w,\theta} \) deems some sequence of observations \( E_t \) to be either impossible or positively certain,
this carries over to the likelihoods via the definition of $H_{\theta}$:

\[
H_{\theta} \cap E_{\ell} = \emptyset \quad \Rightarrow \quad p_{[e_{0}]}(E_{\ell}|H_{\theta}) = 0, \quad (2.16)
\]

\[
H_{\theta} \cap E_{\ell} = H_{\theta} \quad \Rightarrow \quad p_{[e_{0}]}(E_{\ell}|H_{\theta}) = 1. \quad (2.17)
\]

In the case of a continuum of hypotheses, it is difficult to make sense of the above expression for conditional probability. Here it is better to resort to the alternative axiomatisation of probability using conditional probability assignments.

Statistical hypotheses are not purely deterministic. So it is not exactly straightforward to link the elements $H_{\theta}$ in the algebra with likelihoods of these hypotheses, expressed in $p_{[e_{0}]}(E_{\ell}|H_{\theta})$. However, the following argues that we can employ the characteristics of the elements $H_{\theta}$ to derive these likelihoods. Recall that the likelihoods are a function of the selection function $w$ and the vector components $\theta_{qm}$, as determined by the probability model of $h_{\theta}$. For present purposes it therefore suffices to derive the probability vectors only.

*Deriving a probability model.* To derive these probability vectors, we need one assumption on the probability assignment within the hypotheses: the probability distribution over the possible subsequences $e^{m}$ constructed from the sequences $e \in H_{\theta}$ must be uniform. For all sequences $e \in H_{\theta}$, the function $w$ determines which states occur at which positions. Independently of how these states follow up on each other, we can formulate, for each state separately, the assumption that among the subsequences $e^{m}$ that build up the sequences $e$ according to the function $w$ there are no preferred ones. That is, for each state $m$ separately, the subsequences $e^{m}$ are assumed to be equally probable. We may say that this assumption is based on some form of the principle of indifference. The uniform probability within hypotheses expresses that, apart from isolating the hypotheses $H_{\theta}$ in the algebra and thus focusing on specific sets of sequences $e$, we have no further reason to prefer one sequence over another.

This uniformity can be used to derive the likelihoods $\theta_{qm}$. I will not give a complete proof but provide a proof sketch only. The general idea in the sketch derives from ergodicity theory: on the assumption of uniform probability within the hypotheses, the long-run relative frequencies may be used as single-case probabilities. The proof has two main ingredients. First, note that for every $e \in H_{\theta}$, the fractions of results $q$ in the subsequences $e^{m}$ are always $\theta_{qm}$. Second, note that any possible subsequence $e^{m}$ with specific relative frequencies can be constructed by permuting, possibly using an infinite number of operations, one single sequence with those frequencies. Therefore, as a further
specification of the second ingredient, to state that the probability is uniform over the subsequences is the same as saying that all permutations of a single subsequence $e^m$ are equally probable.

Now consider all the possible permutations of the subsequence $e^m$, and look at a specific position $t$ within it. After one permutation, any one of the results $e^m(i)$ may end up in the specific position $t$. But because all permutations are equally probable, all the results $e^m(i)$ are equally probable to end up there. And since, according to the first ingredient, a fraction of $\theta_{qm}$ of the results $e^m(i)$ has the value $q$, the probability of subsequences $e^m$ to have a $q$ at position $t$ is also exactly $\theta_{qm}$. From the characteristics of $H_{w\emptyset}$ and the assumption of a uniform probability over subsequences we have thus derived the values $\theta_{qm}$ for the likelihoods $p_{[e_0]}(Q_{t+1}^T | H_{w\emptyset} \cap E_t)$. We thus obtain the principle of direct probabilities.

An additional reason for frequentist semantics. The foregoing provides additional reasons for adopting a picture in which hypotheses are associated with elements in the algebra $Q$. First, the principle can be formulated in terms of a single probability assignment over an observational algebra. There is no need for second-order probability. Second, the principle of direct probability can be derived from an assumption of uniformity. Now it may be that hardcore subjectivists will not be impressed by this line of argument, since they may not be willing to accept such assumptions of uniformity over the probability assignment in the first place. But such subjectivists are not likely to be moved by the principle of direct probability itself either, certainly not if that principle is taken as an implicit definition of the objectivist notion of chance. For all those in favour of some form of the principle of direct probability, the derivation suggested here may be a natural motivation.

2.4.3 Physical models

The above use of objective probability models must not be confused with another such use, which originates in Polya (1954) and has also been discussed in Kuipers (1978). In this use of models, the objective interpretation does not apply to the probability of observations conditional on the hypotheses, but to the predictions that result from the inductive scheme as a whole. The remainder of this section is devoted to some observations on these so-called physical models.

Polya urn models. As a first example, consider the Carnapian prediction rule $pr_{\lambda,\gamma}$ for two possible observations with $\lambda = 2$ and $\gamma_q = \frac{1}{2}$, which is sometimes
called the straight rule:

$$p(Q_{t+1}^q | E_t) = \frac{t_q + 1}{t + 2}$$  \hspace{1cm} (2.18)

Here $t_q$ is the number of results $q$ in $c_t$. As discussed more elaborately in the following chapters, this rule is suitable for predicting the results generated by a device that produces results with a constant but initially unknown chance.

Interestingly, there is also an objective interpretation of the probabilities generated by the prediction rule. That is, we can construct a physical system that generates the results $q$ with probabilities exactly matching the predictions. Imagine an urn consisting of one blue and one green ball, so that $q \in \{0, 1\}$. If we pick a ball from the urn at random, each of the two colours have a chance of $\frac{1}{2}$ of being picked, which matches the straight rule for $t = 0$. Now let us say that the first ball is green. We then put back this ball into the urn, and add one further green ball, so that there are 3 balls in the urn, namely two green ones and one blue. If we subsequently pick a ball from the urn, the chances of picking a green or a blue one again match the prediction rule. This time, there is a chance of $\frac{2}{3}$ on green, and of $\frac{1}{3}$ on blue. More generally, by always replacing the ball just picked and adding one of the same colour after that, we can take care that the chances on colours keep matching the predictions.

**Petri dish models.** This is just one example of a physical system replicating the predictions generated by some inductive scheme. Many more such systems may be constructed. For example, instead of urns with balls we may imagine drawing strings of beads from a jewellery box and adding beads to these strings, with the observations being the individual beads on the strings. Such systems are useful for replicating so-called partially exchangeable processes, using specific selections of strings of beads. Another system, which is more similar to the balls in Polya urns, is presented by colonies of bacteria in a Petri dish. This specific system is particularly suited for replicating predictions that derive from a partition of hypotheses in a Bayesian scheme, as I will now show.

Consider two colonies of bacteria mixed together in a Petri dish, and let us say that at the start of the investigations the two colonies have equal size. The colonies of bacteria differ in the proportion of certain types of cells, for example, blue and green cells. More in particular, for colony $H_0$ a proportion of $\frac{2}{3}$ is of the blue type, $q = 0$, while the rest is of the green type, $q = 1$. For colony $H_1$, on the other hand, a proportion of $\frac{2}{3}$ is of the green type while the rest is of the blue type. Now imagine that a scientist samples a single cell from the Petri dish at random, and after determining its type, removes bacteria of the other type from the Petri dish. So if the sampled cell is of the green type,
2.5. COMPARING INDUCTIVE SCHEMES

$q = 1$, a proportion of $\frac{1}{3}$ of colony $H_1$ is removed, while $\frac{2}{3}$ of $H_0$ is. After that, the scientist leaves the bacteria to grow back and fill up the Petri dish again. Since the colonies have the same growing rate, colony $H_1$ is twice as large as $H_0$ in the new mixture. Further, after the growth both colonies have restored the respective original proportions between the types, which are again $\frac{1}{3}$ and $\frac{2}{3}$ or vice versa. The scientist may then repeat the whole procedure.

As already suggested by the notation, the determination of the type of a randomly sampled cell in the above experimental setting replicates exactly the predictions resulting from a Bayesian scheme with two hypotheses, $H_0$ and $H_1$, which have likelihoods $\langle \frac{2}{3}, \frac{1}{3} \rangle$ and $\langle \frac{1}{3}, \frac{2}{3} \rangle$ for the possible results $q = 0$ and $q = 1$, respectively. It will be clear that the setting may be generalised to any number of types or colonies. Perhaps surprisingly, taking an infinity of colonies all associated with different proportions $\langle \theta, 1 - \theta \rangle$ comes down to the same system as the Polya urn described above. The next chapter returns in more detail to this equivalence, which is a special case of De Finetti’s representation theorem.

2.5 Comparing inductive schemes

In this last section I consider the two schemes presented in the preceding chapter. I first show that with hypotheses as elements of the observation algebra $Q$, the Bayesian and Carnapian schemes are formally equivalent. The second subsection discusses whether the two schemes allow for the same range of inductive predictions.

Carnapian scheme over a $\sigma$-algebra. Recall from chapter 1 that a Carnapian prediction rule $pr$ comes down to a complete probability assignment $p$ over the algebra $Q_0$. By contrast, when using a frequentist semantics for hypotheses a Bayesian scheme minimally requires an extended algebra $Q = \sigma(Q_0)$. This is because the frequentist hypotheses $H_{w\theta}$ are associated with elements in $Q$ that fall outside the algebra $Q_0$. Fortunately, as proved in Billingsley (1995: 36-41), every probability function over an algebra $Q_0$ can be extended uniquely to a probability function over the $\sigma$-algebra generated by it. So the probability function $p$ over the algebra $Q$ is already implicit to the Carnapian scheme.

There is a considerable conceptual price for the extension of the probability from $Q_0$ to a unique probability over the $\sigma$-algebra $Q$. For one thing, the whole idea of a $\sigma$-algebra seems at variance with empiricist views: it allows for sets such as $H_{w\theta}$, which consist of infinite disjunctions of infinite conjunctions of single observations. Such sets seem entirely unempirical. Moreover, even if we grant the use of a $\sigma$-algebra, the unique extension of the probability over it can
only be derived if the first three Kolmogorov axioms are supplemented with an
axiom on so-called \(\sigma\)-additivity. This axiom states that the probability of an
infinite disjunction of disjoint sets can be written down as an infinite sum of
the probabilities of these sets. Following the discussion in Williamson (1999),
the axiom has a dubitable status for both empiricists and subjectivists. For
example, it is impossible to justify the axiom of \(\sigma\)-additivity by reference to
betting contracts.

Nevertheless, given the conceptual clarity offered by the use of frequentist
hypotheses, I am myself more than willing to pay the conceptual price. In fact,
I find the use of infinite set operations and \(\sigma\)-additivity only marginally more
farfetched than the whole project of finding a formal framework for inductive
inference, such as probabilistic inductive logic. From this point of view, the use
of frequentist hypotheses is a small and profitable investment.

*The razor of Von Mises.* It remains to be seen that the Bayesian scheme can
indeed be defined by a single probability assignment over the algebra \(Q\). I will
now present a more detailed argument for that, and show that it presents a
further reason for the frequentist semantics.

Recall that in the scheme of chapter 1, each hypothesis \(h_j \in \mathcal{H}\) is associated
with a labelled algebra \(\{h_j\} \times Q_0\), so that the space over which we define the
probability functions \(p_{[h_j]}\) is given by \(\mathcal{H} \times Q_0\). With the definition of hypotheses
as elements in \(Q\), we can reform this scheme in two steps. As a first step, we
can extend the probability assignments \(p_{[h_j]}\) to a probability over the extended
algebra \(Q\) assuming \(\sigma\)-additivity. This yields a probability \(p_{[h_j]}\) over each algebra
\(\{h_j\} \times Q\). We are then facing a rather awkward conceptual possibility; we can
employ the hypotheses \(H_j\), as elements in the algebra \(Q\), as arguments in their
own probability assignments: \(p_{[h_j]}(H_j)\). However, for frequentist hypotheses
\(h_j\) the weak law of large numbers entails that the probability of the set \(H_j\),
according to its own probability model, is \(p_{[h_j]}(H_j) = 1\). Thus, in the extension
of \(p_{[h_j]}\) to the algebra \(Q\), all probability is located within the tail event \(H_j \in Q\).

This brings us to the second step in reforming the scheme. If we consider
the extended algebra \(\mathcal{H} \times Q\), it seems that we can merge the algebras along the
range of hypotheses \(h_j\). The sets in the respective algebras \(\{h_j\} \times Q\) that carry
the probability mass, namely the hypotheses \(H_j\), do not overlap. In each of the
hypotheses \(\{h_j\} \times Q\) the probability \(p_{[h_j]}\) is concentrated within the sets \(H_j\),
but these latter sets are mutually exclusive. Because of this, we can harmlessly
compress the range of algebras \(\{h_j\} \times Q\) into a single algebra \(Q\), within which
the hypotheses are simply given by the sets \(H_j\). So the second step in reforming
2.5. COMPARING INDUCTIVE SCHEMES

Figure 2.1: The Bayesian scheme and the Carnapian scheme come down to the same thing. Both determine a probability over the extended observation algebra $\mathcal{Q}$. The Bayesian scheme can be viewed as the microstructure of the Carnapian scheme. Prediction rules result from integral expressions over statistical hypotheses in a Bayesian scheme.

The scheme is that instead of using a range of extended observational algebras $\mathcal{H} \times \mathcal{Q}$, we can make do with a single extended algebra $\mathcal{Q}$.

This is where the razor of Von Mises finds its application. In the Kolmogorovian picture of the preceding chapter, hypotheses $h_j$ are each associated with a separate algebra $\mathcal{Q}_0$. But following the two steps of the foregoing, it seems that we can trim away this abundance of algebras and leave only the single observational algebra $\mathcal{Q}$. This can be done by identifying the probability models with elements in the extended algebra according to the frequentist interpretation. Indeed, the razor of von Mises shaves the long Kolmogorovian beard of algebras.

It follows that the Carnapian and Bayesian schemes are formally equivalent. Both schemes may be defined by a single probability $p$ over the observation algebra $\mathcal{Q}$. The choice of this probability $p$ may be effected by choosing a prediction rule $pr(q, e_t)$, but also by choosing a partition of hypotheses, associated with a range of probability models, along with a prior probability assignment over these hypotheses. As depicted in figure 2.1, the Bayesian scheme thus emerges as a detailed version of the Carnapian scheme, and not as the generalisation which is presented in chapter 1. The Bayesian scheme may be said to present the microstructure underlying the Carnapian scheme.
Reasons for a frequentist semantics. The frequentist semantics makes for a more integrated picture of the Bayesian scheme. First, the definition of statistical hypotheses as elements in the observation algebra avoids the use of second-order probability, thus connecting better to the Kolmogorov axioms. Second, in view of the empiricist roots of inductive logic I think that the use of a single algebra $\mathcal{Q}$ is more natural than the use of $\mathcal{H} \times \mathcal{Q}_0$. Third, it is advantageous that both objective and subjective probabilities find a natural place in the Bayesian scheme, making the discussion on the interpretations of probability somewhat less ardent. The likelihoods may even be derived from frequentist hypotheses at the cost of a uniformity assumption. And finally, as discussed below, it is rather nice that the difference between the Carnapian and the Bayesian scheme can be traced back to an enrichment of the language or algebra that is used in these schemes.

Limits of the Bayesian scheme. The formal equivalence that is derived by means of the frequentist semantics also adds to the urgency of the question why we are considering two schemes in the first place. After all, if they can be written down in the same format, studying one of them will be enough, and simplicity considerations then lead us to the Carnapian scheme. The remainder of this section is concerned with this question. A natural answer is that the two schemes allow us to express different probability assignments over the algebra. It may seem obvious that any Bayesian scheme comes down to some Carnapian scheme: the Bayesian scheme generates predictions, and these predictions can always be summarised in a prediction rule, and thus in a Carnapian scheme. But in reality it proves very difficult to find elegant expressions for the particular prediction rule that corresponds to some Bayesian scheme. This may be a reason for employing a Bayesian scheme after all. To such reasons I will come back more elaborately in chapter 3.

Focusing only on the range of prediction rules and not on the possibility to express these rules in a convenient form, the Carnapian scheme can accommodate any Bayesian scheme. But it is not so obvious that the Bayesian scheme can accommodate any Carnapian prediction rule. Clearly, the equivalence between the Carnapian and the Bayesian scheme becomes trivial if we are allowed to consider any hypothesis we like. We may then simply use one hypothesis $h$ in the Bayesian scheme, and give it likelihoods that correspond to the predictions of the rule that we want to replicate. But there are many prediction rules that do not coincide with a probability model associated with a frequentist hypothesis. As argued above, the class of frequentist hypotheses is rather restricted.
Frankly, I do not know whether prediction rules can always be replicated with a Bayesian scheme using frequentist hypotheses, and I also do not know what kind of argument may eventually settle that matter. But for lack of any such argument towards the affirmative, it seems that Carnapian schemes have a slight advantage over Bayesian schemes. With the use of frequentist hypotheses we run the risk of unknowingly restricting the range of possible prediction rules.

It may be argued that statistical hypotheses that are not frequentist do not deserve consideration in the first place. If indeed there are restrictions imposed by only using frequentist hypotheses, we must accept these simply because there is otherwise no proper interpretation for the probabilities in the scheme. This is indeed a very dogmatic reaction. It seems unwise to make it part and parcel of the Bayesian scheme itself.

*Hypotheses as enrichment of the language.* There is another difference between Carnapian and Bayesian schemes that is significant in the next chapter, and in this thesis more generally. The two schemes are equal in the sense that they both determine a single probability function over $\mathcal{Q}$, but they differ in that they determine it by specifying different sets. To return to one of the points of section 2.3.1, the introduction of hypotheses as tail events in the algebra amounts to an enrichment of the language used to express inductive predictions. With the sets $H_{\omega \theta}$, new events or terms are added to the observational terms already present in the algebra $\mathcal{Q}_0$. This enrichment allows us to specify the input probability of inductive arguments in a different way, which will be seen to have clear conceptual advantages. It will be argued below that the Bayesian scheme offers a more detailed description of inductive predictions, and a more natural grip on the inductive predictions themselves. In particular, the Bayesian scheme allows us to express the projectability assumptions that precede any inductive argument. The Bayesian scheme is therefore more suitable than the Carnapian for conceptualizing, applying, and adapting inductive predictions.