

Lecture Notes 5

October 25, 2015

Scribe: Wei-Chang Lee, Chi-Ning Chou

After explaining the concept of independence, today we can finally introduce the Second Borel-Cantelli Lemma. A lemma that can lately be used to prove one of the most important result in probability theory, the strong law of large numbers. Also, we talk about another fundamental concept in probability, random variable, which let us change our views from set to the real field.

0.1 The Second Borel-Cantelli Lemma

Also called the reverse Borel-Cantelli Lemma.

Theorem 1 Let $A_1, A_2, A_3 \dots A_n \dots \in \mathcal{A}$ be mutually independent. If $\sum_{i=1}^{\infty} P(A_i) = \infty$, then

$$P(A_{n.i.o}) = 1$$

, where $A_{n.i.o} = \limsup_{n \rightarrow \infty} A_n$.

Proof: First Notice that since $\{A_n\}$ are independent, so $P(\bigcap_{k=n}^{n+j} A_k^c) = \prod_{k=n}^{n+j} (1 - P(A_k))$ and $1 - x \leq e^{-x} \forall x \geq 0$ then

$$P(\bigcap_{k=n}^{n+j} A_k^c) \leq e^{-\sum_{k=n}^{n+j} P(A_k)} \forall n$$

And $\sum_{i=1}^{\infty} P(A_i) = \infty$, one has

$$\lim_{j \rightarrow \infty} P(\bigcap_{k=n}^{n+j} A_k^c) \leq \lim_{j \rightarrow \infty} e^{-\sum_{k=n}^{n+j} P(A_k)} = 0$$

From Boole's inequality, it follows that

$$P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c) \leq \sum_{n=1}^{\infty} P(\bigcap_{k=n}^{\infty} A_k^c) = 0$$

Then by De Morgan's rule we have,

$$P(\limsup_{n \rightarrow \infty} A_n) = P(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k) = 1 - P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c) = 1$$

■

So when $\{A_n\}$ are mutually independent, combining both Borel-Cantelli lemma, we have a zero-one law: Borel's zero-one law

$$P(A_{n.i.o.}) \text{ is either 0 or 1 with respect to } \sum_{i=1}^{\infty} P(A_i) \text{ converged or not.}$$

This give us a peek of Kolmogorov's zero-one law.

Theorem 2 Suppose a probability space (Ω, \mathcal{A}, P) and A_n be a sequence of mutually independent σ -algebras contained in \mathcal{A} then the tail events $F \in \bigcap_{n=1}^{\infty} G_n$ where $G_n = \sigma(\bigcup_{k=n}^{\infty} A_k)$ is either zero or one i.e. $P(F) = 1$ or $P(F) = 0$.

Remark 1 A naive way to understand the theorem is to consider if you have a sequence of independent random variables and an event that is invariant if you ignore finitely many of the variables, then the probability of that event is either 0 or 1 .

Remark 2 Independent assumption is really important in the Borel's zero one law. If a event E with $0 < P(E) < 1$ and let a sequence of events $\{A_n\}$ where each $A_i = E$ then it is clear $\{A_n\}$ are not independent at all and $P(A_{n.i.o.}) = P(E)$ which violates the zero one law.

Intuition (Borel Cantelli Lemma)

The Borel Cantelli Lemma enable us to construct probability in infinite many time/games from probability in sequence of events. If we toss a coin or play a independent game infinite many times, and let A_n be events in n th toss/round. We can first study a pattern of coin or strategy in games has how much chance to win in one toss/round then use Borel Cantelli Lemma to extend it to infinite case that certain patterns will definitely happen or our strategy can finally win. An intersting example is given as below.

Example 1 [Infinite Monkey Theorem] Suppose a monkey has a typewriter, he types any alphabet at random. If he never stops typing, then he can almost surely type any Shakespeare's sonnet.

Solve: Consider the alphabets sequence the monkey types, we can divide this infinitely long string to infinitely many chunks, each chunk has length k where k equals to the number of alphabets in given Shakespeare's sonnet. Let $\{E_n\}$ denotes the monkey successfully types the sonnet in the n th chunk, then we have

$$P(E_n) = \frac{1}{26^k} > 0 \text{ so } \sum_{n=1}^{\infty} P(E_n) = \infty$$

And chunks are independent from one another. Using the Second Borel Cantelli Lemma, we get

$$P(E_{n.i.o.}) = 1$$

There are infinite many chunks, the monkey almost surely types Shakespeare's sonnet. ■

Remark 3 In reality, the monkey can not fulfil this since not only did he has certain behavior pattern but also reality is a very big finite number. If the universe is filled with monkeys and they keep typing, the chance to type Hamlet is less than $\frac{1}{10^{183800}}$. The gap between infinite and finite is bigger than every big finite number we can guess.

Example 2 [Strong Law of large number] We neglect the proof but argue how it can be used in proving it. We want to show that

$$P(\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu) = P(\Omega_0) = 1$$

Where $S_n = X_1 + X_2 \dots + X_n$ and $E(X_i) = \mu$ then it is equivalent to there exist $\Omega_0 \in \mathcal{A}$ for every $w \in \Omega_0$ it holds that

$$\lim_{n \rightarrow \infty} |\frac{S_n}{n} - \mu| = 0$$

It suffices to show for any $\epsilon > 0$, $|\frac{S_n}{n} - \mu| > \epsilon$ can occur only a finite number of times. And it can be done by Borel-Cantelli Lemma.

0.2 Random Variables

Before giving formal definition of what is a random variable. Let use first introduce two important terms, Borel sets and Σ -measurable function.

Definition 1 (Borel σ -algebra) Let S be a topological space then $\mathcal{B}(S)$, the Borel σ -algebra on S , is the σ -algebra generated by the family of open subsets of S .

$$\mathcal{B}(S) := \sigma(\text{open sets})$$

And

$$\mathcal{B} := \mathcal{B}(\mathbb{R})$$

Every subsets of \mathbb{R} we meet in everyday use is an element of \mathcal{B} ; but elements of \mathcal{B} can be quite complicated. An easy way to understand \mathcal{B} is by the π system (closed under finite intersection) ξ ,

$$\xi = \pi(\mathbb{R}) := \{(-\infty, x] : x \in \mathbb{R}\}$$

then

$$\mathcal{B} = \mathcal{B}(\mathbb{R}) = \sigma(\xi)$$

To introduce the concepts of Σ -measurable function and random variable, we should first define the preimage of a set function.

Definition 2 (preimage of set function) Let $h : \Omega \rightarrow \mathcal{X} \subset \mathbb{R}$ be a set function. Then the preimage of $S \subset \mathbb{R}$ over h is

$$h^{-1} := \{w \in \Omega : h(w) \in S\}$$

Note that the image of the set function is not necessary a really set. That is, it can also be a single point in \mathbb{R} just as the following definition of Σ -measurable function.

Definition 3 (Σ -measurable function) Suppose that $h : \Omega \rightarrow \mathbb{R}$ then h is called a Σ -measurable function if $h^{-1} : \mathcal{B} \rightarrow \Sigma$, that is, $h^{-1}(S) \in \Sigma \forall S \in \mathcal{B}$. And we write $m\Sigma$ to be the set of all these measurable function on Ω .

Intuition (measurable function)

A set function h is Σ -measurable if the preimages of every Borel set lies in Σ .

Definition 4 (Random variable) Let (Ω, \mathcal{A}) be our sample space and family of events. A random variable X is an element of $m\mathcal{A}$ which will satisfied our random mechanism P . Thus,

$$X : \Omega \rightarrow \mathbb{R}, X^{-1} : \mathcal{B} \rightarrow \mathcal{A}$$

Intuition (Random variable)

Suppose X is a random variable carried by arbitrary probability triple (Ω, \mathcal{A}, P) then we have

$$\mathcal{B} \xrightarrow{X^{-1}} \mathcal{A} \xrightarrow{P} [0, 1]$$

Define μ_x by

$$\mu_x := P \circ X^{-1} \text{ i.e. } \mu_x(B) = P(\{w : X(w) \in B\}) \forall B \in \mathcal{B}$$

And μ_x is actually a probability measure on (R, \mathcal{B}) , then for any abstract probability space we can use X to change it to (R, \mathcal{B}) which we are familiar with.

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (R, \mathcal{B}, \mu_x)$$

and define the function $F_X : R \rightarrow [0, 1]$ as:

$$F_X(c) := \mu_x((-\infty, c]) = P(\{w : X(w) \leq c\}) = P(X \leq c)$$

Now how can we verify any real-valued function X on Ω is indeed a random variable? We first begin from the reverse image of X .

Property 1 (Properties of set function) Let $X : \Omega \rightarrow \mathcal{X} \subset \mathbb{R}$ be a set function, then the following properties hold. (Note that here X is not necessary a random variable)

1. (Close under complementation): $X^{-1}(S^C) = (X^{-1}(S))^C$
2. (Close under union): $X^{-1}(\bigcup_{\alpha \in \Gamma} S_\alpha) = \bigcup_{\alpha \in \Gamma} X^{-1}(S_\alpha)$, where $\{S_\alpha\}_{\alpha \in \Gamma}$
3. (Close under intersection): $X^{-1}(\bigcap_{\alpha \in \Gamma} S_\alpha) = \bigcap_{\alpha \in \Gamma} X^{-1}(S_\alpha)$, where $\{S_\alpha\}_{\alpha \in \Gamma}$

By definition, a random variable should satisfy the condition that the preimage of every Borel set should in the event space \mathcal{A} . However, it's difficult to check since it's hard to enumerate all Borel set and claim the results. Thus, we would like to find a relaxed but necessary condition for a set function to be a random variable. And the following theorem does so.

Theorem 3 (necessary and sufficient condition for random variable) Let $X : \Omega \rightarrow \mathcal{X} \subset \mathbb{R}$ be a set function, then

$$X \text{ is a random variable} \Leftrightarrow \forall x \in \mathbb{R}, \{w : X(w) \leq x\} \in \mathcal{A}$$

The proof is left in Appendix 0.4

0.3 From Set Function to Value Function

Note that random variable is a function that helps us map the event set \mathcal{A} onto the Borel set. The importance here is that now we can do the equivalent operation on **real number** instead of arbitrary σ -algebra. This not only provides us a general and uniform way to play but also gives us the opportunity to operate on an **ordered** set.

Soon, we might wonder can we play with an even more general function: value function instead of just a set function onto real number? And by the construction of random variable, there are two direct value functions that play an important role in probability theory.

Theorem 4 (value functions) *Let X be a random variable w.r.t. (Ω, \mathcal{A}, P) , then we can define*

- *The measure function of X is $\mu_X : \mathcal{B} \rightarrow \mathbf{R}^+$ such that*

$$\forall B \in \mathcal{B}, \mu_X(B) := P\{w : X(w) \in B\}$$

- *The cumulative distribution of X is $F_X : \mathbf{R} \rightarrow \mathbf{R}^+$ such that*

$$\forall x \in \mathbf{R}, F_X(x) := P\{w : X(w) \leq x\}$$

Here, we also have a necessary and sufficient condition for cumulative function. Intuitively, when a function F satisfies the following conditions, then there's a random variable with its unique cumulative distribution being F .

Theorem 5 (necessary and sufficient condition of cumulative distribution) *F is a cumulative distribution iff*

- *(upper and lower bound) $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$*
- *(non-decreasing) $\forall x \leq y, F(x) \leq F(y)$*
- *(right continuous) $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$*

The proof is left in Appendix 0.5.

Intuition (set function and value function)

Careful with the difference of

- **(random variable):** $X : \Omega \rightarrow S \subset \mathbf{R}$
- **(measure function):** $\mu_X : \mathcal{B} \rightarrow \mathbf{R}^+$
- **(cumulative distribution):** $F_X : \mathbf{R} \rightarrow \mathbf{R}^+$

0.4 Proof of the necessary and sufficient condition of random variable

Recall that Theorem 3 provides a necessary and sufficient condition for a set function to be a random variable. Here, we prove the correctness of the theorem.

(\Rightarrow) First, we can see that $\{w : X(w) \leq x\} = X^{-1}((-\infty, x])$. Thus, it's sufficient to show that $(-\infty, x)$ is in Borel set $\forall x$. And this is simple since $(-\infty, x) = \bigcup_{n \in \mathbb{N}} [-n, x] \in \mathcal{B}$.

(\Leftarrow) This direction can be proved in two steps:

1. Show that $\forall a < b, X^{-1}([a, b]) \in \mathcal{A}$.
2. Show that $\xi = \{S : \exists w \in \Omega, X(w) = S\}$ is a σ -algebra.

With these two results, we can see that the image of X contains \mathcal{B} . Thus, X is a random variable. The following show the correctness of these two:

1. $\forall a < b$, consider

$$\begin{aligned} X^{-1}([a, b]) &= X^{-1}((-\infty, b] \setminus (-\infty, a]) = X^{-1}((-\infty, b] \cap (-\infty, a])^C) \\ &= X^{-1}((-\infty, b]) \cap X^{-1}((-\infty, a])^C \\ &= X^{-1}((-\infty, b]) \cap X^{-1}((-\infty, a])^C \in \mathcal{A} \end{aligned}$$

2. Check the three axioms of σ -algebra:

- (a) Clearly, $\emptyset \in \xi$.
- (b) $\forall S \in \xi, X^{-1}(S^C) = X^{-1}(S)^C \in \mathcal{A}$ by Property 1. Thus $S^C \in \xi$.
- (c) $\forall \{S_\alpha\}_{\alpha \in \Gamma} \in \xi, X^{-1}(\bigcup_{\alpha \in \Gamma} S_\alpha) = \bigcup_{\alpha \in \Gamma} X^{-1}(S_\alpha) \in \mathcal{A}$. Thus $\bigcup_{\alpha \in \Gamma} S_\alpha \in \xi$.

0.5 Proof of the necessary and sufficient condition of cumulative function

Take $\Omega = (0, 1)$, $\mathcal{A} = \mathcal{B}(0, 1)$ and P be Lebesgue measure.

Remark 4 If $B = (a, b)$ or $[a, b]$ then $Leb(B) = b - a$, i.e. the measure of an interval is its length.

For $w \in (0, 1)$ define

$$X(w) = \sup\{y : F(y) < w\}$$

We need to show that X is a random variable and the distribution function $F_x = F$. We must first verify $X(w)$ is well-defined:

1. since $w > 0$ and $\lim_{x \rightarrow -\infty} F(x) = 0$, so $\{y : F(y) < w\} \neq \emptyset$.
2. since $w < 1$ and $\lim_{x \rightarrow \infty} F(x) = 1$, so $\{y : F(y) < w\}$ bounded above.

So $X(w) = \sup\{y : F(y) < w\}$ exists in $\mathcal{R} \forall w \in (0, 1)$, $X(w)$ is well defined. We further claim that:

$$\{w : X(w) \leq c\} = \{w : w \leq F(c)\} = (0, F(c)]$$

Then X is a random variable since

$$\{w : X(w) \leq c\} = (0, F(c)] \in \mathcal{B}(0, 1)$$

Aware that $F(c)$ may be 0 or 1 for some $c \in \mathcal{R}$, it is not that trivial to show such c will let $(0, F(c)] \in \mathcal{B}$ but since $\{w : X(w) \leq c\} = \{w : w \leq F(c)\}$ and $w \in (0, 1)$, we have:

$$F(c) = 0, \{w : w \leq F(c)\} = \emptyset \in \mathcal{B} \text{ and } F(c) = 1, \{w : w \leq F(c)\} = \Omega \in \mathcal{B}$$

And

$$F_x(c) = P(X^{-1}(\infty, c)) = \text{Leb}(\{w : X(w) \leq c\}) = \text{Leb}((0, F(c)]) = F(c)$$

So we can define a random variable X on $(\Omega, \mathcal{B}(0, 1), \text{Leb})$ which its distribution function is just given cumulative function. We complete the proof by showing:

$$\{w : X(w) \leq c\} \subset \{w : w \leq F(c)\}, \{w : X(w) \leq c\} \supset \{w : w \leq F(c)\}$$

First observe that

$$F(X(w) + \delta) \geq w, \forall \delta > 0$$

Since if $F(X(w) + \delta) < w$ it implies a contradiction:

$$X(w) + \delta \in \{y : F(y) < w\} \leq X(w)$$

And

$$\lim_{\delta \rightarrow 0} F(X(w) + \delta) \xrightarrow{\text{right-continuous}} F(X(w)) \geq w$$

1. \subset case: $X(w) \leq c \implies F(c) \geq F(X(w)) \geq w$ (non-decreasing) so,

$$\{w : X(w) \leq c\} \subset \{w : w \leq F(c)\}$$

2. \supset case: $F(c) \geq w \implies c$ is an upper bound of $\{y : F(y) < w\}, c \geq \sup\{y : F(y) < w\} = X(w)$

$$\{w : X(w) \leq c\} \supset \{w : w \leq F(c)\}$$

Chi-Ning's version Here, we are going to show that the three conditions in Theorem 5 is necessary and sufficient for a function F to be a cumulative distribution of a random variable.

(Necessary) Let F be the cumulative distribution of random variable X defined on (Ω, \mathcal{A}, P) . Check the three conditions as follow:

1. First, let $A_n = (-\infty, -n]$. As $A_n \rightarrow \emptyset$ when $n \rightarrow \infty$, by the continuity axiom, $\lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} F(-n) = 0$. Since, $\lim_{x \rightarrow -\infty} F(x) \leq \lim_{x \rightarrow -\infty} F(\lfloor x \rfloor) = \lim_{n \rightarrow \infty} F(-n) = 0$, we obtain the lower bound. The upper bound can be obtained with the same technique.
2. If $x \leq y$, then $(-\infty, x] \subseteq (-\infty, y]$. Clearly, $F(x) = P((-\infty, x]) \leq P((-\infty, y]) = F(y)$.

3. Let $\{x_n\}$ be a sequence strictly decreases (approaches) to x_0 , then $A_n := (-\infty, x_n] \rightarrow A = (-\infty, x]$. That is, $F(x) = P(A)$ and $F(x_n) = P(A_n)$, $\forall n$. Now consider $B_n = \mathbb{R} \setminus A_n$, $\forall n$ and $B = \mathbb{R} \setminus A$. For convenience, let $A_0 = \Omega$ and $B_0 = \emptyset$. Then we have

$$\begin{aligned} 1 - P(A) &= P(B) = P(\lim_{n \rightarrow \infty} B_n) = P(\bigcup_{n=1}^{\infty} B_n \setminus B_{n-1}) \\ &= \sum_{n=1}^{\infty} P(B_n \setminus B_{n-1}) = \sum_{n=1}^{\infty} P(A_{n-1}) - P(A_n) \\ &= 1 - \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

Thus, $F(x_0) = P(A) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{x \rightarrow x_0^+} F(x)$.

(sufficient) Let F be a function satisfies the three conditions in Theorem 5. Take $(\Omega = (0, 1), \mathcal{A} = \mathcal{B}(0, 1), P = \mathcal{L}(0, 1))$ as the probability space. Now, we are going to construct a random variable $X : \Omega \rightarrow \mathcal{X} \subset \mathbb{R}$ w.r.t F over (Ω, \mathcal{A}, P) . X is defined as

$$\forall w \in \Omega, X(w) := \inf\{x : w \leq F(x)\}$$

To show that X is indeed the random variable with F as the underlying cumulative distribution. We need to first demonstrate that X is a random variable then show the relation between X and F .

1. By Theorem 3, it's sufficient to show that $\forall x \in \mathbb{R}, X^{-1}((-\infty, x]) \in \mathcal{A}$. Consider $\forall x \in \mathbb{R}$,

$$\begin{aligned} X^{-1}((-\infty, x]) &= \{w : X(w) \leq (-\infty, x]\} = \{w : \inf\{y : w \leq F(y)\} \in (-\infty, x]\} \\ &= \{w : w \leq F(x)\} = (0, F(x)] \in \mathcal{A} \end{aligned}$$

Thus, X is a random variable over (Ω, \mathcal{A}, P) .

2. Check that $\forall x \in \mathbb{R}, F(x) = P(X \leq x)$. Consider $\forall x \in \mathbb{R}$

$$\begin{aligned} P(X \leq x) &= P(\{w : X(w) \leq x\}) = P(\{w : \inf\{y : w \leq F(y)\} \leq x\}) \\ (\because y \leq x \Rightarrow F(y) \leq F(x)) &= P(\{w : \inf\{F(y) : w \leq F(y)\} \leq F(x)\}) \\ &= P(\{w : w \leq F(x)\}) = P((0, F(x)]) \\ (\because P = \mathcal{L}) &= F(x) \end{aligned}$$

That is, F is the cumulative distribution of random variable X .