Introduction to Probability Theory

Peter Lucas
Instituut for Computing and Information Sciences
University of Nijmegen

Linda van der Gaag
Instituut of Infomation and Computing Sciences
Utrecht University

Abstract

Probability theory is one of the earliest methods for associating a measure of uncertainty to expressions concerning its truth. In this paper several notions from probability theory are briefly introduced. In addition, we discuss the problems one encounters in applying this theory in the context of knowledge-based systems.

Contents

1 Probability distribution

The notions that play a central role in probability theory have been developed for the description of experiments. In empirical research a more or less standard procedure is to repeatedly perform a certain experiment under essentially the same conditions. Each performance yields an *outcome* which cannot be predicted with certainty in advance. For many types of experiments, however, one is able to describe the set of all *possible* outcomes. The nonempty set of all possible outcomes of such an experiment is called its *sample space*; it is generally denoted by Ω. In the sequel, we shall only be concerned with experiments having a countable sample space.

Example Consider the experiment of throwing a die. The outcome of the experiment is the number of spots up the die. The sample space of this experiment therefore consists of six elements: $\Omega = \{1, 2, 3, 4, 5, 6\}$ □

A subset e of the sample space Ω of a certain experiment is called an *event*. If upon performance of the experiment the outcome is in e, then it is said that the event e has occurred. In case the event e has not occurred, we use the notation $\bar{e}$, called the *complement* of e. Note that we have $\bar{e} = \Omega \setminus e$. The event that occurs if and only if both events e_1 and e_2 occur, is called the *intersection* of e_1 and e_2, and will be denoted by $e_1 \cap e_2$. The intersection of n events e_i will be denoted by

$$\bigcap_{i=1}^{n} e_i$$

The event occurring if at least one of e_1 and e_2 occurs is called the *union* of e_1 and e_2, and will be denoted by $e_1 \cup e_2$. The union of n events e_i will be denoted by

$$\bigcup_{i=1}^{n} e_i$$

Example Consider the experiment of throwing a die and its associated sample space Ω once more. The subset $e_1 = \{2, 4, 6\}$ of Ω represents the event that an even number of spots has come up the die. The subset $e_2 = \bar{e}_1 = \Omega \setminus e_1 = \{1, 3, 5\}$ represents the event that an odd number of spots has come up. The events e_1 and e_2 cannot occur simultaneously: if event e_1 occurs, that is, if an even number of spots has come up, then it is not possible that in the same throw an odd number of spots has come up. So, the event $e_1 \cap e_2$ cannot occur. Note that the event $e_1 \cup e_2$ occurs in every performance of the experiment. The subset $e_3 = \{3, 6\}$ represents the event that the number of spots that has come up is a multiple of three. Note that the events e_1 and e_3 have occurred simultaneously in case six spots are shown up the die: in that case the event $e_1 \cap e_3$ has occurred. □

Definition 1 *The events* $e_1, \ldots, e_n \subseteq \Omega$, $n \geq 1$, *are called* mutually exclusive *or* disjoint events *if* $e_i \cap e_j = \varnothing$, $i \neq j$, $1 \leq i, j \leq n$.

We assume that an experiment yields an outcome independent of the outcomes of prior performances of the experiment. Now suppose that a particular experiment has been performed N times. If throughout these N performances an event e has occurred n times, the ratio $\frac{n}{N}$ is called the *relative frequency* of the occurrence of event e in N performances of the experiment.

As N increases, the relative frequency of the occurrence of the event e tends to stabilize about a certain value; this value is called the *probability* that the outcome of the experiment is in e, or the probability of event e, for short.

In general, the notions of a probability and a probability function are defined axiomatically.

Definition 2 *Let Ω be the sample space of an experiment. If a number $P(e)$ is associated with each subset $e \subseteq \Omega$, such that*

- $P(e) \geq 0$,
- $P(\Omega) = 1$, *and*
- $P(\bigcup_{i=1}^{n} e_i) = \sum_{i=1}^{n} P(e_i)$, *if e_i, $i = 1, \ldots, n$, $n \geq 1$, are mutually exclusive events,*

then P is called a probability function *on the sample space Ω. For each subset $e \subseteq \Omega$, the number $P(e)$ is called the probability that event e will occur.*

Note that a probability function P on a sample space Ω is a function $P : \wp(\Omega) \rightarrow [0, 1]$.

Example Consider the experiment of throwing a die once more, and its associated sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. The function P such that $P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6}$ is a probability function on Ω. Since the sets $\{2\}$, $\{4\}$, and $\{6\}$ are disjoint, we have according to the third axiom of the preceding definition that $P(\{2, 4, 6\}) = \frac{1}{2}$: the probability of an even number of spots coming up the die, equals $\frac{1}{2}$. □

Theorem 1 *Let Ω be the sample space of an experiment and P a probability function on Ω. Then, for each event $e \subseteq \Omega$, we have*

$$P(\bar{e}) = 1 - P(e)$$

Proof: We have $\Omega = e \cup \bar{e}$. Furthermore, $e \cap \bar{e} = \varnothing$ holds since e and $\bar{e}$ are mutually exclusive events. From the axioms 2 and 3 of the preceding definition we have that $P(\Omega) = P(e \cup \bar{e}) = P(e) + P(\bar{e}) = 1$. □

2 Conditional probabilities and Bayes' theorem

We consider the case in which probability theory is applied in a medical diagnostic expert system. One would like to know for example the probability of the event that a specific patient has a certain disease. For many diseases, the prior probability of the disease occurring in a certain population is known. In the case of a specific patient, however, information concerning the patient's symptoms, medical history, etc. is available that might be useful in determining the probability of the presence of the disease in this specific patient.

So, in some cases we are interested only in those outcomes which are in a given nonempty subset e of the entire sample space which represents the pieces of evidence concerning the final outcome that are known in advance. Let h be the event we are interested in, that is, the hypothesis. Given that the evidence e has been observed, we now are interested in the degree to which this information influences $P(h)$, the prior probability of the hypothesis h. The probability of h given e is defined in the following definition.

Definition 3 *Let Ω be the sample space of a certain experiment and let P be a probability function on Ω. For each $h, e \subseteq \Omega$ with $P(e) > 0$, the* conditional probability *of h given e, denoted by $P(h \mid e)$, is defined as*

$$P(h \mid e) = \frac{P(h \cap e)}{P(e)}$$

A conditional probability $P(h \mid e)$ often is called a *posterior* probability.

The conditional probabilities given a fixed event $e \subseteq \Omega$ with $P(e) > 0$, again define a probability function on Ω since the three axioms of a probability function are satisfied:

- $P(h \mid e) = \dfrac{P(h \cap e)}{P(e)} \geq 0$, since $P(h \cap e) \geq 0$ and $P(e) > 0$;
- $P(\Omega \mid e) = \dfrac{P(\Omega \cap e)}{P(e)} = \dfrac{P(e)}{P(e)} = 1$;
- $P(\bigcup\limits_{i=1}^{n} h_i \mid e) = \dfrac{P((\bigcup_{i=1}^{n} h_i) \cap e)}{P(e)} = \dfrac{P(\bigcup_{i=1}^{n}(h_i \cap e))}{P(e)} = \dfrac{\sum_{i=1}^{n} P(h_i \cap e)}{P(e)} = \sum\limits_{i=1}^{n} \dfrac{P(h_i \cap e)}{P(e)} = \sum\limits_{i=1}^{n} P(h_i \mid e)$, for mutually exclusive events h_i, $i = 1, \ldots, n$, $n \geq 1$.

This probability function is called the *conditional probability function given e*.

In real-life practice, the probabilities $P(h \mid e)$ cannot always be found in the literature or obtained from statistical analysis. The conditional probabilities $P(e \mid h)$, however, often are easier to come by: in medical textbooks for example, a disease is described in terms of the signs likely to be found in a typical patient suffering from the disease. The following theorem now provides us with a method for computing the conditional probability $P(h \mid e)$ from the probabilities $P(e)$, $P(h)$, and $P(e \mid h)$; the theorem may therefore be used to reverse the 'direction' of probabilities.

Theorem 2 *(*Bayes' theorem*) Let P be a probability function on a sample space Ω. For each $h, e \subseteq \Omega$ such that $P(e) > 0$ and $P(h) > 0$, we have:*

$$P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e)}$$

Proof: The conditional probability of h given e is defined as

$$P(h \mid e) = \frac{P(h \cap e)}{P(e)}$$

Furthermore, we have

$$P(e \mid h) = \frac{P(e \cap h)}{P(h)}$$

So,

$$P(e \mid h) \cdot P(h) = P(h \mid e) \cdot P(e) = P(h \cap e)$$

The property stated in the theorem now follows from these observations. □

Example Consider the problem domain of medical diagnosis. Let h denote the hypothesis that a patient is suffering from liver cirrhosis; furthermore, let e denote the evidence that the patient has jaundice. In this case, the prior probability of liver cirrhosis, that is, $P(\textit{liver-cirrhosis})$, is known: it is the relative frequency of the disease in a particular population. If the prior probability of the occurrence of jaundice in the same population, that is, $P(\textit{jaundice})$, is likewise available and if the probability that a patient suffering from liver cirrhosis has jaundice, that is, the conditional probability $P(\textit{jaundice} \mid \textit{liver-cirrhosis})$, is known, then we can compute the probability that a patient showing signs of jaundice suffers from liver cirrhosis, that is, using Bayes' theorem we can compute the conditional probability $P(\textit{liver-cirrhosis} \mid \textit{jaundice})$. It will be evident that the last-mentioned probability is of importance in medical diagnosis. □

To conclude, we define the notions of independence and conditional independence. Intuitively speaking, it seems natural to call an event h independent of an event e if $P(h \mid e) = P(h)$: the prior probability of event h is not influenced by the knowledge that event e has occurred. However, this intuitive definition of the notion of Independence is not symmetrical in h and e; furthermore, the notion is defined this way only in case $P(e) > 0$. By using the definition of conditional probability and by considering the case for n events, we come to the following definition.

Definition 4 *The events* $e_1, \ldots, e_n \subseteq \Omega$ *are (*mutually*)* independent *if*

$$P(e_{i_1} \cap \ldots \cap e_{i_k}) = P(e_{i_1}) \cdot \ldots \cdot P(e_{i_k})$$

for each subset $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$*,* $1 \leq k \leq n$*,* $n \geq 1$*. The events* $e_1, \ldots, e_n$ *are* conditionally independent *given an event* $h \subseteq \Omega$ *if*

$$P(e_{i_1} \cap \ldots \cap e_{i_k} \mid h) = P(e_{i_1} \mid h) \cdot \ldots \cdot P(e_{i_k} \mid h)$$

for each subset $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$*.*

Note that if the events h and e are independent and if $P(e) > 0$, we have that the earlier mentioned, intuitively more appealing notion of Independence

$$P(h \mid e) = \frac{P(h \cap e)}{P(e)} = \frac{P(h)P(e)}{P(e)} = P(h)$$

is satisfied.

3 Application in knowledge-based systems

A typical application of probability theory is for the purpose of diagnostic classification. During the 1960s several research efforts on probabilistic reasoning were undertaken. The systems constructed in this period of time were primarily for (medical) diagnosis. Although these systems did not exhibit any intelligent reasoning behaviour, they may now be viewed as the precursors to the diagnostic systems developed in the 1990s

Let us take a closer look at the tasks of diagnosis and classification. Let $H = \{h_1, \ldots, h_n\}$ be a set of n possible hypotheses, and let $E = \{e_1, \ldots, e_m\}$ be a set of pieces of evidence which may be observed. For ease of exposition, we assume that each of the hypotheses is either true or false for a given case; equally, we assume that each of the pieces of evidence is either true (that is, it is actually observed in the given case) or false. The diagnostic task now is to find a set of hypotheses $h \subseteq H$, called the (differential) *diagnosis*, which most likely accounts for the set of observed evidence $e \subseteq E$. If we have observed a set of pieces of evidence $e \subseteq E$, then we can simply compute the conditional probabilities $P(h \mid e)$ for each subset $h \subseteq H$ and select the set $h' \subseteq H$ with the highest probability. We have mentioned before that since for real-life applications, the conditional probabilities $P(e \mid h)$ often are easier to come by than the conditional probabilities $P(h \mid e)$, generally Bayes' theorem is used for computing $P(h \mid e)$. It will be evident that the task of diagnosis in this form is computationally complex: since a diagnosis may comprise more than one hypothesis out of n possible ones, the number of diagnoses to be investigated, that is, the number of probabilities to be computed, equals 2^n. A simplifying assumption generally made in the systems for probabilistic reasoning developed in the 1960s, is that the hypotheses in H are mutually exclusive and collectively exhaustive. With this assumption, we only have to consider the n singleton hypotheses $h_i \in H$ as separate possible diagnoses. Bayes' theorem can easily be reformulated to deal with this case.

Theorem 3 *(*Bayes' theorem*) Let P be a probability function on a sample space Ω. Let $h_i \subseteq \Omega$, $i = 1, \ldots, n$, $n \geq 1$, be mutually exclusive hypotheses with $P(h_i) > 0$, such that $\bigcup_{i=1}^n h_i = \Omega$ (that is, they are collectively exhaustive). Furthermore, let $e \subseteq \Omega$ such that $P(e) > 0$. Then, the following property holds:*

$$P(h_i \mid e) = \frac{P(e \mid h_i)P(h_i)}{\sum_{j=1}^n P(e \mid h_j)P(h_j)}$$

Proof: Since $h_1, \ldots, h_n$ are mutually exclusive and collectively exhaustive, we have that $P(e)$ can be written as

$$P(e) = P((\bigcup_{i=1}^n h_i) \cap e) = P(\bigcup_{i=1}^n (h_i \cap e)) = \sum_{i=1}^n P(h_i \cap e) = \sum_{i=1}^n P(e \mid h_i) \cdot P(h_i)$$

Substitution of this result in the before-mentioned form of Bayes' theorem yields the property stated in the theorem. □

For a successful application of Bayes' theorem in the form mentioned in the previous theorem, several conditional and prior probabilities are required. For example, conditional probabilities $P(e \mid h_i)$ for every combination of pieces of evidence $e \subseteq E$, have to be available; note that in general, these conditional probabilities $P(e \mid h_i)$ cannot be computed from their 'component' conditional probabilities $P(e_j \mid h_i)$, $e_j \in e$. It will be evident that exponentially many probabilities have to be known beforehand. Since it is hardly likely that for practical applications all these probabilities can be obtained from for example statistical analysis, a second simplifying assumption was generally made in the systems developed in the 1960s: it was assumed that the pieces of evidence $e_j \in E$ are conditionally independent given any hypothesis $h_i \in H$. Under this assumption Bayes' theorem reduces to the following form.

Theorem 4 *(*Bayes' theorem*) Let P be a probability function on a sample space Ω. Let $h_i \subseteq \Omega$, $i = 1, \ldots, n$, $n \geq 1$, be mutually exclusive and collectively exhaustive hypotheses as in the previous theorem. Furthermore, let $e_{j_1}, \ldots, e_{j_k} \subseteq \Omega$, $1 \leq k \leq m$, $m \geq 1$, be pieces* *of evidence such that they are conditionally independent given any hypothesis h_i. Then, the following property holds:*

$$P(h_i \mid e_{j_1} \cap \cdots \cap e_{j_k}) = \frac{P(e_{j_1} \mid h_i) \cdots P(e_{j_k} \mid h_i) P(h_i)}{\sum_{i=1}^{n} P(e_{j_1} \mid h_i) \cdots P(e_{j_k} \mid h_i) P(h_i)}$$

Proof: The theorem follows immediately from the preceding theorem and the definition of conditional independence. □

It will be evident that with the two assumptions mentioned above only $m \cdot n$ conditional probabilities and $n - 1$ prior probabilities suffice for a successful use of Bayes' theorem.

The pioneering systems for probabilistic reasoning constructed in the 1960s which basically employed the last-mentioned form of Bayes' theorem, were rather small-scaled: they were devised for clear-cut problem domains with only a small number of hypotheses and restricted evidence. For these small systems, all probabilities necessary for applying Bayes' theorem were acquired from a statistical analysis of the data of several hundred sample cases. Now recall that in deriving the last-mentioned form of Bayes' theorem several assumptions were made:

- the hypotheses $h_1, \ldots, h_n$, $n \geq 1$, are mutually exclusive;
- the hypotheses $h_1, \ldots, h_n$ furthermore are collectively exhaustive, that is, $\bigcup_{i=1}^{n} h_i = \Omega$;
- the pieces of evidence $e_1, \ldots, e_m$, $m \geq 1$, are conditionally independent given any hypothesis h_i, $1 \leq i \leq n$.

These conditions, which have to be satisfied for a correct use of Bayes' theorem, generally are not met in practice. But, in spite of these (over-)simplifying assumptions underlying the systems from the 1960s, they performed considerably well. Nevertheless, interest in this approach to reasoning with uncertainty faded in the early 1970s. One of the reasons for this decline in interest is that the method informally sketched in the foregoing is feasible only for highly restricted problem domains: for larger domains or domains in which the above-mentioned simplifying assumptions are seriously violated, the method inevitably will become demanding, either computationally or from the point of view of obtaining the necessary probabilities: often a large number of conditional and prior probabilities is needed, thus requiring enormous amounts of experimental data. Bayesian networks address this problem, and researchers in this field have been able to offer solutions to this problem.

4 The likelihood ratio

Instead of probabilities, it is possible to use the equivalent notion of 'odds', which in some circumstances is more convenient.

Definition 5 *Let P be a probability function on a sample space Ω. Furthermore, let $h \subseteq \Omega$ such that $P(h) < 1$. The* prior odds *of the event h, denoted by $O(h)$, is defined as follows:*

$$O(h) = \frac{P(h)}{1 - P(h)}$$

Note that conversely

$$P(h) = \frac{O(h)}{1 + O(h)}$$

In probability theory the notion of conditional or posterior probability is used. The subjective Bayesian method uses the equivalent notion of posterior odds.

Definition 6 *Let P be a probability function on a sample space Ω. Let $h, e \subseteq \Omega$ such that $P(e) > 0$ and $P(h \mid e) < 1$. The* posterior odds *of a hypothesis h, given evidence e, denoted by $O(h \mid e)$, is defined as follows:*

$$O(h \mid e) = \frac{P(h \mid e)}{1 - P(h \mid e)}$$

We introduce another two notions: the positive and the negative likelihood ratios.

Definition 7 *Let P be a probability function on a sample space Ω. Furthermore, let $h, e \subseteq \Omega$ such that $0 < P(h) < 1$ and $P(e \mid hbar) > 0$. The (*positive*)* likelihood ratio *λ, given h and e, is defined by*

$$\lambda = \frac{P(e \mid h)}{P(e \mid \bar{h})}$$

The likelihood ratio λ often is called the *level of sufficiency*; it represents the degree to which the observation of evidence e influences the prior probability of hypothesis h. A likelihood ratio $\lambda > 1$ indicates that the observation of e tends to confirm the hypothesis h; a likelihood ratio $\lambda < 1$ indicates that the hypothesis *hbar* is confirmed to some degree by the observation of e, or in other words that the observation of e tends to disconfirm h. If $\lambda = 1$, then the observation of e does not influence the prior confidence in h.

Definition 8 *Let P be a probability function on a sample space Ω. Let $h, e \subseteq \Omega$ be such that $0 < P(h) < 1$ and $P(e \mid \bar{h}) < 1$. The (*negative*)* likelihood ratio *$\bar{\lambda}$, given h and e, is defined by*

$$\bar{\lambda} = \frac{1 - P(e \mid h)}{1 - P(e \mid \bar{h})}$$

The negative likelihood ratio $\bar{\lambda}$ often is called the *level of necessity.* A comparison of the likelihood ratios λ and $\bar{\lambda}$ shows that from $\lambda > 1$ it follows that $\bar{\lambda} < 1$, and vice versa; furthermore we have $\lambda = 1$ if and only if $\bar{\lambda} = 1$.

5 The odds-likelihood form of Bayes' theorem

The ratios λ and $\bar{\lambda}$ can be viewed as the bounds of an interval in which lies a value indicating the degree to which evidence e, which has been (dis)confirmed to some degree by some prior evidence e', really influences the prior probability of h. This value is called the effective likelihood ratio, and will be denoted by λ'. The ratio λ' is computed from the value $P(h \mid e')$ according to the following definition.

Definition 9 *Let P be a probability function on a sample space Ω, and let O be the corresponding odds as defined in the foregoing. Furthermore, let $h, e' \subseteq \Omega$. The* effective likelihood ratio *λ', given h and e', is defined as follows:*

$$\lambda' = \frac{O(h \mid e')}{O(h)}$$

The effective likelihood ratio λ' lies between λ and $\bar{\lambda}$. λ' will be closer to λ if e has been confirmed to some degree by the observation of the evidence e'; conversely, λ' will be closer to $\bar{\lambda}$ if e has been disconfirmed to some degree by the prior evidence e'.

Recall that in probability theory Bayes' theorem may be used to incorporate evidence. Bayes' theorem, however, may also be used in the form of the odds.

Theorem 5 *Let P be a probability function on a sample space Ω, and let O be the corresponding odds as defined in the foregoing. Let $h,^{e} \subseteq \Omega$. Furthermore, let the likelihood ratio λ be defined as above. Then, the following property holds:*

$$O(h \mid e) = \lambda \cdot O(h)$$

Proof: From Bayes' theorem we have

$$P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e)}$$

For the complement of h we have, again from Bayes' theorem,

$$P(\bar{h} \mid e) = \frac{P(e \mid \bar{h})P(\bar{h})}{P(e)}$$

Dividing the first equation by the second one results in the following equation:

$$\frac{P(h \mid e)}{P(\bar{h} \mid e)} = \frac{P(e \mid h)P(h)}{P(e \mid \bar{h})P(\bar{h})}$$

from which we have

$$\frac{P(h \mid e)}{1 - P(h \mid e)} = \frac{P(e \mid h)}{P(e \mid \bar{h})} \cdot \frac{P(h)}{1 - P(h)}$$

From this observation it follows that $O(h \mid e) = \lambda \cdot O(h)$. □

This alternative form of Bayes' theorem is called *odds-likelihood form* of the theorem.

The theorem stated above concerns the situation where evidence e has been obtained with absolute certainty. In case we have that e has definitely not occurred, that is, in case $\bar{e}$ has been observed with absolute certainty, we obtain a similar formula.

Theorem 6 *Let P be a probability function on a sample space Ω, and let O be the corresponding odds as defined in the foregoing. Let $h, e \subseteq \Omega$. Furthermore, let the negative likelihood ratio $\bar{\lambda}$ be defined as above. Then, the following property holds:*

$$O(h \mid \bar{e}) = \bar{\lambda} \cdot O(h)$$

Again if more than one piece of evidence is available, these must be combined to obtain single measure of uncertainty for h. Again, we first consider the case where all e_i's have been obtained with absolute certainty. It should be evident that by assuming that the e_i's are conditionally independent given h we have that the following property holds:

$$O\left(h \mid \bigcap_{i=1}^{n} e_i\right) = \prod_{i=1}^{n} \lambda_i \, O(h)$$

where $\lambda_i = \frac{P(e_i|h)}{P(e_i|\bar{h})}$. Similarly, for the case where all $\bar{e}_i$'s have been obtained with absolute certainty, we have:

$$O\left(h \mid \bigcap_{i=1}^{n} \bar{e}_i\right) = \prod_{i=1}^{n} \bar{\lambda}_i \, O(h)$$

We have argued before that in general the e_i's (or $\bar{e}_i$'s respectively) will not have been obtained with absolute certainty, but with a probability $P(e_i \mid e_i')$ given some prior observations e_i'. From the probabilities $P(e_i \mid e_i')$ the posterior odds $O(h \mid e_i')$ are obtained from applying the combination function for propagating uncertain evidence. From these posterior odds we then compute the effective likelihood ratios λ_i'. Again under the assumption that the e_i''s are conditionally independent given h we obtain:

$$O\left(h \mid \bigcap_{i=1}^{n} e_i'\right) = \prod_{i=1}^{n} \lambda_i' \, O(h)$$

Since multiplication is commutative and associative, we have that the order in which the evidence is incorporated, will be irrelevant for the resulting uncertainty with respect to h.

Finally, the odds-likelihood form of Bayes' theorem can be very suitably represented in logarithmic form:

$$\begin{aligned}
\log O\left(h \mid \bigcap_{i=1}^{n} e_i\right) &= \log \prod_{i=1}^{n} \lambda_i \, O(h) \\
&= \sum_{i=1}^{n} \lambda_i + \log O(h) \\
&= \sum_{i=0}^{n} \omega_i
\end{aligned}$$

where $\omega_i = \log \lambda_i$, for $i = 1, \ldots, n$, and $\omega_0 = \log O(h)$.