This chapter discusses the use of theoretical distinctions between hypotheses in Bayesian inductive inference. A theoretical distinction is any distinction between hypotheses that is not reflected in a difference in likelihoods. This chapter shows that under certain conditions inductive predictions may benefit from theoretical distinctions, and further that under such conditions the observations can tell theoretical hypotheses apart. Two considerations follow from this main conclusion. First, the puzzle on theoretical hypotheses can be repeated at a higher level, concerning scientific method more generally. This leads to the claim that underdetermination fulfils a function in scientific method. Second, the choice between theoretical hypotheses in science is usually associated with abductive inference. This chapter therefore contains the promise of a Bayesian model of abduction.

The present chapter can again be read entirely independently. However, the technical part of the chapter is rather concise. For a more elaborate treatment I refer to section 1.3. A deeper understanding of statistical hypotheses can be obtained from chapter 2. Chapters 3 and 7 are useful for understanding the relation between theoretical background and inductive predictions more generally.

9.1 Statistical inference using partitions

This section describes a Bayesian scheme for inductive inferences, running from observations and statistical hypotheses to predictions. The prior probability over hypotheses is first updated to the given observations, and the updated probability is subsequently used to generate predictions. The resulting predictions may be also captured directly in prediction rules, but the hypotheses are seen to be useful for expressing knowledge of underlying chance mechanisms.

Bayesian inductive inference. The inductive scheme employs a formal framework of observations and hypotheses. Consider an observation at time \( i \) with a possible result \( q \in \{0,1\} \), denoted \( Q_i^q \), and denote sequences of observations of length \( t \) with \( E_t \). The example of this section supposes the observations
to consist of results of coin tosses. Consider a partition $\mathcal{H}$ of hypotheses $H_\theta$ concerning the chance $\theta$ on tails, $q = 1$, so that

$$p(Q_{t+1}^1|H_\theta \cap E_t) = \theta,$$

(9.1)

and further a prior probability over these hypotheses, $p(H_\theta)d\theta$. The partition, the likelihoods of the hypotheses in it, and the prior probability over the hypotheses together determine the Bayesian scheme.

The observations are the other component that is needed to arrive at inductive predictions. Bayes’ rule can be used to update the prior probability to given observations $E_t$. After updating we obtain a posterior probability,

$$p(H_\theta|E_t)d\theta = \frac{p(E_t|H_\theta)}{p(E_t)}p(H_\theta)d\theta.$$

(9.2)

Predictions follow directly from this posterior by the law of total probability:

$$p(Q_{t+1}^1|E_t) = \int_0^1 p(Q_{t+1}^1|H_\theta \cap E_t) p(H_\theta|E_t) d\theta = \int_0^1 \theta p(H_\theta|E_t) d\theta.$$

(9.3)

This scheme for predictions covers a substantial part of Bayesian statistical inference, as many such inferences are made with models concerning constant chances.

The use of hypotheses. In a sense, the hypotheses $H_\theta$ are already theoretical. They concern the objective chance of an observation, and such chances cannot be translated into finite observational terms. Moreover, the hypotheses can be eliminated from the inference completely. De Finetti’s representation theorem states that the above scheme of hypotheses covers exactly those prediction rules for which the order of the observations in $E_t$ is inessential. Defining $t_q$ as the number of $Q_q^1$ in $E_t$, these rules maybe characterised with

$$p(Q_{t+1}^1|E_t) = pr(t_q, t).$$

(9.4)

Every rule $pr$ corresponds to a specific prior $p(H_\theta)d\theta$ over the hypotheses in the scheme, and vice versa. In particular, if we assume the prior to be a symmetric Dirichlet distribution, we can derive the Carnapian $\lambda$ rules:

$$p(Q_{t+1}^1|E_t) = \frac{t_q + \lambda/2}{t + \lambda} = pr_\lambda(t_q, t).$$

(9.5)

A higher peak in the Dirichlet density $p(H_\theta)$ is reflected in a larger parameter $\lambda$. 


Although the hypotheses in the above inferences can thus be replaced with direct links between observations, there are good reasons for keeping the hypotheses in. First, they express the chance mechanism that is supposed to underlie the observations. For example, if the observations concern coin tosses, we know that the mechanism underlying the observations concerns constant and independent chances. Second, the hypotheses enable us to express further knowledge of the chance mechanisms in a prior probability over them. In the example, a normal coin motivates a prior over these chances that is strongly peaked at \( \frac{1}{2} \), while a coin from a conjurer’s box may have little probability at \( \frac{1}{2} \) and more probability at 0 and 1. In the prior we can thus express knowledge of the chance mechanism that is not incorporated in the statistical hypotheses themselves. It is not always a straightforward matter to incorporate such knowledge in a direct prediction rule.

### 9.2 Duplicate partitions

In what follows the above scheme will be extended by a duplicate partition. The distinction between the two duplicates is therefore entirely theoretical. It is shown that this purely theoretical distinction facilitates the choice of priors. It is further shown that this move makes the two duplicate subpartitions observationally distinguishable after all. In addition, it is seen that there are computational advantages to keeping the two subpartitions distinct.

*A normal or magical coin.* Let me start with the example on coin tosses. Imagine that we are undecided on whether the coin is from a conjurer’s box or from an ordinary wallet. Now both these kinds of coins have an unknown constant chance on tails, \( q = 1 \), so that we may employ the hypotheses \( H_\theta \). However, we have some further knowledge of the mechanism underlying the observations that must somehow be incorporated in the prior: either the coin is most probably fair, having a chance that is close to \( \frac{1}{2} \), or the coin is most probably strongly biased, having a chance that is close to 0 or 1. To incorporate this knowledge, we can now employ an additional partition into the hypotheses \( G_0 \) concerning the normal coin and \( G_1 \) concerning the magical coin.

Both hypotheses \( G_j \) cover exactly the same subpartition, \( G_j = \{g_j\} \times \mathcal{H} \). They are only labelled differently. We can use the likelihoods \( \theta \) for the hypotheses \( g_0 \times H_\theta \) and \( g_1 \times H_\theta \) alike. In terms of these statistical hypotheses, the distinction between the magical coin and the normal coin is therefore not observable. For each hypothesis in the one subpartition, there is a hypothe-
Figure 9.1: Two different priors over the two subpartitions of Bernoulli processes $H_\theta$. The peak prior is associated with the normal coin, the valley prior with the magical coin. Both function are from the class of Dirichlet priors. For $\lambda = 2$ the prior distribution is uniform, for larger values of $\lambda$ the peak gets higher, and for smaller values of $\lambda$ the valley gets deeper.

sis in the other subpartition that has exactly the same likelihoods for all the observations. The partition as a whole is thus underdetermined.

Advantages of a degenerate partition. There is a particular advantage, however, to employing this duplicate partition in the Bayesian scheme. We have separate control of the priors over the subpartitions on the normal and magical coin, $g_0 \times H$ and $g_1 \times H$ respectively. The further knowledge about the two kinds of coins motivates specific forms for the priors in both partial partitions, leading to two separate Carnapian rules, for example with $\lambda = 10$ and $\lambda = \frac{1}{4}$. The priors are illustrated in figure 9.1. Let us say that initially we are undecided between these two, $p(G_0) = p(G_1)$. The rules can then be weighed with the probabilities of the coin’s origin, resulting in a so-called hyper-Carnapian prediction rule:

$$p(Q_{t+1} | E_t) = p(G_0 | E_t) pr_{10}(t_q, t) + p(G_1 | E_t) pr_{1/4}(t_q, t).$$ (9.6)

The idea here is that the probabilities within the two subpartitions $g_0 \cdot H$ and $g_1 \cdot H$ are updated separately, and that the resulting values yielded by the Carnapian rules can function as the likelihoods in an update over the hypotheses $G_0$ and $G_1$.

Interestingly, even while the subpartitions associated with $G_0$ and $G_1$ consist of pairwise identical hypotheses, the differing priors over them cause different aggregated likelihoods of $G_0$ and $G_1$, namely the different Carnapian rules. That is, the two partitions themselves are observationally indistinguishable, but the different expectations over these partitions make the partitions observationally
distinct after all. As a side effect of updating over $g_0 \times \mathcal{H}$ and $g_1 \times \mathcal{H}$, the observations become relevant to the theoretical distinction between hypotheses $G_0$ and $G_1$.

This effect is less magical than it may seem. The distinction between the hypotheses $G_0$ and $G_1$ may not be observational relative to the subpartitions $\mathcal{H}$, but the hypotheses $G_0$ and $G_1$ do have an observational content: a magical coin is much less likely to yield an observed relative frequency of tails of close to $\frac{1}{2}$ than the normal coin. This content is exactly expressed in the differing priors over the partial partitions $g_0 \times \mathcal{H}$ and $g_1 \times \mathcal{H}$. The theoretical distinction simply facilitates the use of these differing priors over the two subpartitions. A further function of the distinction consists in keeping calculations manageable. The function that expresses the combined prior over a single partition $\mathcal{H}$ is naturally the sum of the priors defined over the above subpartitions, as expressed in figure 9.2, but it is much more convenient to update these terms separately. The resulting predictions can not be equated with a single Carnapian rule, and it is not easy to find some other exchangeable direct prediction rule that captures them.

Relations to preceding chapters. Some remarks may connect the present discussion with preceding chapters. Note first that in terms of the frequentist semantics of chapter 2, the hypotheses $G_0$ and $G_1$ are indeed identical. They consist of the very same subpartitions, and thus of the very same sequences of observations. The use of a duplicate partition reminds of the initial Kolmogorov picture of the Bayesian scheme, as presented in chapter 1. Just as the hypothe-

Figure 9.2: The hyper-Carnapian prediction rule can also be expressed in a single prior over one partition $C$. The prior is simply the sum of the two separate priors over the separate subpartitions.
ses $H_j$ in that chapter, the hypotheses $G_j$ are here associated with a complete observational algebra. Secondly, it is notable that the partition of hypotheses is here used to encode specific inductive assumptions in a prior probability assignment. In this sense the chapter has much in common with chapters 4 and 5. While in these chapters the partition of hypotheses is transformed, in the present chapter the partition is duplicated. But in both cases we manipulate the partition of hypotheses in order to access the appropriate prior.

In the case of the above hyper-Carnapian rule, the reader may find that the advantages of the representation of inductive predictions in terms of statistical hypotheses is entirely unhelpful, or even contrived. It may be much more natural to employ the hypotheses $G_j$ with the Carnapian rules as likelihoods. However, in view of chapter 3 there are independent reasons for preferring the partition of statistical hypotheses $H_\theta$, with constant likelihoods, over the single Carnapian prediction rules. Moreover, I feel that a Bayesian statistician may have some problems in making sense of the Carnapian prediction rules in the role of statistical hypotheses. According to chapter 2 they are not even included in the class of such hypotheses. Finally, and in relation to all this, I want to maintain that the use of the hypotheses $H_\theta$ allows us to disentangle two different aspects of the way we deal with the observations of the coin tosses, and that these two aspects are conflated if we use just the hyper-Carnapian rule.

Another set of considerations concerns the role of the knowledge about underlying mechanisms, in this case knowledge about the possible type of the coin. First of all, I am not sure that we can speak of knowledge of the underlying mechanism. In the example of the coin we are perhaps in that position, but in the standard case of scientific investigations the underlying mechanism can at best be a supposition. On the other hand, such suppositions may be used to inform the priors just as well. As a second consideration, and following up on this, it is not in all cases clear how exactly these suppositions determine the form of the prior probability assignment over the subpartitions. The example may suggest that this link is straightforward, but there are many cases in which this is simply not true. The second part of this thesis illustrates that finding a prior probability assignment over hypotheses that encodes certain assumptions on the underlying mechanism is a substantive and all but trivial part of the task of inductive logic.

Summary. Let me summarise the main point of this section. It directs attention to inductive inferences using two duplicate subpartitions, which differ only in the entirely theoretical property that they posit different mechanisms underly-
The use of underdetermination

One of the main messages of this thesis is that hypotheses are useful for expressing suppositions on chance mechanisms in inductive inference, by making accessible, that is, conceptually and computationally manageable, certain classes of prior probabilities. The present chapter argues that this latter usage also applies to entirely theoretical distinctions between hypotheses: the theoretical distinction motivates specific priors, and the distinction is thereby given observational content. In this last section I transfer this insight to scientific method more generally. First I propose a different perspective on the problem of underdetermination. After that I argue that the use of duplicate partitions is holding the promise of a Bayesian model of abductive inference.

Underdetermined statistical inference. The problem of underdetermination is that science, if interpreted as a realist undertaking, is dramatically underdetermined by observation: at first sight it seems that much of the theoretical superstructure of scientific theories cannot be warranted by the observational substructure. The primary challenge for realists is to show that this apparent underdetermination is not harmful to the realist objective of science, where this objective, put crudely, is to present science as an enterprise that successfully aims for the truth. A good example of this reaction is to be found in structural realism as presented in Worrall (1989), Ladyman (1998), and Votsis (2005). However, some realists take on the bigger challenge of showing that underdetermination can in some cases be avoided. They achieve this by providing inference rules such as abduction, which enable us to choose between theoretical superstructures on the basis of explanatory considerations or other theoretical virtues.

By contrast, in the following I stick to the original challenge of showing that underdetermination does not obstruct the realist aims of science. More specifi-
cally, I attempt to show that underdetermined theoretical superstructures have a specific use in statistics. I thus accept that science is underdetermined, but I go on to suggest that it is possible to explain this fact by reference to the methodological use of underdetermination. How this use of theoretical superstructures reflects back on realism I leave for future research.

Recall the claim of the preceding section that a partition that employs purely theoretical distinctions may offer a better grip on statistical analyses of experimental observations. In the example, the hypotheses $G_0$ and $G_1$ cause underdetermination, since we can never tell them apart by observations. But distinguishing them is very useful in the statistical procedure: they facilitate the expression of knowledge or suppositions on underlying mechanisms in priors, and they carve up statistical inference in manageable parts. More generally, we may tentatively say that the use of theoretical distinctions in statistical analyses reveals the advantages of underdetermination. In future research I hope to support this claim with case studies on actual experiment, in which theoretical distinctions are indeed employed to elicit specific conclusions from the observations. The idea is that enriching the observational algebra with theoretical distinctions improves the expressive force of the inductive scheme, and thus the ability to elicit answers from nature and make specific inductive inferences.

**Abduction.** It is important to note that, as a side effect of using theoretical distinctions, it looks as if these distinctions themselves become observational. This is where the use of theoretical distinctions in inductive inference begins to look like abduction. An abductive inference enables us to choose between a number of observationally indistinguishable, and thus theoretical, alternatives on the basis of certain theoretical virtues, for example explanatory force. Now the key insight here is that such a theoretical virtue is also presented in the fact that one of the two priors in the duplicate partition corresponds better to the observations. Recall that the observations have exactly the same impact on the separate hypotheses in each of the two subpartitions. The different impact is entirely due to the difference in the subjectively determined probability over these two subpartitions. We may therefore say that the observations reflect differently on the two subpartitions, exactly because they interact differently with our expectations.

Let me briefly explain these remarks by relating them to empirical equivalence and the nature of observations. Note first that whether two theories are empirically equivalent or not depends on the notion of theory that is employed. If we assume that the theories about the origin of the coin are determined by
the statistical hypotheses that they consist of, in both cases \( \mathcal{H} \), they are indeed empirically equivalent. But if we say that the prior probability assignment is an inherent part of the theories, then the two theories are not empirically equivalent. Furthermore, on the stipulation that the theories are empirically equivalent, whether two theories may or may not be told apart by observations in a sense indicates the nature of these observations. That is, if we take the content of the observations to be the effect they have on the probability assignment on the whole, it may be argued that the observations in the coin example are not entirely empirical. In that case they somehow incorporate theoretical content. If, on the other hand, we stipulate that the theories are empirically distinct in the first place, or if we stipulate that the content of the observations is given by the likelihoods for the observations and by nothing else apart from that, then there is no reason to say that observations have theoretical content.

It will be clear that I prefer the view that observations also convey theoretical content, and that they manage to do so because of the theoretical scheme in which we have chosen to frame them. My main reason for preferring this view is that I think it holds the promise of a Bayesian model of abduction. It provides a formal expression for the position that there is no sharp line between observation and theory, from which the use of observations for deciding over theoretical distinctions is seen to follow. However, the details of this position, which will elaborate the idea that the observations have different content in the context of different theoretical subpartitions, must be left to further research.