PROBABILITIES II (DA_2022)

Andrea Giansanti

Dipartimento di Fisica, Sapienza Università di Roma

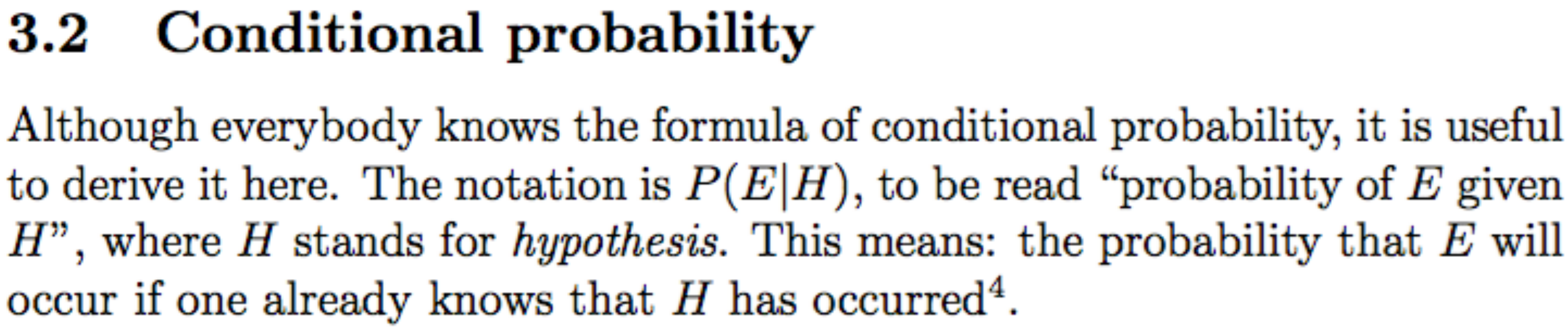
Andrea.Giansanti@roma1.infn.it

DA_2022 Lecture n. 7, Rome 23rd March 2022

DIPARTIMENTO DI FISICA

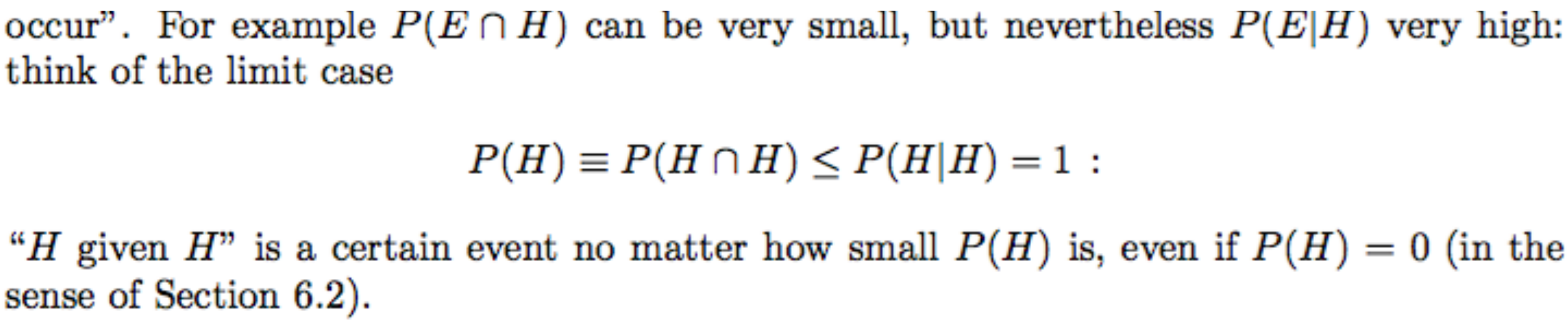
SAPIENZA
UNIVERSITÀ DI ROMA

Let us start, as promised, with the definition of:



3.2 Conditional probability

Although everybody knows the formula of conditional probability, it is useful to derive it here. The notation is $P(E|H)$, to be read "probability of E given H", where H stands for *hypothesis*. This means: the probability that E will occur if one already knows that H has occurred[4].



occur". For example $P(E \cap H)$ can be very small, but nevertheless $P(E|H)$ very high: think of the limit case

$$P(H) \equiv P(H \cap H) \leq P(H|H) = 1 \ :$$

"H given H" is a certain event no matter how small $P(H)$ is, even if $P(H) = 0$ (in the sense of Section 6.2).

From: G. D'Agostini, ArXiv preprint database: https://arxiv.org/pdf/hep-ph/9512295.pdf

MOTIVATION OF LECTURE N. 7

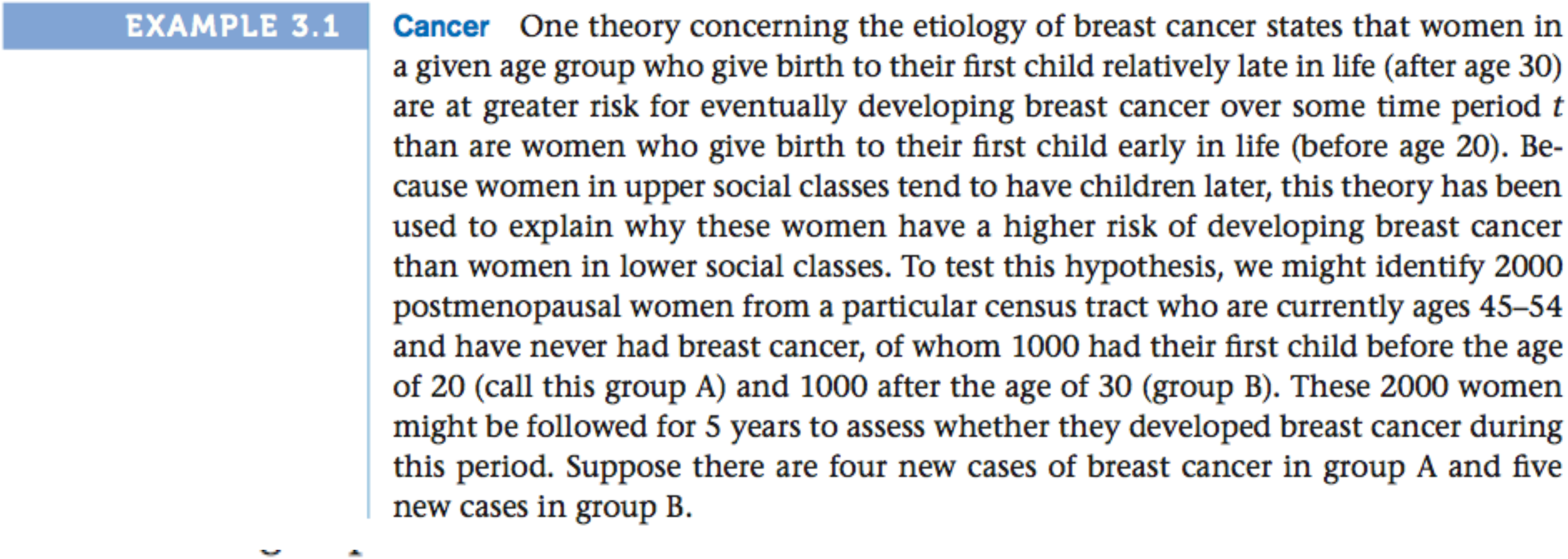
PROBABILITIES II

We shall use chapter 3 by Rosner's textbook to get down-to-Earth and see, in concrete contexts and examples, how are used the concepts we introduced and discussed in the last lecture:

- Uncertainty/decisions
- Events, experiment, probability
- Definitions of probabilities: classic, frequentist, subjective
- Axioms of probabilities
- Events, sets, propositions (logic)
- Venn Diagrams

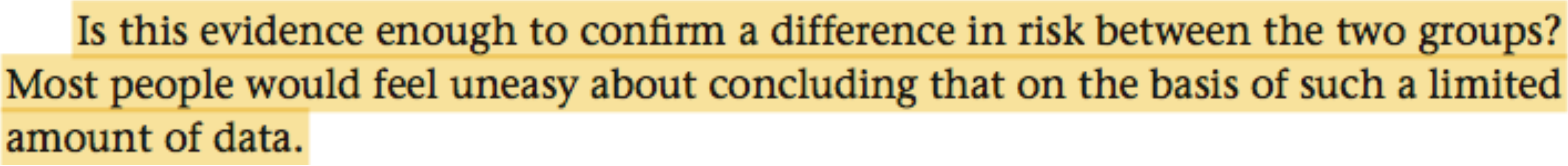
DOWN_TO_EARTH MOTIVATION

Comparing Case/Controls, a general scheme



EXAMPLE 3.1

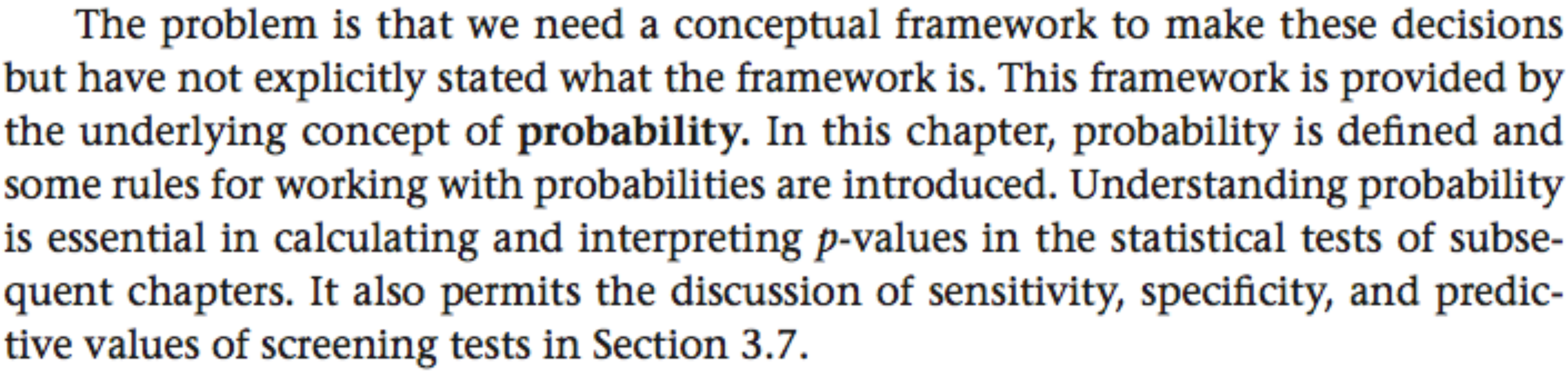
Cancer One theory concerning the etiology of breast cancer states that women in a given age group who give birth to their first child relatively late in life (after age 30) are at greater risk for eventually developing breast cancer over some time period *t* than are women who give birth to their first child early in life (before age 20). Because women in upper social classes tend to have children later, this theory has been used to explain why these women have a higher risk of developing breast cancer than women in lower social classes. To test this hypothesis, we might identify 2000 postmenopausal women from a particular census tract who are currently ages 45–54 and have never had breast cancer, of whom 1000 had their first child before the age of 20 (call this group A) and 1000 after the age of 30 (group B). These 2000 women might be followed for 5 years to assess whether they developed breast cancer during this period. Suppose there are four new cases of breast cancer in group A and five new cases in group B.



Is this evidence enough to confirm a difference in risk between the two groups? Most people would feel uneasy about concluding that on the basis of such a limited amount of data.

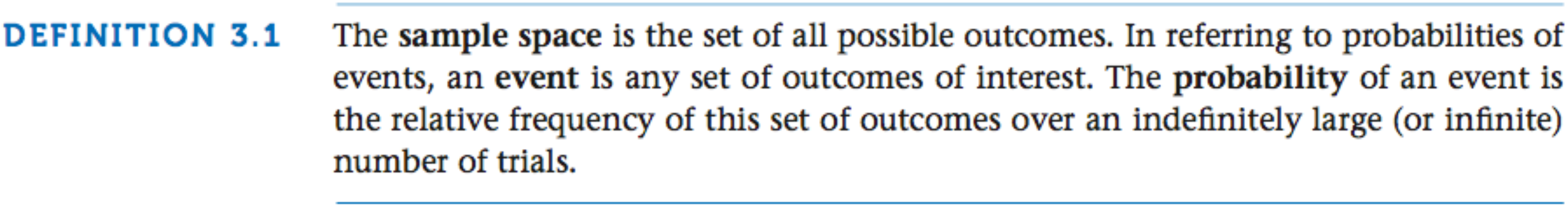
From:Bernard Rosner - Fundamentals of Biostatistics-Brooks Cole (2015), Chapter 3.

MOTIVATION



The problem is that we need a conceptual framework to make these decisions but have not explicitly stated what the framework is. This framework is provided by the underlying concept of **probability.** In this chapter, probability is defined and some rules for working with probabilities are introduced. Understanding probability is essential in calculating and interpreting *p*-values in the statistical tests of subsequent chapters. It also permits the discussion of sensitivity, specificity, and predictive values of screening tests in Section 3.7.

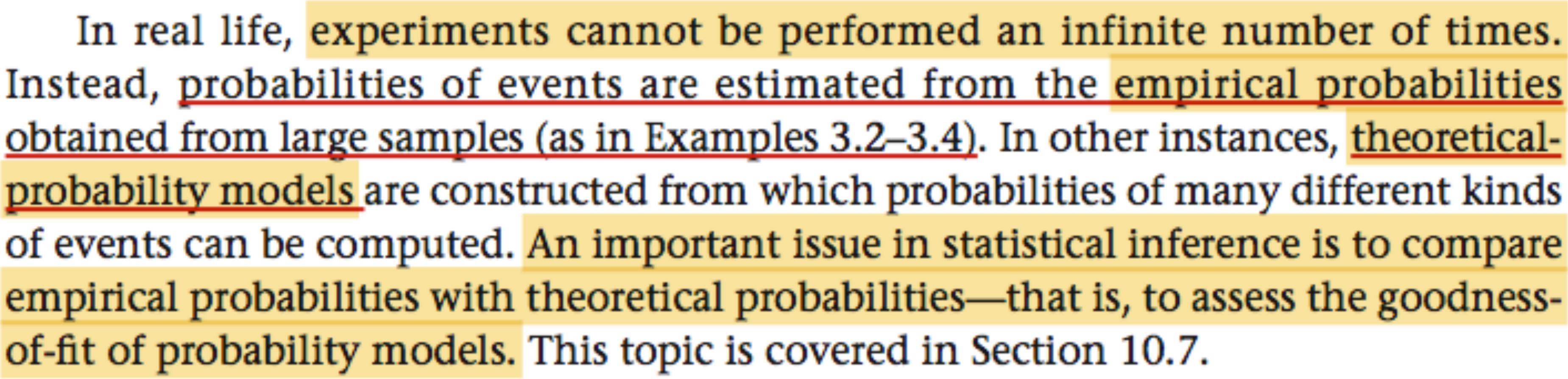
From:Bernard Rosner - Fundamentals of Biostatistics-Brooks Cole (2015), Chapter 3.



DEFINITION 3.1 The **sample space** is the set of all possible outcomes. In referring to probabilities of events, an **event** is any set of outcomes of interest. The **probability** of an event is the relative frequency of this set of outcomes over an indefinitely large (or infinite) number of trials.

Aha! Rosner follows a FREQUENTIST scheme.

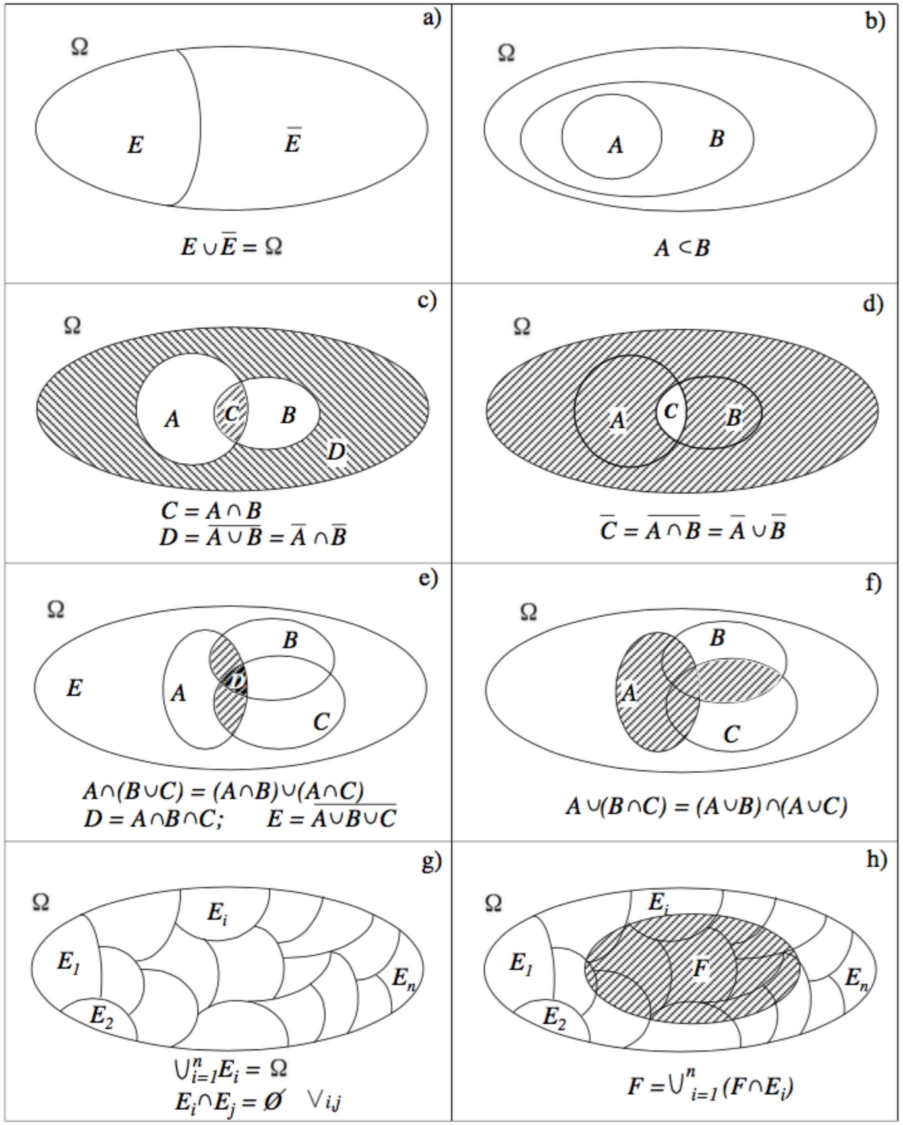
Illustrated by several examples: 3.2, 3.3 and3.4



In real life, experiments cannot be performed an infinite number of times. Instead, probabilities of events are estimated from the empirical probabilities obtained from large samples (as in Examples 3.2–3.4). In other instances, theoretical-probability models are constructed from which probabilities of many different kinds of events can be computed. An important issue in statistical inference is to compare empirical probabilities with theoretical probabilities—that is, to assess the goodness-of-fit of probability models. This topic is covered in Section 10.7.

EVENTS AS SETS I

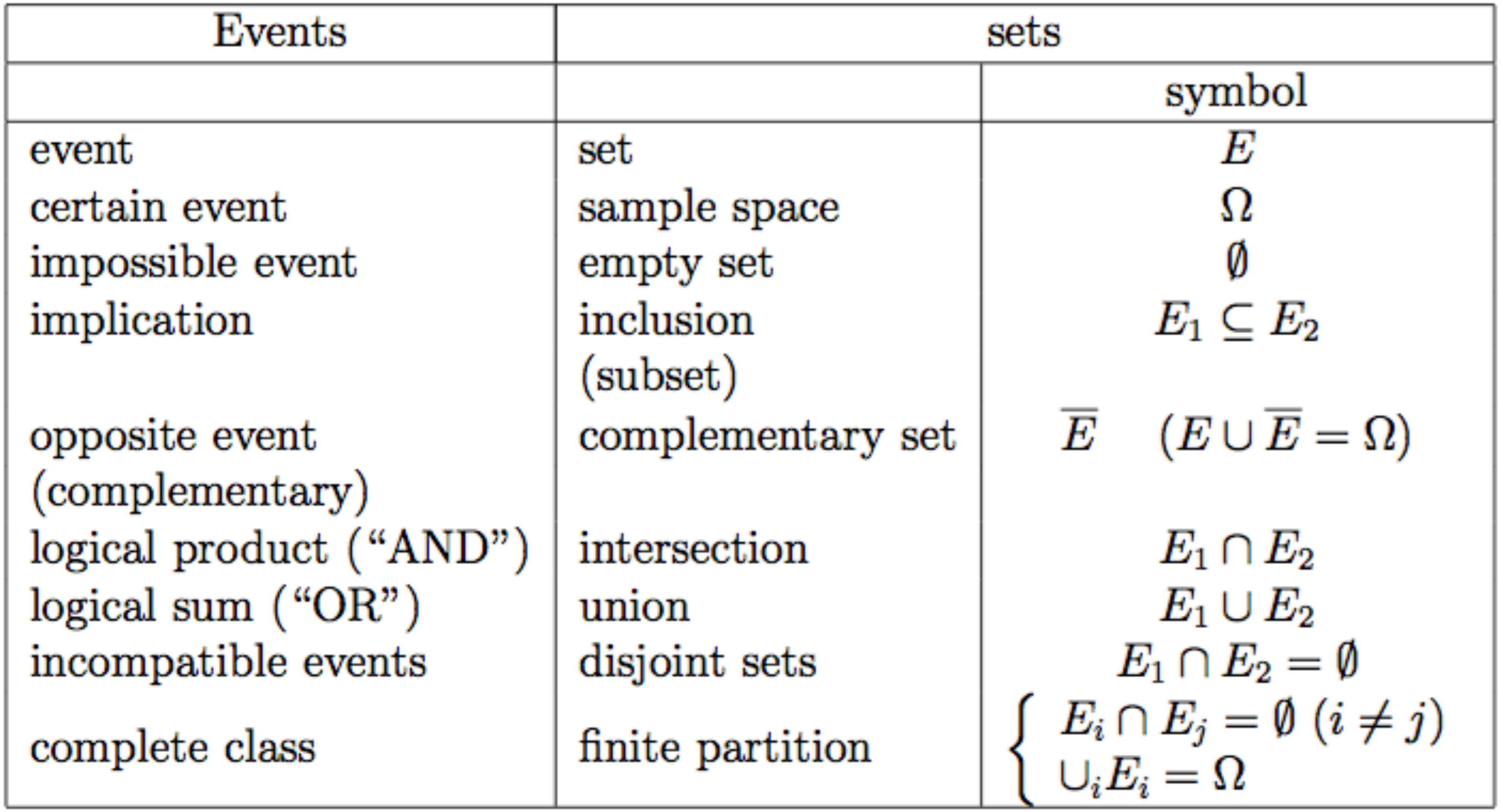
Think about playing darts!



a)
Ω
E
$\bar{E}$
$E \cup \bar{E} = \Omega$
b)
Ω
A
B
$A \subset B$
c)
Ω
A
C
B
D
$C = A \cap B$
$D = \overline{A \cup B} = \bar{A} \cap \bar{B}$
d)
Ω
A
C
B
$\bar{C} = \overline{A \cap B} = \bar{A} \cup \bar{B}$
e)
Ω
E
B
A
D
C
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$D = A \cap B \cap C; \quad E = \overline{A \cup B \cup C}$
f)
Ω
B
A
C
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
g)
Ω
E_i
E_1
E_2
E_n
$\bigcup_{i=1}^{n} E_i = \Omega$
$E_i \cap E_j = \emptyset \quad \forall i,j$
h)
Ω
E_i
E_1
F
E_2
E_n
$F = \bigcup^{n}_{i=1} (F \cap E_i)$

Figure 2: Venn diagrams and set properties.

EVENTS AS SETS II



Events	sets	
		symbol
event	set	E
certain event	sample space	Ω
impossible event	empty set	$\emptyset$
implication	inclusion (subset)	$E_1 \subseteq E_2$
opposite event (complementary)	complementary set	$\overline{E} \quad (E \cup \overline{E} = \Omega)$
logical product (“AND”)	intersection	$E_1 \cap E_2$
logical sum (“OR”)	union	$E_1 \cup E_2$
incompatible events	disjoint sets	$E_1 \cap E_2 = \emptyset$
complete class	finite partition	$\begin{cases} E_i \cap E_j = \emptyset \ (i \neq j) \\ \cup_i E_i = \Omega \end{cases}$

Table 1: Events versus sets.

Since everybody is familiar with the axioms and with the analogy *events* $\Leftrightarrow$ *sets* (see Tab. 1 and Fig. 2) let us remind ourselves of the *rules of probability* in this form:

Axiom 1 $0 \leq P(E) \leq 1$;

Axiom 2 $P(\Omega) = 1$ (a certain event has probability 1);

Axiom 3 $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, if $E_1 \cap E_2 = \emptyset$

From the basic rules the following properties can be derived:

1: $P(E) = 1 - P(\overline{E})$;

2: $P(\emptyset) = 0$;

3: if $A \subseteq B$ then $P(A) \leq P(B)$;

4: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

We also anticipate here a fifth property which will be discussed in section 3.1:

5: $P(A \cap B) = P(A|B) \cdot P(B) = P(A) \cdot P(B|A)$.

EQUATION 3.1

(1) The probability of an event E, denoted by $Pr(E)$, always satisfies $0 \le Pr(E) \le 1$.

(2) If outcomes A and B are two events that cannot both happen at the same time, then $Pr(A \text{ or } B \text{ occurs}) = Pr(A) + Pr(B)$.

DEFINITION 3.2 Two events A and B are **mutually exclusive** if they cannot both happen at the same time.

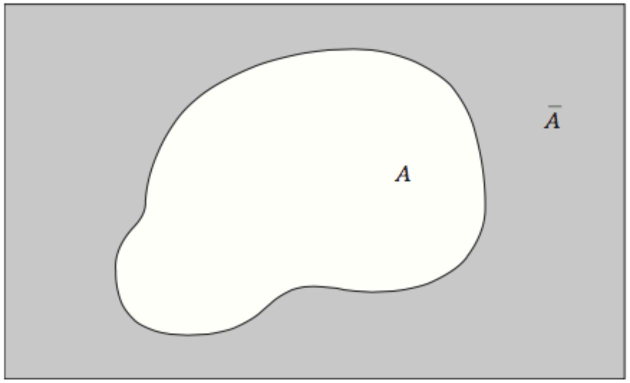
DEFINITION 3.3 The symbol { } is used as shorthand for the phrase “the event.”

DEFINITION 3.4 $A \cup B$ is the event that either A or B occurs, or they both occur.

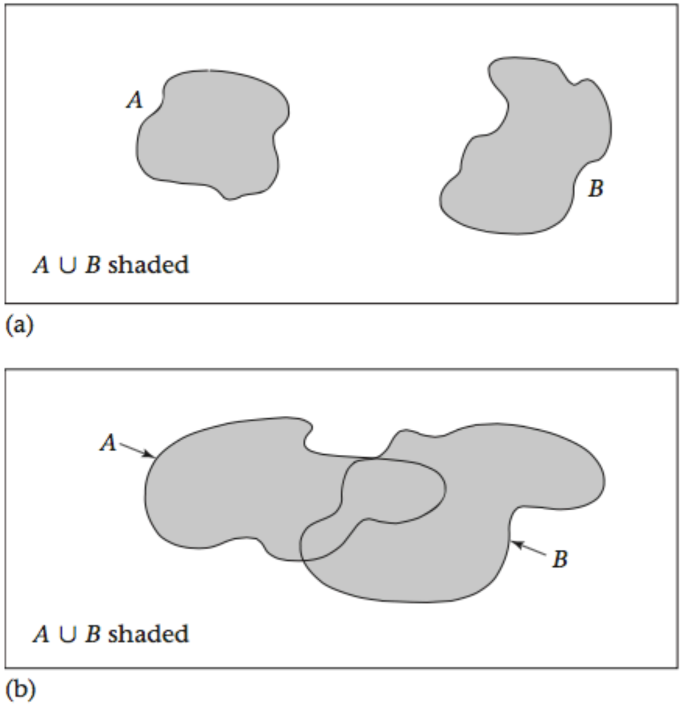
DEFINITION 3.5 $A \cap B$ is the event that both A and B occur simultaneously. $A \cap B$ is depicted diagrammatically in Figure 3.2.

DEFINITION 3.6 $\bar{A}$ is the event that A does not occur. It is called the **complement** of A. Notice that $Pr(\bar{A}) = 1 - Pr(A)$, because $\bar{A}$ occurs only when A does not occur. Event $\bar{A}$ is diagrammed in Figure 3.3.

FIGURE 3.3 Diagrammatic representation of $\bar{A}$

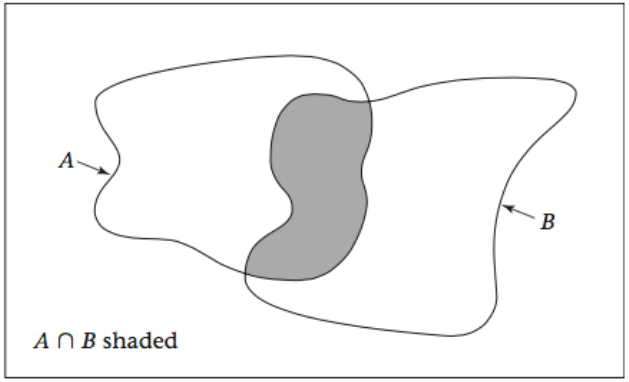


Diagrammatic representation of $A \cup B$: (a) A, B mutually exclusive; (b) A, B not mutually exclusive



Various Venn diagrams

Diagrammatic representation of $A \cap$



Very general equations for probabilities

Multiplication Law of Probability

If $A_1, \ldots, A_k$ are mutually independent events,

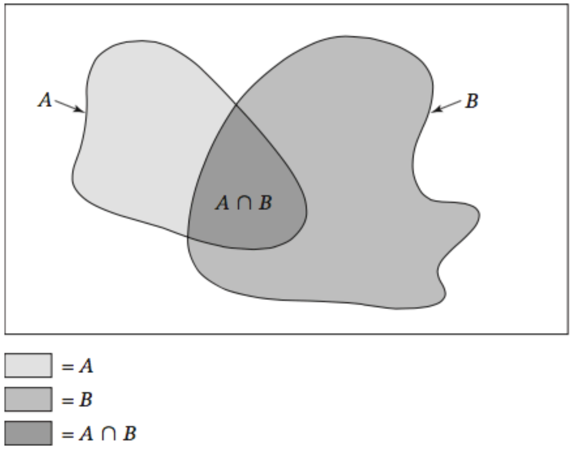
then $Pr\left(A_1 \cap A_2 \cap \ldots \cap A_k\right) = Pr\left(A_1\right) \times Pr\left(A_2\right) \times \ldots \times Pr\left(A_k\right)$

Addition Law of Probability

If A and B are any events,

then $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

Diagrammatic representation of the addition law of probability

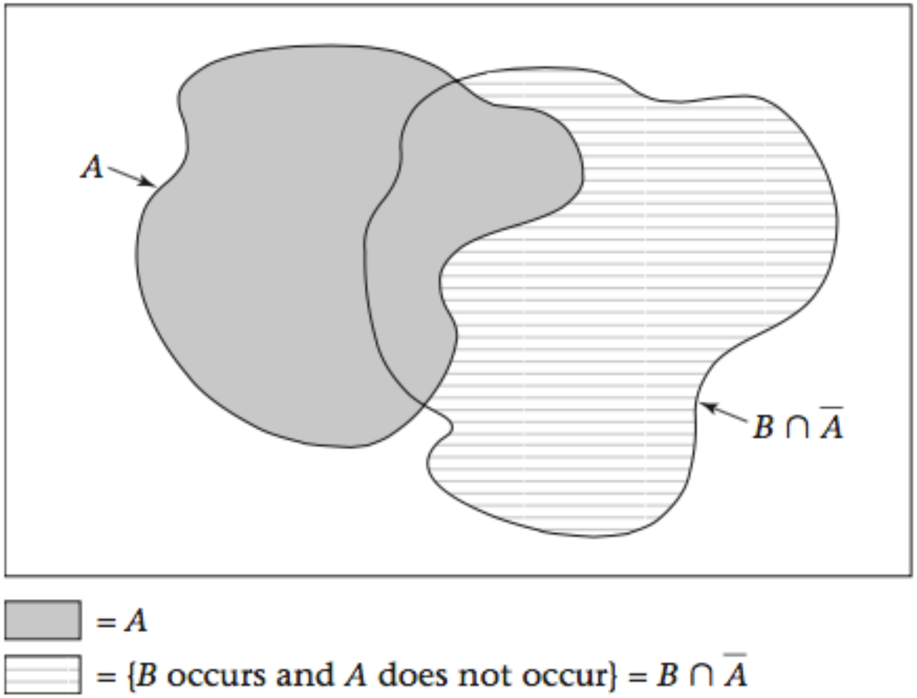


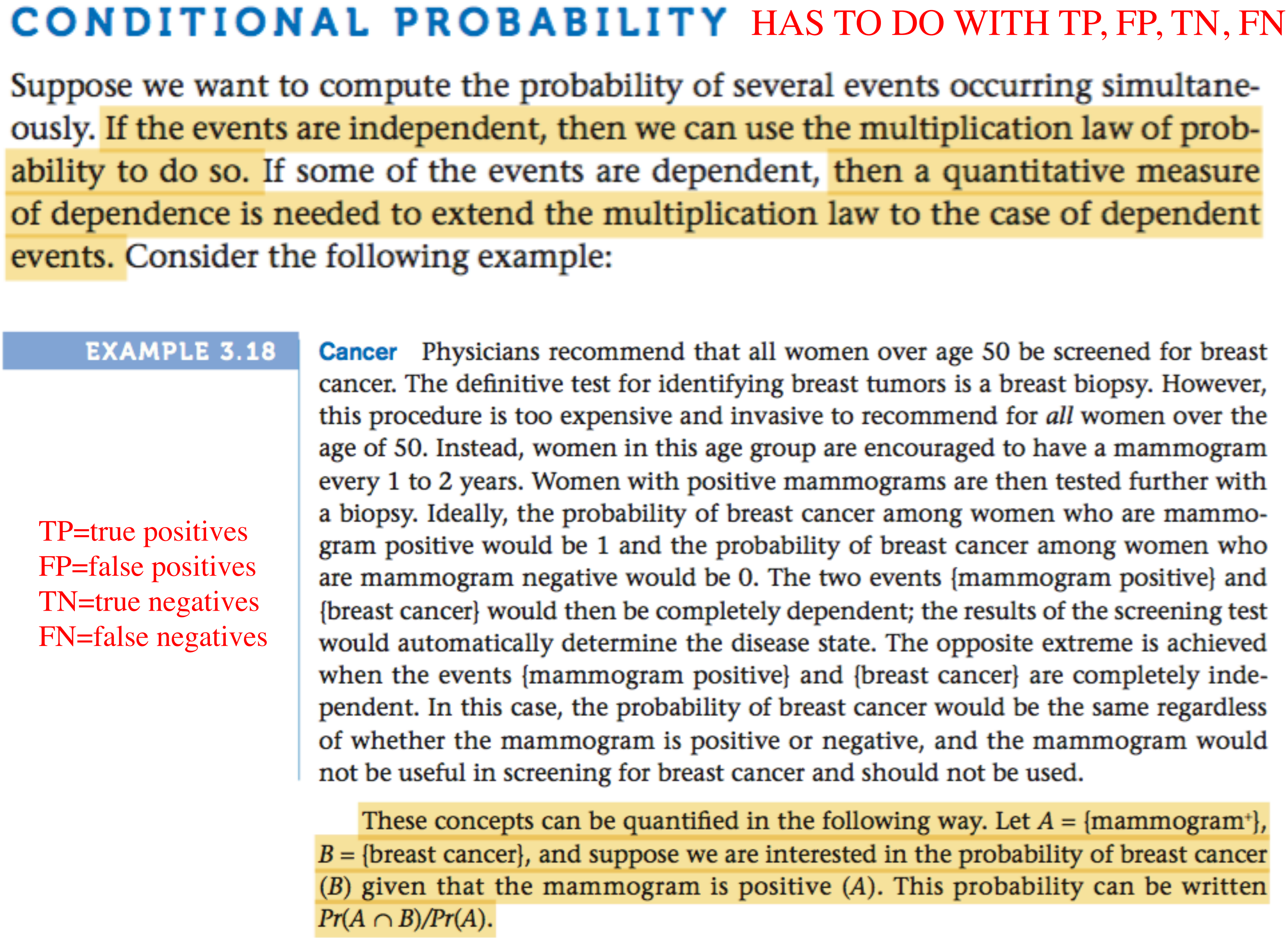
Addition Law of Probability for Independent Events

If two events *A* and *B* are independent, then

$$Pr(A \cup B) = Pr(A) + Pr(B) \times [1 - Pr(A)]$$

Diagrammatic representation of the addition law of probability for independent events





CONDITIONAL PROBABILITY HAS TO DO WITH TP, FP, TN, FN

Suppose we want to compute the probability of several events occurring simultaneously. If the events are independent, then we can use the multiplication law of probability to do so. If some of the events are dependent, then a quantitative measure of dependence is needed to extend the multiplication law to the case of dependent events. Consider the following example:

EXAMPLE 3.18

TP=true positives
FP=false positives
TN=true negatives
FN=false negatives

Cancer Physicians recommend that all women over age 50 be screened for breast cancer. The definitive test for identifying breast tumors is a breast biopsy. However, this procedure is too expensive and invasive to recommend for *all* women over the age of 50. Instead, women in this age group are encouraged to have a mammogram every 1 to 2 years. Women with positive mammograms are then tested further with a biopsy. Ideally, the probability of breast cancer among women who are mammogram positive would be 1 and the probability of breast cancer among women who are mammogram negative would be 0. The two events {mammogram positive} and {breast cancer} would then be completely dependent; the results of the screening test would automatically determine the disease state. The opposite extreme is achieved when the events {mammogram positive} and {breast cancer} are completely independent. In this case, the probability of breast cancer would be the same regardless of whether the mammogram is positive or negative, and the mammogram would not be useful in screening for breast cancer and should not be used.

These concepts can be quantified in the following way. Let A = {mammogram$^+$}, B = {breast cancer}, and suppose we are interested in the probability of breast cancer (B) given that the mammogram is positive (A). This probability can be written $Pr(A \cap B)/Pr(A)$.

DEFINITION 3.9 The quantity $Pr(A \cap B)/Pr(A)$ is defined as the **conditional probability of B given A**, which is written $Pr(B|A)$.

However, from Section 3.4 we know that, by definition of the multiplication law of probability, if two events are independent, then $Pr(A \cap B) = Pr(A) \times Pr(B)$. If both sides are divided by $Pr(A)$, then $Pr(B) = Pr(A \cap B)/Pr(A) = Pr(B|A)$. Similarly, we can show that if A and B are independent events, then $Pr(B|\bar{A}) = Pr(B|A) = Pr(B)$. This relationship leads to the following alternative interpretation of independence in terms of conditional probabilities.

EQUATION 3.5

(1) If A and B are independent events, then $Pr(B|A) = Pr(B) = Pr(B|\bar{A})$.

(2) If two events A, B are dependent, then $Pr(B|A) \neq Pr(B) \neq Pr(B|\bar{A})$ and $Pr(A \cap B) \neq Pr(A) \times Pr(B)$.

DEFINITION 3.10 The **relative risk (*RR*)** of B given A is

$$Pr(B|A)/Pr(B|\bar{A})$$

Notice that if two events A, B are independent, then the ***RR*** is 1. If two events A, B are dependent, then the *RR* is different from 1. Heuristically, the more the dependence between events increases, the further the *RR* will be from 1.

REVIEW QUESTIONS 3A

1. What is the frequency definition of probability?
2. What is the difference between independent and dependent events?
3. What are mutually exclusive events?
4. What is the addition law of probability?
5. What is conditional probability? How does it differ from unconditional probability?
6. What is relative risk? How do you interpret it?

Total-Probability Rule

The conditional $\left(Pr(B \mid A), Pr(B \mid \bar{A})\right)$ and unconditional $(Pr(B))$ probabilities mentioned previously are related in the following way:

EQUATION 3.6

For any events A and B,

$$Pr(B) = Pr(B \mid A) \times Pr(A) + Pr\left(B \mid \bar{A}\right) \times Pr\left(\bar{A}\right)$$

This formula tells us that the unconditional probability of B is the sum of the conditional probability of B given A *times* the unconditional probability of A *plus* the conditional probability of B given A *not* occurring *times* the unconditional probability of A *not* occurring.

To derive this, we note that if the event B occurs, it must occur either with A or without A. Therefore,

$$Pr(B) = Pr(B \cap A) + Pr\left(B \cap \bar{A}\right)$$

From the definition of conditional probability, we see that

$$Pr(B \cap A) = Pr(A) \times Pr\left(B \middle| A\right)$$

and

$$Pr\left(B \cap \bar{A}\right) = Pr\left(\bar{A}\right) \times Pr\left(B \middle| \bar{A}\right)$$

By substitution, it follows that

$$Pr(B) = Pr\left(B \middle| A\right) Pr(A) + Pr\left(B \middle| \bar{A}\right) Pr\left(\bar{A}\right)$$

Stated another way, the unconditional probability of B is a weighted average of the probabilities of B occurring in two mutually exclusive subsets $(A, \bar{A})$, where the weights are the probabilities of the subsets $(Pr \mid A)$, $Pr(\bar{A})$, respectively.

In Equation 3.6 the probability of event B is expressed in terms of two mutually exclusive events A and $\bar{A}$. In many instances the probability of an event B can be determined in more than two mutually exclusive subsets, denoted by $A_1, A_2, \ldots, A_k$.

DEFINITION 3.11 A set of events $A_1, \ldots, A_k$ is exhaustive if at least one of the events must occur.

Assume that events $A_1, \ldots, A_k$ are mutually exclusive and exhaustive; that is, at least one of the events $A_1, \ldots, A_k$ must occur and no two events can occur simultaneously. Thus, exactly one of the events $A_1, \ldots, A_k$ must occur.

EQUATION 3.7

Total-Probability Rule

Let $A_1, \ldots, A_k$ be mutually exclusive and exhaustive events. The unconditional probability of B $(Pr(B))$ can then be written as a weighted average of the conditional probabilities of B given $A_i \left(Pr(B|A_i)\right)$ with weights $= Pr(A_i)$ as follows:

$$Pr(B) = \sum_{i=1}^{k} Pr\left(B|A_i\right) \times Pr(A_i)$$

To show this, we note that if B occurs, then it must occur together with one and only one of the events, $A_1, \ldots, A_k$. Therefore,

$$Pr(B) = \sum_{i=1}^{k} Pr(B \cap A_i)$$

Also, from the definition of conditional probability,

$$Pr(B \cap A_i) = Pr(A_i) \times Pr\left(B|A_i\right)$$

By substitution, we obtain Equation 3.7.

An application of the total-probability rule is given in the following example:

OUTLINE.

- Events, trials, uncertainty, probabilities
- There are only conditional probabilities: P(H|I)
- Relevance of Bayes' theorem: (subjective/objective)
- Shift from frequentist to bayesian methods
- Jorge López Puga, Martin Krzywinski & Naomi Altman, *Bayes' Theorem*
- Nature Methods , 12, 277 (2015).
- Eikosograms (RW Oldford)

https://cran.r-project.org/web/packages/eikosograms/vignettes/Introduction.html

- Probability distributions: discrete/continuous
- Entropy of a probability distribution

- Study reference: Bernard Rossner, Fundamentals of Biostatistics par 3.7

3 Basic definitions

Let us consider just finite sets of events, this is, conceptually, not a big limitation. All the events we shall consider can be , formally, as subsets of a reference container set Ω, which contains every possible outcome of an experiment; e.g.

$$\Omega = \{\text{head, tail}\}$$

in the case of the toss of a coin. $x \in \Omega$ means "x is an element of Ω", or "x is an event, a subset belonging to Ω". We shall associate to each event $x \in \Omega$ a probability $p(x)$, that is a positive measure normalized to 1. In the discrete case, whre Ω is made by N events the set of the $p(x)$ is a set of N non negative numbers $p(x) \geq 0$ (each one associated to one of the $x \in \Omega$) and such that tali che $\sum_{x \in \Omega} p(x) = 1$. In the simple case of a tossed coin we just have two possible events: $x = \text{head}$ e $x = \text{tail}$ and the probability distribution is $p(\text{head}) = 1/2$, $p(\text{tail}) = 1/2$ (for a fair coin).

Here are some elementary properties that can be derived using Venn diagrams of the type shown in figure 2.

$$p(A) \geq 0 \quad , \quad p(\emptyset) = 0 \quad , \quad p(\Omega) = 1$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$A \cap B = \emptyset \Rightarrow p(A \cup B) = p(A) + p(B)$$

Given a set of N events in Ω: $\{A_1, A_2, ..., A_N\}$ they are *mutually exclusive* if the occurrence of one of them precludes the occurence of the rest of the others. In particular, if the N mutually exclusive events are a partition of Ω then $P(A_i) = 1 - P(\cup A_j), with j \neq i)$. N events in $\Omega$are *independent* if the occurrence of each one of them does not interfere with the occurrence of the others; in this case $P(\cap_i A_i) = \prod_i P(A_i)$. Two events that are not independent are said to be correlated and to express the degree of this correlation one introduces conditional probabilities. Correlated events have, quite intuitively, a non empty intersection. Let us then denote with $p(A|B)$ the probability of the occurrence of A, provided that B occurred, that is the conditional probability of A given B. We can consistently express the intersection of two correlated events A and B as:

$$p(A \cap B) = p(B)p(A|B)$$

that is, the probability of the co-occurrence of the correlated events A and B is given by the probability of A times the conditional probability of A given B. One has also:

$$p(A \cap B) = p(A)p(B|A)$$

, it is worth noting also that:

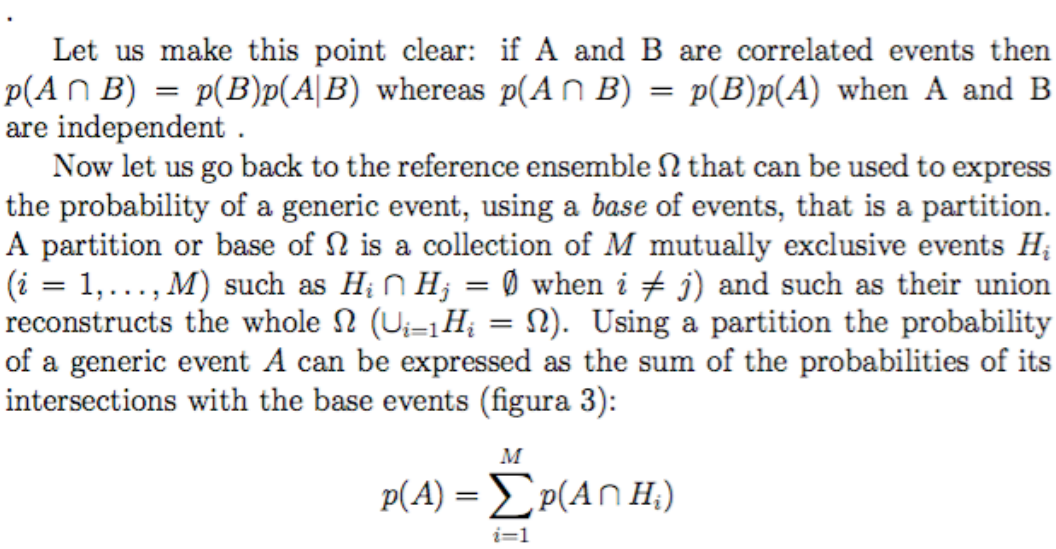
$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{p(B)p(A|B)}{p(A)}$$

and then, in general one has:

$$p(B|A) \neq p(A|B);$$

they are equal just in the case when $p(A) = p(B)$. If the occurrence of A is independent from the occurrence of B, then one has $p(A|B) = p(A)$ and the co-occurrenece of uncorrelated events A and B is just:

$$p(A \cap B) = p(A)p(B)$$



Let us make this point clear: if A and B are correlated events then $p(A \cap B) = p(B)p(A|B)$ whereas $p(A \cap B) = p(B)p(A)$ when A and B are independent .

Now let us go back to the reference ensemble Ω that can be used to express the probability of a generic event, using a *base* of events, that is a partition. A partition or base of Ω is a collection of M mutually exclusive events H_i ($i = 1, \dots, M$) such as $H_i \cap H_j = \emptyset$ when $i \neq j$) and such as their union reconstructs the whole Ω ($\cup_{i=1} H_i = \Omega$). Using a partition the probability of a generic event A can be expressed as the sum of the probabilities of its intersections with the base events (figura 3):

$$p(A) = \sum_{i=1}^{M} p(A \cap H_i)$$

Warning: the degree of correlation of two events g(A,B)=P(A|B)/P(A) is a symmetric notion, whereas
causal relations require asymmetry
Correlation is required for Causation but is not sufficient for

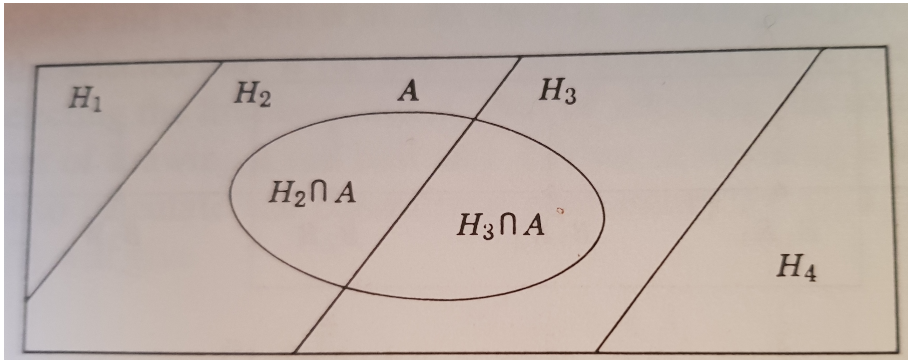


Fig. 11. A partitioning of a sample space.

5 Bayes' Theorem

Let us consider the methodological setting. Suppose you have a fact, an event to consider, E that you want to explain, to interpret, not making use of senses nor by concotting opoinions, but in a possibly transparent way, based on a quantitative analysis. Consider the "total" reference event of the calculus of probability Ω, we have introduced above. Then introduce a proper partition made by parts $\{H_i\}$ of Ω, to be used as a causative base to interpret E. In other words we want to determine the relative correlation of each one of the mutually exclusive H_i events in the partition with the event E. We shall express these correlation through conditional probabilities: of the form: $p(H_i|E)$. Let us start again from the general formula defining conditional probabilities, using events E and H_i: $p(H_i|E)p(E) = p(E \cap H_I) = p(E|H_i)p(H_i)$ and then, isolating $p(H_i|E)$. we get:

$$p(H_i|E) = \frac{p(H_i)p(E|H_i)}{p(E)}$$

,

which is equal to: $\frac{p(H_i)p(E|H_i)}{\sum_j p(E \cap H_j)}$, having used the projection of $p(E)$ over the base $\{H_i\}$, that is: $p(E) = \sum_j p(E \cap H_j)$.

$$\frac{p(H_i)p(E|H_i)}{\sum_j p(H_j)p(E|H_j)}$$

.

Introducing the normalization aka partition function: $Z = \sum_j p(H_j)p(E|H_j)$, we eventually get Bayes' formula in compact form:

$$p(H_i|E) = \frac{1}{Z}\, p(H_i)p(E|H_i)$$

.

Introducing the normalization aka partition function: $Z = \sum_j p(H_j)p(E|H_j)$, we eventually get Bayes' formula in compact form:

$$p(H_i|E) = \frac{1}{Z}\, p(H_i)p(E|H_i)$$

Eikosograms (RW Oldford)

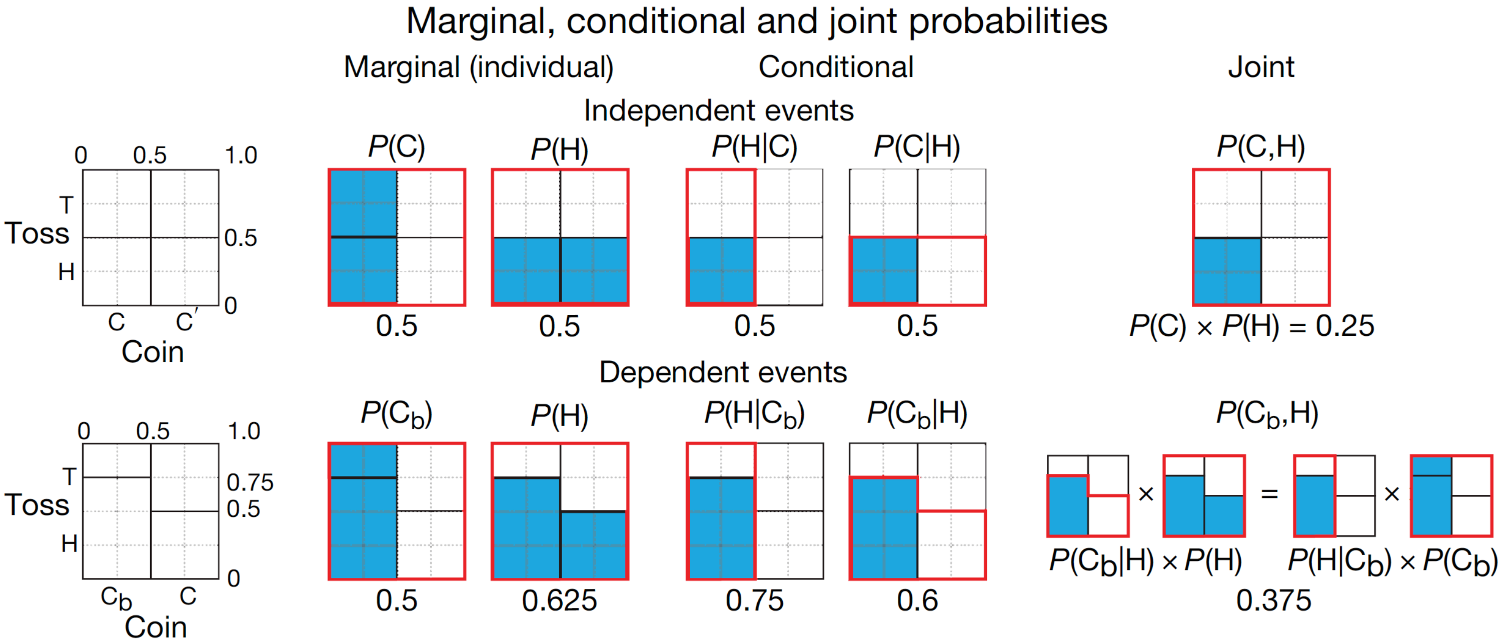


Figure 1 | Marginal, joint and conditional probabilities for independent and dependent events. Probabilities are shown by plots[3], where columns correspond to coins and stacked bars within a column to coin toss outcomes, and are given by the ratio of the blue area to the area of the red outline. The choice of one of two fair coins (C, C′) and outcome of a toss are independent events. For independent events, marginal and conditional probabilities are the same and joint probabilities are calculated using the product of probabilities. If one of the coins, C_b, is biased (yields heads (H) 75% of the time), the events are dependent, and joint probability is calculated using conditional probabilities.

From: N. Altman's Bayes' Theorem

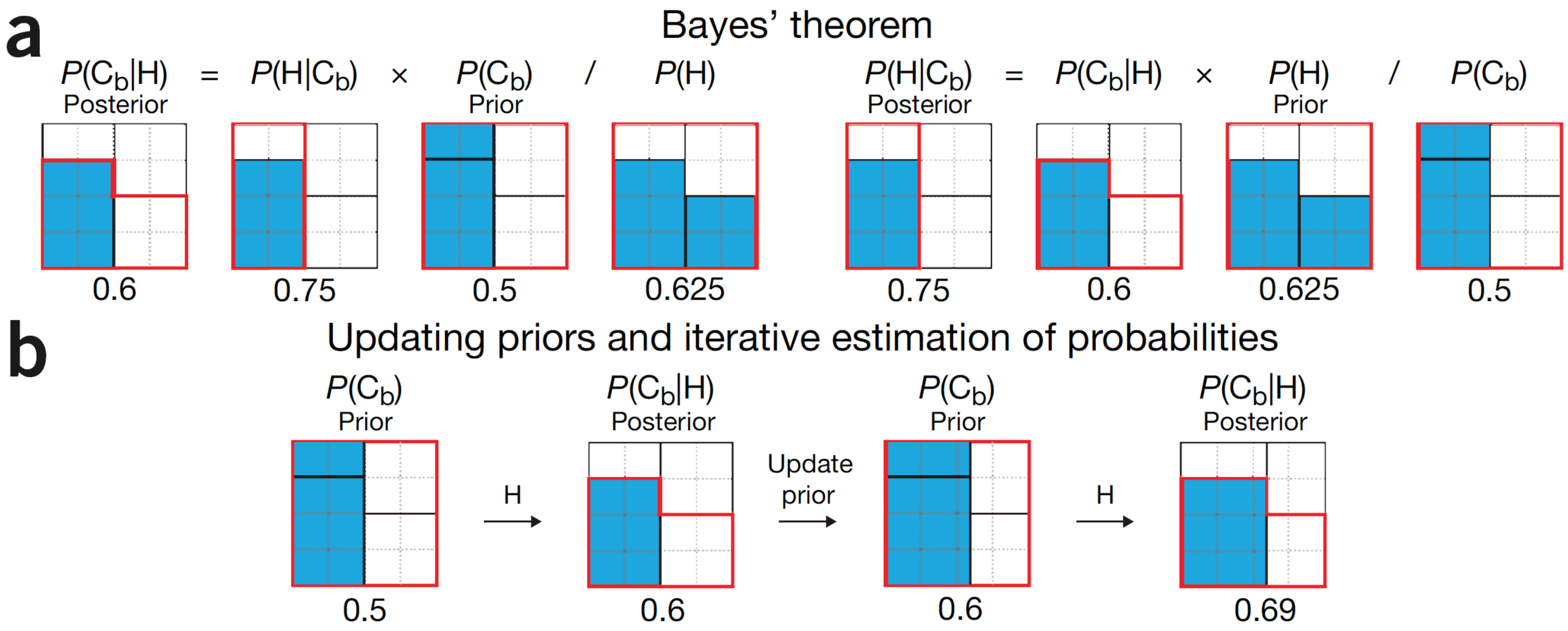


Figure 2 | Graphical interpretation of Bayes' theorem and its application to iterative estimation of probabilities. (**a**) Relationship between conditional probabilities given by Bayes' theorem relating the probability of a hypothesis that the coin is biased, $P(C_b)$, to its probability once the data have been observed, $P(C_b|H)$. (**b**) The probability of the identity of the chosen coin can be inferred from the toss outcome. Observing a head increases the chances that the coin is biased from $P(C_b) = 0.5$ to 0.6, and further to 0.69 if a second head is observed.

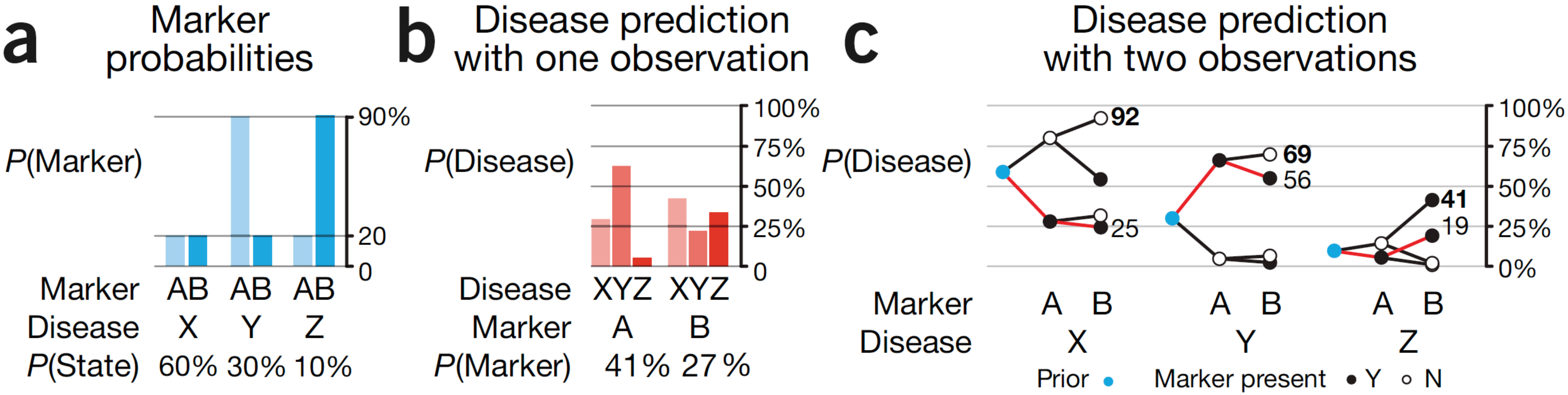
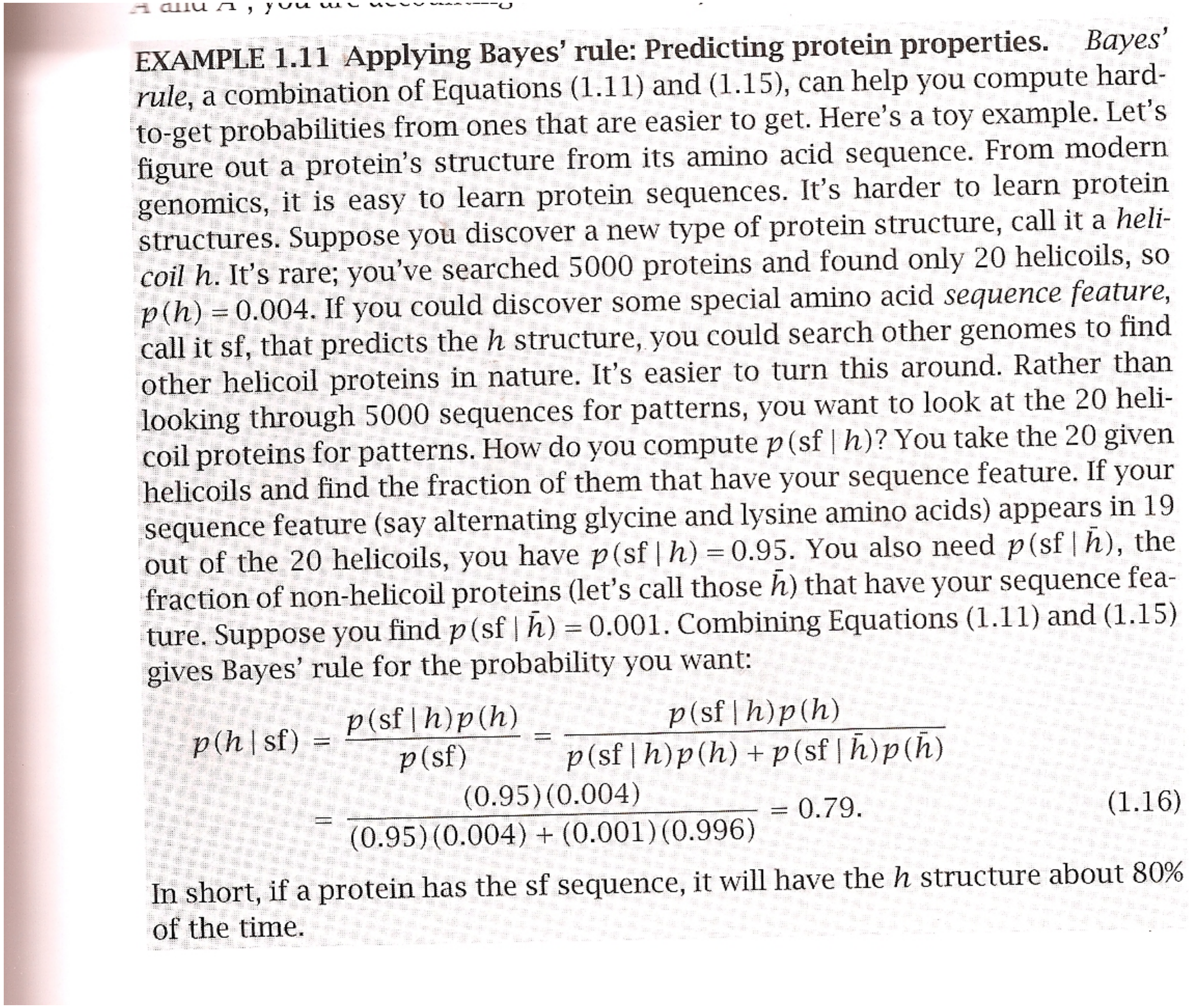


Figure 3 | Disease predictions based on presence of markers. (**a**) Independent conditional probabilities of observing each marker (A, B) given a disease (X, Y, Z) (e.g., $P(\mathrm{A}|\mathrm{Y}) = 0.9$). (**b**) Posterior probability of each disease given a single observation that confirms the presence of one of the markers (e.g., $P(\mathrm{Y}|\mathrm{A}) = 0.66$). (**c**) Evolution of disease probability predictions with multiple assays. For a given disease, each path traces (left to right) the value of the posterior that incorporates all the assay results up to that point, beginning at the prior probability for the disease (blue dot). The assay result is encoded by an empty (marker absent) or a solid (marker present) dot. The red path corresponds to presence of A and B. The highest possible posterior is shown in bold.

The relevance of Bayes’ theorem: see DILL & BROMBERG: EXAMPLE1.11 …BIOINFORMATIC CONTEXT



A and A', you are accounting

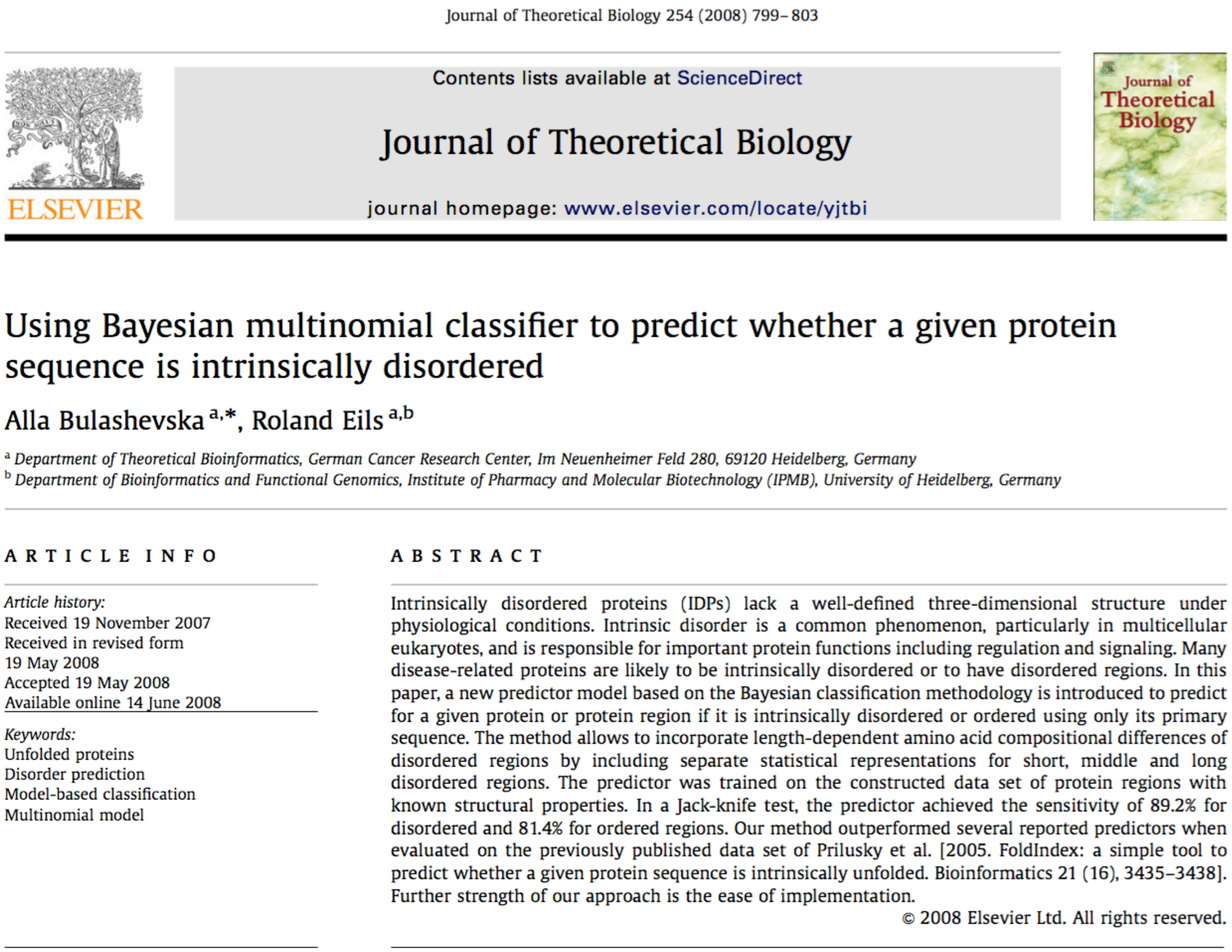
EXAMPLE 1.11 Applying Bayes’ rule: Predicting protein properties. *Bayes’ rule*, a combination of Equations (1.11) and (1.15), can help you compute hard-to-get probabilities from ones that are easier to get. Here’s a toy example. Let’s figure out a protein’s structure from its amino acid sequence. From modern genomics, it is easy to learn protein sequences. It’s harder to learn protein structures. Suppose you discover a new type of protein structure, call it a *helicoil h*. It’s rare; you’ve searched 5000 proteins and found only 20 helicoils, so $p(h) = 0.004$. If you could discover some special amino acid *sequence feature*, call it sf, that predicts the h structure, you could search other genomes to find other helicoil proteins in nature. It’s easier to turn this around. Rather than looking through 5000 sequences for patterns, you want to look at the 20 helicoil proteins for patterns. How do you compute $p(\mathrm{sf} \mid h)$? You take the 20 given helicoils and find the fraction of them that have your sequence feature. If your sequence feature (say alternating glycine and lysine amino acids) appears in 19 out of the 20 helicoils, you have $p(\mathrm{sf} \mid h) = 0.95$. You also need $p(\mathrm{sf} \mid \bar{h})$, the fraction of non-helicoil proteins (let’s call those $\bar{h}$) that have your sequence feature. Suppose you find $p(\mathrm{sf} \mid \bar{h}) = 0.001$. Combining Equations (1.11) and (1.15) gives Bayes’ rule for the probability you want:

$$p(h \mid \mathrm{sf}) = \frac{p(\mathrm{sf} \mid h)p(h)}{p(\mathrm{sf})} = \frac{p(\mathrm{sf} \mid h)p(h)}{p(\mathrm{sf} \mid h)p(h) + p(\mathrm{sf} \mid \bar{h})p(\bar{h})}$$

$$= \frac{(0.95)(0.004)}{(0.95)(0.004) + (0.001)(0.996)} = 0.79. \qquad (1.16)$$

In short, if a protein has the sf sequence, it will have the h structure about 80% of the time.

Realistic example of bayesian methodology



Journal of Theoretical Biology 254 (2008) 799–803

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Using Bayesian multinomial classifier to predict whether a given protein sequence is intrinsically disordered

Alla Bulashevska [a,*], Roland Eils [a,b]

[a] Department of Theoretical Bioinformatics, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
[b] Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology (IPMB), University of Heidelberg, Germany

ARTICLE INFO

Article history:
Received 19 November 2007
Received in revised form
19 May 2008
Accepted 19 May 2008
Available online 14 June 2008

Keywords:
Unfolded proteins
Disorder prediction
Model-based classification
Multinomial model

ABSTRACT

Intrinsically disordered proteins (IDPs) lack a well-defined three-dimensional structure under physiological conditions. Intrinsic disorder is a common phenomenon, particularly in multicellular eukaryotes, and is responsible for important protein functions including regulation and signaling. Many disease-related proteins are likely to be intrinsically disordered or to have disordered regions. In this paper, a new predictor model based on the Bayesian classification methodology is introduced to predict for a given protein or protein region if it is intrinsically disordered or ordered using only its primary sequence. The method allows to incorporate length-dependent amino acid compositional differences of disordered regions by including separate statistical representations for short, middle and long disordered regions. The predictor was trained on the constructed data set of protein regions with known structural properties. In a Jack-knife test, the predictor achieved the sensitivity of 89.2% for disordered and 81.4% for ordered regions. Our method outperformed several reported predictors when evaluated on the previously published data set of Prilusky et al. [2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21 (16), 3435–3438]. Further strength of our approach is the ease of implementation.

Indicators to evaluate methods

$$\text{Sensitivity (or recall)} : S_n = \frac{TP}{TP+FN} = \frac{TP}{N_d} \tag{1}$$

is the number of correctly identified disordered proteins normalized to the total number of disordered proteins in the sample

$$\text{Specificity} : S_p = \frac{TN}{TN+FP} = \frac{TN}{N_o} \tag{2}$$

is the ratio between the number correctly identified ordered proteins and the total number of ordered proteins in the sample;

$$\text{Rate of false positives} : f_p = \frac{FP}{TN+FP} = 1 - S_p \tag{3}$$

is the ratio between the number of ordered proteins predicted as disordered and the total number of ordered proteins in the sample;

$$\text{Accuracy} : ACC = \frac{S_n + S_p}{2} \tag{4}$$

that is the average between sensitivity and specificity. It measures the overall performance of the predictor. Then,

$$\text{Precision (or selectivity)} : \mathrm{Pr} = \frac{TP}{TP+FP} = \frac{TP}{n_d} \tag{5}$$

Study Materials

- Slides
- Puga2015a
- D’Agostini9512295 3.1-3.5
- Eikosograms (RW Oldford)

https://cran.r-project.org/web/packages/eikosograms/vignettes/Introduction.html