Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ETH RISK CENTER

LECTURE NOTES

Fundamentals of Probability and Statistics
Chapter 1: Lecture 1

Spring 2018

Written by marco **b**roccardo
IBK, ETHZ

1 Element of set theory

Set theory is a branch of mathematic and probability theory which enables to tackle class of problems dealing with random phenomena.

1.1 Definitions

Random phenomenon a phenomenon that has more than one possible outcome. The true outcome is unknown until it is observed.

Sample space collection of all possible outcomes of a random phenomenon. Notation: S.

Sample point each element of the sample space. Notation: x.

Event a collection of sample points which represents a subset of the sample space. Notation: E, $E \subseteq S$.

These definitions are typically visualized via Venn's diagram. Figure 1 shows a typical Venn diagram. Observe that the rectangular shape is used for S, the oval shape for E, and the dot for x.

1.2 Operation on Events

Union: given the events E_1 and and E_2, the union event, denoted with $E_1 \cup E_2$, is the event that contains all sample points in either E_1 **or** E_2. Figure 2(i).

Intersection: given the events E_1 and and E_2, the intersection event, denoted with $E_1 \cap E_2$ or simply E_1E_2, is the event that contains the sample points both in E_1 **and** E_2 Figure 2(ii).

These operations obey certain properties:

Commutative property of union: $E_1 \cup E_2 = E_2 \cup E_1$.

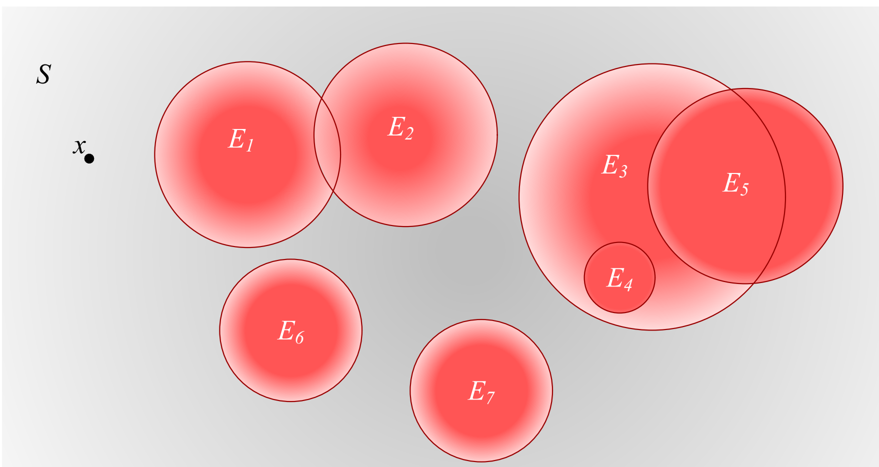

Figure 1: Venn diagram

Commutative property of intersection: $E_1E_2 = E_2E_1$.

Associative property of union: $E_1 \cup (E_2 \cup E_3) = (E_1 \cup E_2) \cup E_3 = E_1 \cup E_2 \cup E_3$.

Associative property of intersection: $E_1(E_2E_3) = (E_1E_2)E_3 = E_1E_2E_3$.

Distributive property: $E_1(E_2 \cup E_3) = (E_1E_2) \cup (E_1E_3) = E_1E_2 \cup E_1E_3$.

Observe that intersection takes precedence over union. I follows, that in the application of the distributive property the intersection operations must be performed prior to the union operations.

1.3 Special Events

Certain event: the event that contains all possible sample points of the sample space. Then, the sample space S is the certain event.

Null Event: the event that contains no sample points. Notation $\varnothing$.

Mutually exclusive events: the events E_1 and E_2 are mutually exclusive when they have no common sample points. Then, $E_1E_2 = \varnothing$.

Collectively exhaustive events: the events $E_1, E_2, ..., E_N$ are collectively exhaustive when their union spans the entire sample space. Then, $E_1 \cup E_2 \cup ... \cup E_N = S$.

Complementary events: the complement of the event E is the event $\bar{E}$ contains all the sample space that are not in in the event E. Then, $E\bar{E} = \varnothing$ and $E \cup \bar{E} = S$, i.e. E and $\bar{E}$ are mutually exclusive and collectively exhaustive.

1.4 De Morgan's rules

Given a series of events $E_1, E_2, ..., E_N$, it easy to prove the following rules:

$$\overline{E_1 \cup E_2 \cup ... \cup E_N} = \bar{E}_1\bar{E}_2...\bar{E}_N, \text{ i.e. } \overline{\bigcup_{n=1}^{N} E_n} = \bigcap_{n=1}^{N} \bar{E}_n, \tag{1}$$

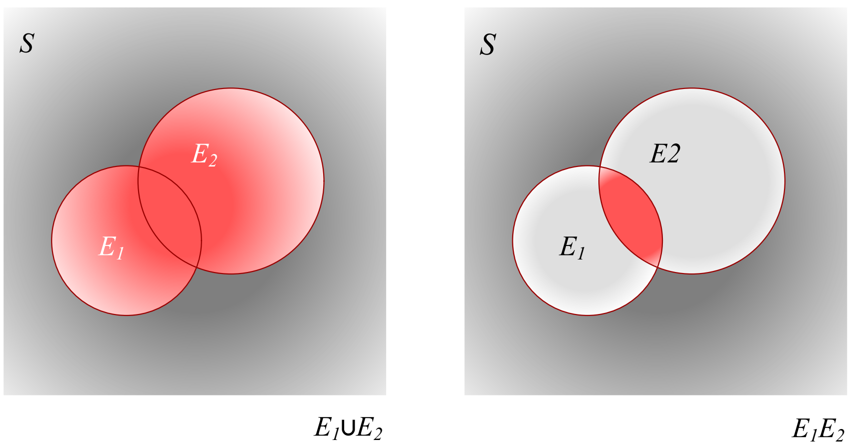

Figure 2: (i) Union and (ii) Intersection Events

and

$$\overline{E_1 E_2 ... E_N} = \bar{E}_1 \cup \bar{E}_2 \cup ... \cup \bar{E}_N, \text{ i.e. } \overline{\bigcap_{n=1}^{N} E_n} = \bigcup_{n=1}^{N} \bar{E}_n. \tag{2}$$

2 Element of Probability theory

The probability of an event E in a sample space S, is a measure, a weight, of the likelihood of occurrence of the event relative to other events in S. Notation: $P(E|S)$ or $P(E)$. Probability theory is based on the following Kolmogorov [1] three axioms:

I. $0 \leq P(E) \leq 1$.

II. $P(S) = 1$.

III. $P(\bigcup_{n=1}^{N} E_n) = \sum_{n=1}^{N} P(E_n)$, for mutually exclusive $E_1, E_2, ..., E_N$.

The following results are derived based on the above axioms:

i. $P(\bar{E}) = 1 - P(E)$.

ii. $P(\varnothing) = 0$.

iii. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 E_2)$.

Common pitfalls: when approaching for the first time probability theory, a common mistake among students is to confuse the null event, $\varnothing$, with the real number 0. Observe that 0 can be a sample point, and an event, it follow that $0 \neq \varnothing$, and $P(\varnothing) \neq P(0)$! Moreover, observe that $P(E) = 0 \nRightarrow E = \varnothing$.

The last rule can be easily generalized for N events, i.e.

$$P(E_1 \cup E_2 \cup ... \cup E_N) = \sum_{n=1}^{N} P(E_n) - \sum_{m=1}^{N} \sum_{n>m}^{N} P(E_n E_m) + \sum_{l=1}^{N} \sum_{m>l}^{N} \sum_{n>m}^{N} P(E_n E_m E_l) - \\ ... + (-1)^{N-1} P(E_1 E_2 ... E_N). \tag{3}$$

Moreover based on (1) we can write

$$P(E_1 \cup E_2 \cup ... \cup E_N) = 1 - P(\overline{E_1 \cup E_2 \cup ... \cup E_N}) = 1 - P(\bar{E}_1 \bar{E}_2 ... \bar{E}_N). \tag{4}$$

2.1 Conditional Probability

Conditional Probability is a simple, yet very profound, concept in Probability theory. We actually argue that this is the most important concept for this class. In fact, it sets the foundation for a good understanding of all the lecture notes, and the basis for a gently introduction to Bayesian statistics. Given two events E_1 and E_2, the conditional probability of E_1 given E_2, denoted with $P(E_1|E_2)$, defines the the probability of observing sample points of E_1, given that

[1] Andrey Nikolaevich Kolmogorov (*1903 †1987) was a Russian and Soviet mathematician who is the father of modern Probability Theory. His contributions spans from abstract mathematics to physics. It is considered one of the greatest minds of the 20th century

the event E_2 has occurred. This is a way of redefining the sample space, since by assuming E_2 has occurred the possible sample points are now confined within E_2. By definition, the conditional probability is defined as

$$\begin{aligned} P(E_2|E_1) &= \frac{P(E_1E_2)}{P(E_1)}, \text{ if } P(E_1 > 0) \\ &= 0, \text{ if } P(E_1 = 0), \end{aligned} \tag{5}$$

or rearranging the last equation as:

$$P(E_1E_2) = P(E_2|E_1)P(E_1), \tag{6}$$

or by using the commutative property as

$$P(E_1E_2) = P(E_1|E_2)P(E_2), \tag{7}$$

For a set of events $E_1, E_2, ..., E_N$ the probability of intersection can be written in different way, depending on the order of the conditioning, e.g.

$$P(E_1E_2...E_N) = P(E_N|E_1...E_{N-1})P(E_{N-1}|E_1...E_{N-2})...P(E_2|E_1)P(E_1), \tag{8}$$

$$= P(E_1|E_2...E_N)P(E_2|E_3...E_N)...P(E_{N-1}|E_N)P(E_N), \tag{9}$$

In contrast to the conditional probability $P(E_2|E_1)$, the unconditional probability $P(E_1)$ is named marginal probability.

2.2 Statistical Independence

Two events E_1 and E_2 are statistically independent, $E_1 \perp\!\!\!\perp E_2$, if

$$P(E_2|E_1) = P(E_2), \tag{10}$$

i.e., the occurrence or the knowledge of E_2 does not affect the probability of occurrence of E_1. It follows that for two statistically independent events

$$P(E_1E_2) = P(E_1)P(E_2). \tag{11}$$

Common pitfalls: often students confuse independence with mutually exclusive. The two notions are mathematically and conceptually very different. Mutually exclusive relates to share sample points between two events, while statically independence relates to the probability properties expressed by (10). It is easy to show that $P(E_1E_2) = P(E_1)P(E_2) \iff E_1 \perp\!\!\!\perp E_2$. However, for a set of events $E_1, E_2, ..., E_N$, statistically independence requires that

$$P\Big(E_n| \bigcap_{n \neq m} E_m\Big) = P(E_n), \tag{12}$$

i.e., the conditional probability of E_n given any set of remaining events must be equal to the unconditional probability of E_n. Equivalently, the events are statistically independent if for any selections of indices the joint probability of the events is equal to the product of their marginal probabilities. For example for three events E_1, E_2, E_3, statistically independence, $E_1 \perp\!\!\!\perp E_2 \perp\!\!\!\perp E_3 \Rightarrow$

$$\begin{aligned} &P(E_1E_2E_3) = P(E_1)P(E_2)P(E_3), \\ &P(E_1E_2) = P(E_1)P(E_2), \\ &P(E_2E_3) = P(E_2)P(E_3), \\ &P(E_3E_1) = P(E_3)P(E_1), \end{aligned} \tag{13}$$

however, $P(E_1E_2E_3) = P(E_1)P(E_2)P(E_3)$ alone $\not\Rightarrow E_1 \perp\!\!\!\perp E_2 \perp\!\!\!\perp E_3$.

2.3 Theorem of Total Probability

Consider a set of mutually exclusive and collectively exhaustive events $E_1, E_2, ...E_N$, i.e. $E_n E_m = \varnothing$ and $\bigcup_{n=1}^{N} E_n = S$. Then, consider an event A, according to the theorem of total probability,

$$P(A) = \sum_{n=1}^{N} P(A|E_n)P(E_n). \tag{14}$$

The proof is given by taking in follow

$$\begin{aligned} P(A) &= P(AE_1 \cup AE_2 \cup ... \cup AE_N) \\ &= \sum_{n=1}^{N} P(AE_n) = \sum_{n=1}^{N} P(A|E_n)P(E_n) \end{aligned} \tag{15}$$

Figure 4 shows a Venn diagram representation of the events. The theorem of total probability is pivotal in risk assessment, since often it is easier to compute conditional probabilities than marginal probabilities.

2.4 Bayes' Theorem

Consider two events A and B. Then, considering the commutative property $P(AB) = P(A|B)P(B) = P(B|A)P(A)$ and re-arranging terms we can write

$$P(B|A) = \frac{P(A|B)}{P(A)} P(B). \tag{16}$$

The last equation is known as Bayes's[2] theorem. This theorem is at the base of Bayesian statistics. It's significance lies in the fact that the probability of event A appears in the unconditional form $P(A)$ on the right, and in its conditional form $P(B|A)$ on the left side. It follows,

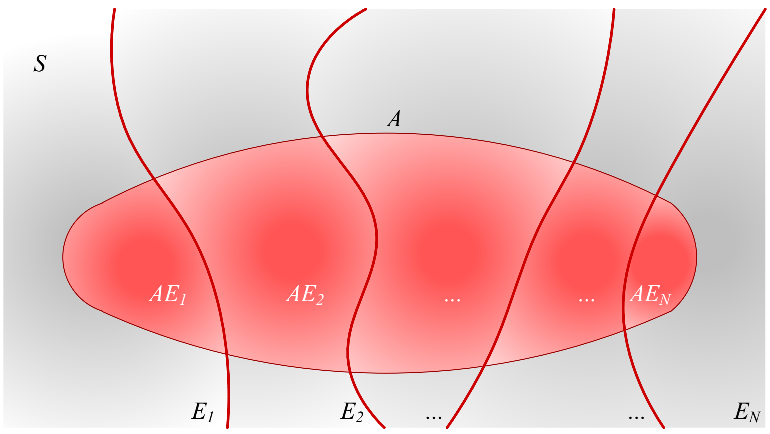

Figure 3: Representation of total probability

[2]Thomas Bayes (*1701 †1761) was a British Presbyterian minister, statistician, and philosopher. It is curious to know that Mr Bayes never published his work on Bayes' theorem. His notes were revised and published by the Welsh philosopher and mathematician Richard Price after Bayes death. However, it must be said that the great French mathematician Pierre-Simon Laplace (*1749 †1827) pioneered and popularised the modern Bayesian probability.

that Bayes' Theorem sets the foundation for updating the probability of an event A when the observation of event B is made.

Given a set of mutually exclusive and collectively exhaustive events $E_1, E_2, ...E_N$, we can write

$$P(E_n|A) = \frac{P(A|E_n)}{P(A)} P(E_n), \tag{17}$$

and by using the total probability theorem (23)

$$P(E_n|A) = \frac{P(A|E_n)}{\sum_{n=1}^{N} P(A|E_n)P(E_n)} P(E_n). \tag{18}$$

All rules of probability apply to conditional probabilities, provided a proper resize of the sample space is made. For example:

$$P(E_1 \cup E_2|A) = P(E_1|A) + P(E_2|A) - P(E_1E_2|A) \tag{19}$$

$$P(E_1E_2|A) = P(E_1|E_2A)P(E_2|A) = P(E_2|E_1A)P(E_1|A), \tag{20}$$

and given a set of mutually exclusive and collectively exhaustive events $E_1, E_2, ...E_N$, we can rewrite (23)

$$P(B|A) = \sum_{n=1}^{N} P(B|E_nA)P(E_n|A). \tag{21}$$

Example I

Suppose that we would like to know the probability of failing of the Bay Bridge, Figure 4, which connects the city of Oakland with the city of San Francisco. The Bay bridge is a series system composed of two spans. The east span connects the city of Oakland with Yerba Buena island, and the west span connects Yerba Buena island with the city of San Francisco. The sample space of the intensity measures of the earthquake are defined as $S_{im} = \{IV, VI, VIII, X\}$, where IV is a low intensity earthquake, VI a medium intensity earthquake, $VIII$ is a strong earthquake, and X is a devastating earthquake. Seismologists and earthquake engineers compute the probability of occurrence of such events. For this case, the values[3]are reported in Table 1.

Let's denote with F_S the system failure event, F_W is the west span failure event, and F_E is the east span failure event. Structural engineers compute the probability of failure of structures conditional to a level of hazard, i.e. $F_W|IM$, and $F_E|IM$. These values[3] are reported in Table 2.

The Bay Bridge system fails (i.e. F_S) if the east span fails (i.e. F_E) OR the west span fails (i.e. F_W). Then, we can write this event as

$$P(F_S) = P(F_W \cup F_E) = P(F_W) + P(F_E) - P(F_WF_E). \tag{22}$$

We can compute $P(F_W)$, $P(F_E)$, and $P(F_EF_W)$, by the total probability theorem. In fact, the earthquake intensities represent a collection of mutually exclusive and collectively exhaustive events. Then, for example, we can write $P(F_W)$ as

$$P(F_W) = \sum_{n=1}^{N} P(F_W|IM_n)P(IM_n). \tag{23}$$

In the same way, we can use the total probability theory to compute F_E and F_EF_W.

[3]Numbers reported here are purely academic.

Problems

i Compute $P(F_W)$, $P(F_E)$, and $P(F_E F_W)$, and $P(F_S)$.

ii Are the events $F_W|IM$ and $F_E|IM$ statistically independent?

iii Are the events F_W and F_E statistically independent? Which conclusions can you draw?

Suppose that we know that the Bay Bridge system has failed; however, we do not know which of the two spans has failed. Given this information we would like to update the probability that the west span has failed. Then, we can use the Bayes' theorem as

$$P(F_W|F_W \cup F_E) = \frac{P(F_W \cup F_E|F_W)}{P(F_W \cup F_E)} P(F_W), \tag{24}$$

it should be clear that $P(F_W \cup F_E|F_W) = 1$ since if we know that the west span has failed the system is surely failed. Then we can write

$$P(F_W|F_W \cup F_E) = \frac{1}{P(F_W \cup F_E)} P(F_W), \tag{25}$$

and finally solve the problem with the values from the previous problems.

Problems

i Compute $P(F_W|F_W \cup F_E)$, $P(F_E|F_W \cup F_E)$. Which conclusions can you draw?

ii What is the probability that the west component ALONE was the cause of failure?

iii Given that the Bay Bridge system failed, what is the probability that $IM = VIII$ (i.e. $P(IM = VIII|F_W \cup F_E)$)?

iv Given that the Bay Bridge system failed, what is the most likely earthquake intensity? Is it an "intuitive" result? Which conclusions can you draw?

Figure 4: Bay Bridge. Source: Wikipedia.

$P(IM = IV)$	$P(IM = VI)$	$P(IM = VIII)$	$P(IM = X)$
0.8	0.15	0.045	0.005

Table 1: IM probabilities

IM	$P(F_W \mid IM = im)$	$P(F_E \mid IM = im)$	$P(F_W F_E \mid IM = im)$
IV	$1.00E{-4}$	$1.00E{-4}$	$1.00E{-8}$
VI	$5.00E{-3}$	$1.00E{-3}$	$5.00E{-6}$
$VIII$	$8.00E{-2}$	$5.00E{-3}$	$4.00E{-4}$
X	$2.00E{-1}$	$5.00E{-2}$	$1.00E{-2}$

Table 2: Conditional probabilities of failure