LECTURE 7

§9. CONDITIONAL PROBABILITY

In this section we ask and answer the following question. Suppose we assign a probability to a sample space and then learn that an event A has occurred. How should we change the probabilities of the remaining events? We call the new probability for an event B the *conditional probability of* B *given* A and denote it by $P(B\,/A)$.

The probability of an event may depend on the occurrence (or nonoccurrence) of another event. If this dependence exists, the associated probability is a conditional probability. In the sample space Ω, the conditional probability $P(B\,/A)$, means the likelihood of realizing an outcome in B assuming that it belongs to A. In other words, we are interested in the event B within the reconstituted sample space A. Hence, with the appropriate normalization, we obtain the conditional probability of B given A.

Example 23. A die is rolled. Let B be the event "a six turns up". Let A be the event "a number greater than 4 turns up". Before the experiment $P(B) = 1/6$. Now we are told that the event A has occurred. This leaves only two possible outcomes: 5 and 6. In the absence of any other information, we would still regard these outcomes to be equally likely, so the probability of B becomes $1/2$ making, $P(B\,/A) = 1/2$.

Definition 8. Let A and B be two events on a sample space Ω, on the subsets of which is defined a probability $P(\cdot)$. The conditional probability of the event B, given the event A, denoted by

$$P(B\,/A) = \frac{P(A \cap B)}{P(A)} \qquad \text{if} \quad P(A) \neq 0 \tag{19}$$

and if $P(A) = 0$, then $P(B\,/A)$ is undefined.

By multiplying both sides of equation (19) by $P(A)$ we obtain

$$P(A \cap B) = P(A) \cdot P(B\,/A) = P(B) \cdot P(A\,/B). \tag{20}$$

This formula is often useful as a tool to enable us in computing the desired probabilities more easily.

Example 24. Suppose that an urn contains 8 red and 4 white balls. We draw 2 balls from the urn without replacement. What is the probability that both drawn balls are red (event A)?

Solution: Let A_1 and A_2 denote, respectively the events that the first and second ball drawn is red. By (20)

$$P(A) = P(A_1 \cap A_2) = P(A_1) \cdot P(A_2\,/A_1\,).$$

It is obvious that $P(A_1) = \dfrac{8}{12}$.
Now given that the first ball selected is red, there are 7 remaining red balls and 4 white balls and so $P(A_2\,/A_1\,) = \dfrac{7}{11}$. The desired probability is

$$P(A) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}.$$

Of course, this probability could also have been computed by the classical definition

$$P(A) = \frac{\binom{8}{2}}{\binom{12}{2}}.$$

A generalization of equation (20) is sometimes referred to as the *multiplication rule*

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1) \cdot P(A_2\,/A_1\,) \cdot P(A_3\,/(A_1 \cap A_2)\,) \cdot \ldots \cdot P(A_n\,/(A_1 \cap A_2 \cap \ldots \cap A_{n-1})\,).$$

To prove the multiplication rule apply the definition of conditional probability to its right–hand side. This gives

$$P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \ldots \cdot \frac{P(\bigcap\limits_{k=1}^{n} A_k)}{P(\bigcap\limits_{k=1}^{n-1} A_k)}$$

and reducing we get $P\left(\bigcap\limits_{k=1}^{n} A_k\right)$.

§10. INDEPENDENCE AND DEPENDENCE

The notions of independent and dependent events play a central role in probability theory. If the events A and B have the property that the conditional probability of B, given A, is equal to the unconditional probability of B, one intuitively feels that event B is statistically independent of A, in the sense that the probability of B having occurred is not affected by the knowledge that A has occurred.

Since $P(B\,/A) = \dfrac{P(A \cap B)}{P(A)}$ we see that B is independent of A if

$$P(A \cap B) = P(A) \cdot P(B). \tag{21}$$

As equation (21) is symmetric in A and B, it shows that whenever B is independent of A, A is also independent of B. We thus have the following definition

Definition 9. Two events A and B are said to be independent if equation (21) holds. Two events A and B that are not independent are said to be dependent.

Example 25. Suppose that we toss 2 dice and suppose that each of the 36 possible outcomes is equally likely to occur, and hence has probability $\dfrac{1}{36}$. Assume further that we observe that the first die is a 4. What is the conditional probability that the sum of the 2 dice equal 8 (event B) given that the first die equals 4 (event A)?

Solution: The event A consists of 6 outcomes

$$\{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}.$$

The event B consists of 5 outcomes

$$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}.$$

The event $A \cap B$ consists of 1 outcome $\{(4,4)\}$, Ω is 36 element set. We obtained

$$P(A) = \frac{1}{6} \qquad P(B) = \frac{5}{36} \qquad P(A \cap B) = \frac{1}{36}$$

and

$$P(B\,/A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{1}{6}.$$

Therefore $P(A \cap B) \neq P(A) \cdot P(B)$ and the events A and B are not independent.

Example 26. A card is selected at random from an ordinary deck of 52 playing cards. If A is the event that the card is an ace and B is the event that it is a club, then A and B are independent. This follows because

$$P(A \cap B) = \frac{1}{52}, \quad P(A) = \frac{4}{52} \quad \text{and} \quad P(B) = \frac{13}{52}.$$

PROPERTIES OF INDEPENDENT EVENTS.

Property 1. If $P(A) \neq 0$ then A and B are independent if and only if

$$P(B\,/A) = P(B).$$

Property 2. If A and B are independent then so are A and $\overline{B}$.

So if A and B are independent then

$$A \text{ and } \overline{B}; \qquad \overline{A} \text{ and } B; \qquad \overline{A} \text{ and } \overline{B}$$

are also independent.

Property 3. If A is independent of B_i, $i = 1, 2$ and $B_1 \cap B_2 = \emptyset$ then A is independent of $B_1 \cup B_2$.

Let us construct an example of three events A_1, A_2, A_3 such that

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2), \qquad P(A_1 \cap A_3) = P(A_1) \cdot P(A_3), \qquad P(A_2 \cap A_3) = P(A_2) \cdot P(A_3),$$

(i. e. A_1, A_2, A_3 are pairwise independent) but

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3).$$

The proofs of Properties 1- 3 we can find in Appendix-5 of the present lecture.

Example 27. Two symmetrical dice are thrown. Let A_1 denote the event that the first die equals 4 and let A_2 be the event that the second die equals 3. A_3 denote the event that the sum of the dice is 7.

We have

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{6},$$

and

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{36}.$$

Therefore A_1, A_2, A_3 are pairwise independent (compare with Example 25).
It is not difficult to verify that

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2).$$

Therefore, the concept of independence becomes more complex for the case of more than two events. We are thus led to the following definition.

Definition 10. The events $A_1, ..., A_n$ are said to be independent if, for any k $(1 \leq k \leq n)$ and for every subset $A_{i_1}, ..., A_{i_k}$ of these events

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

APPENDIX–4.

$P(\cdot\,/B\,)$ IS A PROBABILITY

Conditional Probabilities satisfy all of the properties of ordinary probabilities. This is proved by Theorem 2 which shows that $P(\cdot\,/B\,)$ satisfies the three axioms of a probability.

Theorem 2. *Conditional probability $P(A\,/B\,)$ as a function of event A satisfies the following conditions:*

a) $P(A\,/B\,) \geq 0$ *for any A;*

b) $P(\Omega\,/B\,) = 1$;

c) If events A_i are mutually exclusive, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \Big/ B\right) = \sum_{i=1}^{\infty} P(A_i\,/B\,).$$

Proof: **Condition a) is obvious. Condition b) follows because**

$$P(\Omega\,/B\,) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Condition c) follows since

$$P\left(\bigcup_{i=1}^{\infty} A_i \Big/ B\right) = \frac{P\left(\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \bigcap B\right)}{P(B)} = \frac{P\left(\bigcup\limits_{i=1}^{\infty} (A_i \bigcap B)\right)}{P(B)} = \frac{\sum\limits_{i=1}^{\infty} P(A_i \bigcap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i\,/B\,),$$

where the next–to–last equality follows because $A_i \cap A_j = \emptyset$ implies that $(A_i \cap B) \cap (A_j \cap B) = \emptyset$. The proof is complete.

If we define $P_1(A) = P(A\,/B\,)$ (event B is fixed and $P(B) \neq 0$), then it follows from Theorem 2 that $P_1(\cdot)$ may be regarded as a probability function on the events of the sample space Ω. Hence all of the properties proved for probabilities apply to it.

APPENDIX-5.

Proofs of Properties 1 — 3.

Proof of the Property 1: **Let us prove the necessity. Let A and B be independent, therefore $P(A \cap B) = P(A) \cdot P(B)$. From here**

$$P(B\,/A\,) = \frac{P(A \cap B)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B).$$

Now we suppose that $P(B\,/A\,) = P(B)$. Therefore

$$P(A \cap B) = P(A) \cdot P(B\,/A\,) = P(A) \cdot P(B),$$

i. e. A and B are independent.

Proof of the Property 2: **As A and B are independent then $P(A \cap B) = P(A) \cdot P(B)$. It is not difficult to verify that**

$$A \cap \overline{B} = A \setminus (A \cap B).$$

Hence

$$P(A \cap \overline{B}) = P\,(A \setminus (A \cap B)) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A)[1 - P(B)] = P(A) \cdot P(\overline{B}).$$

The result is proved.

Proof of Property 3: **We have $P(A \cap B_i) = P(A) \cdot P(B_i), i = 1, 2$. Therefore**

$$P(A \cap (B_1 \cup B_2)) = P((A \cap B_1) \cup (A \cap B_2)) = P(A \cap B_1) + P(A \cap B_2) =$$
$$= P(A) \cdot P(B_1) + P(A) \cdot P(B_2) = P(A) \cdot [P(B_1) + P(B_2)] = P(A) \cdot P(B_1 \cup B_2).$$

The proof is complete.

APPENDIX-6.

PROBLEM. *For any two events A and B*

$$|P(A \cap B) - P(A)\,P(B)| \le \frac{1}{4}. \tag{A1}$$

Solution: **We start with the following proposition**

Proposition 1. *For any event A*

$$P(A)\,P(\overline{A}) \le \frac{1}{4}, \tag{A2}$$

where $\overline{A}$ is the complement of A.

Letting $P(A) = x$ and $P(\overline{A}) = 1 - x$ we have to prove that a function

$$f(x) = x\,(1-x), \qquad x \in [0,1]$$

has a maximum at the point $x = \frac{1}{2}$. The proof is obvious.
For any two events A_1 and A_2

$$P(A_1) = P(A_1 \cap A_2) + P(A_1 \cap \overline{A}_2) \tag{A3}$$

The proof immediately follows from the equality

$$A_1 = (A_1 \cap A_2) \bigcup (A_1 \cap \overline{A_2}).$$

Substituting $A_1 = B$ and $A_2 = A$ into the formula (A3) we get

$$P(B) = P(A \cap B) + P(\overline{A} \cap B) \le P(A \cap B) + P(\overline{A}) \tag{A4}$$

Above we also used that P is a monotone function, that is for any $A_1 \subset A_2$ implies

$$P(A_1) \le P(A_2). \tag{A5}$$

Since $\overline{A} \cap B \subset \overline{A}$ which implies that $P(\overline{A} \cap B) \le P(\overline{A})$.
Multiplying both sides (A4) by $P(A)$ we get

$$P(A)\,P(B) \le P(A)\,P(A \cap B) + P(A)\,P(\overline{A}) \le P(A \cap B) + P(A)\,P(\overline{A})$$

here we used that $P(A) \le 1$.
Therefore we obtain the inequality

$$P(A)\,P(B) - P(A \cap B) \le P(A)\,P(\overline{A}).$$

It follows from (A2) that

$$P(A)\,P(B) - P(A \cap B) \le \frac{1}{4}. \tag{A6}$$

Using Property (A5) of Probability we have

$$P(B) \geq P(A \cap B) \quad \textbf{and} \quad P(A) \geq P(A \cap B)$$

which imply that

$$P(A)\,P(B) \geq P(A \cap B)\,P(A \cap B).$$

Therefore

$$P(A)\,P(B) - P(A \cap B) \geq P(A \cap B)\,P(A \cap B) - P(A \cap B) = -P(A \cap B)\,P(\overline{A \cap B}) \geq -\frac{1}{4},$$

where the above equality is obtained by noting

$$P(A \cap B) - 1 = -P(\overline{A \cap B})$$

and (A2).
Thus we obtain two inequalities (compare with (A6))

$$P(A)\,P(B) - P(A \cap B) \leq \frac{1}{4},$$

$$P(A)\,P(B) - P(A \cap B) \geq -\frac{1}{4}.$$

Now (A1) follows from these two inequalities. The proof is complete.

APPENDIX-7.

PROBLEM. **Let A and B be mutually exclusive events of an experiment. Then, when independent trials of this experiment are performed, the event A will occur before the event B with probability**

$$\frac{P(A)}{P(A)+P(B)}.$$

Solution: **If we let C_n denote the event that no A or B appears on the first $n-1$ trials and A appears on the nth trial, then the desired probability p is**

$$p = P(\bigcup_{n=1}^{\infty} C_n) = \sum_{n=1}^{\infty} P(C_n).$$

Now, since $P(A\ \text{ on any trial}) = P(A)$ and $P(B\ \text{ on any trial}) = P(B)$, we obtain, by the independence of trials

$$P(C_n) = [1-(P(A)+P(B))]^{n-1}\,P(A),$$

and thus

$$p = P(A)\sum_{n=1}^{\infty}[1-(P(A)+P(B))]^{n-1} =$$

(using the formula of the sum of terms of geometric progression; the first term equals 1)

$$= P(A)\,\frac{1}{1-(1-(P(A)+P(B)))} = \frac{P(A)}{P(A)+P(B)}.$$