Conditional Distribution and Conditional Expectation

November 16, 2020

1. Conditional Expectation of Discrete RVs

2. Conditional Expectation of Continuous RVs

3. Information and Conditioning

- given an event A

$$p_{X|A}(x) = \frac{P((X = x) \cap A)}{P(A)}$$

- given a RV

$$p_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$E(X|A) = \sum_x x p_{X|A}(x)$$

For any function g

$$E(g(X)|A) = \sum_x g(x) p_{X|A}(x)$$

$$E(X|A) = \sum_x x p_{X|A}(x)$$

For any function g

$$E(g(X)|A) = \sum_x g(x) p_{X|A}(x)$$

$$E(X|Y = y) = \sum_x x p_{X|Y}(x|y)$$

For any function g

$$E(g(X)|Y = y) = \sum_x g(x) p_{X|Y}(x|y)$$

Conditional Expectation Given $Y = y$

$$E(X|Y = y) = \sum_x x p_{X|Y}(x|y)$$

For any function g

$$E(g(X)|Y = y) = \sum_x g(x) p_{X|Y}(x|y)$$

Consider binomial asset pricing model

- $S_0 = 4, u = 2, d = 1/2$
- $p(H) = 1/3, p(T) = 2/3$

Find

1. The conditional expectation of S_2 given that $S_1 = 8$.
2. The conditional expectation of S_3 given that $S_1 = 8$

1. pmf of S_2 given $S_1 = 8$

$x \mid S_1 = 8$	2	16
p	$1/3$	$2/3$

The conditional expectation of S_2 given that $S_1 = 8$

$$E(S_2 \mid S_1 = 8) = 2 \times \frac{1}{3} + 16 \times \frac{2}{3} = \frac{34}{3}$$

2. Do it by yourself

1. pmf of S_2 given $S_1 = 8$

$x \mid S_1 = 8$	2	16
p	$1/3$	$2/3$

The conditional expectation of S_2 given that $S_1 = 8$

$$E(S_2 \mid S_1 = 8) = 2 \times \frac{1}{3} + 16 \times \frac{2}{3} = \frac{34}{3}$$

2. Do it by yourself

1. $A_1, A_2, \ldots, A_n$ be a partition of the sample space

$$E(X) = E(X|A_1)P(A_1) + E(X|A_2)P(A_2) + \cdots + E(X|A_n)P(A_n)$$

For any set B such that $P(BA_i) > 0$ for all i

$$E(X|B) = E(X|BA_1)P(A_1|B) + E(X|BA_2)P(A_2|B) + \cdots + E(X|BA_n)P(A_n|B)$$

2. $$E(X) = \sum_i E(X|Y = y_i)P(Y = y_i)$$

1. $A_1, A_2, \ldots, A_n$ be a partition of the sample space

$$E(X) = E(X|A_1)P(A_1) + E(X|A_2)P(A_2) + \cdots + E(X|A_n)P(A_n)$$

For any set B such that $P(BA_i) > 0$ for all i

$$E(X|B) = E(X|BA_1)P(A_1|B) + E(X|BA_2)P(A_2|B) + \cdots + E(X|BA_n)P(A_n|B)$$

2. $$E(X) = \sum_i E(X|Y = y_i)P(Y = y_i)$$

1. - By total rule
$$p_X(x) = \sum_{i=1}^{n} P(X = x|A_i)P(A_i) = \sum_{i=1}^{n} P_{X|A_i}(x)P(A_i)$$
   -
$$E(X) = \sum_{x} x p_X(x) = \sum_{x} x \sum_{i=1}^{n} P_{X|A_i}(x)P(A_i) = \sum_{i=1}^{n} \sum xP_{X|A_i}(x)P(A_i) = \sum_{i=1}^{n} E(X|A_i)P(A_i)$$
   - Substituting $p_{X|B}$ by $\sum_i p_{X|BA_i}P(A_i|B)$, we have
$$E(X|B) = \sum_{x} x p_{X|B}(x) = \sum_{x} x \sum_{i=1}^{n} P_{X|BA_i}(x)P(A_i|B) = \sum_{i=1}^{n} \sum xP_{X|BA_i}(x)P(A_i|$$
$$= \sum_{i=1}^{n} E(X|BA_i)P(A_i|B)$$
2. Apply (1) for $A_i = (Y = y_i)$

1. - By total rule

$$p_X(x) = \sum_{i=1}^{n} P(X = x|A_i)P(A_i) = \sum_{i=1}^{n} P_{X|A_i}(x)P(A_i)$$

   -

$$E(X) = \sum_x x p_X(x) = \sum_x x \sum_{i=1}^{n} P_{X|A_i}(x)P(A_i) = \sum_{i=1}^{n} \sum x P_{X|A_i}(x)P(A_i) = \sum_{i=1}^{n} E(X|A_i)P(A_i)$$

   - Substituting $p_{X|B}$ by $\sum_i p_{X|BA_i} P(A_i|B)$, we have

$$E(X|B) = \sum_x x p_{X|B}(x) = \sum_x x \sum_{i=1}^{n} P_{X|BA_i}(x)P(A_i|B) = \sum_{i=1}^{n} \sum x P_{X|BA_i}(x)P(A_i|$$

$$= \sum_{i=1}^{n} E(X|BA_i)P(A_i|B)$$

2. Apply (1) for $A_i = (Y = y_i)$

Messages transmitted by a computer in Boston through a data network are destined for New York with probability 0.5, for Chicago with probability 0.3, and for San Francisco with probability 0.2. The transit time X of a message is random. Its mean is 0.05 seconds if it is destined for New York, 0. 1 seconds if it is destined for Chicago, and 0.3 seconds if it is destined for San Francisco. Then

$$E[X] = 0.5 \times 0.05 + 0.3 \times 0.1 + 0.2 \times 0.3 = 0.115$$

seconds

Explain

- There are 3 cases of destination: A_1: New York, A_2: Chicago, A_3: San Francisco
- Summarize information

	Prob $P(A_i)$	Conditional expected transit time $E(X\|A_i)$
A_1	0.5	0.05
A_2	0.3	0.1
A_3	0.2	0.3
		Overall expected transit time $E(X) = 0.5 \times 0.05 + 0.3 \times 0.1 + 0.2 \times 0.3 = 0.115$

You write a software program over and over, and each time there is probability p that it works correctly, independent of previous attempts.
What is the mean and variance of X, the number of tries until the program works correctly?

- $A_1 = (X = 1) =$ "the first try is successful", $A_2 = (X > 1) =$ "the first try is unsuccessful"
- By total expectation theorem,

$$E(X) = E(X|A_1)P(A_1) + E(X|A_2)P(A_2)$$

- $P(A_1) = P(X = 1) = p$, $P(A_2) = P(X > 1) = 1 - P(X = 1) = 1 - p$
- If the first try is successful then $X = 1$ and

$$E(X|A_1) = E(X|X = 1) = 1$$

- If the first try fails $X > 1$ then we have wasted one try, and we are back where we started (i.e. X has the same distribution as $X + 1$)

$$E(X|A_2) = E(X|X > 1) = 1 + E(X)$$

-

$$E(X) = p + (1 - p)(1 + E(X)) \Rightarrow E(X) = \frac{1}{p}$$

- Similar
$$E(X^2) = E(X^2|A_1)P(A_1) + E(X^2|A_2)P(A_2)$$
- $$E(X^2|A_1) = E(X^2|X = 1) = 1$$
- $$E(X^2|A_2) = E(X^2|X > 1) = (1 + E(X))^2$$
- $$E(X^2) = p + (1-p)(1 + E(X))^2 \Rightarrow E(X^2) = \frac{2}{p^2} - \frac{1}{p}$$
- $Var(X) = E(X^2) - (E(X))^2 = \frac{1-p}{p^2}$

A lost tourist arrives at a point with 3 roads. The first road brings him back to the same point after 1 hours of walk. The second road brings him back to the same point after 6 hours of travel. The last road leads to the city after 2 hours of walk. There are no signs on the roads.
Assuming that the tourist chooses a road equally likely at all times.
What is the mean time until the tourist arrives to the city.

- At each time, there are 3 choices for the tourist: A_1: first road, A_2: second road, A_3: last road
- Denote T: time to go to the city
- Summarize information

	Prob $P(A_i)$	Conditional expected transit time $E(X\|A_i)$
A_1	$1/3$	$1+E(T)$
A_2	$1/3$	$6+E(T)$
A_3	$1/3$	2
		Overall expected transit time $E(T)=\frac{1}{3}(1+E(T)+6+E(T)+2)$

It implies that $E(T)=9$ (hours)

Conditional pdf of X given an event A

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(A)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Conditional expectation of X given an event A

$$E(X|A) = \int_{-\infty}^{\infty} x f_{X|A}(x) dx$$

Conditional pdf of X given an event A

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(A)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Conditional expectation of X given an event A

$$E(X|A) = \int_{-\infty}^{\infty} x f_{X|A}(x) dx$$

Total expectation theorem

$$E(X) = \sum_i E(X|A_i)P(A_i)$$

where $A_1, \ldots, A_n$ is a partition of the sample space

The expected value rule

$$E(g(X)|A) = \int_{-\infty}^{\infty} g(x) f_{X|A}(x) dx$$

Total expectation theorem

$$E(X) = \sum_i E(X|A_i)P(A_i)$$

where $A_1, \ldots, A_n$ is a partition of the sample space

The expected value rule

$$E(g(X)|A) = \int_{-\infty}^{\infty} g(x)f_{X|A}(x)dx$$

X has piecewise constant pdf

$$f(x) = \begin{cases} 1/3 & \text{if } 0 \le x < 1 \\ 2/3 & \text{if } 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

$$A_1 = \{X \text{ lies on } [0,1)\} \qquad A_2 = \{X \text{ lies on } [1,2]\}$$

1. Find $f_{X|A_1}$ and $f_{X|A_2}$
2. Expectation and variance of X

Solution

1. $P(X \in A_1) = \int_0^1 f_X(x)dx = \int_0^1 \frac{1}{3}dx = \frac{1}{3}$. So

$$f_{X|A_1} = \begin{cases} \frac{f_X(x)}{P(X \in A_1)} = \frac{1/3}{1/3} = 1 & \text{for } x \in [0, 1) \\ 0 & \text{otherwise} \end{cases}$$

Similar $P(X \in A_2) = \int_1^2 f_X(x)dx = \int_1^2 \frac{2}{3}dx = \frac{2}{3}$. So

$$f_{X|A_2} = \begin{cases} \frac{f_X(x)}{P(X \in A_1)} = \frac{2/3}{2/3} = 1 & \text{for } x \in [1, 2] \\ 0 & \text{otherwise} \end{cases}$$

Both are uniform RVs on $[0, 1)$ and $[1, 2]$ respectively.

1. $P(X \in A_1) = \int_0^1 f_X(x)dx = \int_0^1 \frac{1}{3}dx = \frac{1}{3}$. So

$$f_{X|A_1} = \begin{cases} \frac{f_X(x)}{P(X \in A_1)} = \frac{1/3}{1/3} = 1 & \text{for } x \in [0,1) \\ 0 & \text{otherwise} \end{cases}$$

Similar $P(X \in A_2) = \int_1^2 f_X(x)dx = \int_1^2 \frac{2}{3}dx = \frac{2}{3}$. So

$$f_{X|A_2} = \begin{cases} \frac{f_X(x)}{P(X \in A_1)} = \frac{2/3}{2/3} = 1 & \text{for } x \in [1,2] \\ 0 & \text{otherwise} \end{cases}$$

Both are uniform RVs on $[0,1)$ and $[1,2]$ respectively.

2. Remind if $U \hookrightarrow U([a,b])$ then $E(U) = \frac{a+b}{2}$ and $E(U^2) = \frac{a^2+ab+b^2}{3}$

$$E(X|A_1) = \frac{1}{2} E(X^2|A_1) = \frac{1}{3}$$
$$E(X|A_2) = \frac{3}{2} E(X^2|A_1) = \frac{7}{3}$$

Hence

$$E(X) = E(X|A_1)P(A_1) + E(X|A_2)P(A_2) = \frac{1}{2}.\frac{1}{3} + \frac{3}{2}.\frac{2}{3} = \frac{7}{6}$$
$$E(X^2) = E(X^2|A_1)P(A_1) + E(X^2|A_2)P(A_2) = \frac{1}{3}.\frac{1}{3} + \frac{7}{3}.\frac{2}{3} = \frac{15}{9}$$

So

$$Var(X) = E(X^2) - (EX)^2 = \frac{15}{9} - \left(\frac{7}{6}\right) = \frac{11}{36}$$

2. Remind if $U \hookrightarrow U([a,b])$ then $E(U) = \frac{a+b}{2}$ and $E(U^2) = \frac{a^2+ab+b^2}{3}$

$$E(X|A_1) = \frac{1}{2} E(X^2|A_1) = \frac{1}{3}$$
$$E(X|A_2) = \frac{3}{2} E(X^2|A_1) = \frac{7}{3}$$

Hence

$$E(X) = E(X|A_1)P(A_1) + E(X|A_2)P(A_2) = \frac{1}{2}.\frac{1}{3} + \frac{3}{2}.\frac{2}{3} = \frac{7}{6}$$
$$E(X^2) = E(X^2|A_1)P(A_1) + E(X^2|A_2)P(A_2) = \frac{1}{3}.\frac{1}{3} + \frac{7}{3}.\frac{2}{3} = \frac{15}{9}$$

So

$$Var(X) = E(X^2) - (EX)^2 = \frac{15}{9} - \left(\frac{7}{6}\right) = \frac{11}{36}$$

The metro train arrives at the station near your home every quarter hour starting at 6:00 AM. You walk into the station every morning between 7:10 and 7:30 AM, with the time in this interval being a uniform random variable. What is the PDF of the time you have to wait for the first train to arrive?

Solution

- X: arrival time, uniform $[7:10, 7:30]$
- Y: waiting time
- $A = \{7:10 \leq X \leq 7:15\} =$ { you board the 7:15 train }
- $B = \{7:15 < X \leq 7:30\} =$ { you board the 7:30 train }

$$f_Y(y) = f_{Y|A}(y)P(A) + f_{Y|B}P(B) = \frac{1}{5}.\frac{1}{4} + \frac{1}{15}.\frac{3}{4} = \frac{1}{10}, \text{ for } 0 \leq y \leq 5$$

and

$$f_Y(y) = f_{Y|A}(y)P(A) + f_{Y|B}P(B) = 0.\frac{1}{4} + \frac{1}{15}.\frac{3}{4} = \frac{1}{20}, \text{ for } 5 < y \leq 15$$

Conditional expectation of X given $Y = y$

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

Total expectation theorem

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} E(X|Y=y) f_Y(y)dx & \text{if } Y \text{ is a continuous RV} \\ \sum_y E(X|Y=y)P(Y=y) & \text{if } Y \text{ is a discrete RV} \end{cases}$$

The expected value rule

$$E(g(X)|Y=y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx$$

and

$$E(h(X,Y)|Y=y) = \int_{-\infty}^{\infty} h(x,y) f_{X|Y}(x|y)dx$$

Conditional expectation of X given $Y = y$

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

Total expectation theorem

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y) dx & \text{if } Y \text{ is a continuous RV} \\ \sum_y E(X|Y = y) P(Y = y) & \text{if } Y \text{ is a discrete RV} \end{cases}$$

The expected value rule

$$E(g(X)|Y = y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx$$

and

$$E(h(X, Y)|Y = y) = \int_{-\infty}^{\infty} h(x, y) f_{X|Y}(x|y)dx$$

Example

Let X_1 and X_2 be jointly normal random variables with parameters μ_1, σ_1, μ_2, σ_2, and ρ. The joint pdf is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{1-\rho^2}\left(\frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x_2-\mu_2)^2}{2\sigma_2^2} - 2\rho\frac{(x_1-\sigma_1)(x_2-\sigma_2)}{\sigma_1\sigma_2}\right)}$$

X_1 is a normal distribution $\mathcal{N}(\mu_1, \sigma_1^2)$ with pdf

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}$$

Condition pdf of X_2 given $X_1 = x_1$ is

$$f_{X_2|X_1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)} = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{\frac{\left(x-\left(\mu_2+\sigma_2\frac{x_1-\mu_1}{\sigma_1}\right)\right)^2}{2\sigma_2^2(1-\rho^2)}}$$

Given $X_1 = x_1$,

$$X_2|(X_1 = x_1) \sim \mathcal{N}\left(\mu_2 + \sigma_2 \frac{x_1 - \mu_1}{\sigma_1}, \sigma_2^2(1 - \rho^2)\right)$$

So $E(X_2|X_1 = x_1) = \mu_2 + \sigma_2 \frac{x_1 - \mu_1}{\sigma_1}$

Another way to find conditional distribution of X_2 given $X_1 = x_1$

Using the construction

$$\begin{cases} X_1 = \mu_1 + \sigma_1 Z_1 \\ X_2 = \mu_2 + \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2) \end{cases}$$

Given $X_1 = x_1$, we have

$$Z_1 = \frac{x_1 - \mu_1}{\sigma_1}$$

Then given $X_1 = x_1$

$$X_2 = \mu_2 + \sigma_2 \left(\rho \frac{x_1 - \mu_1}{\sigma_1} + \sqrt{1-\rho^2} Z_2 \right) = \mu_2 + \sigma_2 \frac{x_1 - \mu_1}{\sigma_1} + \sigma_2 \sqrt{1-\rho^2} Z_2$$

Hence

$$E_{X_2|X_1}(x_2|x_1) = E\left(\mu_2 + \sigma_2 \left(\rho \frac{x_1 - \mu_1}{\sigma_1} + \sqrt{1-\rho^2} Z_2 \right) \right) = \mu_2 + \sigma_2 \frac{x_1 - \mu_1}{\sigma_1}$$

Example - The Expectation of the Sum of a Random Number of Random Variables

Suppose that the expected number of accidents per week at an industrial plant is four. Suppose also that the numbers of workers injured in each accident are independent random variables with a common mean of 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of injuries during a week?

- N:the number of accidents then $E(N) = 4$
- X_i: the number injured in the ith accident, $i = 1, 2, ...,$ then $E(X-i) = 2$
- the total number of injuries can be expressed as

$$Z = \sum_{i=1}^{N} X_i$$

and

$$E(Z) = \sum_{n} E(Z|N = n)P(N = n)$$

- $$E(Z|N = n) = E\left(\sum_{i=1}^{N} X_i|N = n\right) = E\left(\underbrace{\sum_{i=1}^{n} X_i}_{\text{independent of } N} \quad |N = n\right) = E\left(\sum_{i=1}^{n} X_i\right) = 4n$$

$$E(Z) = \sum_{n} 4nP(N = n) = 4\sum_{n} nP(N = n) = 4E(N) = 4 \times 2 = 8$$

- $X_1, X_2, \ldots$ are i.i.d with same distribution as X
- N: discrete RV taking value in $0, 1, 2, \ldots$ and independent of X_i

then

$$E\left(\sum_{i=1}^{N} X_i\right) = E(X)E(N)$$

In the binomial model, Ω: set of all possible outcomesof three coin tosses

- $\mathcal{F}_0 = \{\emptyset, \Omega\}$: you are told nothing about the coin tosses
- $\mathcal{F}_1 = \sigma(A_T, A_H)$: you are told the result of the first toss
- $\mathcal{F}_2 = \sigma(A_{TT}, A_{TH}, A_{HT}, A_{HH})$: you are told the first two tosses
- $\mathcal{F}_3$ = the set of all subset of Ω: you are told all three coin tosses
- $\sigma(S_2) = \sigma(S_2 \leq x)$: knowledge from price information at period 2 or information learnt from S_2

Definition

Let X be a RV defined on a nonempty sample space Ω and $\mathcal{G}$ be a σ - algebra of subsets of Ω. If $\sigma(X) \subset \mathcal{G}$ then we say that X is $\mathcal{G}$ - measurable

Interpretation

- X is $\mathcal{G}$ - measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X
- If X is $\mathcal{G}$ - measurable then the information in $\mathcal{G}$ is also sufficient to determine the value of $f(X)$ for any Borel measurable function f. In other words, $f(X)$ is also $\mathcal{G}$ - measurable

Definition

Let X be a RV defined on a nonempty sample space Ω and $\mathcal{G}$ be a σ - algebra of subsets of Ω. If $\sigma(X) \subset \mathcal{G}$ then we say that X is $\mathcal{G}$ - measurable

Interpretation

- X is $\mathcal{G}$ - measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X
- If X is $\mathcal{G}$ - measurable then the information in $\mathcal{G}$ is also sufficient to determine the value of $f(X)$ for any Borel measurable function f. In other words, $f(X)$ is also $\mathcal{G}$ - measurable

Definition

Let X be a RV defined on a nonempty sample space Ω and $\mathcal{G}$ be a σ - algebra of subsets of Ω. If $\sigma(X) \subset \mathcal{G}$ then we say that X is $\mathcal{G}$ - measurable

Interpretation

- X is $\mathcal{G}$ - measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X
- If X is $\mathcal{G}$ - measurable then the information in $\mathcal{G}$ is also sufficient to determine the value of $f(X)$ for any Borel measurable function f. In other words, $f(X)$ is also $\mathcal{G}$ - measurable

S_2 is $\mathcal{F}_2$ - measurable but not $\mathcal{F}_1$ measurable.
??? S_2 is $\mathcal{F}_3$: measurable

- Let $\Omega, \mathcal{F}, P$ be a probability space.
- $\mathcal{H}$, $\mathcal{G}$: sub - σ - algebra of $\mathcal{F}$
- X, Y: random variables on $\Omega, \mathcal{F}, P$

1. $\mathcal{H}$ and $\mathcal{G}$ are independent if

$$P(AB) = P(A)P(B) \text{ for all } A \in \mathcal{H}, B \in \mathcal{G}$$

2. X is independent of $\mathcal{G}$ is $\sigma(X)$ and $\mathcal{G}$ are independent
3. X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent

- $\mathcal{G}_3$: information from the 3rd toss $=\sigma(A_{..T}, A_{..H})$ with

$$A_{..T} = \{HHT, HTT, THT, TTT\}$$

  and

$$A_{..H} = \{HHH, HTH, THH, TTH\}$$

- $\mathcal{G}_2$: information from the 2nd toss $=\sigma(A_{.T}, A_{.H})$ with

$$A_{.T} = \{HTH, HTT, TTH, TTT\}$$

  and

$$A_{.H} = \{HHH, HHT, THH, THT\}$$

then

- $\mathcal{G}_3$ and $\mathcal{G}_2$ are independent
- $Y_3 = \frac{S_3}{S_2}$: factor at period 3 is independent of $\mathcal{G}_2$ and $\mathcal{F}_2$

Let $\Omega, \mathcal{F}, P$ be a probability space, $\mathcal{G}$ be a sub - σ - algebra of $\mathcal{F}$ and X be a random variable. The conditional expectation of X given $\mathcal{G}$, denoted $E(X|\mathcal{G})$, is any random variable that satisfies

1. **(Measurability)** $E(X|\mathcal{G})$ is $\mathcal{G}$ - measurable
2. **(Partial averaging)**

$$\int_A E(X|\mathcal{G})dP(w) = \int_A X(w)dP(w), \ \forall A$$

If $\mathcal{G} = \sigma(Y)$ then we general write $E(X|Y)$ rather then $E(X|\sigma(Y))$

- $E(X|\mathcal{G})$ is an estimate of X based on the information in $\mathcal{G}$
- Property (1) guarantees that the value of estimate $E(X|\mathcal{G})$ can be determined from the information in $\mathcal{G}$
- Property (2) ensure that $E(X|\mathcal{G})$ is indeed an estimate of X: partial - average over the "small" set of $\mathcal{G}$ is a good estimator of X

$\mathcal{H} \subset \mathcal{G}$: sub - σ - algebra of $\mathcal{F}$ in the probability space $(\Omega, \mathcal{F}, P)$

1. **(Linearity of conditional expectation)**

$$E(c_1X + c_2Y|\mathcal{G}) = c_1E(X|\mathcal{G}) + c_2E(Y|\mathcal{G})$$

2. **(Taking out what is know)** If X is $\mathcal{G}$ - measurable then

$$E(XY|\mathcal{G}) = XE(Y|\mathcal{G})$$

3. **(Iterated conditioning)**

$$E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H})$$

4. **(Independence)** If X is independent of $\mathcal{G}$ then

$$E(X|\mathcal{G}) = E(X)$$

Denote $E(X|Y) = E(X|\sigma(Y))$ then

$$E(X|Y) = g(Y)$$

where $g(y) = E(X|Y = y)$

Considering a binomial asset pricing model with $S_0 = 4$, $p = 1/3$ and $q = 2/3$.
Compare

$$E(S_2 + S_3|S_1)$$

and

$$E(S_2|S_1) + E(S_3|S_1)$$

- S_1 takes two values 8 and 2
- Given $S_1 = 8$

$$E(S_2|S_1 = 8) = \frac{2}{3}(16) + \frac{1}{3}(4) = 12$$

$$E(S_3|S_1 = 8) = \left(\frac{2}{3}\right)^2 (32) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)(8) + \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)(8) + \left(\frac{1}{3}\right)^2 (2) = 18$$

So

$$E(S_2|S_1 = 8) + E(S_3|S_1 = 8) = 12 + 18 = 30$$

- Given $S_1 = 8$, (S_2, S_3) takes pair values $(16, 32)$, $(16, 8)$, $(4, 8)$, $(4, 2)$. So

$$
\begin{aligned}
E(S_2 + S_3 | S_1 = 8) = {} & \left(\frac{2}{3}\right)^2 (16 + 32) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)(16 + 8) \\
& + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)(4 + 8) + \left(\frac{1}{3}\right)^2 (4 + 2) = 30
\end{aligned}
$$

- Hence $E(S_2 + S_3 | S_1 = 8) = E(S_2 | S_1 = 8) + E(S_3 | S_1 = 8) = 30$
- Similary $E(S_2 + S_3 | S_1 = 2) = E(S_2 | S_1 = 2) + E(S_3 | S_1 = 2) = 7.5$
- Regardless the outcome of S_1, we have

$$
E(S_2 + S_3 | S_1) = E(S_2 | S_1) + E(S_3 | S_1)
$$

Compare

$$E(S_1S_2|S_1)$$

and

$$S_1E(S_2|S_1)$$

Compare

$$E(S_3|S_1)$$

and

$$E(E(S_3|(S_1, S_2))|S_1)$$

Compare

$$E\left(\frac{S_2}{S_1}|S_1\right)$$

and

$$E\left(\frac{S_2}{S_1}\right)$$