Conditional Probability and Independence

Reuy-Lin Sheu

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

November, 2023

Conditional Probability

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

- Suppose a random experiment is conducted, and $\omega \in \Omega$ is the outcome. Let $A,\ B \in \mathcal{F}$ be two events. The probability that $\omega \in A$ is $P(A)$.
- Suppose ω is not known to you, but you are told a partial information that $\omega \in B$. Conditional on the information that $\omega \in B$, the probability that $\omega \in A$ becomes

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)} \in [0, 1].$$

- In general, the information whether B has just occurred may well change our way of betting on the event A. So, $P(A|B) \neq P(A)$.
- When $P(A|B) = P(A)$, in which case we say A is independent of B. The additional information from learning that B happened does not permit any inference about the probability for the occurrence of A. That is, A is independent of B *iff*

$$P(A \cap B) = P(A)P(B).$$

Independence

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

- Notice that, whether two sets are independent depends on the probability measure P.
- For example, in the two-stage binomial asset pricing model, recall that the probability $p \in [0,1]$ for a single toss H and $q = 1 - p$ for T with which we define the probability P in $(\Omega, \mathcal{F}, P)$ as
$$P(HH) = p^2, P(HT) = pq, p(TH) = pq, P(TT) = q^2.$$
- Let $A = \{HH, HT\}$ be the event that the stock price moves up in the first trading, whereas $B = \{HT, TH\}$ be the event that "one up and one down" with $P(B) = 2pq$.
- Since $P(A)P(B) = 2p^2q$ and $P(A \cap B) = P(HT) = pq$, the two events A and B are independent if and only if $2p^2q = pq$. That is, $p = 0.5 \vee p = 1(q = 0) \vee p = 0$.
- When $p = 0.5$, the probability of B that "one up and one down" is 0.5. Suppose you are told that in the first day the stock price actually went up (i.e. Event A happened), the probability of $P(B|A)$, which is now the probability of "down" in the second day, is still 0.5. So, events A and B are independent.

Independence

Let $(\Omega, \mathcal{F}, P)$ be a probability space and

$$P(A) = p, P(B) = 2pq, P(A)P(B) = 2p^2q, P(A \cap B) = pq.$$

- Suppose now the probability measure P changes to $p = 0.01$. The chance for the stock price to go up is doomed. It is very likely to end up with "two downs" in a two-days trading section.
- In this case, the probability of B, one up and one down, is still quite small. It is $P(B) = 2pq = 2 \cdot 0.01 \cdot 0.99 = 0.0198$.
- However, if you are told a surprise that in the first day the stock price indeed went up, now you know $B = \{HT, TH\}$ (one up and one down) becomes very highly probable. In fact, $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{pq}{p} = q = 0.99$.
- With the probability measure specified by $p = 0.01$, not only do we see that $P(B|A) \neq P(B)$, but also $P(B|A) \gg P(B)$ that knowing A greatly change the way we think of the probability of B to happen.

Independence

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

- We first observe that, when $P(A) = 0$, then A is independent of any $B \in \mathcal{F}$ since $B \cap A \subset A \Rightarrow 0 \leq P(B \cap A) \leq P(A) = 0$. Therefore, $P(A \cap B) = 0 = P(A)P(B)$.
- Now assume that $P(A) \neq 0, P(B) \neq 0$. It is obvious that, if A is independent of B, then B is also independent of A. That is,

$$\begin{aligned} P(A|B) = P(A) \quad &\Leftrightarrow \quad \frac{P(A \cap B)}{P(B)} = P(A) \\ &\Leftrightarrow \quad \frac{P(A \cap B)}{P(A)} = P(B) \Leftrightarrow P(B|A) = P(B) \end{aligned}$$

- Moreover, if A and B are independent, from

$$P(A|B^c) = \frac{P(AB^C)}{P(B^c)} = \frac{\overbrace{1 - P(A^c)}^{P(A)} - \overbrace{P(AB)}^{P(A)P(B)}}{1 - P(B)} = P(A),$$

  we know A is independent of B^c.
- Immediately, if A and B are independent, B is also independent of A^c and also there is B^c independent of A^c.

Independence

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

- Moreover, any event $A \in \mathcal{F}$ must be independent of Ω and also of $\Omega^c = \emptyset$.
- This is because $P(A \cap \Omega) = P(A) = P(A)\underbrace{P(\Omega)}_{=1}$ is always true.
- In other words, suppose that A and B are independent, any pair of events $(A', B'),\ A' \in \sigma(A),\ B' \in \sigma(B)$ is also independent.
- Let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. We say that $\mathcal{G}$ and $\mathcal{H}$ are *independent* if every set in $\mathcal{G}$ is independent of every set in $\mathcal{H}$. That is, $P(A \cap B) = P(A)P(B),\ \ \forall A \in \mathcal{G},\ B \in \mathcal{H}$.
- Therefore, two events $A, B \in \mathcal{F}$ are independent if and only if the two sub-sigma-algebras $\sigma(A)$ and $\sigma(B)$ are independent.
- Notice that, for an event $A \in \mathcal{F}$, it must not be independent of its own complement A^c since $P(A \cap A^c) = P(\emptyset) = 0 \neq P(A)P(A^c)$, unless $P(A) = 0 \vee 1$.

Independence of σ-algebras (Exercise)

- Toss a coin twice with $\Omega_2 = \{HH, HT, TH, TT\}$ and $\mathcal{F}_2$ (with $|\mathcal{F}_2| = 16$) is the σ-algebra consisting of all information up to time 2.
- Let the probability of tossing a H in the first trial is $p_1 \in (0, 1)$ and that for the second trial is $p_2 \in (0, 1)$. Define the probability P as

$$P(HH) = p_1 p_2, P(HT) = p_1(1-p_2), p(TH) = (1-p_1)p_2, P(TT) = (1-p_1)(1-p_2).$$

- Then, $(\Omega_2, \mathcal{F}_2, P)$ is a probability space.
- Let $\mathcal{G} = \{\emptyset, \Omega_2, \{HH, HT\}, \{TH, TT\}\} \subset \mathcal{F}_2$ be the sub-σ-algebra consisting of information of the first toss; while $\mathcal{H} = \{\emptyset, \Omega_2, \{HH, TH\}, \{HT, TT\}\} \subset \mathcal{F}_2$ be the sub-σ-algebra consisting of information of the second toss.
- Show that the two σ-algebras $\mathcal{G}, \mathcal{H}$ are independent under the specified probability.
- If we change the probability P of $(\Omega_2, \mathcal{F}_2, P)$ to become $P(HH) = \frac{1}{9}, P(HT) = \frac{2}{9}, p(TH) = \frac{1}{3}, P(TT) =??$. Show that, the two sub-σ-algebras $\mathcal{G}, \mathcal{H}$ become dependent.

Independence of random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

- We say that two random variables X and Y are *independent* if the σ-algebras they generate, $\sigma(X)$ and $\sigma(Y)$, are independent.
- For example, in the probability space of the binomial asset pricing model with three independent tosses of the same coin, the two random variables S_2 and $\frac{S_3}{S_2}$ are independent.
- The σ-algebra generated by S_2 is $\mathcal{F}_2$

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

- On the other hand, the random variable S_3 takes on values uS_2 if the third toss is H; and dS_2 otherwise. So, $\frac{S_3}{S_2}$ takes only two values u, d. Then,

$$\sigma(\frac{S_3}{S_2}) = \{\emptyset, \Omega_3, A_{..H}, A_{..T}\}$$

- It is easy to see the independence by the following expression:

$$P(A_{HH} \cap A_{..H}) = P(\{HHH\}) = p^3 = \underbrace{P(A_{HH})}_{=p^2} \underbrace{P(A_{..H})}_{=p}$$

Independence of random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

- Suppose X and Y are independent random variables. Let g and h are Borel measurable functions from $(\mathbb{R}, \mathcal{B})$ to $(\mathbb{R}, \mathcal{B})$. Then, $g(X)$ and $h(Y)$ are also independent random variables.
- This is so because $\sigma(g(X)) \subset \sigma(X)$ and $\sigma(h(Y)) \subset \sigma(Y)$.
- For example, if X and Y are independent random variables, then X^2 and e^Y are also independent to each other.
- Information gets suppressed after operations and becomes less informative. For example, suppose $X(\omega) = 1$ if $\omega = H$ and $X = -1$ if $\omega = T$. Then, $X^2(\omega) = 1, \forall \omega \in \{H, T\}$.
- In this example, X^2 provides less information than X does:

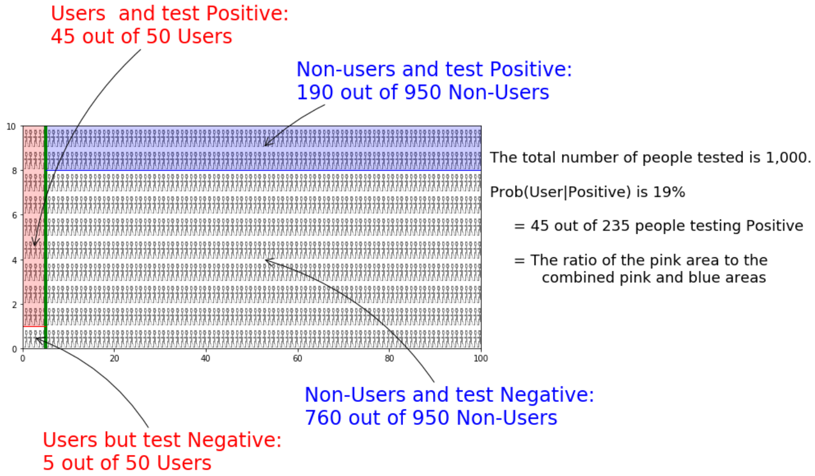
$$\sigma(X^2) = \{\emptyset, \Omega\} \subset \sigma(X) = \{\emptyset, \Omega, \{H\}, \{T\}\}$$

Formulae based on conditional probability - Bayes' theorem

- Conditional probabilities are sometimes given, or can be easily determined especially in sequential random experiments.
- For example, in Drug testing, a particular test for whether someone has been using cannabis is 90% sensitive, meaning the true positive rate (TPR)$= 0.90 = \frac{45}{50}$. It means that the probability of a positive test result, conditioned on the individual truly being positive, is 0.9.
- In addition, the drug test is also 80% specific, meaning true negative rate (TNR) $= 0.80 = \frac{760}{950}$. Therefore, the test correctly identifies 80% of non-use for non-users. The probability of a negative test result, conditioned on the individual truly being negative, is 80%.
- The prevalence, the probability of a random person who uses cannabis, is $5\% = \frac{50}{1000}$.
- We would like to know the probability that a random person who tests positive is really a cannabis user? That is, we would like to know the conditional probability $P(\mathrm{User}|\mathrm{Positve})$.

Formulae based on conditional probability - Bayes' theorem

- There are $45 + 190 = 235$ persons who are tested positive; among which 45 persons are true users. Then, $P(\mathrm{User}|\mathrm{Positve}) = \frac{45}{235} \approx 19.15\%$.



Formulae based on conditional probability - Bayes' theorem

- Bayes' theorem is stated mathematically as the following equation:

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

- In the above example, we have

$$P(\mathrm{User}|\mathrm{Positve}) = \frac{\overbrace{P(\mathrm{Positve}|\mathrm{User})}^{\mathrm{TPR}=0.9} \quad \overbrace{P(\mathrm{User})}^{\mathrm{Prevalence}=0.05}}{P(\mathrm{Positve})} = \frac{0.045}{P(\mathrm{Positve})}.$$

- As for those tested positive, they may be contributed from either users or non-users. That is,

$$\begin{aligned} P(\mathrm{Positve}) &= P(\mathrm{Positve} \wedge \mathrm{User}) + P(\mathrm{Positve} \wedge \mathrm{NonUser}) \\ &= P(\mathrm{Positve}|\mathrm{User})P(\mathrm{User}) + P(\mathrm{Positve}|\mathrm{NonUser})P(\mathrm{NonUser}) \\ &= 0.9 \cdot 0.05 + (1 - \mathrm{TNR} = 0.2) \cdot 0.95 = 0.235 \end{aligned}$$

- Therefore, $P(\mathrm{User}|\mathrm{Positve}) = \frac{0.045}{0.235} \approx 0.1915$, which is much higher than the prevalence rate 0.05.

Formulae based on conditional probability- Bayes' theorem (Theorem 4.2 at page 28 in the textbook)

- In general, we have the following more general form of Bayes' theorem: Let $(\Omega, \mathcal{F}, P)$ be a probability space. Assume that $F_1, F_2, \ldots, F_n$ are pairwise disjoint events in $\mathcal{F}$ and that $F_1 \cup F_2 \cup \ldots \cup F_n = \Omega$. Each event F_j, $j = 1, 2, \ldots, n$ is often known as a hypothesis, while $P(F_j)$ its prior probability.
- In the above drug testing example, F_1 is the hypothesis set of drug users with the prevalence 0.5 as its prior probability, while F_2 is the complement of hypothesis F_1.
- Let $A \in \mathcal{F}$ be an evidence event. Then, the conditional probability $P(F_j|A)$ is the posterior probability of the hypothesis F_j given the evidence A.
- In the example, A is the event of people who are tested positive. Given the evidence that a person is tested positive, the posterior probability that the person is indeed a cannabis user can be formulated as $P(F_1|A)$.

Formulae based on conditional probability- Bayes' theorem (Theorem 4.2 at page 28 in the textbook)

- Bayes' theorem says that, the posterior (conditional) probability of the hypothesis F_j to be true, given the evidence A can be computed through the set of (assumed) prior probability $P(F_i), i = 1, 2, \ldots, n$ on all the hypotheses; and the conditional probabilities for the evidence A to happen, given the various hypotheses.
- Namely,

$$P(F_j|A) = \frac{P(F_j \cap A)}{P(A)} = \frac{P(A|F_j)P(F_j)}{P(A|F_1)P(F_1) + \cdots + P(A|F_n)P(F_n)}$$

- In the denominator of the formula, since $\Omega = F_1 \cup F_2 \cup \ldots \cup F_n$ is partitioned into various hypothesis, we have

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P\left(A \cap (\cup_{j=1}^{n} F_j)\right) \\ &= P\left(\cup_{j=1}^{n} (A \cap F_j)\right) \\ &= P(A \cap F_1) + P(A \cap F_2) + \cdots + P(A \cap F_n) \\ &= P(A|F_1)P(F_1) + P(A|F_2)P(F_2) + \cdots + P(A|F_n)P(F_n) \end{aligned}$$

Joint distribution (general setting)

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X and Y be two random variables on it. The induced measure by X and Y are $\mathcal{L}_X$ and $\mathcal{L}_Y$ respectively.

- Consider the random vector

$$(X, Y) : (\Omega, \mathcal{F}, P) \to (\mathbb{R} \times \mathbb{R}, \sigma(\mathcal{B} \times \mathcal{B})).$$

  The induced measure by the pair of random vector is defined by

$$\mathcal{L}_{(X,Y)}(C) = P\{\omega \in \Omega |\ (X(\omega), Y(\omega)) \in C\},\ \forall C \in \sigma(\mathcal{B} \times \mathcal{B}).$$

- In particular, we should consider a "rectangular" type of C in $\mathbb{R}^2$ that $C = A \times B,\ A, B \in \mathcal{B}$. In this case,

$$\{\omega \in \Omega | (X(\omega), Y(\omega)) \in A \times B\} = \{\omega | X(\omega) \in A\} \bigcap \{\omega | Y(\omega) \in B\}.$$

- Not every $2D$ borel set $\sigma(\mathcal{B} \times \mathcal{B})$ is rectangular. For example, the union of two rectangles is usually not a rectangle.

Joint distribution (discrete case) (page 70 in the textbook)

Suppose X and Y be two discrete r.v.'s on $(\Omega, \mathcal{F}, P)$ with discrete range spaces $\mathcal{R}(X)$ and $\mathcal{R}(Y)$ respectively.

- The random vector $(X, Y) : (\Omega, \mathcal{F}, P) \to (\mathbb{R} \times \mathbb{R}, \sigma(\mathcal{B} \times \mathcal{B}))$ has the joint probability mass function defined on $\mathcal{R}(X) \times \mathcal{R}(Y)$, denoted by, $\forall (x, y) \in \mathcal{R}(X) \times \mathcal{R}(Y)$,

$$\begin{aligned} p_{(X,Y)}(x, y) &= P(X = x, Y = y) \\ &= P(\omega \in \Omega | X(\omega) = x, Y(\omega) = y) = \mathcal{L}_{(X,Y)}(\{x\} \times \{y\}) \end{aligned}$$

  where $C = \{x\} \times \{y\}$ is the most important special "rectangle" type of Borel sets in the discrete case.

- Let $A \in \sigma(\mathcal{B} \times \mathcal{B})$ be a joint Borel set. Then,

$$P((X, Y) \in A) = P(\omega \in \Omega | (X(\omega), Y(\omega)) \in A) = \sum_{(x,y) \in A} p_{(X,Y)}(x, y)$$

- The *marginal probability functions* of the r. vector (X, Y) are

$$P(X = x) = \sum_{y \in \mathcal{R}(Y)} P(X = x, Y = y); \ P(Y = y) = \sum_{x \in \mathcal{R}(X)} P(X = x, Y = y)$$

Joint distribution (discrete case) (Example 7.1 at page 70 in the textbook)

An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random and let X be the number of red balls; and Y be the number of white balls.

- We first notice that the sample space Ω contain all possible $C_3^{10} = 120$ selection of 3 balls from a basket of 10 balls.
- The r.v. X sends every selection $\omega \in \Omega$ to the number of red balls. It is easy to see that the range space of X is $\mathcal{R}(X) = \{0, 1, 2\}$.
- As for the r.v. Y, $Y(\omega)$ is the number of white balls in a selection of 3 balls $\omega \in \Omega$. Thus, $\mathcal{R}(X) = \{0, 1, 2, 3\}$.
- The random vector (X, Y) sends an $\omega \in \Omega$ to a point in $\mathbb{R}^2$. Specifically, $(X, Y)(\omega) = (X(\omega), Y(\omega)) \in \{0, 1, 2\} \times \{0, 1, 2, 3\} \subset \mathbb{R}^2$
- The joint probability mass function is defined on $\forall (x, y) \in \{0, 1, 2\} \times \{0, 1, 2, 3\}$ for the probability that both $X = x$ and $Y = y$ to happen simultaneously.

Joint distribution (discrete case) (Example 7.1 at page 70 in the textbook)

An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random and let X be the number of red balls; and Y be the number of white balls.

- For example, $(0,0) \in \{0,1,2\} \times \{0,1,2,3\}$. The joint probability for both $X=0$ and $Y=0$ to happen simultaneously is

$$p_{(X,Y)}(0,0) = P(\omega \in \Omega | X(\omega) = 0, Y(\omega) = 0) = \mathcal{L}_{(X,Y)}(\{0\} \times \{0\}) = \frac{1}{120}$$

  since there is only one possible choice: green, green, green that has no red nor white balls.
- Another example: $(1,0) \in \{0,1,2\} \times \{0,1,2,3\}$. Then,

$$p_{(X,Y)}(1,0) = P(\omega \in \Omega | X(\omega) = 1, Y(\omega) = 0) = \mathcal{L}_{(X,Y)}(\{1\} \times \{0\}) = \frac{2 \cdot 3}{120}$$

- For $(X,Y) = (2,0)$, $p_{(X,Y)}(2,0) = \frac{3}{120}$.
- The marginal probability $P(Y=0)$, the probability that a 3-balls selection contains no white ball is

$$P(Y=0) = P(X=0, Y=0) + P(X=1, Y=0) + P(X=2, Y=0) = \frac{10}{120}.$$

Joint distribution (discrete case) (Example 7.1 at page 70 in the textbook)

An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random and let X be the number of red balls; and Y be the number of white balls.

- The following table shows that joint probability mass function and two marginal probability mass functions.

$y \backslash x$	0	1	2	$P(Y=y)$
0	$1/120$	$2 \cdot 3/120$	$3/120$	$10/120$
1	$5 \cdot 3/120$	$2 \cdot 5 \cdot 3/120$	$5/120$	$50/120$
2	$10 \cdot 3/120$	$10 \cdot 2/120$	0	$50/120$
3	$10/120$	0	0	$10/120$
$P(X=x)$	$56/120$	$56/120$	$8/120$	1

- From this table, we can also compute

$$P(X=2|X \geq Y) = \frac{P(X=2, X \geq Y)}{P(X \geq Y)} = \frac{3+5}{1+6+3+30+5} = \frac{8}{45}.$$

Joint distribution (continuous case) (page 71-72 in the textbook)

Suppose X and Y be two continuous r.v.'s on $(\Omega, \mathcal{F}, P)$ with range spaces $\mathcal{R}(X)$ and $\mathcal{R}(Y)$ respectively.

- The random vector $(X, Y) : (\Omega, \mathcal{F}, P) \to (\mathbb{R} \times \mathbb{R}, \sigma(\mathcal{B} \times \mathcal{B}))$ has the joint probability density function $f_{(X,Y)}(x, y) \geq 0$ defined on $\mathcal{R}(X) \times \mathcal{R}(Y)$, so that, $\forall S \in \sigma(\mathcal{B} \times \mathcal{B})$,

$$\begin{aligned} P((X, Y) \in S) &= P\left(\omega \in \Omega | X(\omega) = x, Y(\omega) = y, (x, y) \in S\right) \\ &= \iint_S f(x, y) dx \wedge dy \end{aligned}$$

- The two marginal densities of the random vector (X, Y), which are densities of X and Y, can be computed through the joint probability density $f_{(X,Y)}(x, y)$ as

$$f_X(x) = \int_{y \in \mathcal{R}(Y)} f(x, y) dy; \;\; f_Y(y) = \int_{x \in \mathcal{R}(X)} f(x, y) dx$$

Joint distribution (continuous case) (Example 7.8 at page 73 in the textbook)

- The simplest joint density function for a pair of continuous r.v.'s X and Y on $(\Omega, \mathcal{F}, P)$ is that (X, Y) is jointly uniform over $S = \mathcal{R}(X) \times \mathcal{R}(X)$.
- It has the following joint probability density function

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{1}{\mathrm{area}(S)}, & \text{if } (x, y) \in S; \\ 0, & \text{otherwise.} \end{cases}$$

- For example, $S = (x, y) : \ 0 \le y \le x \le 1$. Then, the random vector (X, Y) jointly uniform over S has the joint probability density

$$f_{(X,Y)}(x, y) = \begin{cases} 2, & \text{if } 0 \le y \le x \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

- The two marginal density functions are

$$f_X(x, y) = \int_0^x 2 \cdot dy = 2x; \quad f_Y(x, y) = \int_y^1 2 \cdot dx = 2(1 - y).$$

Joint distribution (continuous case) (Example 7.9 at page 74 in the textbook)

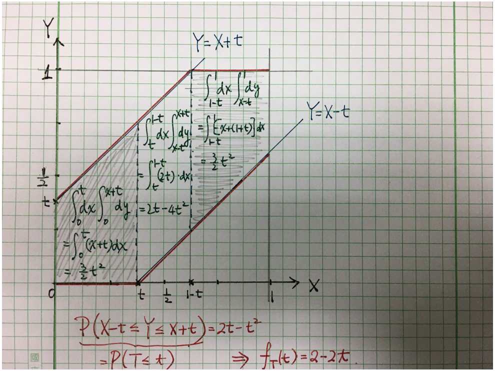
- Mr. and Mrs. Smith agree to meet at a specific location "between 5 and 6 p.m.". Assume that they both arrive there at a random time between 5 and 6 and that their arrivals are independent.
- What is the density function for the time one of the,m have to wait for the other?
- We may assume that X be the time when Mr. Smith arrives, while Y being the time when Mrs. Smith arrives with the time unit 1 hour. The assumption imply that the random vector (X, Y) is uniform on $S = [0,1] \times [0,1]$.
- We would like to find the probability density for $T = |X - Y|$, which has possible values in $[0,1]$.
- Then, for $t \in [0,1]$,

$$\begin{aligned} P(T \leq t) &= P(|X - Y| \leq t) \\ &= P(-t \leq Y - X \leq t) \\ &= P(X - t \leq Y \leq X + t) \end{aligned}$$

Joint distribution (continuous case) (Example 7.9 at page 74 in the textbook)

- To compute $P(T \leq t) = P(X - t \leq Y \leq X + t)$ for $t \in [0, 1]$, we divide into two cases: $0 \leq t \leq \frac{1}{2}$ and $\frac{1}{2} \leq t \leq 1$.
- When $0 \leq t \leq \frac{1}{2}$, we have $t \leq 1 - t$ and

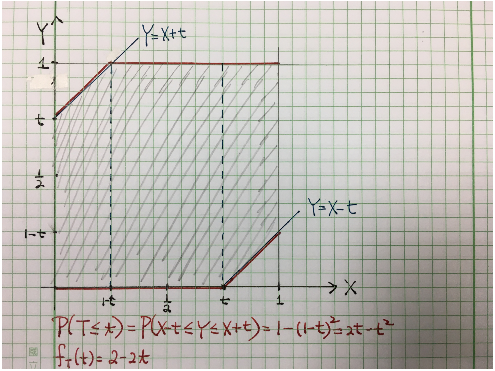
$$P(X - t \leq Y \leq X + t) = \begin{cases} P(0 \leq Y \leq X + t), \text{ if } 0 \leq X \leq t; \\ P(X - t \leq Y \leq X + t), \text{ if } t \leq X \leq 1 - t; \\ P(X - t \leq Y \leq 1), \text{ if } 1 - t \leq X \leq 1. \end{cases}$$



Joint distribution (continuous case) (Example 7.9 at page 74 in the textbook)

- When $\frac{1}{2} \leq t \leq 1$, we have $1 - t \leq t$ and

$$P(X - t \leq Y \leq X + t) = \begin{cases} P(0 \leq Y \leq X + t), & \text{if } 0 \leq X \leq 1 - t; \\ P(0 \leq Y \leq 1), & \text{if } 1 - t \leq X \leq t; \\ P(X - t \leq Y \leq 1), & \text{if } t \leq X \leq 1. \end{cases}$$



Expectation with Joint distribution (page 87 in the textbook)

- Suppose X and Y are two r.v.'s defined on $(\Omega, \mathcal{F}, P)$ with range spaces $\mathcal{R}(X)$ and $\mathcal{R}(Y)$ respectively, and g is a two-variable real-valued function defined on $\mathcal{R}(X) \times \mathcal{R}(Y)$.
- Then, $g(X, Y) : (\Omega, \mathcal{F}, P) \xrightarrow{(X,Y)} \mathcal{R}(X) \times \mathcal{R}(Y) \xrightarrow{g} \mathbb{R}$ can be calculated for its expectation

$$E(g(X, Y)) = \iint_{\mathcal{R}(X)\times\mathcal{R}(Y)} g(x, y) f_{(X,Y)}(x, y) dx \wedge dy$$

- When (X, Y) is instead a discrete pair with joint probability mass function $p_(X, Y)(x, y)$, then

$$E(g(X, Y)) = \sum_{(x,y)\in\mathcal{R}(X)\times\mathcal{R}(Y)} g(x, y) p_{(X,Y)}(x, y).$$

Expectation with Joint distribution (discrete case) (Example 8.1 at page 87 in the textbook)

Assume that 2 among 5 items are defective. Put the items in a random order and inspect them one by one. Let X be the number of inspections needed to find the first defective item; and Y be the number of additional inspections needed to find the second defective item. Compute $E|X-Y|$.

- We first construct a table for the joint p. m. f. of (X, Y).

joint p. m. f. of (X, Y) $(g(i,j) = \|i-j\|)$				
$X = i \setminus Y = j$	1	2	3	4
1	0.1(0)	0.1(1)	0.1(2)	0.1(3)
2	0.1(1)	0.1(0)	0.1(1)	0
3	0.1(2)	0.1(1)	0	0
4	0.1(3)	0	0	0

- For example, in the first column, $p(X=1, Y=1) = \frac{2}{5} \cdot \frac{1}{4} = 0.1$. $p(X=2, Y=1) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = 0.1$. $p(X=3, Y=1) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = 0.1$. $p(X=4, Y=1) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = 0.1$.
- $E[g(X, Y)] = E|X-Y| = 14 \cdot 0.1 = 1.4$

Expectation with Joint distribution (continuous case) (Example 8.2 at page 87 in the textbook)

Assume that (X, Y) is a random point in the triangle

$$S = \{(x, y) : x, y \geq 0, x + y \leq 1\}.$$

Compute EX and $E(XY)$.

- The joint probability density on S is

$$f_{(X,Y)}(x, y) = \begin{cases} 2, & \text{if } (x, y) \in S; \\ 0, & \text{otherwise.} \end{cases}$$

- We first compute the density of X through the joint p.d. of (X, Y) by $f_X(x) = \int_{y \in \mathcal{R}(Y)} f(x, y) dy = \int_0^{1-x} 2 \cdot dy = 2(1 - x)$.
- Then, $EX = \int_0^1 x \cdot 2(1 - x) \cdot dx = \frac{1}{3}$.
- Moreover,

$$\begin{aligned} E(g(X, Y)) &= E(XY) \\ &= \int_S 2g(x, y) dx \wedge dy \\ &= \int_0^1 dx \int_0^{1-x} 2xy dx dy = \frac{1}{12}. \end{aligned}$$

Joint distribution and independence

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X and Y be two random variables on it. The induced measure by X and Y are $\mathcal{L}_X$ and $\mathcal{L}_Y$ respectively.

- Then, X and Y are independent if and only if, for any "rectangular" type of C in $\mathbb{R}^2$ such that $C = A \times B,\ A, B \in \mathcal{B}$.

$$\begin{aligned}
\underbrace{\mathcal{L}_{(X,Y)}(A \times B)}_{\text{joint dist.}} &= P(\{\omega \in \Omega | (X(\omega), Y(\omega)) \in A \times B\}) \\
&= P(\{X \in A\} \cap \{Y \in B\}) \\
&= P\{X \in A\} P(Y \in B) \\
&= \underbrace{\mathcal{L}_X(A)}_{\text{margianl dist. of } X} \cdot \underbrace{\mathcal{L}_Y(B)}_{\text{margianl dist. of } Y} .
\end{aligned}$$

- In other words, for independent random variables X and Y, the *joint distribution* represented by the measure $\mathcal{L}_{(X,Y)}$ factors into the product of the *marginal distributions* represented by the measures $\mathcal{L}_X$ and $\mathcal{L}_Y$.

Independence of random variables

X and Y are independent if and only if $\forall C = A \times B,\ A, B \in \mathcal{B}$,

$$\mathcal{L}_{(X,Y)}(A \times B) = \mathcal{L}_X(A)\mathcal{L}_Y(B).$$

- Let $\mathbf{1}_A$ and $\mathbf{1}_B$ be indicator functions on two Borel sets A and B respectively.
- Then, consider for independent random variables X and Y

$$\begin{aligned} E[(\mathbf{1}_A \circ X)(\mathbf{1}_B \circ Y)] &= \int_\Omega (\mathbf{1}_A \circ X)(\mathbf{1}_B \circ Y)(\omega) dP(\omega) \\ &= P(\{X \in A\} \cap \{Y \in B\}) \\ &= P(\{X \in A\})\, P(\{Y \in B\}) \\ &= E[\mathbf{1}_A \circ X] E[\mathbf{1}_B \circ Y]. \end{aligned}$$

- (Theorem 8.2 at page 91 in the textbook) By standard procedure taking the limit of all the step functions via MCT(monotone convergence theorem), we have, for independent random variables X and Y; and for any two Borel functions g and h,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$