Lecture Note #03 of
Econometrics I & Advanced Econometrics I (2013SY)

TAKEUCHI, Yoshiyuki

April 24, 2013

§ 2. Probability and Probability Space

§ 2.1 Family Class and Measure (cont'd)

Theorem 2.1.4

Let $(\Omega, \mathcal{B}, P)$ be a probability space $(i = 1, \cdots, \infty)$.
For $A,\ B,\ C_i,\ D_i \in \mathcal{B}$, the following properties are satisfied.

(a) $\quad P(\emptyset) = 0$

(b) $\quad P(\sum_{i=1}^{n} C_i) = \sum_{i=1}^{n} P(C_i), \qquad C_i \cap C_j = \emptyset \ (i \neq j)$ (finite addtivity)

(c) $\quad P(A^C) = 1 - P(A)$

(d) $\quad A \subset B \Rightarrow P(A) \leq P(B)$

(e) $\quad P(A) \leq 1$

(f) $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (additive theorem)

(g) $\quad D_n \subset D_{n+1} \quad n = 1, 2, \cdots \ \Rightarrow \ P(D_n) \uparrow P(\bigcup_{n=1}^{\infty} D_n)$

(h) $\quad D_n \supset D_{n+1} \quad n = 1, 2, \cdots \ \Rightarrow \ P(D_n) \downarrow P(\bigcap_{n=1}^{\infty} D_n)$

(i) $\quad P(\bigcup_{i=1}^{n} D_i) \leq \sum_{i=1}^{n} P(D_i) \qquad n = 1, 2, \cdots, \infty$

(j) $\quad$ If D_n is monotone, $\lim P(D_n) = P \lim(D_n)$

§ 2.2 Conditional Probability

Definition 2.2.1

Suppose $(\Omega, \mathcal{B}, P)$. For A , $B \in \mathcal{B}$ $\quad P(A) > 0$, we define

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

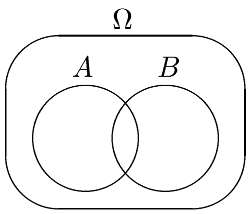
as a conditional probability of the event B given (or conditioning) the event A.

Theorem 2.2.2

The conditonal probability $P(B \mid A)$ is satisfied the following properties;

(1) $P(B \mid A) \geq 0$,

(2) $A \cap \Omega^* = A$, if $\Omega^* = A$. Therefore, we obtain $P(\Omega^* \mid A) = \frac{P(A)}{P(A)} = 1$.



The sample space is restricted to A.

Theorem 2.2.3

(1) For A, $B \in \mathcal{B}$, we obtain

(i) $P(A \cap B) = P(A)P(B \mid A)$ $\quad$ for $P(A) > 0$,

(ii) $P(A \cap B) = P(B)P(A \mid B)$ $\quad$ for $P(B) > 0$.

(2) For a sequence of events $A_1, A_2, \ldots, A_n$ such that $P(\bigcap_{i=1}^{k} A_i) > 0$ $(k = 1, 2, \ldots, n-1)$, we obtain

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_2 \cap A_1) \cdots P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right).$$

Now, let A, $B \in \Omega$. As $A^C \in \Omega$ and $A \cap A^C = \emptyset$, we can represent B as

$$B = (A \cap B) \cup (A^C \cap B).$$

And using the fact that $(A \cap B) \cap (A^C \cap B) = \emptyset$, we get

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^C \cap B) \\ &= P(B \mid A)P(A) + P(B \mid A^C)P(A^C). \end{aligned}$$

To generalize this result, we can obtain the following theorem.

Theorem 2.2.4

Suppose $A_1, A_2, \dots, A_n \in \Omega,\ B \in \Omega$.
If $A_i \cap A_j = \emptyset \quad {}^{\forall} i \neq j$ (i.e. disjoint) and $\quad \Omega = \sum_{i=1}^{n} A_i$ are satisfied, we obtain

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i).$$

If $P(A) > 0$ and $P(B) > 0$, we get

$$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

by Theorem 2.2.3. With this result and Theorem 2.2.4, we can obtain the following theorem.

Theorem 2.2.5 Bayes' Theorem

Let a mutually disjoint sequence of events $A_1, \dots, A_n$ be satisfied $\sum_{i=1}^{n} A_i = \Omega$. Then, for any event such that $P(B) > 0$,

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}, \qquad k = 1, 2, \dots, n$$

is satisfied.

The Bayes statistics uses this relationship in fundamental concepts.

- $P(A_i)$: prior probability （「事前確率」 in Japanese）
- $P(B \mid A_i)$: likelihood, generating system （「尤度」 in Japanese）
- $P(A_i \mid B)$: posterior probability （「事後確率」 in Japanese）

§ 2.3 Independence of events

Definition 2.3.1

Let $A, B \in \mathcal{B}$. For events A, B, we say that events A and B are mutually **independent** if

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

is satisfied.

When A and B are independent each other, we can obtain

$$\begin{aligned} P(B) &= P(B \mid A), \\ P(A) &= P(A \mid B) \end{aligned}$$

by Theorem 2.2.3. In this case, the conditional expectation equals the unconditional expectation, that is to say it does not rely on the condition.

< Note > Difference between independence and disjointness.

$$\begin{array}{ll} disjointness & P(A \cap B) = P(\emptyset) = 0 \\ independence & P(A \cap B) = P(A) \cdot P(B) \end{array}$$

For $P(A), P(B) > 0$, A and B are not independent if they are disjoint.

Theorem 2.3.2

Let $A, B \in \mathcal{B}$. If events A and B are independent, then A^C and B^C are independent.
(Proof)

$$\begin{aligned} P(A^C \cap B^C) &= 1 - P(A \cup B) \\ &= 1 - (P(A) + P(B) - P(A \cap B)) \end{aligned}$$

Since A and B are independent,

$$P(A \cap B) = P(A) \times P(B).$$

Hence,

$$\begin{aligned} P(A^C \cap B^C) &= 1 - \{P(A) + P(B)\} + P(A) \cdot P(B) \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^C) \times P(B^C). \end{aligned}$$

Therefore, A^C and B^C are independent. ∎

Independence of more than two events

(1) Independence of a sequence of finite events

Independence of events $A_1, A_2, \cdots, A_n$ are defined as

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) &= P\left(\bigcap_{i=1}^{n} A_i\right) \\ &= \prod_{i=1}^{n} P(A_i). \end{aligned}$$

Theorem 2.3.3

If events $A_1, A_2, \cdots, A_n$ are (jointly) independent, then any two events A_i and A_j $(i \neq j)$ are mutually independent.

(remark) Converse of the theorem is not true. Give you an example (Bernstein's example) below.

Suppose we have four cards with 3-digit numbers such that {110, 101, 011, 000}. Let A_i be the event that ith-digit of the card is 1 when a card is randomly chosen. In this example, we have

$$P(A_1) = 1/2, \; P(A_2) = 1/2, \;\; \text{and} \;\; P(A_3) = 1/2.$$

and

$$P(A_1 \cap A_2) = 1/4 = P(A_1)P(A_2),$$

$$P(A_2 \cap A_3) = 1/4 = P(A_2)P(A_3),$$

$$\text{and} \;\; P(A_1 \cap A_3) = 1/4 = P(A_1)P(A_3).$$

However,

$$P(A_1 \cap A_2 \cap A_3) = 0 \neq P(A_1)P(A_2)P(A_3) = 1/8.$$

(2) Independence of a sequence of infinite events

Suppose a sequence of coutable (infinite) events.

Definition 2.3.4

Let $A_j \in \mathcal{B}$. $A_1, A_2, \cdots$ is independent if for every subsequence A_{ji} $(i = 1, 2, \cdots, n)$ are independent, that is to say,

$$P\left(\bigcap_{i=1}^{n} A_{ji}\right) = \prod_{i=1}^{n} A_{ji}.$$

Theorem 2.3.5

If $A_1, A_2, \cdots$ are independent, then we get

$$P\left(\bigcap_{j=1}^{\infty} A_j\right) = \prod_{j=1}^{\infty} P(A_j),$$
$$P\left(\bigcap_{j=1}^{\infty} A_j^C\right) = \prod_{j=1}^{\infty} P(A_j^C).$$

(Proof)
$B_n \equiv \bigcap_{j=1}^{n} A_j$ is a decreasing sequence. By Theorem 1.2.1 and complete additivity, we obtain

$$P\left(\bigcap_{j=1}^{\infty} A_j\right) = P(\lim B_n) = \lim P(B_n) = \prod_{j=1}^{\infty} A_j.$$

∎

Theorem 2.3.6 Borel-Cantelli Lemma

(a) For any sequence of events $A_n \in \mathcal{B} \quad (n = 1, 2, \cdots)$, if $\sum_{n=1}^{\infty} P(A_n) < \infty$ is satisfied, then we get

$$P(\limsup A_n) = 0.$$

(b) Suppose an sequence of events $A_n \in \mathcal{B} \quad n = 1, 2, \cdots$ is independent. if $\sum_{n=1}^{\infty} P(A_n) = \infty$ is satisfied, then we get

$$P(\limsup A_n) = 1.$$

This lemma is important for handling the convergence theorem.

References

Stępniak, C., (2007), "Bernstein's examples on independent events," *The College Mathematical Journal*, 38 (2), pp. 140–142.