Probability Level M (STATS5024)

Xin Li

School of Mathematics and Statistics, University of Glasgow

2020–2021

Chapter 1

Foundations of Probability

1.1 What probability means

The usefulness of probability theory

Probability theory is a mathematical system to permit consistent reasoning in the presence of uncertainty (lack of complete information about a system). Its uses are therefore wide!

Statistical inference

Probability underpins the science of Statistical Inference (or Statistics), which provides methods for analysing and modelling results obtained in many different areas of research. The properties of random samples are used to calculate the precision with which characteristics of populations may be estimated.

Genetics

Offspring inherit characteristics from their parents by random processes that can be described by probability models. Human parents who are believed to have a high risk of passing on an inherited disorder might be offered counselling, based on probability calculations, to help them with family planning decisions.

Epidemiology

Epidemic diseases are spread from person to person but there is a random element to their behaviour. Sometimes an encounter with a diseased individual causes a previously uninfected individual to become infected, but this is not always the outcome. Probability models are used to describe how an epidemic disease is spreading and to forecast its future course.

Finance

A life table shows, for each age, the probability that a person of that age will die before his or her next birthday. Life tables are used to work out the probability of surviving to any particular age, or the remaining life expectancy for people at different ages. Actuaries use these probabilities to work out fair premiums for life assurance and to assess the adequacy of pension funds.

Experiments, outcomes and events

Definitions

- experiment: any process by which information is obtained (a much more general definition than the usual scientific one)
- replicate, or trial: a single performance of an experiment
- outcome: the information recorded as the result of one replicate of an experiment
- random experiment, or stochastic experiment: an experiment which has more than one possible outcome, even when performed under identical conditions, and where it is not known beforehand which of the possible outcomes will result when the experiment is next performed

Probability is all about building mathematical models for the outcomes of (random) experiments.

Example 1

Here are descriptions of four random experiments.

(a) Record the category of treatment need for the next patient triaged at the local accident and emergency unit. Triage is a system for sorting injured people into groups based on their need for, or likely benefit from, immediate medical treatment. Each patient's need of treatment is categorised as either 'Very urgent' or 'Urgent' or 'Standard'.

(b) Record the number of hits on the School of Mathematics and Statistics main web page in the 24-hour period starting at midnight tonight.

(c) Record the maximum level of atmospheric SO_2 (in parts per million, or ppm) at a measuring site beneath the Glasgow Central railway bridge tomorrow morning between 08.00 and 09.00. (The global average is estimated to be about 1 ppm.)

(d) For the next pair of twins born alive at the Princess Royal Maternity Hospital, Glasgow, record whether the twins are monozygotic (identical, MZ) or dizygotic (non-identical, DZ) and the sex of each twin.

Definitions

- **sample space**: a set, often denoted S, that contains all the possible outcomes of a random experiment; S might be finite, countable or uncountable
- **countable set**: an infinite set whose elements may be listed as $\{s_1, s_2, \ldots\}$, e.g., the natural numbers, the whole numbers
- **uncountable set**: an infinite set that is not countable, e.g., the real numbers, an interval of the real line
- **event**: a subset of S, i.e., a collection of outcomes
- **simple event**: an event with a single outcome
- **compound event**: not a simple event

Example 1 revisited

(a)
- The **finite** sample space
  $S =$
- The simple event
  $E =$ 'need of treatment is very urgent' $=$
- The compound event
  $F =$ 'need of treatment is not very urgent' $=$

(b)
- A **countable** sample space
  $S =$
- The event $E =$ 'there are at least 2000 hits'
  $E =$
- The event $F =$ 'there are at least 1000 hits'
  $F =$
- The event $G =$ 'number of hits is divisible by 100'
  $G =$

(c)
- A **uncountable** sample space
  $S =$
- The event $E =$ 'the SO_2 level is at most 1 ppm'
  $E =$
- The event $F =$ 'the SO_2 level is at least 1 ppm'
  $F =$
- The event $G =$ 'the SO_2 level is above 3 ppm'
  $G =$

(d) The **finite** sample space

The outcome of an experiment is not necessarily numeric; see experiments (a) and (d) above. An outcome may consist of more than one piece of information; see experiment (d) above.

S itself is an event, and it is certain to occur. The empty set, $\emptyset = \{\}$, is counted as a subset of S; it represents an event that is impossible.

When S is finite or countable, all possible subsets of S are considered to be events. When S is uncountable, there are subsets of S that we do not wish to class as events (as it is impossible to assign probabilities to them consistently). If we let $\mathcal{E}$ denote the class of all events (i.e., all subsets of S to which we can assign probabilities) then $\mathcal{E}$ is known as a σ-algebra. $\mathcal{E}$ is defined in such a way that complements, unions and intersections of events in $\mathcal{E}$ are also events in $\mathcal{E}$. As far as this course is concerned, we will not discuss subsets of S that are not in $\mathcal{E}$ so we will not consider this matter further. If you are interested in finding out more, please consult a book on measure theory such as: P. Billingsley (2008), *Probability and Measure*, 3rd Edition, Wiley.

Suppose E and F are events, that is they are subsets of the sample space, S.

- S (universal set): an event that is certain

- $\emptyset$ (empty set): an event that is impossible
- E' (complement): E does not occur
- $E \cup F$ (union): E occurs or F occurs (or both occur)
- $E \cap F$ (intersection): both E and F occur

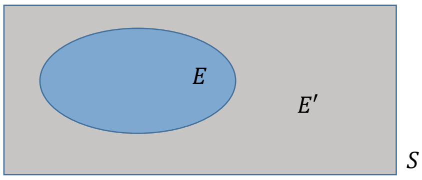

Figure 1.1: Venn diagram showing E and E'

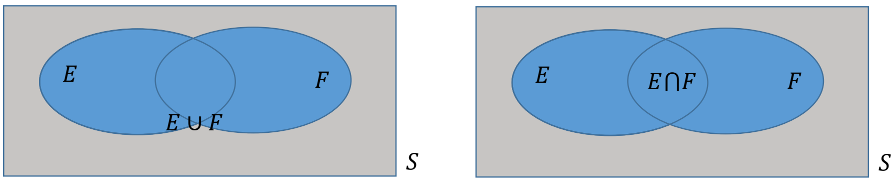

Figure 1.2: Venn diagram showing $E \cup F$ and $E \cap F$

- $E \cap F = \emptyset$: E and F are **mutually exclusive** or **disjoint** events
- Corollary: E and E' are always disjoint
- $E \subseteq F$ (subset): if E occurs, F is certain to occur
- If $E \subseteq F$, then $F \setminus E$ $(= F \cap E')$ is the event that consists of all the outcomes in F that are not in E

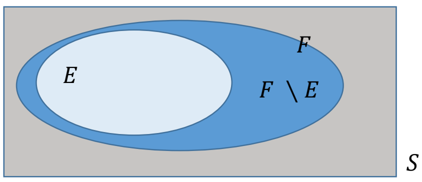

Figure 1.3: Venn diagram showing $F \setminus E$ when $E \subseteq F$

- Note $F = E \cup (F \setminus E) \quad [= E \cup (F \cap E')]$
- E and $F \setminus E$ are disjoint events, and F is said to be the **disjoint union** of E and $F \setminus E$.
- In general, for any two events E and F, F is the disjoint union of $F \cap E$ and $F \cap E'$:

$$F = (F \cap E) \cup (F \cap E').$$

Laws of set theory

Let A, B and C denote three events. Laws of set theory:

- Commutative: $A \cup B = B \cup A$ and $A \cap B = B \cap A$
- Associative: $(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C = A \cap (B \cap C)$
- Distributive: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- De Morgan: $(A \cup B)' = A' \cap B'$ and $(A \cap B)' = A' \cup B'$

Example 1 revisited

(a) $F = E'$, so $E \cup F = S$ and $E \cap F = \emptyset$

(b)
- $E \subseteq F$, so
- $F \setminus E =$
- $E \cap G =$
- $F \cap G =$
- $E \cap F \cap G =$

The interpretation of a probability

The probability of an event E, denoted $P(E)$, quantifies the uncertainty about whether or not E will occur when the underlying random experiment is carried out. Higher probabilities are assigned to events that are more likely to occur.

Conventionally, probability is measured in the range $[0, 1]$, with S being assigned probability 1 and $\emptyset$ being assigned probability 0. As we shall see later in the course, some events with probability 1 are not certain and some events with probability 0 are not impossible. An event with probability 1 is said to occur *almost surely*. An event with probability 0 is said to occur *almost never*.

Mathematicians, statisticians, philosophers and others interested in probability do not all agree about how to interpret the probability of an event. There are three main interpretations, which we shall discuss in the following simple context.

Example 2

Suppose that a standard die is rolled and the score on the uppermost face of the die is recorded. $S = \{1, 2, 3, 4, 5, 6\}$. Let E be the event, 'the score is at least 5'.

Equally-likely outcomes

Every outcome in S might be assumed to be equally likely to occur and assigned the probability $\frac{1}{6}$ so that E has probability $\frac{2}{6} = \frac{1}{3}$. This probability is a statement about the symmetry of the underlying experiment. A major problem is that most random experiments do not have this symmetry.

Relative frequency

If the event E occurs on n_E out of n replicates of an experiment, then the relative frequency of E is n_E/n. The probability of E might be defined to be the long-term, or limiting, value of this proportion when the experiment is carried out a very large number of times. So the statement $P(E) = \frac{1}{3}$ means that, on average, a score of 5 or 6 will occur on one-third of all rolls of the die. A problem with the frequentist definition of probability is that it is tied conceptually to experiments which may be repeated endlessly under identical conditions (none?).

Subjective probability

In recent times, the view has become popular that a probability is better treated as a statement of personal belief so probability is not tied to experiments that can be replicated infinitely. The statement that $P(E) = \frac{1}{3}$ is now seen as a statement of one person's current belief about the die or the experiment or, perhaps most compellingly, a statement of their ignorance about the experiment. In extending the range of probability to include one-off events, the subjectivist school has broken the link between a frequentist probability and objective reality.

Whichever of these views of the meaning of a probability they favour, almost all probabilists are willing to use the same mathematical methods for manipulating probabilities within a probability model. The next section will lay the foundations for all such calculations.

1.2 Axioms and rules of probability

Kolmogorov's axioms

Consider the sample space S. There are restrictions on the probabilities that we can assign to events in S in order to make sure that the probabilities we assign to different events are consistent with each other. A set of probabilities will give consistent results if it obeys the axioms of probability stated by Kolmogorov (1933).

The axioms of probability

(1) $0 \leqslant P(E) \leqslant 1$, for any event $E \subseteq S$

(2) $P(S) = 1$

(3) If E, F are disjoint events, then $P(E \cup F) = P(E) + P(F)$

(4) If $E_1, E_2, \ldots$ are disjoint events, then

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i).$$

At first sight, it might seem that Axiom 4 is redundant given Axiom 3. In fact, Axiom 3 can only be extended by induction to a finite collection of events $E_1, \ldots, E_k$, whereas Axiom 4 can refer to an infinite collection of events. On the other hand, Axiom 3 can be recovered from Axiom 4 by putting $E_1 = E$, $E_2 = F$ and $E_3 = E_4 = \cdots = \emptyset$.

Example 3

An experiment has k equally-likely outcomes, $s_1, s_2, \ldots, s_k$. For any event E, $P(E)$ is defined to be $n(E)/k$, where $n(E)$ is the number of different outcomes in E. Show that these probabilities satisfy the axioms of probability.

Solution

Rules of probability

If, every time we wished to work out the probability of a specific event, we had to work directly from the axioms we would find this extremely tedious. In fact, we can easily derive from the Axioms other general results that are more useful to us; these are commonly known as rules of probability. .

Example 4

Let $E \subseteq S$ be any event. Prove that $P(E') = 1 - P(E)$.

Solution

- $E \cup E' = S \implies P(E \cup E') = P(S) = 1$ (using Axiom 2)
- E, E' disjoint $\implies P(E) + P(E') = P(E \cup E')$ (using Axiom 3)
- So, $P(E') = P(E \cup E') - P(E) = 1 - P(E)$

Example 5

Prove that $P(\emptyset) = 0$.

Solution

- $P(\emptyset) = P(S') = 1 - P(S) = 1 - 1 = 0$

Example 6

$E, F \subseteq S$ such that $E \subseteq F$. Prove that $P(E) \leqslant P(F)$.

Solution

- $0 \leqslant P(F \setminus E)$ (by Axiom 1)
- $F = E \cup (F \setminus E)$, a disjoint union
- So, $P(F) = P(E) + P(F \setminus E) \implies P(F) \geqslant P(E)$

Example 7

Let $E, F \subseteq S$ be any two events (not necessarily disjoint). Prove that: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Solution

- The event $E \cup F$ is the union of three disjoint events:
  - $E \cap F$ (the outcomes that are common to both E and F)
  - $E \setminus (E \cap F)$ (the outcomes that are in E but not also in F)
  - $F \setminus (E \cap F)$ (the outcomes that are in F but not also in E)
- $P(E \cup F) = P(E \cap F) + P\big(E \setminus (EF)\big) + P\big(F \setminus (EF)\big)$ (Axiom 4) [1]
- But E is itself a disjoint union $E = (E \cap F) \cup \big(E \setminus (E \cap F)\big)$
- So: $P(E) = P(E \cap F) + P\big(E \setminus (E \cap F)\big) \implies P\big(E \setminus (E \cap F)\big) = P(E) - P(E \cap F)$ [2]
- Similarly: $P(F) = P(E \cap F) + P\big(F \setminus (E \cap F)\big) \implies P\big(F \setminus (E \cap F)\big) = P(F) - P(E \cap F)$ [3]
- Substituting [2] and [3] into [1] gives $P(E \cup F) = P(E \cap F) + [P(E) - P(E \cap F)] + [P(F) - P(E \cap F)] = P(E) + P(F) - P(E \cap F)$
- Note: This rule of probability reduces to Axiom 3 when E and F are disjoint because, in that case, $P(E \cap F) = P(\emptyset) = 0$

Example 8

Let $E_1, E_2, E_3 \subseteq S$ be any three events. Prove that

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) \\ &\quad - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) \\ &\quad + P(E_1 \cap E_2 \cap E_3) \end{aligned}$$

Solution

Alternative solution

It is left as a tutorial example to complete the following sketch proof.

- Show that $E_1 \cup E_2 \cup E_3$ is the disjoint union of $E_1 \cap E_2 \cap E_3$, $E_1 \cap E_2 \cap E_3'$, $E_1 \cap E_2' \cap E_3$, $E_1' \cap E_2 \cap E_3$, $E_1 \cap E_2' \cap E_3'$, $E_1' \cap E_2' \cap E_3$ and $E_1' \cap E_2 \cap E_3'$.
- Write $P(E_1 \cup E_2 \cup E_3)$ in terms of the probabilities of these events.
- Regroup the separate probabilities to obtain the result.

1.3 Conditional probability

Definition of conditional probability

Suppose that E and F are events in a sample space, S. In general, knowing that E has occurred (or assuming that E will occur) might change our assessment of the probability that F will occur.

Definition

Suppose that $E, F \subseteq S$ are events such that $P(E) > 0$. The conditional probability of F given that E occurs is defined to be:

$$P(F|E) = \frac{P(E \cap F)}{P(E)}.$$

Example 9

A bag contains 100 balls that are identical apart from colour. 67 of them are black and the other 33 red. Consider the experiment of drawing balls from the bag at random and without replacement.

(a) Find the probability that the first ball drawn is black.

(b) Find the probability that the first ball drawn is black and the second ball drawn is red.

(c) Given that the first ball drawn is black, find the conditional probability that the second ball drawn is red.

Solution

Let E = 'first ball drawn is black' and F = 'second ball drawn is red'.

(a) $P(E) = \frac{67}{100} = 0.67$

(b)
- Number of different (equally-likely) ways of drawing 2 balls $= 100 \times 99$
- Number of different ways of drawing a black ball then a red one $= 67 \times 33$
- So $P(E \cap F) = \frac{67 \times 33}{100 \times 99} = 0.2233$

(c)

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{(67 \times 33)/(100 \times 99)}{67/100} = \frac{33}{99} = \frac{1}{3}.$$

(Given that E occurs, 99 balls remain in the bag, 33 of which are red.)

The multiplication theorem of probability

Suppose that $E, F \subseteq S$ are events such that $P(E) > 0$. Then:

$$P(E \cap F) = P(F|E)P(E).$$

Example 10

An investor is considering investing in shares ("equities") and gold. A financial expert has told the investor that, in the coming year, there is probability 0.6 that shares will rise in value, probability 0.8 that gold will rise in value, and probability 0.5 that both will rise in value.

(a) What is the probability that at least one of these assets will rise in value?

(b) What is the probability that gold will rise in value but shares will not?

(c) What is the conditional probability that shares will rise in value given that gold rises in value?

Solution

Let E = 'shares will rise in value' and F = 'gold will rise in value'. We are given that $P(E) = 0.6$, $P(F) = 0.8$, $P(E \cap F) = 0.5$. So:

Example 10 (continued)

The same investor now considers investing in property as well. According to the expert, given that shares rise in value in the year ahead, then property will rise in value with probability 0.9.

(d) What is the probability that both shares and property will rise in value in the year ahead?

(e) What is known about the probability that property will rise in value?

Solution

Let G = 'property will rise in value'. Then:

Independence

The independence of events is one of the most crucial concepts in probability. Informally, we say that E and F are independent events if the occurrence or non-occurrence of F does not change the probability of E (and vice versa).

Definition

Suppose that $E, F \subseteq S$ are events. Then E and F are defined to be independent if

$$P(E \cap F) = P(E) \times P(F).$$

Suppose that E and F are independent according to this definition. Assume that $P(E)$ is greater than 0, so that $P(F|E)$ is defined. Then

$$P(E \cap F) = P(E) \times P(F) \Leftrightarrow \frac{P(E \cap F)}{P(E)} = P(F) \Leftrightarrow P(F|E) = P(F).$$

In other words, when $P(E) > 0$, E and F are independent if and only if the conditional probability of F given E is the same as the unconditional probability of F. Similarly, if $P(F) > 0$, then E and F are independent if and only if the conditional probability of E given F is the same as the unconditional probability of E. Intuitively, these are precisely the relationships we want the formal definition of independence to express.

In practice, it is very rare to discover that two events are independent by working out separately the probabilities of E, F and $E \cap F$ and showing that they are related by the above formula. Almost invariably, independence is a model assumption justified by knowledge of the experimental conditions. When events are independent, even apparently complex probabilities become relatively easy to calculate.

Example 11

Two 'fair', standard dice are rolled and the scores on the dice are added to give a total score. Find the probability that the total score is 7.

Solution

We can extend the concept of independence to more than 2 events. For example, the 3 events E, F and G are said to be independent if:

- the events are pairwise independent (i.e., E and F are independent, E and G are independent, F and G are independent), and
- $P(E \cap F \cap G) = P(E) \times P(F) \times P(G)$.

Notice that both these conditions are required in the definition, since neither necessarily implies the other (see Tutorial Examples).

Example 12

Three fair standard dice are rolled. Find the probability that:

(a) the same score is obtained on all three dice;

(b) different scores are obtained on all three dice;

(c) the same score is obtained on two of the dice, but a different score on the third.

Solution

$$S = \{(x, y, z) : x = 1, 2, \ldots, 6; y = 1, 2, \ldots, 6; z = 1, 2, \ldots, 6\}$$

The scores on the three dice are independent, so the $6^3 = 216$ outcomes in S are all equally likely, with probability $\frac{1}{216}$.

(a)
- $E_1 =$ 'same score on all three dice' $=$
- $\implies P(E_1) =$

(b)
- $E_2 =$ 'different scores on all three dice'
- number of possible values for $x =$
- number of different values for $y =$
- number of different values for $z =$
- $\implies$ number of outcomes in $E_2 =$
- $P(E_2) =$

(c)
- $E_3 =$ 'same score on two dice, different score on third die'

Suppose that $E, F \subseteq S$ are disjoint events which have non-zero probabilities. Then E and F cannot be independent. This is intuitively obvious, since knowing that E has occurred tells us that F cannot occur. Formally, $P(E \cap F) = 0$ whereas $P(E), P(F) > 0$, and so $P(E \cap F) \neq P(E) \times P(F)$.

Table 1.1 summarises our results about disjoint and independent events.

Partitions and total probability

For any events $E, F \subseteq S$, we have already shown that:

$$F = (F \cap E) \cup (F \cap E') \implies P(F) = P(F \cap E) + P(F \cap E').$$

Using the multiplication theorem of probability, assuming that $P(E) > 0$ and $P(E') > 0$, this implies that

$$P(F) = P(F|E)P(E) + P(F|E')P(E').$$

The law of total probability is an extension of this result.

Table 1.1: Events E and F such that $P(E) > 0$ and $P(F) > 0$.

E, F disjoint
$P(E \cap F) = 0$
$P(E \cup F) = P(E) + P(F)$
$P(E\|F) = 0$
$P(F\|E) = 0$

E and F not disjoint	
E, F independent	E, F not independent
$P(E \cap F) = P(E)P(F)$	$P(E \cap F) = P(E\|F)P(F) = P(F\|E)P(E)$
$P(E \cup F) = P(E) + P(F) - P(E)P(F)$	$P(E \cup F) = P(E) + P(F) - P(E \cap F)$
$P(E\|F) = P(E)$	$P(E\|F) = P(E \cap F)/P(F)$
$P(F\|E) = P(F)$	$P(F\|E) = P(F \cap E)/P(E)$

Definition

A collection of events $E_1, E_2, \ldots$ forms a partition of the sample space S if:

- $\bigcup_i E_i = S$
- $E_i \cap E_j = \emptyset$, when $i \neq j$
- $P(E_i) > 0$ for each i

In particular, if E is any event with $0 < P(E) < 1$, E and E' partition the sample space.

The law of total probability

Suppose that the events $E_1, E_2, \ldots$ partition the sample space S. Let F be any event in S. Then:

$$P(F) = \sum_i P(F|E_i)P(E_i).$$

Proof

It follows from the distributive laws of set theory (Section 1.1) that:

$$F = F \cap S = F \cap \left(\bigcup_i E_i \right) = \bigcup_i (F \cap E_i).$$

Since the events E_i $(i = 1, 2, \ldots)$ are disjoint, the events $F \cap E_i$ $(i = 1, 2, \ldots)$ are also disjoint. So,

$$\begin{aligned} P(F) &= \sum_i P(F \cap E_i) \qquad \text{(Axiom 4)} \\ &= \sum_i P(F|E_i)P(E_i) \qquad \text{(Multiplication theorem)} \end{aligned}$$

Example 13

Helicobacter pylori is a bacterium that is harmful to human beings, by lodging in the lining of the stomach and intestine causing gastritis and peptic ulcers. A combination drug therapy has been developed for eradicating *Helicobacter pylori*. The effectiveness of this treatment depends on whether or not a patient is resistant to the effect of a particular chemical compound (Metronidazole), but a patient's resistance status is not routinely determined before beginning treatment to eradicate *Helicobacter pylori*. It is estimated that this treatment successfully eradicates *Helicobacter pylori* in 75% of resistant patients and in 95% of non-resistant patients.

Suppose that, in a certain population of individuals infected with *Helicobacter pylori*, 25% are resistant. In what proportion of patients from this population is the treatment unsuccessful?

Solution

Example 13 (continued)

If a patient's treatment is unsuccessful, what is the probability that that patient is resistant? In order to answer this question, we require to use Bayes' Theorem.

Bayes' Theorem

Suppose that the events $E_1, E_2, \ldots$ partition the sample space S. Let F be any event in S, with $P(F) > 0$. Then:

$$P(E_j|F) = \frac{P(F|E_j)P(E_j)}{\sum_i P(F|E_i)P(E_i)} \qquad (j = 1, 2, \ldots).$$

Proof

By the multiplication theorem of probability, $P(F \cap E_j) = P(F|E_j)P(E_j)$ and $P(F \cap E_j) = P(E_j \cap F) = P(E_j|F)P(F)$. So $P(E_j|F)P(F) = P(F|E_j)P(E_j)$. That is,

$$P(E_j|F) = \frac{P(F|E_j)P(E_j)}{P(F)} = \frac{P(F|E_j)P(E_j)}{\sum_i P(F|E_i)P(E_i)}.$$

(The last line follows from the law of total probability.)

Example 13 (continued)