STATISTICS

Module 2 *of*
MATHEMATICS AND STATISTICS

Marina Santacroce

marina.santacroce@unibocconi.it
Department of Decision Sciences, Bocconi University

L5. Conditional probability and Bayes formula.

Outline

Introduction

Conditional probabilities are useful

- when we update the probability of events according to the information;
- easy way for computing probabilities of complex events.

Example: Two (green and red) dice are rolled. The sample space is $S = \{(i,j); i,j : 1,\dots,6\}$. Consider the event E={*sum is* 8}.

$\Rightarrow \mathbb{P}(E) = \frac{\#(E)}{\#(S)} = \frac{5}{36}$.

If the event red die shows 6 has occurred, how will the probability of E be modified?

Let $F = \{\textit{red die is 6}\}$ and denote by $\mathbb{P}(E|F)$ the "*conditional probability of E given F*"

Intuition: If F has occurred, the event E can occur if and only if $E \cap F$ occurs. Equivalently, $\mathbb{P}(E|F)$ will be the the probability that (6,2) occurs out of the 6 outcomes of $F = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$.

$\Rightarrow \mathbb{P}(E|F) = \frac{\#(E\cap F)}{\#(F)} = \frac{1}{6}$.

Conditional Probability

Def. Consider two events E and F with $\mathbb{P}(F) > 0$. The conditional probability of E given F is

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

Example: A fair coin is flipped twice. Compute the probability of E={*2 heads*} conditionally to

a) F={ *head at first flip*}

b) G={*at least one head*}.

▷ $S = \{(H, H), (T, H), (H, T), (T, T)\}$ are the equally likely 4 possible outcomes.

$E = \{(H, H)\}$, $F = \{(H, T), (H, H)\}$ e $G = \{(H, H), (T, H), (H, T)\}$.

$\Rightarrow$ a) $\mathbb{P}(E|F) = \frac{\mathbb{P}(E\cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}((H,H))}{\mathbb{P}(\{(H,H),(H,T)\})} = \frac{\#((H,H))}{\#(\{(H,H),(H,T)\})} = \frac{1}{2}$

$\Rightarrow$ b) $\mathbb{P}(E|G) = \frac{\mathbb{P}(E\cap G)}{\mathbb{P}(G)} = \frac{\mathbb{P}((H,H))}{\mathbb{P}(\{(H,H),(H,T),(T,H)\})} = \frac{1}{3}$

By the definition we immediately have

$$\mathbb{P}(E \cap F) = \mathbb{P}(F)\mathbb{P}(E|F),$$

which is used to compute the probability of the intersection.

Let us generalize to n events.

The Multiplication Rule

Prop. Let $\mathbb{P}(E_1 \cap E_2 \cap \ldots E_{n-1}) > 0$, then

$$\mathbb{P}(E_1 \cap E_2 \cap \ldots E_n) = \mathbb{P}(E_1)\,\mathbb{P}(E_2|E_1)\,\mathbb{P}(E_3|E_1 \cap E_2) \cdots \mathbb{P}(E_n|E_1 \cap E_2 \ldots E_{n-1}).$$

Proof follows from the definition of conditional probability. Note that $\mathbb{P}(E_1 \cap E_2 \cap \ldots E_{n-1}) > 0$ *is sufficient for all conditional probabilities appearing in the formula to be well defined.*

Example: An urn contains 8 red and 4 white balls. Three balls are drawn without replacement, what is the probability $E =${*2 red and 1 white*} occurs? Denote by R_i (W_i) *"red (white) at the i-th draw"*.

$\mathbb{P}(E) = \mathbb{P}((W_1 R_2 R_3) \cup (R_1 W_2 R_3) \cup (R_1 R_2 W_3))$ given the three events are mutually exclusive,

$\mathbb{P}(E) = \mathbb{P}(W_1 R_2 R_3) + \mathbb{P}(R_1 W_2 R_3) + \mathbb{P}(R_1 R_2 W_3)$.

$\mathbb{P}(W_1 R_2 R_3) = \mathbb{P}(W_1)\mathbb{P}(R_2|W_1)\mathbb{P}(R_3|W_1 R_2) = \frac{4}{12}\frac{8}{11}\frac{7}{10} = \frac{28}{165}$

Besides, the three events have the same probability $\Rightarrow$ $\mathbb{P}(E) = 3\frac{28}{165} = \frac{28}{55}$

$\mathbb{P}(\cdot|F)$ is a probability measure.

Prop. Given a sample space S, a family of events, a probability $\mathbb{P}$ and an event F with $\mathbb{P}(F) > 0$, the conditional probability $\mathbb{P}(\cdot|F)$ is a *probability measure* on the sample space. S.

▷ Let us check $\mathbb{P}(\cdot|F)$ satisfies the axioms:

A.1 for any E, $0 \leq \mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \leq 1$

A.2 $\mathbb{P}(S|F) = \frac{\mathbb{P}(S \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F)}{\mathbb{P}(F)} = 1$

A.3 If $\{E_n, n \geq 1\}$ is a sequence of mutually exclusive events, then $\mathbb{P}(\bigcup_{n:1}^{\infty} E_n|F) = \sum_{n:1}^{\infty} \mathbb{P}(E_n|F)$.

By the definition of conditional probability,

$$\mathbb{P}(\bigcup_{n:1}^{\infty} E_n|F) = \frac{\mathbb{P}((\bigcup_{n:1}^{\infty} E_n) \cap F)}{\mathbb{P}(F)}.$$

By the distributive law and axiom A.3 of $\mathbb{P}$ (note that $(E_n \cap F) \cap (E_m \cap F)) = \emptyset$ if $n \neq m$), we compute the numerator

$$\mathbb{P}(\bigcup_{n:1}^{\infty}(E_n \cap F)) = \sum_{n:1}^{\infty} \mathbb{P}(E_n \cap F).$$

Dividing by $\mathbb{P}(F)$, the result follows.

Partition rule

We have for any events E, F, with $0 < \mathbb{P}(F) < 1$

$$\mathbb{P}(E) = \mathbb{P}(F)\mathbb{P}(E|F) + \mathbb{P}(F^c)\mathbb{P}(E|F^c).$$

- We can write E as union of two mutually exclusive events

$$E = E \cap (F \cup F^c) = (E \cap F) \cup (E \cap F^c)$$

$$\mathbb{P}(E) = \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) = \mathbb{P}(F)\mathbb{P}(E|F) + \mathbb{P}(F^c)\mathbb{P}(E|F^c).$$

This can be generalized to the following case.

Prop. Consider the events E, $F_1, \ldots F_n$, with $\mathbb{P}(F_i) > 0$ and $\cup_{i:1}^n F_i = S$ e $F_i \cap F_j = \emptyset$ for any $i \neq j$. Then,

$$\mathbb{P}(E) = \sum_{i:1}^{n} \mathbb{P}(F_i)\mathbb{P}(E|F_i),$$

(*weighted average* of $\mathbb{P}(E|F_i)$ with weights $\mathbb{P}(F_i)$).

The well known Bayes formula is byproduct of the *partition rule*.

Bayes Formula

Prop. Consider a partition of S, $F_1, \dots F_n$, with $\mathbb{P}(F_i) > 0$ and $\cup_{i:1}^n F_i = S$ e $F_i \cap F_j = \emptyset$ for any $i \neq j$ and an event E, $\mathbb{P}(E) > 0$. Then, for any F_j , $j \in \{1, \dots, n\}$

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(F_j)\mathbb{P}(E|F_j)}{\sum_{i:1}^{n} \mathbb{P}(F_i)\mathbb{P}(E|F_i)}$$

- the F_js stand for alternative "hypothesis",
- $\mathbb{P}(F_j)$ are known as *priors*,
- $\mathbb{P}(F_j|E)$ the *a posteriori probability* (given E).

Example: An insurance company believes that people can be divided into 2 classes: those who are accident prone and those who are not. Their statistics shows that an accident-prone person will have an accident within a year with probability 0.4, while this probability is 0.2 for a person belonging to the other class. If the accident prone persons are the 30% of the population, what is the probability that a new policyholder will have an accident within a year? If he had, what is the probability that he is an accident prone person?

Monty Hall Problem: Goats and cars

Three doors A,B,C are closed. Behind one of them there is a Ferrari, whereas the others hide two goats. The player selects one door, say A for sake of simplicity. The TV presenter, who knows where the car is, tells her/him that the car is not behind door B and gives him the possibility to revise her/his choice (i.e. that is to select door C). Should the player revise her/his choice?