An introduction to probability theory

Reuy-Lin Sheu

Department of Mathematics, National Cheng Kung University,
Tainan, Taiwan

September 14th, 2023

Probability Theory is a mathematical model

- The mathematical theory of probability involves with conceptual experiments such as tossing a fair coin for 100 times; throwing three dice and so forth.
- In a real experiment when a coin is tossed, it does not necessarily fall heads or tails; it might roll away or even stand on its edge.
- However, in probability theory, "tossing a coin" is a mathematical model in which we agree that "head" or "tail" are the only possible "outcomes" to be observed. Nothing more!
- Therefore, "outcome" in probability theory is more like a hypothetical term than something real that might possibly happen in an experiment.

A σ-algebra in Probability space

- Let Ω be a sample space consists of certain outcomes $\omega \in \Omega$. A σ-algebra is a collection $\mathcal{G}$ of subsets of Ω such that
  - (i) $\emptyset \in \mathcal{G}$;
  - (ii) If $A \in \mathcal{G}$, then its complement $A^c \in \mathcal{G}$;
  - (iii) Let $\{A_i\}_{i=1}^{\infty}$ is a sequence of sets in $\mathcal{G}$, then $\cup_{i=1}^{\infty} A_i$ also belongs to $\mathcal{G}$.
  - (iii)' The above definitions also imply that, if $\{A_i\}_{i=1}^{\infty}$ is a sequence of sets in $\mathcal{G}$, then $\{A_i^c\}_{i=1}^{\infty}$ is a sequence of sets in $\mathcal{G}$ (by (ii)). By (iii), we know $(\cup_{i=1}^{\infty} A_i^c) \in \mathcal{G}$. Again, by (ii), $(\cup_{i=1}^{\infty} A_i^c)^c = \cap_{i=1}^{\infty} A_i \in \mathcal{G}$.
- In other words, a σ-algebra is a set of "subsets of Ω" such that, it contains $\emptyset$, Ω and it is closed under (a) the complement operation of a set; as well as closed under (b) countable unions; and (c) countable intersections.
- Elements in the σ-algebra are called the "events."
- In general, there are many σ-algebras associated with a sample space Ω, among which the most trivial one is $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Examples of σ-algebras

- Let $\Omega_3 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ be the sample space (background space) for tossing a coin three times.
- The trivial σ-algebra is $\mathcal{F}_0 = \{\emptyset, \Omega_3\}$.
- The next simplest σ-algebra associated with Ω_3 is the σ-algebra $\mathcal{F}_1$ consisting of information of only the first toss.
- Let us define

$$\begin{aligned} A_H &\triangleq \{HHH, HHT, HTH, HTT\} = \{H \text{ on the first toss }\} \\ A_T &\triangleq \{THH, THT, TTH, TTT\} = \{T \text{ on the first toss }\} \end{aligned}$$

- Then, it is easy to see that $\mathcal{F}_1 = \{\emptyset, \Omega_3, A_H, A_T\}$ is a σ-algebra which matters only the first coin toss.

σ-algebra $\mathcal{F}_2$ consisting of information up to the second toss.

- Similarly, we can define

$$
\begin{aligned}
A_{HH} &\triangleq \{HHH, HHT\} = \{HH \text{ on the first two tosses }\} \\
A_{HT} &\triangleq \{HTH, HTT\} = \{HT \text{ on the first two tosses }\} \\
A_{TH} &\triangleq \{THH, THT\} = \{TH \text{ on the first two tosses }\} \\
A_{TT} &\triangleq \{TTH, TTT\} = \{TT \text{ on the first two tosses }\}
\end{aligned}
$$

- Then, the information of the first toss is $A_H = A_{HH} \cup A_{HT}$ and $A_T = A_{TH} \cup A_{TT}$. The information of the second toss is $A_{HH} \cup A_{TH}$ and $A_{HT} \cup A_{TT}$. There are also information about two tosses are the same: $A_{HH} \cup A_{TT}$; two tosses are different: $A_{HT} \cup A_{TH}$; at least one tail in two tosses: A_{HH}^c; at least one head in two tosses: A_{TT}^c; etc.
- All types of information generated up to the second toss form the following σ-algebra:

$$
\begin{aligned}
\mathcal{F}_2 = \{ \; & \emptyset, \Omega, A_{HH}, A_{HT}, A_{TH}, A_{TT}, \\
& A_H, A_T, A_{HH} \cup A_{TH}, A_{HT} \cup A_{TT}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, \\
& A_{HH}^c, A_{HT}^c, A_{TH}^c, A_{TT}^c \}.
\end{aligned}
$$

Complete information arising from all three coin tosses in Ω_3

- Notice that the σ-algebra $\mathcal{F}_2$ can be viewed as the set of all subsets of $\{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}$. That is,

$$\mathcal{F}_2 = 2^{\{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}}.$$

- In this sense, we say that the σ-algebra $\mathcal{F}_2$ is generated by the subsets $\{A_{HH}, A_{HT}, A_{TH}, A_{TT}\} \subset \Omega_3$ (and their complements, countable unions and countable intersections).
- The total number of elements in $\mathcal{F}_2$ is $2^4 = 16$.
- The complete information for all three coin tosses in Ω_3 is the σ-algebra $\mathcal{F}_3$ that contains all subsets of Ω_3. That is,

$$\mathcal{F}_3 = 2^{\Omega_3}, \quad |\mathcal{F}_3| = 2^8 = 256.$$

- Notice that

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3.$$

The Borel σ-algebra and a random variable

- The Borel σ-algebra, or just called Borel-algebra, denoted by $\mathcal{B}(\mathbb{R})$, is the smallest σ-algebra "generated" by all open intervals in $\mathbb{R}$.
- Members in $\mathcal{B}(\mathbb{R})$ are called Borel sets.
- Let us emphasize that, the word "generate" means to take all the complements, countable unions and countable intersections.
- The Borel-algebra thus consists of all types of sets, including all open intervals (a, b) in $\mathbb{R}$; all open half lines $(a, \infty) = \bigcup_{n=1}^{\infty}(a, a+n)$, $(-\infty, a)$; all closed intervals $[a, b]$, closed half lines; half-open half-closed intervals $(a, b]$; and any singleton set $\{a\}$.
- Let Ω be a sample space and $\mathcal{F} = 2^{\Omega}$ be the largest σ-algebra consisting of all the subsets of Ω. A random variable X is a real value function $X: \ (\Omega, \mathcal{F}) \ \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

σ-algebras generated by random variables

- The σ-algebra $\sigma(X)$ generated by X is defined by

$$\sigma(X) = \{X^{-1}(B) : \ B \text{ is a Borel set in } \mathbb{R}\}$$

  where $X^{-1}(B) = \{\omega \in \Omega : \ X(\omega) \in B\}$ is the pre-image of B.

- For example, in the binomial asset pricing model, the stock prices $S_1, \ S_2, \ S_3$ at time $1, 2, 3$, respectively, are random variables on $(\Omega_3, \mathcal{F}_3 = 2^{\Omega_3})$.

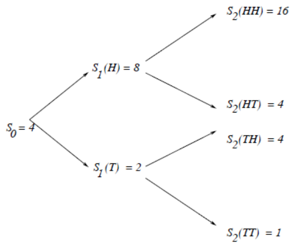
- S_1 takes values only on $\{uS_0, dS_0\}$. S_2 takes values on $\{u^2S_0, udS_0, d^2S_0\}$. S_3 takes values on $\{u^3S_0, u^2dS_0, ud^2S_0, d^3S_0\}$.

- Suppose $S_0 = 4, u = 2, d = 0.5$. Then, S_1 takes values on $\{8, 2\}$. If we take the Borel set as $(-\infty, 0)$, Obviously, $S_1^{-1}((-\infty, 0)) = \emptyset$.

- Moreover, $S_1^{-1}((0, 3)) = A_T = \{THH, THT, TTH, TTT\}$, and $S_1^{-1}((4, 30]) = A_H, \ S_1^{-1}(\{2, 4, 6, 8\}) = \Omega_3$.

- It is not difficult to see that the σ-algebra $\sigma(S_1)$ generated by S_1 is exactly $\mathcal{F}_1$.

σ-algebras generated by random variables



- For S_2, it generates $\sigma(S_2) \subset \mathcal{F}_2 \subset \mathcal{F}_3$

$$\sigma(S_2) = \{\emptyset, \Omega, A_{HH}, A_{TT}, A_{HH}^c, A_{TT}^c, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}\}.$$

$$\begin{aligned}\mathcal{F}_2 = \quad \{ \quad & \emptyset, \Omega, A_{HH}, A_{HT}, A_{TH}, A_{TT}, \\ & A_H, A_T, A_{HH} \cup A_{TH}, A_{HT} \cup A_{TT}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, \\ & A_{HH}^c, A_{HT}^c, A_{TH}^c, A_{TT}^c\}.\end{aligned}$$

- We say that X is $\mathcal{G}$-measurable if $\sigma(X) \subset \mathcal{G}$. Therefore, S_2 is both $\mathcal{F}_2$-measurable and $\mathcal{F}_3$-measurable, but is not $\mathcal{F}_1$-measurable.

Probability measure P on $(\Omega, \mathcal{F})$

- Let Ω be a sample space and the σ-algebra $\mathcal{F}$ contains some partial information of the experiment.
- A probability measure P is a real value function on the σ-algebra $\mathcal{F}$ so that measure $P(A)$ for $A \in \mathcal{F}$ gives a quantitative description for the chance of event A to happen.
- A probability measure P on $(\Omega, \mathcal{F})$ must satisfy
  1. $P(A) \in [0, 1], \ \forall A \in \mathcal{F}$;
  2. $P(\Omega) = 1$;
  3. if $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k).$$

Probability measure P on $(\Omega, \mathcal{F})$

- In most of the occasions, we do not (cannot) specify the probability measure for every event in the sigma algebra.
- Rather, we only specify the probability measure for certain events, but require to compute the probability for others that might interest us.
- For example, in the probability space $(\Omega_3, \mathcal{F}_3)$, we only define $P(H) = \frac{1}{3}, P(T) = \frac{2}{3}$. Then, by assuming the independence[1] among different coin tosses, we can compute

$$\begin{aligned} P(HHH) &= (\frac{1}{3})^3,\ P(HHT) = P(HTH) = P(THH) = (\frac{1}{3})^2(\frac{2}{3}) \\ P(TTT) &= (\frac{2}{3})^3,\ P(TTH) = P(THT) = P(HTT) = (\frac{1}{3})^2(\frac{2}{3}) \end{aligned}$$

- Moreover, we can compute, using definition 3., e.g.,

$$\begin{aligned} P(A_H) &= P(\{HHH, HHT, HTH, HTT\}) \\ &= P(HHH) + P(HHT) + P(HTH) + P(HTT) \\ &= (\frac{1}{3})^3 + 2(\frac{1}{3})^2(\frac{2}{3}) + (\frac{1}{3})(\frac{2}{3})^2 = \frac{1}{3}. \end{aligned}$$

[1] to be elaborated soon

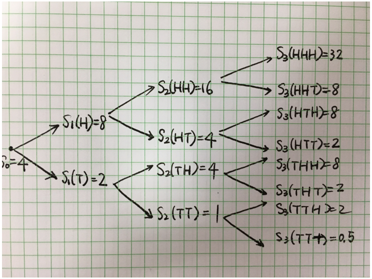
Induced measure on Borel sets by a random variable

- Let $(\Omega, \mathcal{F}, P)$ be probability space and $X: (\Omega, \mathcal{F}, P) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable on Ω.
- We can utilize P on $\mathcal{F}$ to define an induced measure $\mathcal{L}_X$ by X on the Borel algebra $\mathcal{B}(\mathbb{R})$ by

$$\mathcal{L}_X(B) = P(X^{-1}(B)) = P(X \in B).$$

- For example, in the binomial asset pricing model with $S_0 = 4$, $u = 1/d = 2$ and $P(H) = \frac{1}{3}$, we can use the stock price at time 3, S_3, to induce the following measure $\mathcal{L}_{S_3}$ on Borel sets

Induced measure on Borel sets by a random variable

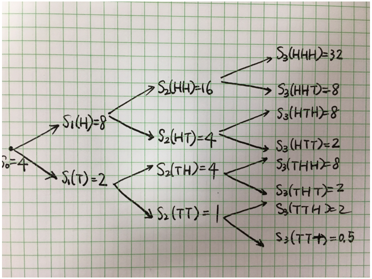


- The induced measure $\mathcal{L}_{S_3}$ evaluates $\emptyset$ as 0, while the whole $\mathbb{R}$ has a total measure of 1 under $\mathcal{L}_{S_3}$:

$$\mathcal{L}_{S_3}(\emptyset) = P(S_3 \in \emptyset) = P(\emptyset) = 0; \; \mathcal{L}_{S_3}(\mathbb{R}) = P(S_3 \in \mathbb{R}) = P(\Omega_3) = 1;$$

- For other Borel sets, $\mathcal{L}_{S_3}$ gives a measure which might be quite different for the usual Borel measure:
  $\mathcal{L}_{S_3}((8,50)) = P(HHH) = (\frac{1}{3})^3$; and
  $\mathcal{L}_{S_3}([8,50)) = P(\{HHH, HHT, HTH, THH\}) = \frac{1}{3}^3 + 3\frac{1}{3}^2 \cdot \frac{2}{3} = \frac{7}{27}$.

Induced measure represented by a cumulative distribution function



- A common way to record the information from the induced measure is to give the *cumulative distribution function* $F_{S_3}(x)$ defined as $F_{S_3}(x) = P(S_3 \in (-\infty, x])$

$$F_{S_3}(x) = P(S_3 \leq x) = \begin{cases} 0, & \text{if } x < 0.5, \\ \frac{8}{27}, & \text{if } 0.5 \leq x < 2, \\ \frac{20}{27}, & \text{if } 2 \leq x < 8, \\ \frac{26}{27}, & \text{if } 8 \leq x < 32, \\ 1, & \text{if } 32 \leq x. \end{cases}$$

Homework Exercise: Prob. Review #1

- The same random variable can have many different distributions. Suppose we set $p = q = \frac{1}{2}$. Then, the same random variable S_3 induces a different measure (from the one induced by $p = \frac{1}{3}$ in the above example) on the Borel sets. Please write down the cumulative distribution function $F_{S_3}(x)$ in this case.

- Two different random variables can have the same distribution. Consider the following two securities:
  (A) An European call with strike price 14 expiring at time 2. The value of this call option at time 2 is a random variable $(S_2 - 14)^+$.
  (B) An European put with strike price 3 expiring at time 2, which allows the buyer to sell the stock at price 3 to the agent. The value of this put option at time 2 is a random variable $(3 - S_2)^+$.

  Show that, with the probability for one toss set at $p = q = \frac{1}{2}$, and $S_0 = 4$, $u = 1/d = 2$, the two options have the same cumulative distribution function.

The End for the section of Probability Theory Review

Thank you for listening!

Any question?