Applied Statistics

for Computer Science BSc, Exam

Probability Theory and Mathematical Statistics

for Computer Science Engineering BSc, Term grade

István Fazekas
University of Debrecen

2020/21 fall

This work was supported by the construction EFOP-3.4.3-16-2016-00021. The project was supported by the European Union, co-financed by the European Social Fund.

Main topics

1. Probability theory
2. Statistics

Mathematical tools: combinatorics, calculus

Computer tool: Matlab

Book:

Yates, Goodman:

Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers

Lecture 3

Kolmogorov's axioms

Conditional probability

Independence of events

Kolmogorov's axioms of probability

For infinite probability spaces we need additivity for countably infinite events.

Non-negative

$$P(A) \geq 0 \quad \text{for any event } A. \tag{1}$$

Normed

$$P(\Omega) = 1. \tag{2}$$

Countably additive (σ-additive)

$$\boxed{P(A_1 + A_2 + \cdots) = P(A_1) + P(A_2) + \cdots}, \tag{3}$$

if $A_1, A_2, \ldots$ are pairwise exclusive events.

Properties of the probability

Countably additivity implies finitely additivity: if $A_1, A_2, \dots, A_n$ are pairwise exclusive, then

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n). \tag{4}$$

Hint: first show that $P(\emptyset) = 0$.

Therefore the previously given properties of the probability remain true.

Notation.

$(\Omega, \mathcal{A}, P)$ is the Kolmogorov's probability space, where
$\Omega \neq \emptyset$ is the sample space,
$\mathcal{A}$ is the family of events,
P is the probability.

Continuity of the probability

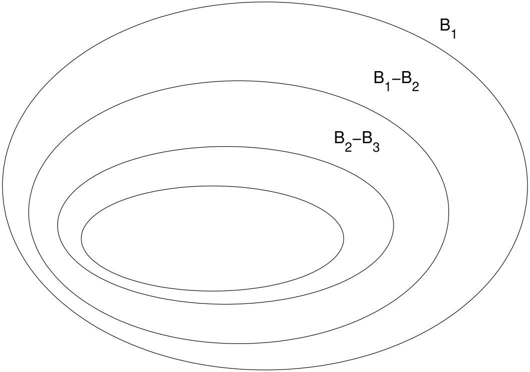

Figure: Decreasing sequence of events

Continuity of the probability

Theorem. Assume that P is non-negative and normed. Then P is countably additive if and only if it is finitely additive and continuous, that is

$$\lim_{n\to\infty} P(B_n) = 0\,,$$

if $B_i,\ i = 1, 2, \ldots$ are events so that $B_1 \supseteq B_2 \supseteq \ldots$ and $\bigcap_{i=1}^{\infty} B_i = \emptyset\,.$

Countable probability spaces

Any countably infinite probability space can be described as follows.

$$\Omega = \{\omega_1, \omega_2, \dots\},$$

$$P(A) = \sum_{\omega_i \in A} p_i \tag{5}$$

for any event A, where $p_1, p_2, \dots$ is a given discrete distribution, that is they are non-negative numbers with

$$\sum_{i=1}^{\infty} p_i = 1.$$

Discrete distribution

$p_1, p_2, \ldots$ is called a discrete (probability) distribution (or mass function), if the numbers p_i are non-negative and

$$\sum_{i=1}^{\infty} p_i = 1.$$

Example.
Let $p_k = e^{-\lambda}\lambda^k/k!\,,\ k = 0, 1, 2, \ldots,$ where $\lambda > 0$ is a constant. Using $\sum_{k=0}^{\infty} \lambda^k/k! = e^{\lambda}$ one can prove that $p_1, p_2, \ldots$ is a distribution. It is called Poisson distribution with parameter λ.

Geometric way to calculate probability

Let G be a subset of $\mathbb{R}^n$ $(n = 1, 2, 3)$.
Choose a point uniformly at random from G-re.
Let $A \subseteq G$.
Then the probability that the point is inside A is

$$\boxed{P(A) = \lambda(A)/\lambda(G)},$$

where λ is the length, area or volume if we are on the line, plane or space, respectively (we assume $0 < \lambda(G) < \infty$).

Geometric way to calculate probability

Example. Choose a point from the square 2×2. Denote by X its distance from the nearest side. Calculate $P(X < a)$.

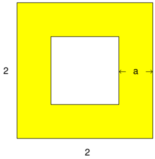

Figure: The track with width a.

Obviously $P(X < a) = 0$, if $a < 0$, and $P(X < a) = 1$, if $a > 1$.
If $0 < a \leq 1$, then the favourable part is a track with width a.
Its area is $4 - (2 - 2a)^2$.
Therefore
$P(X < a) = (4 - (2 - 2a)^2)/4 = 1 - (1 - a)^2 = 2a - a^2$, $0 < a \leq 1$.

Conditional probability

Let A and B be events.
Repeat the experiment n-times.
We are interested in A restricted to those cases when B occurs.
Then the restricted relative frequency of A given B is

$$\frac{k_{AB}}{k_B} = \left(\frac{k_{AB}}{n}\right) \bigg/ \left(\frac{k_B}{n}\right) ,$$

where k_{AB} and k_B denote the frequencies of AB and B, respectively.
But $\left(\frac{k_{AB}}{n}\right) \bigg/ \left(\frac{k_B}{n}\right)$ is around $P(AB)/P(B)$. So we have
Definition. Let A and B events, assume $P(B) > 0$.
Then the conditional probability of A given B is

$$\boxed{P(A|B) = P(AB)/P(B)}$$

Hungarian lottery

On the lottery ticket we should mark 5 numbers out of 90.
At a lottery draw officially is drawn 5 numbers.
If our 5 numbers are the same as the 5 numbers drawn, the we get a huge amount of money.
What is the probability that we hit the jackpot?

$$P(A) = \frac{1}{\binom{90}{5}}$$

We are watching on TV the lottery draw.
The first 4 numbers drawn are marked on our lottery ticket.
Now the fifth number will be drawn.
What is the probability that we will hit the jackpot?

$$P(A|B) = \frac{P(AB)}{P(B)} = \left(1 \Big/ \binom{90}{5}\right) : \left(\binom{5}{4} \Big/ \binom{90}{4}\right) = \frac{1}{86}.$$

Partition of the sample space

The sequence of events $A_1, A_2, \ldots$ is called a partition if they are pairwise exclusive and their union is the whole sample space. Obviously

$$P(A_1) + P(A_2) + \cdots = 1.$$

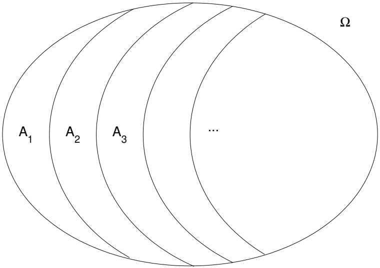

Figure: A partition

Total probability theorem

Let $B_1, B_2, \ldots$ be a partition of the sample space. Assume that $P(B_i) > 0$ for all i. Then for any event A we have

$$\boxed{P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots}\ .$$

Proof. Use the additivity of the probability and the definition of the conditional probability.

Bayes theorem

Bayes formula. Let A and B have positive probabilities. Then

$$\boxed{P(B|A) = P(A|B)\,P(B)/P(A)}\;.$$

Theorem. Let A be an event,
$B_1, B_2, \ldots$ a partition of the sample space,
$P(A) > 0$, $P(B_i) > 0$, $i = 1, 2 \ldots$.
Then for any i we have

$$\boxed{P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}}$$

Proof. Use Bayes formula and total probability theorem.

An application of total probability and Bayes theorems

In a store there are 100 TV sets. 50 of them are of type A, 30 of them are of type B, and 20 are of type C. It is known that 1% of type A TV sets are defective, 2% of type B TV sets are defective, and 3% of type C TV sets are defective.

(a) A TV set is chosen randomly. What is the probability that it is defective?

(b) A TV set is chosen randomly. It turns out that it is defective. What is the probability that it is of type A?

Solution.

(a) Use total probability theorem.

$$P\,(\text{defective}) = 0.01 \cdot 0.5 + 0.02 \cdot 0.3 + 0.03 \cdot 0.2 = 1.7\%$$

(b) Use Bayes theorem

$$P(\text{type} \quad \text{A}|\text{defective}) = \frac{0.01 \cdot 0.5}{0.01 \cdot 0.5 + 0.02 \cdot 0.3 + 0.03 \cdot 0.2} \approx 30\%.$$

Independence of two events

Definition. We say that A is independent of B if the occurrence of B does not affect the probability of A. That is

$$P(A|B) = P(A). \tag{6}$$

However, this definition is not symmetric and we should assume that $P(B) > 0$.
In the following definition the roles of A and B are symmetric.
Definition. We say that A and B are independent if

$$P(AB) = P(A)P(B). \tag{7}$$

Proposition. If $P(B) > 0$, then equations (6) and (7) are equivalent. If $P(B) = 0$, then equation (7) is true for any A.
Proof. Apply the definition of the conditional probability.

Independence of two events

Example. Toss two coins. Let A be the event that the first one shows H, and B be the event that the second one shows T. Then

$$P(AB) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B)$$

so A and B are independent.
Exercise. Let $P(A) = 0$. Show that A and any B are independent.
Hint. $0 = 0$
Exercise. Let A and B be independent. Show that A and $\overline{B}$ are independent.
Exercise. Let A and any B be independent. Show that $P(A)$ is either 1 or 0.
Hint. $P(AA) = P(A)P(A)$

Pairwise independence

Definition. We say that the events $A_1, A_2, \ldots$ are pairwise independent if any two of them are independent, that is

$$P(A_iA_j) = P(A_i)P(A_j), \quad i \neq j.$$

Remark. The pairwise independence of A, B, C means that

$$P(AB) = P(A)P(B),$$
$$P(AC) = P(A)P(C), \; P(BC) = P(B)P(C) \; .$$

Independence

Definition. We say that the events $A_1, A_2, \ldots, A_n$ are independent if for any $k = 1, 2, \ldots, n$ and for any combinations $i_1, \ldots, i_k$ of the numbers $1, 2, \ldots, n$ we have

$$P(A_{i_1} A_{i_2} \ldots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

An infinite family of events is called independent, if any of its finite subfamilies is independent.

Remark. The independence of A, B, C means that

$$P(AB) = P(A)P(B),$$
$$P(AC) = P(A)P(C),\ P(BC) = P(B)P(C)$$

and

$$P(ABC) = P(A)P(B)P(C).$$

Pairwise independence and independence

Proposition. Independence implies pairwise independence but not vice versa.

Example. Choose a point randomly from the unit square. Let A, B and C be the events that the point is chosen from the subsets below.

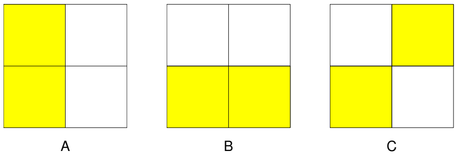

Figure: The events A, B and C.

Then $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(AB) = P(AC) = P(BC) = \frac{1}{4}$, $P(ABC) = \frac{1}{4}$. So they are pairwise independent but not independent.