Author: Dr. Shi Guo Email: guoshi1984@hotmail.com

The reason of the article is to provide a basic review of several key concepts in measure theory. Each concept is illustrated with an example so that one can easily understand.

1 σ Algebra

a. Sigma algebra definition
Given a non-empty set Ω, a sigma algebra is a collection of all the subsets of Ω that
1) Include empty set and whole set
2) Include the complement of any element in the sigma algebra
3) Is closed under countable union

b. Sigma algebra example by tossing a coin
The procedure to find out the sigma algebra is to enumerate all the subsets under the whole set Ω. We now see an example.

We first toss a coin 0 time, there is no outcome, so $\Omega = \{\emptyset\}$. And the σ algebra contains an empty set only.
$F_0 = \{\emptyset\}$
In this case, it is trivial to check 1) 2) and 3)

We then toss the coin once, the outcome is either Head(H) or Tail(T)
Check 1) Ω = H, T
Enumerate all of the subsets of Ω, we get
F_1 = {0, Ω, H, T}
Check 2) $\emptyset^c = \Omega$, $\Omega^c = \emptyset$, H_c = T in F_1, T_c = H in F_1
Check 3) $H \cup T = \Omega$ in F_1
So we confirm
F_1 = 0, Ω, H, T

We then toss the coin twice
Check 1) Ω = { HH, HT, TH, TT }
Enumerate all the subsets of Ω, we get
$F_2 = \{\emptyset, \Omega, HH, HT, TH, TT,$
$HH \cup HT, HH \cup TH, HH \cup TT, HT \cup TH, HT \cup TT, TH \cup TT,$
$HH \cup HT \cup TH, HH \cup HT \cup TT, HH \cup TH \cup TT, HT \cup TH \cup TT, \}$
It is easy to check 2), for example $HH^c = HT \cup TH \cup TT$ in F_2, $HT^c = HH \cup TH \cup TT$ in F_2, $TH^c = HH \cup HT \cup TT$ in F_2, $TT^c = HH \cup HT \cup TH$ in F_2
The rest check is ignored.
3) is easy to check too. So we confirm
$F_2 = \{\emptyset, \Omega, HH, HT, TH, TT,$
HH^c, HT^c, TH^c, TT^c
$HH \cup HT, HH \cup TH, HH \cup TT, HT \cup TH, HT \cup TT, TH \cup TT,$
$TT^c, TH^c, HT^c, HH^c\}$

c. Why define sigma algebra?
On top of the sigma algebra, we can define the probability, because the object that probability measure takes is the sigma algebra.

2 Filtration

Consider a sequence of coin toss
For the first toss, we get F_1
For the first and second toss, we get F_2
For the first n tosses, we get F_n
The collection of sigma algebra F_1, F_2 F_n is called a Filtration.

3 Random variable

a. Definition
A random variable is function from Ω to R, with the property that for every Borel subset B of R, its inverse image from the subset of Ω is in σ-algebra F.

b. Example
Consider 3 toss case, H with prob p, T with prob q
Def. random variable S $S_0(w_0) = 4$ for all ω

$$\begin{aligned} S_{n+1}(w_{n+1}) =& 2S_n(w_n) \text{ if } w_{n+1} = H \\ & \frac{1}{2}S_n(w_n) \text{ if } w_{n+1} = T \end{aligned}$$

so
$S_0(w_1w_2w_3) = 4$ for all w_i
$S_1(w_1w_2w_3) = 8$ if w_1 = H
$S_1(w_1w_2w_3) = 2$ if w_1 = T
$S_2(w_1w_2w_3) = 16$ if $w_1 = w_2 = H$
$S_2(w_1w_2w_3) = 4$ if $w_1 \neq w_2$
$S_2(w_1w_2w_3) = 1$ if $w_1 = w_2 = T$

4 σ Algebra Generated by a Random Variable and Measurable Function

Give consider a random variable S: Ω to R, for every open set in R, the collection of their inverse image forms an sigma algebra, and it is called the sigma algebra generated by S. And S is called F-measurable. The concept measurable is not very intuitive to understand. An easy way to understand this is S is completely determined by F, then S is F measurable.

5 Conditional Expectation

a. Definition

The conditional expectation is a random variable that satisfies the two following conditions

1) $E[X|\mathcal{G}]$ is $\mathcal{G}$ measurable, which means the value of $E[X|\mathcal{G}]$ is completely determined by $\mathcal{G}$

2) $\int_A E[X|\mathcal{G}](w)dP(w) = \int_A X(w)dP(w)$ for all A which belongs to $\mathcal{G}$

b. Example to understand 2)

Consider 3 toss case, H with prob p, T with prob q

Define random variable S

$S_0(w) = 4$ for all w

$S_{n+1}(w) = 2S_n(w)$ if $w_{n+1} = H$

$S_{n+1}(w) = \frac{1}{2}S_n(w)$ if $w_{n+1} = T$

Expectation of 3 tosses random variable S_3 give the first two is HH

$$
\begin{aligned}
&E_2(S_3|HH) = pS_3(HHH) + qS_3(HHT)\\
&E_2(S_3|HT) = pS_3(HTH) + qS_3(HTT)\\
&E_2(S_3|TH) = pS_3(THH) + qS_3(THT)\\
&E_2(S_3|TT) = pS_3(TTH) + qS_3(TTT)\\
&E_2(S_3|HH)P(HH) = prob(HHH)S_3(HHH) + prob(HHT)S_3(HHT)\\
&E_2(S_3|HT)P(HT) = prob(HTH)S_3(HTH) + prob(HTT)S_3(HTT)\\
&E_2(S_3|TH)P(TH) = prob(THH)S_3(THH) + prob(THT)S_3(THT)\\
&E_2(S_3|TT)P(TT) = prob(HTH)S_3(TTH) + prob(TTT)S_3(TTT)
\end{aligned}
$$

This confirms def 2), for A = HH or HT or TH or TT

$\int E_2(S_3|\mathcal{G})(w)dP(w) = \int_A X(w)dP(w)$

c. Properties

1) The conditional expectation is a random variable. Because the value is dependent on $\mathcal{G}$.

2) If X is $\mathcal{G}$ measurable, then $E[X|\mathcal{G}] = X$.

3) If X is $\mathcal{G}$ measurable $E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$, this is to take out what is known.

4) If X is independent of $\mathcal{G}$, $E[X|\mathcal{G}] = EX$

To understand 2), 3) and 4), consider two extreme cases

Define random variable S

$S_0(w) = 4$ for all w

$S_{n+1}(w) = 2S_n(w)$ if $w_{n+1} = H$

$S_{n+1}(w) = \frac{1}{2}S_n(w)$ if $w_{n+1} = T$

Then a condition expectation can be defined as

$E[S_n|F_t] = E[S_n|\omega_1, \omega_2, ..., \omega_t]$

If t=n, then $E[S_n|F_n] = S_n$, this is because when F_n is known, then S_n is known, there is nothing to average. This corresponds to Property 2) and 3)

If t=0, then $E[S_n|F_0] = E[S_n]$, this is because F_0 provides no restriction to average S_n, the conditional expectation needs to average all possible cases, it is a general expectation. This corresponds to Property 4).

5) If G is a subset of H
$E[E[X|G|H]] = E[X|H]$

6 Law of Large Numbers

a. Weak law of large number
Suppose X_1, X_2,..., X_n are iid, and u is the expectation.
$lim_{n\to\infty} Pr(|\bar{X} - u| > >\epsilon) = 0$

b. Strong law of large number
$Pr(lim_{n\to\infty} \bar{X} = u) = 1$

c. Difference
In weak case, $|X - u| > \epsilon$ can happen infinite times, however, in strong case, it does not. There exist in certain case where X_n converges in weak case but does not converge in strong case. An example would be a series of X_n that is conditionally convergent, which means the series does not converge absolutely, and by rearranging terms, the series converges to a different value. For example, if X be random variable following geometric distribution with probability 0.5. Then the expectation of a new random variable $2^X(-1)^X X^{-1}$ is

$$\begin{aligned} E[2^X(-1)^X X^{-1}] = \sum_1^\infty \frac{(-1)^x}{x} \\ = -1 + \frac{1}{2} - \frac{1}{3}... \\ = -ln2 \end{aligned}$$

By rearranging the terms,

$$\begin{aligned} &-1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} \\ =&(-1 + \frac{1}{2}) + \frac{1}{4} + (-\frac{1}{3} + \frac{1}{6}) + \frac{1}{8} \\ =& -\frac{1}{2}ln2 \end{aligned}$$

Therefore, this is conditionally convergent, meaning it satisfies the weak law not the strong law.