Handout 1 - Conditional Expectation Part 1

Charlie Qu

Sep 23, 2021

1. Basic concepts in probability

Example 1.1. We consider the next international A-level soccer game between Argentina and Brazil. Outcome of Argentina's next game against Brazil can be: win, tie, or defeat.

Definition 1.1. (Sample space). The sample space Ω is the set of all possible outcomes of an experiment.

Example 1.2 In Example 1.1, the sample space is

$$\Omega = \{win, defeat, tie\}$$

Definition 1.2. (Event). An event is any collection of possible outcomes of an experiment, i.e., any subset of sample space Ω.

Example 1.3. In Example 1.1, the following are examples of events.

$$\emptyset, \{win, defeat, tie\}, \{defeat, tie\}, \{defeat\}, \{win, tie\}$$

Remark 1.1.: the power set of Ω is the collection of all possible subsets of the sample space $Omega$.

Definition 1.3. (σ-algebra). A σ-algebra of the sample space Ω, denoted by $\mathcal{F}$, is a collection of subsets of Ω that satisfies the following properties.
• $\Omega \in \mathcal{F}$.
• If $A \in \mathcal{F}$, $A^c \in \mathcal{F}$,where $A^c \equiv \Omega/A$.
• If $A_1, A2, ... \in \mathcal{B}$, then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{B}$, ($\mathcal{B}$ is closed under countable union).

Definition 1.4. (Measurable space). A measurable space is a couple $(\Omega, \mathcal{F})$, where Ω is the sample space and $\mathcal{F}$ is a σ-algebra defined on Ω. A subset $A \subseteq \Omega$ is said to be a measurable set if and only if $A \in \mathcal{F}$.

Example 1.4. The following are possible σ-algebra that can be defined in Example 1.1.

$$\mathcal{F}_1 = \{\emptyset, \{win, defeat, tie\}\}$$

$$\mathcal{F}_2 = \{\emptyset, \{win, defeat, tie\}, \{win\}, \{defeat, tie\}\}$$

$$\mathcal{F}_3 = \{\emptyset, \{win, defeat, tie\}, \{win\}, \{defeat, tie\}, \{defeat\}, \{win, tie\}\}$$

The first example is the smallest possible σ-algebra (also known as the trivial σ-algebra) and the last one is the largest possible σ-algebra(i.e. the power set of Ω, or total Ω). If we combine the sample space with any of the three examples of σ-algebra we form a measurable space.

In plain language, the σ-algebra in a measurable set is composed of all the events in that set which we can measure, i.e., which we can compute probabilities to. Anything event excluded from the σ-algebra will not be measurable, i.e., out of the class of events we can compute probabilities to.

Definition 1.5. (Probability measure). Let $(\Omega, \mathcal{F})$ be a measurable space. The probability measure (or probability function or, simply, probability) is a real function P: $\mathcal{F} \to \mathbb{R}$ that satisfies the following axioms:

• **Axiom 1 (non-negative probability):** For every $E \in \mathcal{F}, P(E) \geq 0$.
• **Axiom 2 (unit total measure):** $P(\Omega) = 1$.
• **Axiom 3:** (σ-additivity): If $\{E_i\}_{i=1}^{\infty}$ is a collection of events in $\mathcal{F}$ that are pair-wise disjoint (i.e. $E_i \cap E_j = \emptyset$

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

Example 1.5. Consider the measurable space $(\Omega, \mathcal{F})$, where

$$\Omega = \{win, defeat, tie\}$$

$$\mathcal{F} = \{\emptyset, \{win\}, \{defeat, tie\}, \{win, defeat, tie\}\}$$

We have already verified that $\mathcal{F}$ is a σ-algebra for Ω. Let the function $\mathcal{F} \to \mathbb{R}$ be such that:

$$P(\{win\}) = 0.8$$

Using the axioms of probability, we can deduce the probability of the remaining measurable events.

$$P(\{win, defeat, tie\}) = 1, P(\emptyset) = 0, P(\{defeat, tie\}) = 0.2$$

Here are some key propostions for a probability measure
Proposition 1.1. $P(\emptyset) = 0$.
Proposition 1.2. $P(\bigcup_{i=1}^{\infty} E_i) = \bigcup_{i=1}^{\infty} P(E_i)$.
Proposition 1.3. For proper $A, B \in \mathcal{F}$.
a. $P(A^c) = 1 - P(A)$.
b. $P(A) \in [0, 1]$.
c. If $A \subseteq B$, then $P(A) \leq P(B)$.
d. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

2. Conditional Probability

The conditional probability is a probability measure under the condition that a certain event have occurred. As soon as we know that some event has occurred, the relevant probability space changes. In consequence with this, the relevant probability measure changes to the conditional probability, which we now define.

Definition 2.1.(Conditoinal Probability). Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $B \in \mathcal{F}$ such that $P(B) > 0$. Then, for all $A \in \mathcal{F}$ the conditional probability of the event A given B, denoted A|B, is defined as:

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

As the name indicates, the conditional probability P (?|B) is a probability function defined in a new measurable space. We now explain this further. Our original probability space was $(\Omega, \mathcal{F}, P)$ and we learn that B has occurred. This piece of information requires us to update the relevant probability space. In particular, the new relevant measurable space becomes $(\mathcal{B}, \mathcal{F}_{\mathcal{B}})$ where
• The new sample space is $\mathcal{B}$.
• The new σ-algebra is $\mathcal{F}_{\mathcal{B}}$ is defined as follows. By definition, $\mathcal{A}_{\mathcal{B}} \in \mathcal{F}_{\mathcal{B}}$ if and only if $\mathcal{A}_{\mathcal{B}} = A \cap B$ for some $A \in \mathcal{F}$.

We confirm this by verifying that P(.|B) is satisfies all the axioms of probability. As required by the definition, we assume that P(B)>0, then
• **Axiom 1 (non-negative probability):** For every $E \in \mathcal{F}_{\mathcal{B}}, P(E|B) = P(E \cap B)/P(B) \geq 0$.
• **Axiom 2 (unit total measure):** $P(B|B) = 1$.

• **Axiom 3** (σ-additivity): If $\{E_i\}_{i=1}^{\infty}$ is a collection of events in $\mathcal{F}$ that are pair-wise disjoint (i.e. $E_i \cap E_j = \emptyset$), then

$$P(\bigcup_{i=1}^{\infty} E_i|B) = \sum_{i=1}^{\infty} P(E_i|B)$$

Definition 2.2. (Independence) Let $(\Omega, \mathcal{F}, P)$ be a probability space. Two measurable events A, B are independent if and only if:

$$P(A|B) = P(A)P(B)$$

Corollary 2.1.. If A and B are independent and P (B) > 0,

$$P(A|B) = P(A)$$

Remark 2.1.: Law of total probability.
Remark 2.2.: Bayes' Theorem.

Definition 2.4. (Measurable function) Let $(\Omega_1, \mathcal{F_\infty})$,$(\Omega_1, \mathcal{F_\in})$ be two measurable spaces. We say that a function $X : \Omega_1 \to \Omega_2$ is $\mathcal{F_\infty}/\mathcal{F_\in}$-measurable(or simply measurable) if and only if:

$$foranyB \in \mathcal{F_\in} \Rightarrow \{\omega \in \Omega_1 : X(\omega) \in B\} \in \mathcal{F_\infty}$$

.

Example 2.1. In the soccer example, recall that $\Omega_1 = \{win, defeat, tie\}$, then $X : \Omega_1 \to \Omega_2$ such that $\Omega_2 = \{3, 0, 1\}$ is an RV. We can check its measurability by its definitoin easily.

Definition 2.5. (RV-Real-valued). Let $(\Omega, \mathcal{F}, P)$ be a probability. A random variable is a function $X : \Omega \to \mathbb{R}$ that is measurable.

3. Characterizing the distribution of a random variable

Definition 3.1. (CDF) The cumulative distribution function (CDF) of a random variable $X : \Omega \to \mathbb{R}$ is given by:

$$F_X(x) = P(X \leq x)$$

Definition 3.2. (PDF) The probability density function (PDF) of a continuous random variable X is the function that satisfies:

$$F_X(x) = \int_{-\infty}^{x} P(t)dt$$

For joint PDF:

$$F_{X,Y}(x, y) = P(X \leq x \cap Y \leq y)$$

Remark 3.1.: the conversion between CDF and PDF.

4. Condiontioinal expectation

Definition 4.1. (Conditional distribution) The conditional distribution {Y|X=x} is the distribution of the random variable Y given that we already know that another random variable, X, is equal to a specific value x.
(1)The conditional density (PMF or PDF) is defined as follows:

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

for any $y \in \mathbb{R}$.

(2) The conditional CDF is given as

$$F_{Y|X}(y|x) = \frac{\partial F_{X,Y}(x,y)/\partial x}{f_X(x)}$$

(3) Whenever the conditional density is defined, the conditional CDF can be computed from it in the usual way

$$F_{Y|X=x}(y) = \int_{-\infty}^{y} f_{Y|X=x}(s)ds$$

Granger causality and predicability. In its most general sense, we might say that Y_{t+h} is predictable using X_t if the conditional distribution of Y_{t+h} given X_t is di?erent from the unconditional distribution of Yt+h: Let
(1) Conditional distribution: $Y_{t+1}|X_t = x \sim F_{Y|X}$.
(2) Unconditional distribution: $Y_{t+1} \sim F_Y$.
Then we say that "X_t Granger-causes Y_{t+1} in distribution" if

$$F_Y(y) \neq F_{Y|X}(y|x)$$

,for some (x,y).

Remark 4.1.: the conditional distribution here is a different from the definitive conditional distribution we talked about before. As for Definition 4.1, in order to have a proper conditional distribution, the joint and marginal distribution both should proper defined first. However, for Granger causality case, the joint they are not required.

Definition 4.2 (Expectation)

$$E(X) = \int_{-\infty}^{\infty} f_X(x)xdx$$

Definition 4.3 (Conditional expectation). Let X, Y be random variables, he conditional expectation of {Y|X=x} is the expectation of the conditional distribution, i.e.,

$$E(Y|X=x) = \int y f_{Y|X=x}(y)dy$$

Remark 4.2.: E(Y|X = x) is a (deterministic) function of x. It is possible to show that this function is measurable as well. **Proposition 4.1.** Let X, Y, Z be random variables, a and b are real numbers, g is a real function. Then we have
(i) $E(a|Y) = a$.
(ii) $E(aX + bZ|Y) = aE(X|Y) + bE(Z|Y)$.
(iii) E(X|Y)=E(X) if and only if X and Y are independent.
(iv) $E(Xg(Y)|Y) = g(Y)E(X|Y)$. In Particular, $E(g(Y)|Y) = g(Y)$.
(v) $E(X|Y, g(Y)) = E(X|Y)$.

Theorem 4.1 (Law of iterated expectations). Let X, Y be random variables, then,

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

Furthermore, $E(E(Y|X,Z)|X) = E(Y|X)$.

Theorem 4.1 (Law of iterated expectations). Let I_{t-1}, I_t be the information set of a previous period and current period, so that $I_{t-1} \subseteq I_t$, the LIE can be revised as

$$E(E(Y_{t+1}|I_t)|I_{t-1}) = E(Y|I_{t-1})$$

Remark 4.3.: This can be understood as the prediction of a dumb person on the prediction of a smart person, is not different from his own pure prediction.

Definition 4.4 (Mean Independence).Let X, Y be random variables Then, Y is mean independent of X if and only if:

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

Proposition 4.2. (Independence ⇒ mean independence) Let X, Y be random variables. If Y is independent of X and E(Y) exists, Y is mean independent of X. The reverse implication is not true.

Proposition 4.3. (Independence ⇒ both mean independence relationships). If Y is independent of X then:
(a) if E(Y) exists, Y is mean independent of X and
(b) if E(X) exists, X is mean independent of Y.