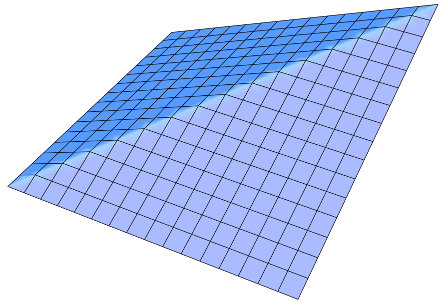

STAT 240: Probability (Advanced Level)

Prof. Christopher Small • Fall 2013 • University of Waterloo
TR 1:00–2:20PM in MC 4063

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is https://rob-zimmerman.github.io/coursenotes.

Contents

Editorial Note The course textbook is *Probability Through Problems* (2001) by Marek Capiński and Tomasz Zastawniak. The vast majority of the problems in these notes come directly from that text. Several problems were left incomplete during lecture; I have tried to complete them all here.

Lecture 1 — Tuesday, September 10, 2013

1. Modelling Random Experiments

Set Theory and Random Experiments

Probability theory is young, at least compared with geometry. It has developed over the last few centuries, and its rules are not always intuitive. How do we formalize them?

We begin with some basic set-theoretic notation. The **sample space**, denoted by Ω, is the set of all distinct **outcomes** of a random experiment. For example, if we toss a fair coin twice, then $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$ contains all possible pairs of faces. An element $\omega \in \Omega$ is one outcome of the experiment; for example, (H,H) means that both tosses landed heads. We write $\emptyset$ for the empty set.

If A and B are sets, then $A \subseteq B$ means that A is a subset of B, with equality allowed. We can also define subsets by properties: if Q is a property, then the corresponding subset is $\{\omega \in \Omega \mid Q(\omega) \text{ is true}\}$. The power set 2^A is the set of all subsets of A, including $\emptyset$ and A itself. We denote the number of elements in A by $\#A$ or $|A|$.

The union of A and B is $A \cup B$. More generally,

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup \dots.$$

The intersection of A and B is $A \cap B$, and

$$\bigcap_{n=1}^{\infty} A_n = A_1 \cap A_2 \cap \dots.$$

If $\mathscr{A}$ is a collection of sets (i.e., a set of sets), we write $\cup\mathscr{A}$ for the union over all sets in $\mathscr{A}$, and we write $\cap\mathscr{A}$ for the intersection over all sets of $\mathscr{A}$.

Example 1.1 Let $A = \{1, 2, 3\}$. Then

$$2^A = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, A\}.$$

Set $\mathscr{A} = 2^A$. Then $\cup\mathscr{A} = A$ and $\cap\mathscr{A} = \emptyset$. We see that $\mathscr{A} = \{A_1, A_2, \ldots, A_8\}$, and then $\cup\mathscr{A} = \bigcup_{n=1}^{8} A_n$.

We need a carefully defined **random experiment**. We also need a set Ω whose elements ω are the distinct outcomes of that random experiment. Every property Q defined on the outcomes of the random experiment determines a subset of Ω, and that subset is called an **event**. Thus

$$A = \{\omega \in \Omega \mid Q(\omega) \text{ is true}\}$$

is an event. For each event $A \subseteq \Omega$, its **probability** $\mathbb{P}(A)$ is a number in $[0, 1]$.

Lecture 2 — Thursday, September 12, 2013

Let's specify the outcomes and probabilities in several situations.

Example 2.1 (Problem 1.6) From a pack of 52 cards, we draw 2.

Solution. Define $S = \{\text{spades, hearts, diamonds, clubs}\}$, $D = \{A, 2, 3, \ldots, 10, J, Q, K\}$. We have $\#S = 4$ and $\#D = 13$. There are two cases: if we draw two cards with replacement, then our sample space is

$$\Omega = (S \times D) \times (S \times D) = \{[(s_1, d_1), (s_2, d_2)] \mid s_j \in S, d_j \in D \text{ for } j = 1, 2\}.$$

We have $\#\Omega = 52^2 = 2704$, and $\mathbb{P}(\{\omega\}) = \frac{1}{2704}$. In general, $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$ for $A \subset \Omega$. If we draw two cards without replacement, then construct the **diagonal set** $\Delta = \{[(s, d), (s, d)] \mid s \in S, d \in D\}$, and let $\Omega' = \Omega \setminus \Delta$. Then $\#\Omega' = 2652$, $\mathbb{P}(\{\omega\}) = \frac{1}{2652}$, and $\mathbb{P}(A) = \frac{\#A}{\#\Omega'}$ for $A \subseteq \Omega'$. $\square$

Example 2.2 (Problem 1.8) We play a series of chess games. The winner is the one who first scores three points, where one point is awarded for a win and draws do not count.

Solution. We need to set this up carefully. Consider the possible lengths of the series. We have $\Omega_{3a} = \{(W, W, W)\}$ and $\Omega_{3b} = \{(L, L, L)\}$. We also have $\Omega_{4a} = \{(\underbrace{\ldots}_{2W,1L}, W)\}$, so $\#\Omega_{4a} = 3$, and $\Omega_{4b} = \{(\underbrace{\ldots}_{2L,1W}, L)\}$, so $\#\Omega_{4b} = 3$. Finally, $\Omega_{5a} = \{(\underbrace{\ldots}_{2W,2L}, W)\}$, so $\#\Omega_{5a} = \binom{4}{2} = 6$, and $\Omega_{5b} = \{(\underbrace{\ldots}_{2L,2W}, L)\}$, so $\#\Omega_{5b} = \binom{4}{2} = 6$. Therefore,

$$\Omega = \Omega_{3a} \cup \Omega_{3b} \cup \cdots \cup \Omega_{5b},$$

and hence $\#\Omega = 20$. However, we should ask whether $\mathbb{P}(\{\omega\}) = \frac{1}{20}$ is actually reasonable. Is (W, W, W) as likely as (L, L, W, W, W)? We always need to make some simplifying assumptions. $\square$

Exercise 2.3 (Problem 1.9) Propose a set of outcomes and measures for the situation in which we keep throwing a coin until it lands heads up.

2. Classical Probability Spaces

Counting Identities

Suppose Ω is finite, and define $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$. Then probability reduces to combinatorics; we are essentially solving counting problems.

Example 2.4 (Problem 2.1) How many subsets of an n-element set are there?

Solution. There are 2^n such subsets. It suffices to prove this for the set $\{1, 2, \ldots, n\}$. We define a function $f_A : \{1, 2, \dots, n\} \to \{T, F\}$ for each $A \subset \{1, 2, \dots, n\}$, where

$$f_A(i) = \begin{cases} T & \text{if } i \in A \\ F & \text{if } i \notin A \end{cases}$$

The correspondence $A \mapsto f_A$ is a bijection between the subsets of $\{1, \dots, n\}$ and the functions from $\{1, \dots, n\}$ to $\{T, F\}$. There are 2^n such functions: we have two choices for each of $f_A(1), \dots, f_A(n)$. Hence there are 2^n subsets. $\square$

Example 2.5 (Problem 2.2) How many ways can we order an n-element set?

Solution. There are $n!$ such orderings. Let the underlying set be S with $\#S = n$. To form an ordering of S, choose the first element in n ways, then the second element in $n-1$ ways (since one element is already used), then the third in $n-2$ ways, and so on, until the last position, which has 1 choice. Hence the total number of orderings is

$$n(n-1)(n-2) \dots 2 \cdot 1 = n!.$$

$\square$

Example 2.6 (Problem 2.3) How many k-element subsets of an n-element set are there?

Solution. The number of such subsets is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!}.$$

$\square$

Note that

$$\binom{n}{m} = 0$$

if $m > n$. The same equation holds if $m < 0$, because

$$\binom{n}{m} = \binom{n}{n-m}.$$

These statements are equivalent. We can even define

$$\binom{1/2}{k} = \frac{(1/2)(-1/2)\dots(1/2-k+1)}{k!},$$

but that goes beyond the present question.

Example 2.7 (Problem 2.4) How many one-to-one mappings are there from an n-element set to an m-element set?

Solution. If $0 \le n \le m$, the number of such one-to-one mappings is

$$m(m-1)\dots(m-n+1) = \frac{m!}{(m-n)!} = m^{(n)}.$$

If $n > m$, there are no one-to-one mappings by the pigeonhole principle. □

Example 2.8 (Problem 2.5) How many mappings are there from an n-element set to an m-element set?

Solution. There are m^n such mappings. □

Example 2.9 (Problem 2.6) Prove that

$$2^n = \sum_{k=0}^{n} \binom{n}{k}.$$

Solution. Let Ω be an n-element set. The number of subsets of Ω is the sum over all possible subset sizes, so

$$\sum_{k=0}^{n} (\text{\# of subsets of size } k) = 2^n = \sum_{k=0}^{n} \binom{n}{k}$$

as claimed. □

Example 2.10 (Problem 2.7: Vandermonde's identity) Prove that

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}.$$

Lecture 3 — Tuesday, September 17, 2013

Last time we discussed Problem 2.7, the Vandermonde identity,

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}.$$

It is worth remembering. One proof is to consider a club with m men and n women and ask how many k-person executive committees can be formed. On the one hand, the answer is simply

$$\binom{m+n}{k}.$$

On the other hand, if exactly i men are chosen, then we must choose i men from the m available and $k-i$ women from the n available, so the number of such committees is

$$\binom{m}{i}\binom{n}{k-i}.$$

Summing over all possible values of i gives the identity.

This identity is important in probability. Indeed, we have

$$\sum_{i=0}^{k} \frac{\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}} = 1.$$

If we denote the fraction by $p(i)$, then

$$\sum_{i=0}^{k} p(i) = 1,$$

which defines the **hypergeometric distribution**.

Example 3.1 (Problem 2.8) Ten people are randomly seated at a round table. What is the probability that a particular couple will sit next to each other?

Solution (by a student). Fix the location of one person. Then the probability that the second person is seated just to the left or right of the first is $\frac{1}{9} + \frac{1}{9}$, so the desired probability is $\frac{2}{9}$. □

Example 3.2 (Problem 2.10) Two dice are thrown. What is the probability that the total number of dots that appear is...

1. equal to 7?
2. equal to 3?
3. greater than 5?
4. an even number?

Solution. Let $\Omega = \{(i, j) \mid i, j \in \{1, \dots, 6\}\}$.

For (1), the probability is

$$\frac{\#\{(i,j) \mid i + j = 7\}}{\#\Omega} = \frac{1}{36} \sum_{i=1}^{6} \#\{(i, 7-i)\} = \frac{6}{36} = \frac{1}{6}.$$

For (2), the probability is

$$\frac{\#\{(i,j) \mid i + j = 3\}}{\#\Omega} = \frac{\#\{(1,2),(2,1)\}}{36} = \frac{2}{36} = \frac{1}{18}.$$

For (3), the probability is

$$\frac{\#\{(i,j) \mid i + j > 5\}}{\#\Omega} = \frac{26}{36} = \frac{13}{18}.$$

For (4), the probability is

$$\mathbb{P}((\text{odd, odd}) \cup (\text{even, even})) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

□

Example 3.3 (Problem 2.12: The birthday problem) What is the probability that among 25 people, at least 2 have their birthday on the same day of the year?

Solution. In the birthday problem,

$$\begin{aligned}\mathbb{P}(\text{all 25 birthdays are different}) &= \frac{365}{365} \cdot \frac{364}{365} \cdots \frac{(365-25+1)}{365} \\ &= \frac{{}_{365}P_{25}}{365^{25}} \approx 0.4313.\end{aligned}$$

Therefore, $\mathbb{P}(\text{at least 2 birthdays coincide}) = 1 - 0.4313 \approx 0.5687$. □

Example 3.4 (Problem 2.21) From a pack of 52 cards, we draw one by one. What is the probability that an ace appears for the first time on the fifth draw?

Solution (by a student). $\#\Omega = {}_{52}P_5$ and there are 48 cards that are not aces. Moreover,

$$\#A = 48 \cdot 47 \cdot 46 \cdot 45 \cdot 4 = {}_{48}P_4 \cdot 4.$$

Thus the probability is $\frac{\#A}{\#\Omega} \approx 0.05989$. □

Example 3.5 (Problem 2.13) Among $t = 60$ lottery tickets, $w = 20$ win prizes. We buy $b = 6$ tickets. What is the probability that exactly $g = 2$ are winning tickets? Generalize to arbitrary t, w, b, g.

Solution. We want to choose exactly g winning tickets among the w winners, and exactly $b - g$ losing tickets among the $t - w$ losers. The total number of ways to choose any b tickets from t is $\binom{t}{b}$. Thus, the probability is

$$\mathbb{P}(\text{exactly 2 winners}) = \frac{\binom{w}{g}\binom{t-w}{b-g}}{\binom{t}{b}},$$

which is the hypergeometric probability from Problem 2.7. For the given values, that probability is

$$\frac{\binom{20}{2}\binom{40}{4}}{\binom{60}{6}} \approx 0.3468.$$

$\square$

Lecture 4 — Thursday, September 19, 2013

Example 4.1 (Problem 2.16) A regular bridge deck of 52 cards is dealt among four players. I have four spades. The opponents have shown that they have 8 hearts. What is the probability that my partner has at least 3 spades?

Solution. I have 13 cards, 4 of which are spades. So there are $52 - 13 = 39$ cards held by the others. Among these 39 cards, 8 are hearts held by my opponents. Thus there are $39 - 8 = 31$ cards left, consisting of 9 spades and 22 other cards. My partner has 13 cards. The probability that my partner has exactly i spades is

$$\mathbb{P}(\text{partner has } i \text{ spades}) = \frac{\binom{9}{i}\binom{22}{13-i}}{\binom{31}{13}}.$$

Therefore, the required probability is the following expression:

$$\mathbb{P}(\text{partner has} \geq 3 \text{ spades}) = 1 - \left[\frac{\binom{9}{0}\binom{22}{13}}{\binom{31}{13}} + \frac{\binom{9}{1}\binom{22}{12}}{\binom{31}{13}} + \frac{\binom{9}{2}\binom{22}{11}}{\binom{31}{13}}\right].$$

Note that "hearts" just means "not a spade". That is the important part! $\square$

Example 4.2 (Problem 2.22) How likely is it that ABRACADABRA appears when the letters A, A, A, A, A, B, B, C, D, R, R are shuffled randomly?

Solution (by a student). We have $\#\Omega = 11!$ and

$$\#A = 5! \cdot 2! \cdot 2! \cdot 1! \cdot 1! = 480.$$

Therefore,

$$\mathbb{P}(\text{ABRACADABRA}) = \frac{\#A}{\#\Omega} = \frac{480}{11!} = \frac{1}{83160}.$$

The corresponding **multinomial coefficient** is $\frac{11!}{5! \cdot 2! \cdot 1! \cdot 1! \cdot 2!} = \binom{11}{5, 2, 1, 1, 2}$. □

3. Fields

Fields of Events

We need to formalize the concept of an event. An event is a subset of Ω to which we can attach a probability, and the collection of events must form a field.

Definition 4.3 A family $\mathscr{F}$ of subsets of $\Omega \neq \emptyset$ is called a **field** if:

1. $\Omega \in \mathscr{F}$,
2. $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$,
3. if $A, B \in \mathscr{F}$, then $A \cup B \in \mathscr{F}$.

Proposition 4.4 (Problems 3.1–3.3) We have the following basic examples of fields.

1. $\mathscr{F} = \{\emptyset, \Omega\}$ is a field; it is called the **trivial field**.
2. The family of all subsets of a finite $\Omega \neq \emptyset$ is a field.
3. If $A \subseteq \Omega$, then $\mathscr{F} = \{\emptyset, A, A^c, \Omega\}$ is a field.

Proof. For ((1)), let $\mathscr{F} = \{\emptyset, \Omega\}$. Then $\Omega \in \mathscr{F}$. Also, $\Omega^c = \emptyset \in \mathscr{F}$ and $\emptyset^c = \Omega \in \mathscr{F}$, so $\mathscr{F}$ is closed under complements. Finally, unions of members of $\mathscr{F}$ remain in $\mathscr{F}$: $\emptyset \cup \emptyset = \emptyset$, $\emptyset \cup \Omega = \Omega$, and $\Omega \cup \Omega = \Omega$. Hence $\mathscr{F}$ is a field.

For ((2)), take $\mathscr{F} = 2^\Omega$, the family of all subsets of Ω. We have $\Omega \in \mathscr{F}$. If $E \in \mathscr{F}$, then $E \subseteq \Omega$, so $E^c = \Omega \setminus E \subseteq \Omega$ and therefore $E^c \in \mathscr{F}$. If $E, F \in \mathscr{F}$, then $E \cup F \subseteq \Omega$, so $E \cup F \in \mathscr{F}$. Thus 2^Ω is a field.

For ((3)), let $\mathscr{F} = \{\emptyset, A, A^c, \Omega\}$ with $A \subseteq \Omega$. We have $\Omega \in \mathscr{F}$. Complements stay in $\mathscr{F}$ since $\emptyset^c = \Omega$, $\Omega^c = \emptyset$, $(A^c)^c = A$, and A^c is already listed. For unions, check all possibilities among the four sets: unions with $\emptyset$ return the other set, unions with Ω return Ω, $A \cup A = A$, $A^c \cup A^c = A^c$, and $A \cup A^c = \Omega$. All of these are in $\mathscr{F}$, so $\mathscr{F}$ is a field. □

Example 4.5 (Problem 3.4) Let $\Omega = \{1, 2, 3, 4\}$. Which of the following is a field?

1. $\mathscr{F}_1 = \{\emptyset, \{1, 2\}, \{3, 4\}\}$
2. $\mathscr{F}_2 = \{\emptyset, \Omega, \{1\}, \{2, 3, 4\}, \{1, 2\}, \{3, 4\}\}$
3. $\mathscr{F}_3 = \{\emptyset, \Omega, \{1\}, \{2\}, \{1, 2\}, \{3, 4\}, \{2, 3, 4\}, \{1, 3, 4\}\}$

Solution. For ((1)), $\Omega \notin \mathscr{F}_1$, so $\mathscr{F}_1$ is not a field.

For ((2)), we have $\{1\}, \{3, 4\} \in \mathscr{F}_2$, but $\{1\} \cup \{3, 4\} = \{1, 3, 4\} \notin \mathscr{F}_2$. So $\mathscr{F}_2$ is not closed under unions and is not a field.

For ((3)), observe that $\mathscr{F}_3$ consists of all unions of the three disjoint blocks $\{1\}$, $\{2\}$, and $\{3, 4\}$. Complements and unions of such sets are again unions of these blocks, so $\mathscr{F}_3$ is a field. For example,

$$\emptyset^c = \Omega, \qquad \Omega^c = \emptyset, \qquad \{1\}^c = \{2, 3, 4\},$$
$$\{2\}^c = \{1, 3, 4\}, \qquad \text{and} \qquad \{1, 2\}^c = \{3, 4\}.$$

□

Exercise 4.6

(Problem 3.5) Let $\Omega = (0, 1)$. Determine whether any of the following families of sets is a field:

$$\mathscr{F}_1 = \{\emptyset, (0,1), (0,1/2), (1/2,1)\}$$
$$\mathscr{F}_2 = \{\emptyset, (0,1), (0,1/2), [1/2,1), (0,2/3], (2/3,1)\}$$
$$\mathscr{F}_3 = \{\emptyset, (0,1), (0,2/3), [2/3,1)\}$$

(Problem 3.6) Let $\Omega = [0,1]$. Adding as few sets as possible, complete the family of sets $\{\emptyset, [0,1/2), \{1\}\}$ to obtain a field.

(Problem 3.7) Let $\Omega = \{1,2,3\}$. Complete $\{\{2\},\{3\}\}$ to obtain a field, adding as few sets as possible.

Properties of Fields

Proposition 4.7 (Problem 3.8) If $\mathscr{F}$ is a field, then:

1. $\emptyset \in \mathscr{F}$.
2. If $A, B, C \in \mathscr{F}$, then $A \cup B \cup C \in \mathscr{F}$. So fields are closed under finite unions.
3. If $A, B, C \in \mathscr{F}$, then $A \cap B \cap C \in \mathscr{F}$. So fields are also closed under finite intersections.

Proof. For ((1)), $\emptyset = \Omega^c$. Since $\Omega \in \mathscr{F}$ by ((1)), apply ((2)) to obtain $\Omega^c = \emptyset \in \mathscr{F}$.

For ((2)), the field axiom gives the result for two sets. If a union of n members of $\mathscr{F}$ belongs to $\mathscr{F}$, taking its union with one more member shows that a union of $n+1$ members does too.

For ((3)), the complement and union axioms give

$$A \cap B = (A^c \cup B^c)^c \in \mathscr{F}.$$

The same induction now handles any finite number of intersections. $\square$

Thus fields are closed under all finite unions and intersections.

Lecture 5 — Tuesday, September 24, 2013

Recall that $\mathscr{F}$ is a field of events when it satisfies the following three axioms:

1. $\Omega \in \mathscr{F}$,
2. $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$,
3. $A, B \in \mathscr{F} \Rightarrow A \cup B \in \mathscr{F}$.

Example 5.1 (Problem 3.13) If $A, B \in \mathscr{F}$, show that $A \setminus B \in \mathscr{F}$.

Solution. Since $B \in \mathscr{F}$, we have $B^c \in \mathscr{F}$ by Axiom 2. We also have $A \in \mathscr{F}$, so $A \cap B^c \in \mathscr{F}$. But $A \cap B^c = A \setminus B$, and therefore $A \setminus B \in \mathscr{F}$. $\square$

Example 5.2 (Problem 3.14) If $A, B \in \mathscr{F}$, show that their **symmetric difference**

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

also belongs to $\mathscr{F}$.

Solution. Since $A, B \in \mathscr{F}$, we have $A \setminus B \in \mathscr{F}$ and $B \setminus A \in \mathscr{F}$ by Example 5.1. Also, A field is closed under unions, so $(A \setminus B) \cup (B \setminus A) \in \mathscr{F}$. Therefore $A \triangle B \in \mathscr{F}$. $\square$

Example 5.3 (Problem 3.15) If $\mathscr{F}_1$ and $\mathscr{F}_2$ are fields of subsets of Ω, then so is $\mathscr{F}_1 \cap \mathscr{F}_2$. For example, if $\mathscr{F}_1 = \{\emptyset, A, A^c, \Omega\}$ and $\mathscr{F}_2 = \{\emptyset, B, B^c, \Omega\}$, where $A \notin \{B, B^c\}$ and neither field is trivial, then $\mathscr{F}_1 \cap \mathscr{F}_2 = \{\emptyset, \Omega\}$.

Solution. We check the axioms. First, $\Omega \in \mathscr{F}_1 \cap \mathscr{F}_2$ because $\Omega \in \mathscr{F}_1$ and $\Omega \in \mathscr{F}_2$. Next, suppose $E \in \mathscr{F}_1 \cap \mathscr{F}_2$. Then $E \in \mathscr{F}_1$ and $E \in \mathscr{F}_2$. Since both $\mathscr{F}_1$ and $\mathscr{F}_2$ are fields, it follows from the second field axiom that $E^c \in \mathscr{F}_1$ and $E^c \in \mathscr{F}_2$. So $E^c \in \mathscr{F}_1 \cap \mathscr{F}_2$. Finally, if $E, G \in \mathscr{F}_1 \cap \mathscr{F}_2$, then $E \cup G$ belongs to each field and hence to their intersection. $\square$

Example 5.4 (Problem 3.16) Find two fields whose union is not a field.

Solution. Let $\Omega = \{1, 2, 3\}$. Define $\mathscr{F}_1 = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}$ and $\mathscr{F}_2 = \{\emptyset, \Omega, \{2\}, \{1, 3\}\}$. Then the union of fields is not a field. Indeed, $\{1\}, \{2\} \in \mathscr{F}_1 \cup \mathscr{F}_2$, but $\{1, 2\} \notin \mathscr{F}_1 \cup \mathscr{F}_2$. $\square$

Example 5.5 (Problem 3.17) Is $\mathscr{F} = \{A \subseteq \Omega \mid A \text{ is finite}\}$ a field?

Solution. If Ω is finite, then every subset of Ω is finite, so

$$\mathscr{F} = 2^{\Omega},$$

which is a field.

If Ω is infinite, then $\Omega \notin \mathscr{F}$ because Ω is not finite. This already violates the first field axiom. Equivalently, take any finite $A \in \mathscr{F}$; then A^c is infinite, so $A^c \notin \mathscr{F}$, which violates closure under complements.

Therefore $\mathscr{F}$ is a field if and only if Ω is finite. $\square$

Example 5.6 (Problem 3.18) Let $I_0 = \emptyset$ and $I_n = \{1, \dots, n\}$ for $n \geq 1$. Is

$$\mathscr{F} = \{I_n\}_{n=0}^{\infty} \cup \{\mathbb{N} \setminus I_n\}_{n=0}^{\infty}$$

a field?

Solution. No. Consider $I_1, I_3^c \in \mathscr{F}$. But $I_1 \cup I_3^c = \{1, 4, 5, 6, \dots\} \notin \mathscr{F}$. $\square$

Example 5.7 (Problem 3.19) Let Ω be a possibly infinite set. Show that

$$\mathscr{F} = \{A \subseteq \Omega \mid A \text{ is finite or } A^c \text{ is finite}\}$$

is a field.

Solution. First, $\Omega^c = \emptyset$ is finite, so $\Omega \in \mathscr{F}$. Next, suppose $A \in \mathscr{F}$. There are two cases. If A is finite, then A^c is cofinite and $(A^c)^c = A$ is finite. If A is infinite, then A^c must be finite. In either case, $A^c \in \mathscr{F}$. If $A, B \in \mathscr{F}$ are both finite, then $A \cup B$ is finite. Otherwise at least one, say A, is cofinite, and

$$(A \cup B)^c = A^c \cap B^c \subseteq A^c$$

is finite. Thus $A \cup B \in \mathscr{F}$. $\square$

Atoms

Definition 5.8 Suppose $\mathscr{F}$ is a field. A set $A \in \mathscr{F}$ is called an **atom** if it is minimal in the following sense: $A \neq \emptyset$, and whenever $B \subseteq A$ with $B \in \mathscr{F}$, we must have either $B = A$ or $B = \emptyset$.

Example 5.9

1. Let $\Omega = \{1, 2, 3\}$ and $\mathscr{F} = \{\emptyset, \Omega, \{1\}, \{2, 3\}, \{2\}, \{1, 3\}, \{3\}, \{1, 2\}\}$. The atoms of $\mathscr{F}$ are $\{1\}, \{2\}, \{3\}$.
2. Let $\Omega = (0, 1]$ and $\mathscr{F} = \{\emptyset, \Omega, (0, 2/3], (2/3, 1]\}$. The atoms of $\mathscr{F}$ are $(0, 2/3]$ and $(2/3, 1]$.

Example 5.10 (Problem 3.22) What are the atoms of $\mathscr{F} = 2^\Omega$ where Ω is finite?

Solution. The atoms are $\{\omega\}$ for $\omega \in \Omega$. There are no others. $\square$

Example 5.11 (Problem 3.23) Show that different atoms are disjoint.

Solution. Suppose that the distinct atoms A and B satisfy $A \cap B \neq \emptyset$. Let $C = A \cap B$. Then $C \in \mathscr{F}$ and $C \subseteq A$. Since A is an atom, either $C = \emptyset$ or $C = A$. We excluded the first case, so $C = A$ and hence $A \subseteq B$. Reversing the roles of A and B gives $B \subseteq A$. Thus $A = B$, contradicting their being distinct. $\square$

Example 5.12 (Problem 3.24) Suppose $\mathscr{F}$ is finite. Then every $A \in \mathscr{F}$ such that $A \neq \emptyset$ has a subset that is an atom.

Solution. Suppose A is as given. Let $\omega \in A$. Define $B = \bigcap\{D \in \mathscr{F} \mid \omega \in D\}$. Then $B \in \mathscr{F}$ by closure under finite intersections. Also, B is an atom. To see this, note that $\omega \in B$, so $B \neq \emptyset$. If $C \subseteq B$ with $C \in \mathscr{F}$, we will show that $C = \emptyset$ or $C = B$. Because A is one of the sets in the defining intersection, $B \subseteq A$.

Case 1: $\omega \in C$. Then C is one of the sets in the intersection $\bigcap_{\omega \in D} D$. So $B \subseteq C$. Since also $C \subseteq B$, have $B = C$.

Case 2: $\omega \notin C$. Then consider argument (Case 1) above, with C replaced by $B \setminus C$. So $B \setminus C = B$. Then $C = \emptyset$. □

Corollary 5.13 If $\mathscr{F}$ is finite and $A_1, A_2, \ldots, A_n$ are the atoms, then $\bigcup_{i=1}^n A_i = \Omega$ and $A_i \cap A_j = \emptyset$ for $i \neq j$. Thus the atoms of $\mathscr{F}$ **partition** Ω.

Proof. Distinct atoms are disjoint by Problem 3.23. Given $\omega \in \Omega$, apply the construction in Problem 3.24 to $A = \Omega$ and this particular ω; it produces an atom containing ω. Hence every point of Ω lies in an atom, so the atoms cover Ω. □

Exercise 5.14 (Problem 3.25) Show that if $\mathscr{F}$ is a finite field, then each event $A \in \mathscr{F}$ is the union of finitely many atoms. Show that this is a unique representation: that is, the set of atoms $\{A_1, \ldots, A_k\}$ such that $A = A_1 \cup \cdots \cup A_k$ is unique for each event $A \in \mathscr{F}$.

Example 5.15 (Problem 3.26) Show that if $\mathscr{F}$ is finite, then $\#\mathscr{F} = 2^n$ for some n.

Solution. Let $A_1, \ldots, A_n$ be the atoms of $\mathscr{F}$. From Corollary 5.13, these atoms are pairwise disjoint and

$$\bigcup_{i=1}^{n} A_i = \Omega.$$

If an atom A_i meets $E \in \mathscr{F}$, then the nonempty event $A_i \cap E \subseteq A_i$ must equal A_i. Thus every atom that meets E is contained in E; because the atoms partition Ω, E is exactly the union of those atoms.

For each subset $I \subseteq \{1, \ldots, n\}$, define
$$E_I = \bigcup_{i \in I} A_i.$$
Then $E_I \in \mathscr{F}$, so this gives a map from subsets I to events in $\mathscr{F}$. This map is injective because the atoms are disjoint, and it is surjective because every event is a union of atoms. Hence it is a bijection.

Therefore $\mathscr{F}$ has exactly as many elements as the power set of $\{1, \ldots, n\}$, so
$$\#\mathscr{F} = 2^n.$$
$\square$

Lecture 6 — Thursday, September 26, 2013

4. Finitely Additive Probability

Definition 6.1 Let $\mathscr{F}$ be a field of subsets of $\Omega \neq \emptyset$, and let $\mathbb{P} : \mathscr{F} \to [0, 1]$ be a function. Such a function is called a **finitely additive probability measure** if

1. $\mathbb{P}(\Omega) = 1$, and
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for disjoint A and B; that is, when $A \cap B = \emptyset$.

We call $\mathbb{P}(A)$ the **probability** of the event A.

Examples and Counterexamples

Example 6.2 (Problem 4.1) Let $\Omega = \{1, 2\}$ and let $\mathscr{F} = \{\emptyset, \Omega, \{1\}, \{2\}\}$. Define $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\{1\}) = 1/3$, and $\mathbb{P}(\{2\}) = 2/3$. This is a finitely additive probability measure.

Example 6.3 (Problem 4.2) Let $\Omega = \{1, 2, 3, 4\}$ and let $\mathscr{F} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$. Define $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 2$, $\mathbb{P}(\{1, 2\}) = 5/8$, and $\mathbb{P}(\{3, 4\}) = 5/8$. This is not a finitely additive probability measure: it violates both normalization and finite additivity, since

$$\mathbb{P}(\Omega) = 2 \neq \frac{10}{8} = \mathbb{P}(\{1, 2\}) + \mathbb{P}(\{3, 4\}).$$

Example 6.4 (Problem 4.8) Suppose $\mathscr{F}$ is finite, with $\#\mathscr{F} = 2^n$, and let $A_1, \dots, A_n \in \mathscr{F}$ be the atoms. Every $A \in \mathscr{F}$ can be written as $A = A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}$, a union of atoms. To build a finitely additive probability measure $\mathbb{P}$, choose $\alpha_1, \alpha_2, \dots, \alpha_n \geq 0$ such that $\alpha_1 + \cdots + \alpha_n = 1$. Then define $\mathbb{P}(A) = \alpha_{i_1} + \cdots + \alpha_{i_k}$ whenever $A = A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}$.

As a special case, if $\Omega = \{\omega_1, \dots, \omega_n\}$ is finite and $\mathscr{F} = 2^\Omega$, then the atoms are the singletons $A_i = \{\omega_i\}$.

As an even more special case, if

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = \frac{1}{n} = \frac{1}{\#\Omega},$$

then $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$ in agreement with Chapter 2.

Example 6.5 (Problem 4.9) Prove $\mathbb{P}(\emptyset) = 0$.

Solution. Since $A \cup \emptyset = A$, we have $\mathbb{P}(A \cup \emptyset) = \mathbb{P}(A) + \mathbb{P}(\emptyset)$. Therefore $\mathbb{P}(A) = \mathbb{P}(A) + \mathbb{P}(\emptyset)$, which implies $\mathbb{P}(\emptyset) = 0$. □

Example 6.6 (Problem 4.10) Prove that if $A \subseteq B$, then $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.

Solution. Since $A \subseteq B$, we may write $B = (B \setminus A) \cup A$, and this union is disjoint. Therefore $\mathbb{P}(B \setminus A) + \mathbb{P}(A) = \mathbb{P}(B)$ by Axiom ((2)), so $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. □

Example 6.7 (Problem 4.11) Prove $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Solution. In Problem 4.10, let $B = \Omega$. Then $\mathbb{P}(\Omega \setminus A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A)$ by Axiom ((1)). □

Example 6.8 (Problem 4.12: Monotonicity) If $A \subseteq B$, prove $\mathbb{P}(A) \le \mathbb{P}(B)$.

Solution. We have

$$\mathbb{P}(B) = \mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \ge \mathbb{P}(A).$$

□

Example 6.9 (Problem 4.13) Suppose that A, B, and C are pairwise disjoint. Then $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$.

Solution. Since A, B, and C are pairwise disjoint, we have

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) = \emptyset.$$

So $A \cup B$ and C are disjoint, and finite additivity gives

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}((A \cup B) \cup C) = \mathbb{P}(A \cup B) + \mathbb{P}(C).$$

Also $A \cap B = \emptyset$, so

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Substituting,

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C).$$

□

By induction, Example 6.9 generalizes immediately to finitely many disjoint sets $A_1, A_2, \ldots, A_n$:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

This property is called **finite additivity**.

Example 6.10 (Problem 4.15) Show that $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

Solution. Write $A \cup B = A \cup (B \setminus A)$. Then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ by Axiom ((2)). Since $B \setminus A \subseteq B$, Problem 4.12 gives $\mathbb{P}(B \setminus A) \leq \mathbb{P}(B)$. Therefore $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$. □

Example 6.11 (Problem 4.16) Give an example where $\mathbb{P}(A \cup B) < \mathbb{P}(A) + \mathbb{P}(B)$.

Solution. Choose $A = B = \Omega$. Then

$$\mathbb{P}(A \cup B) = 1 < 2 = \mathbb{P}(A) + \mathbb{P}(B).$$

□

Example 6.12 (Problem 4.17) Show that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Solution. The proof is by induction on n. For $n = 1$, the statement is

$$\mathbb{P}(A_1) \leq \mathbb{P}(A_1),$$

which is true. Now, assume it holds for some $n \geq 1$, i.e.

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Then
$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right) \leq \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}(A_{n+1}),$$
by Example 6.10. Applying the induction hypothesis gives
$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) + \mathbb{P}(A_{n+1}) = \sum_{i=1}^{n+1} \mathbb{P}(A_i).$$
So the result holds for $n+1$, and therefore for all $n \in \mathbb{N}$. $\square$

Exercise 6.13

(Problem 4.18: Inclusion-exclusion formula) Show that
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(Problem 4.19) Find an expression for $\mathbb{P}(A \cup B \cup C)$ along the same lines as Problem 4.18.

(Problem 4.21) Suppose that $\mathbb{P}(A) = 1/2$ and $\mathbb{P}(B) = 2/3$. Show that
$$\frac{1}{6} \leq \mathbb{P}(A \cap B) \leq \frac{1}{2},$$
and give examples to demonstrate that both upper and lower bounds can be attained.

(Problem 4.22) Suppose that $\mathbb{P}(A) = 1/2$ and $\mathbb{P}(B) = 1/3$. Show that
$$\frac{1}{2} \leq \mathbb{P}(A \cup B) \leq \frac{5}{6},$$
and give examples to demonstrate that both upper and lower bounds can be attained.

Lecture 7 — Thursday, October 03, 2013

5. σ-Fields

σ-Fields

To start with, here are some useful facts to recall about countability.

Recall that a set is **countable** if it is finite or can be put in bijection with $\mathbb{N}$. If $A_1, A_2, \dots, A_n$ are countable, then $\bigcup_{i=1}^n A_i$ is countable as well, and in fact $\bigcup_{i=1}^\infty A_i$ is still countable. On the other hand, $2^{\aleph_0}$ is not countable, and $\#2^{\mathbb{N}} = \#\mathbb{R}$.

We are now turning to Chapter 5 on σ-fields, where the symbol σ usually refers to countable operations.

Definition 7.1 A collection $\mathscr{F}$ of subsets of Ω is called a ***σ*-field** on Ω if:

1. $\Omega \in \mathscr{F}$.
2. If $A \in \mathscr{F}$, then $A^c \in \mathscr{F}$.
3. If $A_1, A_2, \dots \in \mathscr{F}$, then $\bigcup_{i=1}^\infty A_i \in \mathscr{F}$.

Example 7.2 (Problem 5.1) Show that a σ-field is a field.

Solution. We only need to check one axiom for a field. Suppose $\mathscr{F}$ is a σ-field. We must show that $A, B \in \mathscr{F}$ implies that $A \cup B \in \mathscr{F}$. Suppose $A, B \in \mathscr{F}$. Choose

$$A_1 = A, A_2 = B, A_3 = B, \dots, A_n = B, \dots.$$

Then $A \cup B = \bigcup_{i=1}^\infty A_i \in \mathscr{F}$. $\square$

Example 7.3 (Problem 5.2) If a field $\mathscr{F}$ is finite, then it is a σ-field.

Solution. Conditions ((1)) and ((2)) are already field axioms. If $A_1, A_2, \dots \in \mathscr{F}$, then only finitely many distinct sets can appear because $\mathscr{F}$ is finite. Thus there are indices $i_1, \dots, i_m$ such that

$$\bigcup_{i=1}^\infty A_i = \bigcup_{j=1}^m A_{i_j} \in \mathscr{F}.$$

$\square$

Example 7.4 (Problem 5.3) Show that 2^Ω is a σ-field.

Proof. Condition ((1)) holds because $\Omega \in 2^\Omega$. Condition ((2)) holds because $A \subseteq \Omega$ implies $A^c \subseteq \Omega$. Condition ((3)) holds because any union of subsets of Ω is again a subset of Ω. $\square$

Example 7.5 (Problem 5.4) Is the family $\mathscr{F}$ of all finite subsets of Ω and their complements $\mathscr{F} = \{A \subseteq \Omega \mid A \text{ finite or } A^c \text{ is finite}\}$ always a σ-field?

Solution. No. For a counterexample, take $\Omega = \mathbb{N}$ and consider the set of even numbers,

$$A = \{2\} \cup \{4\} \cup \dots.$$

Then neither A nor $\Omega \setminus A$ is finite. Since A is a countable union of singletons from $\mathscr{F}$ but $A \notin \mathscr{F}$, Condition ((3)) fails. $\square$

Example 7.6 (Problem 5.5) Is the family consisting of all countable subsets of Ω and their complements always a σ-field?

Solution. Yes. First, $\Omega^c = \emptyset$, and $\emptyset$ is countable, so Condition ((1)) holds. Next, if $A \in \mathscr{F}$, then either A is countable or A^c is countable, so Condition ((2)) follows immediately. Finally, if every A_i is countable, then $\bigcup_{i=1}^\infty A_i$ is countable. If at least one A_i has countable complement, then

$$\Omega \setminus \bigcup_{i=1}^\infty A_i \subseteq \Omega \setminus A_i,$$

so the complement of the union is countable. Therefore Condition ((3)) also holds. $\mathscr{F}$ is called the **countable-cocountable σ-field**. $\square$

Example 7.7 (Problem 5.6) Suppose that $\mathscr{F}$ is a σ-field on $\Omega = [0, 1]$ and that $\left[\frac{1}{n+1}, \frac{1}{n}\right] \in \mathscr{F}$ for all $n \in \mathbb{N}$. Show that the following sets belong to $\mathscr{F}$.

1. $\{0\} \in \mathscr{F}$.

2. $\{\frac{1}{n} \mid n = 2, 3, \dots\} \in \mathscr{F}$.
3. $(\frac{1}{n}, 1] \in \mathscr{F}$.
4. $(0, \frac{1}{n}] \in \mathscr{F}$ for all n.

Solution. For ((1)), define

$$I_n = \left[\frac{1}{n+1}, \frac{1}{n}\right], \quad n \geq 1.$$

By assumption, each $I_n \in \mathscr{F}$. Also,

$$(0, 1] = \bigcup_{n=1}^{\infty} I_n,$$

so $(0, 1] \in \mathscr{F}$ by Condition ((3)). Since $\Omega = [0, 1]$,

$$\{0\} = \Omega \setminus (0, 1] = (0, 1]^c \in \mathscr{F}.$$

For ((2)), for each $n \geq 2$,

$$\left[\frac{1}{n+1}, \frac{1}{n}\right] \cap \left[\frac{1}{n}, \frac{1}{n-1}\right] = \left\{\frac{1}{n}\right\}.$$

Each interval is in $\mathscr{F}$. Because σ-fields are fields by Example 7.2, finite intersections of sets in $\mathscr{F}$ are also in $\mathscr{F}$, so each singleton $\{1/n\}$ is in $\mathscr{F}$. Therefore,

$$\left\{\frac{1}{n} \mid n = 2, 3, \dots\right\} = \bigcup_{n=2}^{\infty} \left\{\frac{1}{n}\right\} \in \mathscr{F},$$

by Condition ((3)).

For ((3)), fix $n \geq 1$. If $n = 1$, then $(1, 1] = \emptyset \in \mathscr{F}$. If $n \geq 2$, then

$$\left(\frac{1}{n}, 1\right] = \bigcup_{k=1}^{n-1} \left[\frac{1}{k+1}, \frac{1}{k}\right].$$

Each interval on the right is in $\mathscr{F}$, and finite unions are in $\mathscr{F}$ by Condition ((3)), so $(1/n, 1] \in \mathscr{F}$.

For ((4)), fix $n \geq 1$. Then

$$\left(0, \frac{1}{n}\right] = \bigcup_{k=n}^{\infty} \left[\frac{1}{k+1}, \frac{1}{k}\right].$$

Again each interval belongs to $\mathscr{F}$, so by Condition ((3)) we get $(0, 1/n] \in \mathscr{F}$. $\square$

Lecture 8 — Tuesday, October 08, 2013

Generated σ-Fields and Borel Sets

Example 8.1 (Problem 5.7) Let $\mathscr{F}$ be a σ-field on Ω. If $A_1, A_2, \dots \in \mathscr{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathscr{F}$.

Solution. Suppose $A_1, A_2, \dots \in \mathscr{F}$. By Condition ((2)), $A_1^c, A_2^c, \dots \in \mathscr{F}$. By Condition ((3)), we have $\bigcup_{i=1}^{\infty} A_i^c \in \mathscr{F}$. Then

$$\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i \in \mathscr{F}.$$

$\square$

Let Ω and $\tilde{\Omega}$ be sets, and let $X : \tilde{\Omega} \to \Omega$ be any function.

$$\tilde{\Omega} \xrightarrow{X} \Omega$$

For $A \subseteq \Omega$, define the **inverse image** (or **preimage**) of A under X to be the set

$$X^{-1}(A) = \{\tilde{\omega} \in \tilde{\Omega} \mid X(\tilde{\omega}) \in A\},$$

The basic properties are $X^{-1}(\Omega) = \tilde{\Omega}$, $X^{-1}(A^c) = X^{-1}(A)^c$, and

$$X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n).$$

Proposition 8.2 If $\mathscr{F}$ is a σ-field on Ω, then

$$\tilde{\mathscr{F}} = \{X^{-1}(A) : A \in \mathscr{F}\}$$

is a σ-field on $\tilde{\Omega}$.

Proof. We check the three axioms.

1. Since $\Omega \in \mathscr{F}$, we have $X^{-1}(\Omega) = \tilde{\Omega} \in \tilde{\mathscr{F}}$.
2. Suppose $\tilde{A} \in \tilde{\mathscr{F}}$. Then $\tilde{A} = X^{-1}(A)$ for some $A \in \mathscr{F}$. Since $A^c \in \mathscr{F}$, it follows that
$$\tilde{A}^c = (X^{-1}(A))^c = X^{-1}(A^c) \in \tilde{\mathscr{F}}.$$
3. If $\tilde{A}_n = X^{-1}(A_n) \in \tilde{\mathscr{F}}$ for every n, then
$$\bigcup_{n=1}^{\infty} \tilde{A}_n = X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) \in \tilde{\mathscr{F}},$$
because $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$.

□

Example 8.3 (Problem 5.9) Suppose $\mathscr{F}$ is a σ-field on Ω, and $A \subseteq \Omega$. Show that $\tilde{\mathscr{F}} = \{A \cap B \mid B \in \mathscr{F}\}$ is a σ-field on A.

Solution. We need to check the axioms. Since $\mathscr{F}$ is a σ-field, $\Omega \in \mathscr{F}$. Then $A = A \cap \Omega \in \tilde{\mathscr{F}}$. Next, suppose $\tilde{B} \in \tilde{\mathscr{F}}$. We can write $\tilde{B} = A \cap B$ for some $B \in \mathscr{F}$. Since $\mathscr{F}$ is a σ-field, we have $B^c \in \mathscr{F}$. Then $A \setminus \tilde{B} = A \cap B^c \in \tilde{\mathscr{F}}$, as required. Finally, suppose $\tilde{B}_1, \tilde{B}_2, \cdots \in \tilde{\mathscr{F}}$. Then there exist $B_1, B_2, \cdots \in \mathscr{F}$ such that $B_i \cap A = \tilde{B}_i$ for all $i \in \mathbb{N}$. Therefore,

$$\bigcup_{i=1}^{\infty} \tilde{B}_i = \bigcup_{i=1}^{\infty} (B_i \cap A) = A \cap \left(\bigcup_{i=1}^{\infty} B_i\right).$$

Since $\bigcup_{i=1}^{\infty} B_i \in \mathscr{F}$ by Condition ((3)), it follows that $\bigcup_{i=1}^{\infty} \tilde{B}_i \in \tilde{\mathscr{F}}$. □

Example 8.4 (Problem 5.10) Let $\tilde{\Omega}$ and Ω be arbitrary sets, and let $X : \tilde{\Omega} \to \Omega$ be any function. Show that if $\tilde{\mathscr{F}}$ is a σ-field on $\tilde{\Omega}$, then

$$\mathscr{F} = \{A \subseteq \Omega \mid X^{-1}(A) \in \tilde{\mathscr{F}}\}$$

is a σ-field on Ω.

Solution. We check the three axioms. Since $X^{-1}(\Omega) = \tilde{\Omega} \in \tilde{\mathscr{F}}$, we have $\Omega \in \mathscr{F}$. If $A \in \mathscr{F}$, then

$$X^{-1}(A^c) = X^{-1}(A)^c \in \tilde{\mathscr{F}},$$

so $A^c \in \mathscr{F}$. Finally, if $A_1, A_2, \dots \in \mathscr{F}$, then

$$X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n) \in \tilde{\mathscr{F}}.$$

Hence $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$, proving that $\mathscr{F}$ is a σ-field. $\square$

Definition 8.5 Let $\mathcal{A}$ be any collection of subsets of Ω. Define $\mathscr{F}_{\mathcal{A}} = \bigcap\{\mathscr{F} \mid \mathscr{F} \text{ is a } \sigma\text{-field on } \Omega \text{ and } \mathcal{A} \subseteq \mathscr{F}\}$. Then $\mathscr{F}_{\mathcal{A}}$ is called the **minimal $\boldsymbol{\sigma}$-field generated by $\boldsymbol{\mathcal{A}}$**, or simply the **$\boldsymbol{\sigma}$-field generated by $\boldsymbol{\mathcal{A}}$**.

The defining family is nonempty because it contains 2^{Ω}, and an arbitrary intersection of σ-fields on Ω is again a σ-field. Thus $\mathscr{F}_{\mathcal{A}}$ is indeed the smallest σ-field containing $\mathcal{A}$.

Example 8.6 Let $\Omega = \mathbb{R}$. Fix $a \in \mathbb{R}$ and let $\mathcal{A} = \{(-\infty, a]\}$. Then $\mathscr{F}_{\mathcal{A}} = \{\emptyset, (-\infty, a], (a, \infty), \mathbb{R}\}$.

Example 8.7 Let $\Omega = \mathbb{R}$ and let $\mathcal{A} = \{(-\infty, a], (-\infty, b]\}$, where $b > a$. Then

$$\mathscr{F}_{\mathcal{A}} = \{\emptyset, \mathbb{R}, (-\infty, a], (-\infty, b], (a, \infty), (b, \infty), (-\infty, a] \cup (b, \infty), (a, b]\}.$$

Everything in there is there by necessity.

Example 8.8 Let $\Omega = \mathbb{R}$ and let

$$\mathcal{A} = \{(a, b) \mid -\infty \le a \le b \le \infty\}.$$

What is $\mathscr{F}_{\mathcal{A}}$? Since $[a, b] = \bigcap_{n=1}^{\infty}(a - 1/n, b + 1/n)$, we have $[a, b] \in \mathscr{F}_{\mathcal{A}}$. The Cantor set is in $\mathscr{F}_{\mathcal{A}}$. All of those nice fractals are in $\mathscr{F}_{\mathcal{A}}$. Also, $\mathbb{Q} \in \mathscr{F}_{\mathcal{A}}$, since

$$\mathbb{Q} = \bigcup_{i=1}^{\infty}\{q_i\} = \bigcup_{i=1}^{\infty}[q_i, q_i].$$

The irrationals $\mathbb{R} \setminus \mathbb{Q}$ are in there. The real transcendental numbers are in there.

Everything is in there!

Actually no, that is not true. Assuming the Axiom of Choice, it can be shown that not *every* subset of $\mathbb{R}$ is in $\mathscr{F}_{\mathcal{A}}$. The sets that are in $\mathscr{F}_{\mathcal{A}}$ are called the **Borel sets**, and there are *lots* of them. To work with volumes, areas, and probability, we need to use the Borel sets. Otherwise we could encounter pathologies such as the Banach–Tarski paradox.

Lecture 9 — Thursday, October 10, 2013

Recall that if $\mathcal{A}$ is a collection of subsets of Ω, then

$$\mathscr{F}_{\mathcal{A}} = \bigcap_{\substack{\mathscr{F} \text{ is a } \sigma\text{-field} \\ \mathcal{A} \subseteq \mathscr{F}}} \mathscr{F}$$

is the smallest σ-field containing $\mathcal{A}$. In particular, when $\Omega = \mathbb{R}$ and $\mathcal{A}$ is the collection of open intervals, $\mathscr{F}_{\mathcal{A}}$ is the collection of Borel sets. There is also a (much) larger class of sets, namely the **Lebesgue measurable sets**, that contains the Borel sets; however, those are not our focus here.

6. Countably Additive Probability

Countably Additive Probability Measures

Definition 9.1 Let $\mathscr{F}$ be a σ-field on Ω. We call a function $\mathbb{P} : \mathscr{F} \to [0, 1]$ a **probability measure** if the two conditions hold:

1. $\mathbb{P}(\Omega) = 1$, and
2. For any sequence $A_1, A_2, \cdots \in \mathscr{F}$ of pairwise disjoint sets (i.e., $A_i \cap A_j = \emptyset$ when $i \neq j$), we have
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$
(**countable additivity**).

This is a generalization of the finitely additive probability measures from Chapter 4.

Example 9.2 (Problem 6.1) Every probability measure is finitely additive. That is, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ whenever $A \cap B = \emptyset$.

Solution. First take $A_i = \emptyset$ for every i. Countable additivity gives

$$\mathbb{P}(\emptyset) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset),$$

which forces $\mathbb{P}(\emptyset) = 0$. Now suppose that $A \cap B = \emptyset$ and define

$$A_1 = A, A_2 = B, A_3 = A_4 = \cdots = \emptyset.$$

This sequence is pairwise disjoint, so

$$\mathbb{P}(A \cup B) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \cdots = \mathbb{P}(A) + \mathbb{P}(B).$$

$\square$

Example 9.3 (Problem 6.5) Let $\Omega = \mathbb{N}$ and let $\mathscr{F} = 2^{\Omega}$, the set of all subsets of Ω. Define $\mathbb{P}(\{i\}) = 2^{-i}$ for $i \in \mathbb{N}$. Can $\mathbb{P}$ be extended to a probability measure on $\mathscr{F}$?

Solution. Yes. For any $A \subseteq \mathbb{N}$, define

$$\mathbb{P}(A) = \sum_{i \in A} 2^{-i}.$$

This series is absolutely convergent, so it is well defined. It extends the original definition because

$$\mathbb{P}(\{i\}) = \sum_{j \in \{i\}} 2^{-j} = 2^{-i}.$$

To check that this is a probability measure, note first that

1. $$\mathbb{P}(\Omega) = \sum_{i \in \mathbb{N}} 2^{-i} = \sum_{i=1}^{\infty} 2^{-i} = 1.$$

2. If $A_1, A_2, \ldots$ are pairwise disjoint subsets of $\mathbb{N}$, then each integer belongs to at most one A_n, and therefore
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{i \in \bigcup_n A_n} 2^{-i} = \sum_{n=1}^{\infty} \sum_{i \in A_n} 2^{-i} = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$
The rearrangement is valid because all terms are nonnegative.

$\square$

Example 9.4 (Problem 6.8: Continuity from below) If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, then

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

Proof. Define $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, and in general $B_n = A_n \setminus A_{n-1}$. We can think of them as disjoint annuli. The decomposition into disjoint annuli is shown in Figure 1. Then

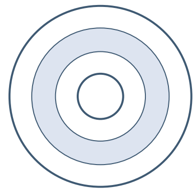

FIGURE 1

$$\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k$$

and

$$\bigcup_{k=1}^{n} B_k = \bigcup_{k=1}^{n} A_k = A_n.$$

Then

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(B_k) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

$\square$

Example 9.5 (Problem 6.9: Continuity from above) If $A_1 \supseteq A_2 \supseteq \dots$, then

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Solution. Define $B_n = A_1 \setminus A_n$ for $n \geq 1$. Since $A_1 \supseteq A_2 \supseteq \dots$, we have

$$B_1 \subseteq B_2 \subseteq \dots.$$

Therefore, by continuity from below (Example 9.4),

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Now

$$\bigcup_{n=1}^{\infty} B_n = A_1 \setminus \bigcap_{n=1}^{\infty} A_n,$$

so

$$\mathbb{P}\left(A_1 \setminus \bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_1 \setminus A_n).$$

Using $\mathbb{P}(A_1 \setminus C) = \mathbb{P}(A_1) - \mathbb{P}(C)$ whenever $C \subseteq A_1$, we get

$$\mathbb{P}(A_1) - \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \big(\mathbb{P}(A_1) - \mathbb{P}(A_n)\big).$$

Hence,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

$\square$

Example 9.6 (Problem 6.10: Countable subadditivity) For any $A_1, A_2, \ldots$, show that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Solution. Define

$$C_1 = A_1 \qquad \text{and} \qquad C_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k, \quad n \geq 2.$$

Then $C_n \subseteq A_n$ for each n, and $C_1, C_2, \ldots$ are pairwise disjoint. Also,

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} C_n.$$

By countable additivity on the disjoint family (C_n),

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} C_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(C_n).$$

Since $C_n \subseteq A_n$, monotonicity gives $\mathbb{P}(C_n) \leq \mathbb{P}(A_n)$ for each n. Therefore,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(C_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

$\square$

Lecture 10 — Tuesday, October 15, 2013

7. Conditional Probability and Independence

Conditional Probability

Definition 10.1 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $B \in \mathscr{F}$ be an event such that $\mathbb{P}(B) \neq 0$. Define the **conditional probability of A given B** by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ for all } A \in \mathscr{F}.$$

Warning Conditional probability is *not* symmetric: in general, $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$.

In the special case where Ω is finite, $\mathbb{P}$ is **uniform**, and $\emptyset \neq B \subseteq \Omega$, we define

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

Example 10.2 (Problem 7.3) Two fair dice are rolled.

1. What is the probability that the sum is at least 5?
2. Suppose we know that at least one die shows 1. How likely is it that the sum is at least 5?

Solution. For ((1)), for $2 \leq j \leq 12$,

$$\mathbb{P}(\text{sum is } j) = \frac{6 - |7 - j|}{36}.$$

Hence,

$$\mathbb{P}(\text{sum is at least } 5) = 1 - \mathbb{P}(\text{sum} \in \{2,3,4\}) = 1 - \left(\frac{1}{36} + \frac{2}{36} + \frac{3}{36}\right) = \frac{30}{36} = \frac{5}{6}.$$

For ((2)), let A be the event that the sum is at least 5, and let B be the event that at least one die shows "1". Then

$$B = \{(1,j) \mid j = 1, \ldots, 6\} \cup \{(i,1) \mid i = 1, \ldots, 6\} = B_1 \cup B_2.$$

Therefore, we obtain

$$\mathbb{P}(B) = \mathbb{P}(B_1) + \mathbb{P}(B_2) - \mathbb{P}(B_1 \cap B_2) = \frac{6}{36} + \frac{6}{36} - \frac{1}{36} = \frac{11}{36}.$$

The intersection is

$$A \cap B = \{(1,4), (1,5), (1,6), (4,1), (5,1), (6,1)\},$$

so $\mathbb{P}(A \cap B) = 6/36 = 1/6$. Therefore

$$\mathbb{P}(A|B) = \frac{6/36}{11/36} = 6/11.$$

□

Exercise 10.3 (Problem 7.4) You have heard that an old friend of yours has two children, one of whom is a girl, but you do not know the sex of the other child. How likely is it that it is a boy?

Example 10.4 (Problem 7.5) Show that $\mathbb{P}(A|\Omega) = \mathbb{P}(A)$.

Solution. We compute

$$\mathbb{P}(A|\Omega) = \frac{\mathbb{P}(A \cap \Omega)}{\mathbb{P}(\Omega)} = \frac{\mathbb{P}(A)}{1} = \mathbb{P}(A).$$

□

Example 10.5 (Problem 7.6) If $B \subseteq A$ and $\mathbb{P}(B) \neq 0$, then $\mathbb{P}(A \mid B) = 1$.

Solution. Since $B \subseteq A$, we have $A \cap B = B$. Therefore

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

□

Example 10.6 (Problem 7.7) If $A \cap B = \emptyset$ and $\mathbb{P}(B) \neq 0$, then $\mathbb{P}(A \mid B) = 0$.

Solution.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = 0.$$

□

Example 10.7 (Problem 7.8) If $A \cap B = \emptyset$ and $\mathbb{P}(C) \neq 0$, then

$$\mathbb{P}(A \cup B \mid C) = \mathbb{P}(A \mid C) + \mathbb{P}(B \mid C).$$

Solution. We compute

$$\begin{aligned}
\mathbb{P}(A \cup B|C) &= \frac{\mathbb{P}((A \cup B) \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}((A \cap C) \cup (B \cap C))}{\mathbb{P}(C)} \\
&= \frac{\mathbb{P}(A \cap C) + \mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} + \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} \\
&= \mathbb{P}(A|C) + \mathbb{P}(B|C).
\end{aligned}$$

□

By induction, Example 10.7 shows that conditional probability measures are finitely additive in their first argument. In fact, they are genuine probability measures:

Example 10.8 (Problem 7.9) Show that $\mathbb{P}(\cdot \mid B) : \mathscr{F} \to \mathbb{R}$, given by $A \mapsto \mathbb{P}(A \mid B)$, is a probability measure on $\mathscr{F}$.

Solution. We need to show that $0 \leq \mathbb{P}(A \mid B) \leq 1$, that $\mathbb{P}(\Omega \mid B) = 1$, and that if $A_1, A_2, \dots \in \mathscr{F}$

are disjoint, then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B).$$

First, for any $A \in \mathscr{F}$,

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

Since $\mathbb{P}(B) > 0$ and $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, it follows that

$$0 \leq \mathbb{P}(A \mid B) \leq 1.$$

Next,

$$\mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

Finally, let $A_1, A_2, \dots \in \mathscr{F}$ be pairwise disjoint. Then $(A_i \cap B)_{i \geq 1}$ is also pairwise disjoint, so by countable additivity of $\mathbb{P}$,

$$\mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B).$$

Dividing by $\mathbb{P}(B) > 0$ gives

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B).$$

Therefore $A \mapsto \mathbb{P}(A \mid B)$ is a probability measure on $\mathscr{F}$. □

Example 10.9 (Problem 7.10) Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $B \in \mathscr{F}$ with $\mathbb{P}(B) > 0$. Then $(B, \{A \cap B \mid A \in \mathscr{F}\}, \mathbb{P}(\cdot \mid B))$ is a probability space.

Solution. Let

$$\mathscr{F}_B := \{A \cap B \mid A \in \mathscr{F}\}.$$

By Problem 5.9, $\mathscr{F}_B$ is a σ-field on B.

Now define a function on $\mathscr{F}_B$ by

$$\mathbb{P}_B(C) := \mathbb{P}(C \mid B), \quad C \in \mathscr{F}_B.$$

This is well defined because each $C \in \mathscr{F}_B$ is itself an event in $\mathscr{F}$ (indeed $C = A \cap B$ for some $A \in \mathscr{F}$), so the conditional probability $\mathbb{P}(C \mid B)$ is already defined.

By Problem 7.9, the map $E \mapsto \mathbb{P}(E \mid B)$ is a probability measure on $\mathscr{F}$. Therefore its restriction to the sub-σ-field $\mathscr{F}_B \subseteq \mathscr{F}$ is also a probability measure. In particular,

$$\mathbb{P}_B(B) = 1,$$

and for pairwise disjoint $C_1, C_2, \dots \in \mathscr{F}_B$,

$$\mathbb{P}_B\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mathbb{P}_B(C_i).$$

Hence $(B, \mathscr{F}_B, \mathbb{P}_B)$ is a probability space, i.e.

$$(B, \{A \cap B \mid A \in \mathscr{F}\}, \mathbb{P}(\cdot \mid B))$$

is a probability space. □

Example 10.10 (Problem 7.11) If $\mathbb{P}(A \cap B) \neq 0$, prove that

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B).$$

Solution. The right-hand side equals

$$\mathbb{P}(A) \cdot \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \cdot \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(A \cap B)} = \mathbb{P}(A \cap B \cap C).$$

□

Example 10.11 (Problem 7.14) Suppose $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. Prove that

$$\mathbb{P}(A \mid B) \geq \mathbb{P}(A) \text{ if and only if } \mathbb{P}(B \mid A) \geq \mathbb{P}(B).$$

Solution. Since $\mathbb{P}(B) > 0$, the inequality $\mathbb{P}(A \mid B) \geq \mathbb{P}(A)$ is equivalent to

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A)\mathbb{P}(B).$$

Since $\mathbb{P}(A) > 0$, this in turn is equivalent to

$$\frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \geq \mathbb{P}(B),$$

which is precisely $\mathbb{P}(B \mid A) \geq \mathbb{P}(B)$. □

Lecture 11 — Thursday, October 17, 2013

Example 11.1 (Problem 7.16) A student must sit an examination consisting of 3 questions selected randomly from a list of 100 questions. To pass, the student must answer all three questions. What is the probability of passing if the student knows the answers to 90 questions on the list?

Solution. First, without using conditional probability, we obtain

$$\mathbb{P}(3 \text{ correct}) = \frac{\binom{90}{3}}{\binom{100}{3}}.$$

Now use conditional probability. Let A_i be the event that the ith question is answered correctly, for $i = 1, 2, 3$. Then

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap A_3) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2) \\ &= \left(\frac{90}{100}\right) \cdot \frac{89}{99} \cdot \frac{88}{98}.\end{aligned}$$

This is the same answer as before after simplification. □

Example 11.2 (Problem 7.17) Mrs. Jones has made a steak-and-kidney pie for her two sons. Eating more than half of the pie will give indigestion to anyone. While she is away, the older son takes a random amount, uniformly distributed over $[0, 1]$. Then the younger son takes a random amount of what remains, uniformly distributed over what is left. Mrs. Jones returns and finds that more than half of the pie is gone. What is the probability that neither son gets indigestion?

Warning This one is difficult, especially given that we have not yet discussed probability densities.

Solution. We begin with the sample space in Figure 2. Let S_1 denote the amount eaten by the

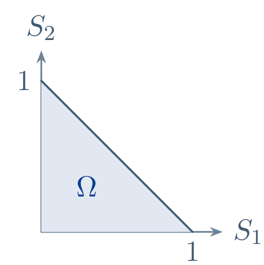

FIGURE 2

older son and let S_2 denote the amount eaten by the younger son. Then S_1 is uniformly distributed on $[0, 1]$, and S_2 is uniformly distributed on $[0, 1 - S_1]$. Let B be the event that more than 1/2 of the pie was eaten; this event is shaded in Figure 3. This gives the denominator. Let A be the

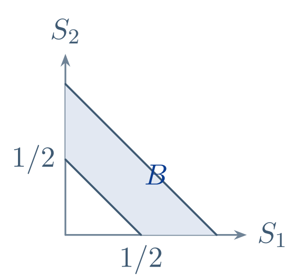

FIGURE 3

event that neither son gets indigestion; the intersection $A \cap B$ is shaded in Figure 4. We now compute the required probabilities by double integration. Consider a small box of area $\mathrm{d}x \cdot \mathrm{d}y$.

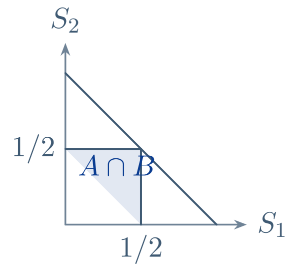

FIGURE 4

The probability that a point falls into that box is approximately

$$\begin{aligned}
&\mathbb{P}(x \leq S_1 \leq x + \mathrm{d}x \text{ and } y \leq S_2 \leq y + \mathrm{d}y) \\
&\quad = \mathbb{P}(x \leq S_1 \leq x + \mathrm{d}x) \cdot \mathbb{P}(y \leq S_2 \leq y + \mathrm{d}y \mid x \leq S_1 \leq x + \mathrm{d}x) \\
&\quad = \mathrm{d}x \cdot \frac{\text{height of box}}{\text{height of vertical strip}} \cong \mathrm{d}x \cdot \mathrm{d}y \frac{1}{1-x}.
\end{aligned}$$

Therefore, the denominator is

$$\begin{aligned}
\mathbb{P}(B) &= 1 - \mathbb{P}(B^c) \\
&= 1 - \int_0^{1/2} \int_0^{1/2-x} \frac{1}{1-x} \,\mathrm{d}y \,\mathrm{d}x \\
&= 1 - \int_0^{1/2} \left[\int_0^{1/2-x} \frac{1}{1-x} \,\mathrm{d}y \right] \mathrm{d}x \\
&= \log \sqrt{2e}
\end{aligned}$$

and the numerator is

$$\mathbb{P}(A \cap B) = \int_0^{1/2} \left[\int_{1/2-x}^{1/2} \frac{1}{1-x} \,\mathrm{d}y \right] \mathrm{d}x = \log 2 - \frac{1}{2}.$$

Hence, the conditional probability is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\log 2 - \frac{1}{2}}{\log \sqrt{2e}} = \frac{2\log 2 - 1}{1 + \log 2}.$$

$\square$

Exercise 11.3

(Problem 7.18) Groping about in the dark, we open one of two drawers in a chest and pick up an item of clothing at random. What is the probability that it is a sock if one drawer contains 6 socks and 6 underpants, and the other contains 2 socks and 4 handkerchiefs?

(Problem 7.19) Let $\Omega = H_1 \cup H_2$ and $H_1 \cap H_2 = \emptyset$. Assuming that $\mathbb{P}(A \mid H_1)$, $\mathbb{P}(A \mid H_2)$, $\mathbb{P}(H_1) \neq 0$, and $\mathbb{P}(H_2) \neq 0$ are known, find an expression for $\mathbb{P}(A)$.

The Law of Total Probability

Example 11.4 (Problem 7.20) Let $H_1, \ldots, H_n$ be pairwise disjoint events such that $\Omega = H_1 \cup \cdots \cup H_n$ and $\mathbb{P}(H_i) \neq 0$ for $i = 1, \ldots, n$. Show that, for any event A,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid H_i)\mathbb{P}(H_i).$$

Solution.

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\big(A \cap (H_1 \cup \cdots \cup H_n)\big).$$

Since the H_i partition Ω, the sets $A \cap H_1, \ldots, A \cap H_n$ are disjoint, so

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap H_i) = \sum_{i=1}^{n} \mathbb{P}(A \mid H_i) \cdot \mathbb{P}(H_i).$$

$\square$

Example 11.5 (Problem 7.21: Law of total probability) Suppose that $H_1, H_2, \ldots$ partition Ω and $\mathbb{P}(H_i) > 0$ for every i. Then

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid H_i) \cdot \mathbb{P}(H_i).$$

Solution. The proof works essentially the same way as in the previous problem. We have $A = \cup_{i=1}^{\infty}(A \cap H_i)$ and the sets $A \cap H_i$ are disjoint, so by countable additivity and the definition of conditional probability, we have

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap H_i) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid H_i) \cdot \mathbb{P}(H_i).$$

$\square$

Example 11.6 (Problem 7.22) A survey is carried out to estimate the proportion of bathroom singers. Each respondent rolls a die in secret and must answer NO if the roll is 1 and YES if the roll is 6, regardless of the truth; if the roll is 2, 3, 4, or 5, the respondent must answer truthfully. Suppose the observed probability of a YES answer is $\frac{2}{3}$. What is the probability that a person is a bathroom singer?

Solution. Let p denote the probability of being a bathroom singer. Using the total probability formula, we have

$$\begin{aligned} \mathbb{P}(\text{YES}) &= \sum_{i=1}^{6} \mathbb{P}(\text{YES} \mid \text{roll} = i) \cdot \mathbb{P}(\text{roll} = i) \\ &= 0\left(\frac{1}{6}\right) + 4p\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right). \end{aligned}$$

Therefore, we obtain

$$\frac{2}{3} = \frac{1}{6}(1 + 4p),$$

so $p = \frac{3}{4}$. $\square$

Theorem 11.7 (Bayes formula: Problem 7.27) Let $H_1, H_2, \ldots$ partition Ω such that $\mathbb{P}(H_n) \neq 0$ for any $n \in \mathbb{N}$ and suppose that $\mathbb{P}(A) \neq 0$. Then

$$\mathbb{P}(H_i \mid A) = \frac{\mathbb{P}(A \mid H_i) \cdot \mathbb{P}(H_i)}{\sum_{n \geq 1} \mathbb{P}(A \mid H_n) \cdot \mathbb{P}(H_n)}.$$

This is a central formula in Bayesian reasoning.

Proof. Since $\mathbb{P}(A) \neq 0$, conditional probability is defined and

$$\mathbb{P}(H_i \mid A) = \frac{\mathbb{P}(H_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap H_i)}{\mathbb{P}(A)}.$$

Now, use the definition

$$\mathbb{P}(A \cap H_i) = \mathbb{P}(A \mid H_i)\mathbb{P}(H_i),$$

which is valid because $\mathbb{P}(H_i) \neq 0$ by assumption. Substituting into the previous display gives

$$\mathbb{P}(H_i \mid A) = \frac{\mathbb{P}(A \mid H_i)\mathbb{P}(H_i)}{\mathbb{P}(A)}.$$

Finally, by the law of total probability,

$$\mathbb{P}(A) = \sum_{n \geq 1} \mathbb{P}(A \mid H_n)\mathbb{P}(H_n).$$

Replacing $\mathbb{P}(A)$ in the denominator yields

$$\mathbb{P}(H_i \mid A) = \frac{\mathbb{P}(A \mid H_i)\mathbb{P}(H_i)}{\sum_{n \geq 1} \mathbb{P}(A \mid H_n)\mathbb{P}(H_n)}.$$

□

Lecture 12 — Tuesday, October 22, 2013

Example 12.1 (Problem 7.30) In the survey from Problem 7.22, suppose the probability of being a bathroom singer is $\frac{3}{4}$. How likely is it that a person who answered YES is indeed a bathroom singer? How likely is it that a person who answered NO is a bathroom singer?

Solution. Let H_1 be the event that a person is a bathroom singer, let H_2 be the complementary event, and let A be the event that the person answered YES. Then

$$\mathbb{P}(H_1 \mid A) = \frac{\mathbb{P}(A \mid H_1) \cdot \mathbb{P}(H_1)}{\mathbb{P}(A \mid H_1)\mathbb{P}(H_1) + \mathbb{P}(A \mid H_2)\mathbb{P}(H_2)} = \frac{(5/6)(3/4)}{(5/6)(3/4) + (1/6)(1/4)} = \frac{15}{16}.$$

Note that this is greater than $\frac{3}{4}$, so answering yes increases the likelihood that the person is a bathroom singer.

For a NO answer,

$$\mathbb{P}(H_1 \mid A^c) = \frac{(1/6)(3/4)}{(1/6)(3/4) + (5/6)(1/4)} = \frac{3}{8}.$$

Thus a NO answer lowers that likelihood from 3/4 to 3/8. □

Independence of Events

Definition 12.2 Events A and B are said to be **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Example 12.3 (Problem 7.32) Suppose that $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$. Show that the following conditions are equivalent:

1. $\mathbb{P}(A) = \mathbb{P}(A \mid B)$,
2. $\mathbb{P}(B) = \mathbb{P}(B \mid A)$,
3. $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Solution. We prove

$$((1)) \iff ((3)) \iff ((2)).$$

For $((1)) \Rightarrow ((3))$, start from

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

Since $\mathbb{P}(B) \neq 0$, multiply by $\mathbb{P}(B)$ to get

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

For $((3)) \Rightarrow ((1))$, if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

then dividing by $\mathbb{P}(B) \neq 0$ gives

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

For $((2)) \Leftrightarrow ((3))$, the same argument with A and B interchanged yields

$$\mathbb{P}(B) = \mathbb{P}(B \mid A) \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

because $\mathbb{P}(A) \neq 0$.

Hence all three statements are equivalent. $\square$

Example 12.4 (Problem 7.34) Show that for any events A and B, the following are equivalent:

1. A and B are independent,
2. A^c and B are independent,
3. A^c and B^c are independent.

Solution. We prove

$$((1)) \iff ((2)) \iff ((3)).$$

For $((1)) \Rightarrow ((2))$, assume

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

Then

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B),$$

so A^c and B are independent.

For $((2)) \Rightarrow ((1))$, assume

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B).$$

Since $B = (A \cap B) \cup (A^c \cap B)$ (disjoint union),

$$\mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A^c)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B),$$

so A and B are independent.

For $((2)) \Rightarrow ((3))$, assume

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B).$$

Because $A^c = (A^c \cap B) \cup (A^c \cap B^c)$ disjointly,

$$\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c) - \mathbb{P}(A^c \cap B) = \mathbb{P}(A^c) - \mathbb{P}(A^c)\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B^c),$$

so A^c and B^c are independent.

For ((3)) $\Rightarrow$ ((2)), assume

$$\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)\mathbb{P}(B^c).$$

Again from $A^c = (A^c \cap B) \cup (A^c \cap B^c)$ disjointly,

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c) - \mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c) - \mathbb{P}(A^c)\mathbb{P}(B^c) = \mathbb{P}(A^c)\mathbb{P}(B),$$

so A^c and B are independent.

Hence all three statements are equivalent. $\square$

Example 12.5 (Problem 7.35) Let $\Omega = \{1, 2, 3, 4\}$ with uniform probability, and let $A = \{1, 2\}$. List all $B \subseteq \Omega$ such that A and B are independent.

Solution. Since $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, we have

$$\frac{\#(A \cap B)}{\#\Omega} = \frac{\#A}{\#\Omega} \cdot \frac{\#B}{\#\Omega} \implies \frac{\#(A \cap B)}{4} = \frac{2}{4} \cdot \frac{\#B}{4}.$$

Therefore $2 \cdot \#(A \cap B) = \#B$. There are three cases. If $\#(A \cap B) = 0$, then $\#B = 0$, so $B = \emptyset$. If $\#(A \cap B) = 1$, then $\#B = 2$, so

$$B \in \{\{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}.$$

If $\#(A \cap B) = 2$, then $\#B = 4$, so $B = \Omega$. $\square$

The following is a number-theoretic extension of Example 12.5.

Example 12.6 (Problem 7.37) Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, let $\mathbb{P}(\{\omega\}) = 1/n$ for $\omega \in \Omega$, and let $A, B \subset \Omega$ be independent. Show that if A has i elements, then B must have

$$j = k\frac{n}{\gcd(i, n)}$$

elements, where $k \in \{0, 1, \ldots, \gcd(n, i)\}$.

Proof. Let $|A| = i$, $|B| = j$, and $|A \cap B| = m$. Under the uniform measure,

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{i}{n}, \qquad \mathbb{P}(B) = \frac{\#B}{\#\Omega} = \frac{j}{n}, \qquad \text{and} \qquad \mathbb{P}(A \cap B) = \frac{\#(A \cap B)}{\#\Omega} = \frac{m}{n}.$$

Since A and B are independent,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \implies \frac{m}{n} = \frac{i}{n} \cdot \frac{j}{n},$$

so $m = \frac{ij}{n}$. Because m is an integer, $n \mid ij$. Now, write $d = \gcd(i, n)$, $i = di_1$, $n = dn_1$, with $\gcd(i_1, n_1) = 1$. The divisibility $n \mid ij$ becomes

$$dn_1 \mid di_1 j \implies n_1 \mid i_1 j.$$

Since $\gcd(i_1, n_1) = 1$, it follows that $n_1 \mid j$. Hence for some integer k,

$$j = kn_1 = k\frac{n}{d} = k\frac{n}{\gcd(i, n)}.$$

Finally, $0 \le j \le n$ implies

$$0 \le k\frac{n}{d} \le n \implies 0 \le k \le d,$$

so

$$k \in \{0, 1, \ldots, d\} = \{0, 1, \ldots, \gcd(i, n)\}.$$

$\square$

Example 12.7 (Problem 7.39) Let A be an event. Prove that the following are equivalent:

1. A is independent of every event B,
2. $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Solution. We have A independent of itself if and only if $\mathbb{P}(A) \in \{0, 1\}$. Indeed, if A is independent of itself, then

$$\mathbb{P}(A \cap A) = \mathbb{P}(A)^2,$$

so $\mathbb{P}(A) = \mathbb{P}(A)^2$, and hence $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Conversely, if $\mathbb{P}(A) = 0$, then $A \cap B \subseteq A$ implies $\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A)\mathbb{P}(B)$. If $\mathbb{P}(A) = 1$, then

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) = \mathbb{P}(B \cap A),$$

because $\mathbb{P}(A^c) = 0$. Hence $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ in either case. □

Example 12.8 (Problem 7.40) A fair coin is tossed three times. Define events

$$A = \{\text{tosses 1 and 2 differ}\},$$
$$B = \{\text{tosses 2 and 3 differ}\}, \qquad \text{and} \qquad C = \{\text{tosses 3 and 1 differ}\}.$$

Show that $\mathbb{P}(A) = \mathbb{P}(A \mid B) = \mathbb{P}(A \mid C)$, but $\mathbb{P}(A) \neq \mathbb{P}(A \mid B \cap C)$.

Solution. We have $\mathbb{P}(A) = 1/2$. Also,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/4}{1/2} = 1/2$$

and

$$\mathbb{P}(A \mid C) = \frac{1/4}{1/2} = 1/2.$$

However,

$$\mathbb{P}(A \mid B \cap C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \frac{0}{1/4} = 0 \neq 1/2.$$

Note that $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C)$, and $\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C)$, but $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$. □

Definition 12.9 Events A, B, and C are **independent** if every pair of events is independent and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$.

Example 12.10 (Problem 7.43) Let A, B, and C be independent events. Show that

1. A and $B \cap C$ are independent,
2. A and $B \cup C$ are independent,
3. A and $B \setminus C$ are independent.

Solution. For ((1)),

$$\mathbb{P}(A \cap (B \cap C)) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \mathbb{P}(A) \cdot \mathbb{P}(B \cap C),$$

by the definition of independence, so A and $B \cap C$ are independent.

For ((2)),

$$\begin{aligned}
\mathbb{P}(A \cap (B \cup C)) &= \mathbb{P}[(A \cap B) \cup (A \cap C)] \\
&= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) - \mathbb{P}(A \cap B \cap C) \\
&= \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \\
&= \mathbb{P}(A)[\mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C)] \\
&= \mathbb{P}(A)\mathbb{P}(B \cup C).
\end{aligned}$$

For ((3)),

$$\begin{aligned}
\mathbb{P}(A \cap (B \setminus C)) &= \mathbb{P}((A \cap B) \setminus (A \cap B \cap C)) \\
&= \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B \cap C) \\
&= \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \\
&= \mathbb{P}(A)\left(\mathbb{P}(B) - \mathbb{P}(B \cap C)\right) \\
&= \mathbb{P}(A)\mathbb{P}(B \setminus C).
\end{aligned}$$

□

Lecture 13 — Thursday, October 24, 2013

Definition 13.1 For $n \geq 2$, the events $A_1, A_2, \ldots, A_n$ are said to be **independent** if for every subset $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$, where $k \geq 2$, we have

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k}).$$

There are $2^n - n - 1$ conditions to check.

Example 13.2 (Problem 7.47: Grouping lemma) Let $A_1, \ldots, A_n, B_1, \ldots, B_m$ be $n+m$ independent events. Let $C \in \mathscr{F}_{\{A_1,\ldots,A_n\}}$ and $D \in \mathscr{F}_{\{B_1,\ldots,B_m\}}$. Then C and D are independent.

Proof. First observe that replacing any of a finite independent family of events by its complement preserves independence. Indeed, if E is one event and G is an intersection of any selection of the others, then

$$\mathbb{P}(E^c \cap G) = \mathbb{P}(G) - \mathbb{P}(E \cap G) = \mathbb{P}(G) - \mathbb{P}(E)\mathbb{P}(G) = \mathbb{P}(E^c)\mathbb{P}(G).$$

Repeating this argument handles any number of complements.

The atoms of $\mathscr{F}_{\{A_1,\ldots,A_n\}}$ are the nonempty sets of the form

$$C_{\boldsymbol{\varepsilon}} = \bigcap_{i=1}^{n} A_i^{(\varepsilon_i)}, \qquad A_i^{(1)} = A_i, \qquad A_i^{(0)} = A_i^c,$$

and the analogous intersections $D_{\boldsymbol{\delta}}$ are the atoms of $\mathscr{F}_{\{B_1,\ldots,B_m\}}$. The preceding observation gives

$$\mathbb{P}(C_{\boldsymbol{\varepsilon}} \cap D_{\boldsymbol{\delta}}) = \mathbb{P}(C_{\boldsymbol{\varepsilon}})\mathbb{P}(D_{\boldsymbol{\delta}})$$

for every pair of atoms. Every C and D in the two generated σ-fields is a finite disjoint union of its respective atoms. Summing the displayed identity over those two finite collections yields

$$\mathbb{P}(C \cap D) = \mathbb{P}(C)\mathbb{P}(D),$$

so C and D are independent. □

8. Random Variables and their Distributions

Random Variables

Definition 13.3 Let $X : \Omega \to \mathbb{R}$ be a function where $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space. We say that X is a **random variable** if $\{\omega \in \Omega \mid a < X(\omega)\} \in \mathscr{F}$ for every $a \in \mathbb{R}$.

Note that $\{\omega \mid a < X(\omega)\} = X^{-1}((a, \infty))$. It is more convenient to write $\{X > a\}$ for this event, and more generally $\{X \in B\} = X^{-1}(B)$. Likewise,

$$\{X = a\} = \{\omega \in \Omega \mid X(\omega) = a\}.$$

A comma in this notation means "and," that is, intersection.

We can thus rewrite Definition 13.3: $X\colon \Omega \to \mathbb{R}$ is a random variable if and only if $\{X > a\} \in \mathscr{F}$ for all $a \in \mathbb{R}$.

Probabilities involving random variables are usually written with an abuse of notation as, for example,

$$\mathbb{P}(X > a) = \mathbb{P}(\{X > a\}) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) > a\}).$$

Example 13.4 (Problem 8.1) Show that $X : \Omega \to \mathbb{R}$ is a random variable if and only if $\{X \in B\} \in \mathscr{F}$ for every Borel set $B \subseteq \mathbb{R}$.

Solution. *Forward implication.* Suppose first that $\{X \in B\} \in \mathscr{F}$ for every Borel set B. Then X is a random variable, because $\{X > a\} = \{X \in (a, \infty)\}$, and (a, ∞) is a Borel set.

Reverse implication. Conversely, suppose that X is a random variable. We must show that $\{X \in B\} \in \mathscr{F}$ for every Borel set B. Let

$$\mathscr{L} = \{A \subseteq \mathbb{R} \mid \{X \in A\} \in \mathscr{F}\}.$$

By Problem 5.10, $\mathscr{L}$ is a σ-field. By assumption, $(a, \infty) \in \mathscr{L}$ for every $a \in \mathbb{R}$. For any open interval,

$$(a, b) = \bigcup_{n=1}^{\infty} \left[(a, \infty) \cap \left(b - \frac{1}{n}, \infty \right)^c \right],$$

so $(a, b) \in \mathscr{L}$. Thus $\mathscr{L}$ is a σ-field containing all open intervals. Since the Borel sets form the smallest such σ-field, we obtain $\mathscr{B} \subseteq \mathscr{L}$. Therefore $\{X \in B\} \in \mathscr{F}$ for every Borel set B. $\square$

Example 13.5 (Problem 8.3) Two dice are rolled. Let X be the larger of the two numbers shown. Compute $\mathbb{P}(\{X \in [2, 4]\})$.

Solution. Let

$$\Omega = \{(i, j) \mid 1 \leq i, j \leq 6\},$$

and let X be the maximum of the two numbers rolled. Thus $X\colon \Omega \to \mathbb{R}$, and for example $X((2, 3)) = 3$. We want to compute

$$\mathbb{P}(\{X \in [2, 4]\}) = \mathbb{P}(2 \leq X \leq 4).$$

We have

$$\{2 \le X \le 4\} = \{X = 2\} \cup \{X = 3\} \cup \{X = 4\}.$$

Also,

$$\{X = j\} = \{X \le j\} \setminus \{X \le j - 1\}.$$

Let X_i be the outcome of the ith roll, so $X = \max(X_1, X_2)$. Then

$$\mathbb{P}(X \le j) = \mathbb{P}(X_1 \le j, X_2 \le j) = \mathbb{P}(X_1 \le j) \cdot \mathbb{P}(X_2 \le j) = \left(\frac{j}{6}\right)\left(\frac{j}{6}\right) = \frac{j^2}{36}.$$

Hence $\mathbb{P}(X = j) = (\frac{j}{6})^2 - (\frac{j-1}{6})^2$, and so

$$\begin{aligned}\mathbb{P}(2 \le X \le 4) &= \left[\left(\tfrac{2}{6}\right)^2 - \left(\tfrac{1}{6}\right)^2\right] + \left[\left(\tfrac{3}{6}\right)^2 - \left(\tfrac{2}{6}\right)^2\right] + \left[\left(\tfrac{4}{6}\right)^2 - \left(\tfrac{3}{6}\right)^2\right] \\ &= \frac{15}{36}.\end{aligned}$$

□

Definition 13.6 A function m defined on the Borel sets of $\mathbb{R}$ which satisfies

$$m([a, b]) = m((a, b)) = m([a, b)) = m((a, b]) = b - a$$

and

$$m\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i)$$

for pairwise disjoint $A_1, A_2, \ldots$ is called **Lebesgue measure**.

Example 13.7 (Problem 8.4) Let $\Omega = [0, 1]$ with Borel sets and Lebesgue measure. Find $\mathbb{P}(\{X \in [0, \frac{1}{2})\})$ for $X(\omega) = \omega^2$.

Solution. Restated, let $\Omega = [0, 1]$, with Borel subsets of $[0, 1]$. We want the uniform probability measure on $[0, 1]$, namely $\mathbb{P}([a, b]) = b - a$, which is Lebesgue measure restricted to $[0, 1]$. Let $X(\omega) = \omega^2$. For $0 \le a < 1$,

$$\{X > a\} = \{\omega^2 > a\} = \{\omega > \sqrt{a}\} = (\sqrt{a}, 1] \in \mathscr{B}.$$

For $a < 0$ this event is Ω, and for $a \geq 1$ it is $\emptyset$. Thus X is a random variable. Now

$$\mathbb{P}(X \in [0, 1/2)) = \mathbb{P}(0 \leq X < 1/2) = \mathbb{P}(X < 1/2) = \mathbb{P}(\omega^2 < 1/2) = \mathbb{P}\left(\omega < \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}.$$

$\square$

Example 13.8 (Problem 8.5) Let X be the number of tosses of a fair coin up to and including the first toss showing heads. Find $\mathbb{P}(X \in 2\mathbb{N})$.

Solution. We want the probability that the number of tosses required to obtain the first head is even. We have

$$\mathbb{P}(X \in 2\mathbb{N}) = \sum_{n=1}^{\infty} \mathbb{P}(\{X = 2n\}) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{2n} = \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^{n} = \frac{1/4}{1 - 1/4} = \frac{1}{3}.$$

$\square$

Lecture 14 — Tuesday, October 29, 2013

Distributions and Distribution Functions

Example 14.1 (Problem 8.6) Show that if X is a constant function, then X is a random variable with respect to any σ-field.

Solution. Let $X(\omega) = c$ for all $\omega \in \Omega$. We use the fact that if $\mathscr{F}$ is a σ-field on Ω, then both Ω and $\emptyset$ are elements of $\mathscr{F}$. Then

$$\{X > a\} = \{\omega \mid X(\omega) > a\} = \begin{cases} \emptyset & c \leq a \\ \Omega & c > a. \end{cases}$$

Hence $\{X > a\} \in \mathscr{F}$ for all real values a. Therefore X is a random variable. $\square$

Example 14.2 (Problem 8.7) Let $\Omega = \{1,2,3,4\}$ and $\mathscr{F} = \{\emptyset, \Omega, \{1\}, \{2,3,4\}\}$. Is $X(\omega) = 1+\omega$ a random variable with respect to $\mathscr{F}$? If not, give an example of a nonconstant function that is.

Solution. Choose $a = 3$. Then

$$\{X > a\} = \{\omega \mid X(\omega) > 3\} = \{\omega \mid 1+\omega > 3\} = \{\omega \mid \omega > 2\} \notin \mathscr{F}.$$

Therefore this is not a random variable. For a nonconstant example, take $X(1) = 1$ and

$$X(2) = X(3) = X(4) = 2.$$

$\square$

Random variables are like mRNA: they transfer information from the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to the real line (the mRNA analogy is somewhat loose). If $X\colon \Omega \to \mathbb{R}$ is a random variable, then we can make the codomain into a probability space $(\mathbb{R}, \mathscr{B}, \mathbb{P}_X)$, where $\mathscr{B}$ is the family of Borel sets of $\mathbb{R}$ and $\mathbb{P}_X$ is the probability measure defined by

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\}).$$

Example 14.3 (Problem 8.14) Let X be a random variable. Show that the map

$$\mathbb{P}_X : B \mapsto \mathbb{P}(\{X \in B\})$$

is a probability measure on the Borel σ-field of $\mathbb{R}$.

Solution. We need to check the axioms. For any Borel set $B \in \mathscr{B}$, we have $\mathbb{P}_X(B) = \mathbb{P}(X \in B) \in [0,1]$. Also,

$$\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1.$$

For σ-additivity, suppose that $B_1, B_2, \dots$ are disjoint Borel subsets of $\mathbb{R}$. Then $\{X \in B_1\}, \{X \in B_2\}, \dots$ are disjoint events. Each $\{X \in B_i\}$ lies in $\mathscr{F}$ by Problem 8.1, and they are disjoint because

$$\{X \in B_i\} \cap \{X \in B_j\} = \{X \in B_i \cap B_j\} = \{X \in \emptyset\} = \emptyset$$

for $i \neq j$. Therefore

$$\mathbb{P}_X\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}\left(X \in \bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty}\{X \in B_i\}\right) = \sum_{i=1}^{\infty} \mathbb{P}(X \in B_i) = \sum_{i=1}^{\infty} \mathbb{P}_X(B_i).$$

□

Definition 14.4 We call the function $\mathbb{P}_X : \mathscr{B} \to [0,1]$ defined by $\mathbb{P}_X(B) = \mathbb{P}(\{X \in B\})$ the **distribution** of X.

Definition 14.5 Random variables, possibly defined on different probability spaces, with the same distribution are said to be **identically distributed**. In particular, $X_1, X_2, \dots, X_n \colon \Omega \to \mathbb{R}$ are identically distributed if

$$\mathbb{P}_{X_1} = \mathbb{P}_{X_2} = \cdots = \mathbb{P}_{X_n}.$$

Example 14.6 (Problem 8.15) Find an example of two different random variables X and X' with the same distribution.

Example. Take $\Omega = \{1,2,3,4\}$, $\mathscr{F} = 2^{\Omega}$, and $\mathbb{P}(\{i\}) = 1/4$. Define X_1 by $X_1(1) = 1$ and $X_1(i) = 0$ for $i \neq 1$. Define X_2 by $X_2(2) = 1$ and $X_2(i) = 0$ for $i \neq 2$. □

Definition 14.7 The **distribution function** (or **cumulative distribution function**) is the function $F_X \colon \mathbb{R} \to [0,1]$ defined by

$$F_X(x) = \mathbb{P}_X((-\infty, x]).$$

Example 14.8 (Problem 8.25) Let X be the number of tosses of a fair coin up to and including the first toss showing heads. Compute and plot the distribution function F_X.

Solution. Let $\Omega = \{H, TH, TTH, TTTH, \dots\}$. We can take $\mathscr{F} = 2^{\Omega}$. Also,

$$\mathbb{P}(\underbrace{T\dots T}_{n-1 \text{ times}} H) = \frac{1}{2^n}$$

and

$$X(\underbrace{T\ldots T}_{n-1 \text{ times}} H) = n.$$

Thus $\mathbb{P}(X = n) = \frac{1}{2^n}$, and

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(\{X \in (-\infty, x]\}) = \mathbb{P}(\{X \leq x\}) = \sum_{n=1}^{\lfloor x \rfloor} \frac{1}{2^n}$$
$$= \begin{cases} 0 & \text{if } x < 1, \\ 1 - \frac{1}{2^n} & \text{if } n \leq x < n+1,\ n \in \mathbb{N}. \end{cases}$$

A sketch of F_X appears as the right-continuous step function in Figure 5:

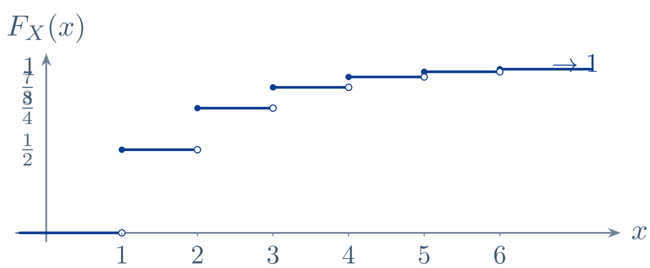

FIGURE 5

□

Example 14.9 (Problem 8.33) Find F_X if $\Omega = [0,1] \times [0,1]$ has the uniform area measure and $X(\omega)$ is the distance from $\omega \in \Omega$ to the nearest edge of the square.

Solution. Let $(x, y) \in [0,1] \times [0,1]$. Then $\omega = (x, y)$, and

$$\mathscr{F} = \{\text{everything you know and love}\} = \mathscr{B}(\Omega),$$

assuming you do not know and love any nonmeasurable subsets of $\mathbb{R}^2$. For $x < 0$, $\mathbb{P}(X \leq x) = 0$. For $x \geq 1/2$, $\mathbb{P}(X \leq x) = 1$. For $x \in [0, 1/2)$, $\mathbb{P}(X > x) = (1 - 2x)^2$, so $F_X(x) = 1 - (1 - 2x)^2$.

That is,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 4x - 4x^2, & 0 \leq x < 1/2, \\ 1, & x \geq 1/2. \end{cases}$$

$\square$

When working out a distribution function, make sure to cover all inputs of the function.

Proposition 14.10 The basic properties of a distribution function are as follows:

1. **Problem 8.26 (monotonicity)**: If $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_2)$.
2. **Problem 8.27**: $\lim_{x\to\infty} F_X(x) = 1$.
3. **Problem 8.28**: $\lim_{x\to-\infty} F_X(x) = 0$.
4. **Problem 8.29 (right-continuity)**: $\lim_{y\to x^+} F_X(y) = F_X(x)$.

Proof. For ((1)), if $x_1 \leq x_2$, then

$$(-\infty, x_1] \subseteq (-\infty, x_2].$$

Hence, by monotonicity of probability,

$$F_X(x_1) = \mathbb{P}_X((-\infty, x_1]) \leq \mathbb{P}_X((-\infty, x_2]) = F_X(x_2).$$

For ((2)), define $A_n := \{X \leq n\}$. Then $A_n \uparrow \Omega$, so by continuity from below (Problem 6.8),

$$\lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\Omega) = 1.$$

Since F_X is nondecreasing by ((1)), this implies

$$\lim_{x\to\infty} F_X(x) = 1.$$

For ((3)), define $B_n := \{X \leq -n\}$. Then $B_n \downarrow \emptyset$, so by continuity from above (Problem 6.9),

$$\lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(\emptyset) = 0.$$

Again using nondecreasingness from ((1)), we obtain

$$\lim_{x\to-\infty} F_X(x) = 0.$$

For ((4)), fix $x \in \mathbb{R}$ and let $y_n := x + \frac{1}{n}$. Then $y_n \downarrow x$, and

$$\{X \le y_n\} \downarrow \{X \le x\}.$$

Applying continuity from above (Problem 6.9) gives

$$\lim_{n\to\infty} F_X\left(x + \frac{1}{n}\right) = \lim_{n\to\infty} \mathbb{P}(X \le y_n) = \mathbb{P}(X \le x) = F_X(x).$$

If $x < y \le x + 1/n$, monotonicity gives

$$F_X(x) \le F_X(y) \le F_X(x + 1/n).$$

The squeeze theorem therefore upgrades the sequential calculation to $\lim_{y\to x^+} F_X(y) = F_X(x)$. □

Exercise 14.11

(Problem 8.31) Prove that F_X is continuous at y if and only if $\mathbb{P}_X(\{y\}) = 0$.

(Problem 8.32) Let $a < b$. Find expressions for the following probabilities in terms of F_X:

(a) $\mathbb{P}_X((a, b])$

(b) $\mathbb{P}_X((a, b))$

(c) $\mathbb{P}_X([a, b])$

(d) $\mathbb{P}_X([a, b))$

(Problem 8.32) Two towns A and B are 50 miles apart. Car 1 leaves town A and car 2 leaves town B independently of one another. The departure time of each car is random and uniformly distributed between 12 and 1 o'clock. The speed of each car is 100 miles per hour and they travel directly to meet one another. Denote by X the distance between the meeting point and A. Find F_X.

Lecture 15 — Thursday, October 31, 2013

Discrete Random Variables

Recall that

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X((-\infty, x]).$$

The function F_X is nondecreasing, satisfies $\lim_{x\to\infty} F_X(x) = 1$ and $\lim_{x\to-\infty} F_X(x) = 0$, and is right-continuous.

Definition 15.1 A random variable X is called **discrete** if there is a finite or countable set $C \subseteq \mathbb{R}$ such that $\mathbb{P}_X(C) = 1$. Another way to say this is $\mathbb{P}(X \in C) = 1$.

Definition 15.2 Let X be discrete with $\mathbb{P}_X(C) = 1$, where C is countable. The **mass function** (or **probability mass function** or **probability function**) $f_X : C \to [0, 1]$ is defined by

$$f_X(x) = \mathbb{P}_X(\{x\}).$$

Example 15.3 (Problem 8.38) A random variable X has the **binomial distribution** $\mathrm{Bin}(n, p)$, where $n \in \{1, 2, \ldots\}$ and $p \in [0, 1]$, if

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Here and in the other discrete mass functions, we use the standard convention $0^0 = 1$ at degenerate parameter values. Verify that such a random variable is discrete.

Solution. Let $C = \{0, 1, 2, \ldots, n\}$, which is clearly countable. Then

$$\mathbb{P}_X(C) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$$

by the binomial theorem. □

Definition 15.4 Let A be an event. An **indicator random variable** $\mathbb{1}_A\colon \Omega \to \mathbb{R}$ is defined by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}.$$

Example 15.5 (Problem 8.39) Show that the indicator function $\mathbb{1}_A$ of an event A has the binomial distribution $\text{Bin}(1, p)$ with $p = \mathbb{P}(A)$.

Solution. Let $C = \{0, 1\}$. Then

$$f_{\mathbb{1}_A}(0) = 1 - p = \binom{1}{0} p^0 (1-p)^1$$

and

$$f_{\mathbb{1}_A}(1) = \binom{1}{1} p^1 (1-p)^0.$$

This is the **Bernoulli distribution**. If $A_1, A_2, \ldots$ are independent events with $\mathbb{P}(A_i) = p$ for all i, then $\mathbb{1}_{A_1}, \mathbb{1}_{A_2}, \ldots$ are called **Bernoulli trials**. $\square$

Definition 15.6 If X is discrete with $\mathbb{P}_X(C) = 1$ and $f_X(x) > 0$ for all $x \in C$, where C is countable, then we call C the **support set** of X.

Example 15.7 (Problem 8.42) A biased coin with probability of heads p and tails $1 - p$ is flipped n times. Let X denote the number of heads obtained. What is the distribution of X?

Solution. Let X be the number of heads obtained. Then X is discrete, and $C = \{0, 1, 2, \ldots, n\}$ is a countable carrier for its distribution. If $0 < p < 1$, this entire set is the support; for $p = 0$ or $p = 1$, the support is $\{0\}$ or $\{n\}$, respectively. Also,

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in C.$$

The factor $p^k(1-p)^{n-k}$ is the probability of any particular sequence of tosses with k heads and $n - k$ tails, while $\binom{n}{k}$ counts the number of such sequences. Thus this is the binomial distribution

$\mathrm{Bin}(n, p)$. If

$$\mathbb{1}_{A_j} = \begin{cases} 1 & \text{if } H \text{ occurs on toss } j, \\ 0 & \text{if } T \text{ occurs on toss } j, \end{cases}$$

then $X = \mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \cdots + \mathbb{1}_{A_n}$. □

Example 15.8 (Problem 8.43) A biased coin with probability of heads $p > 0$ and tails $1 - p$ is flipped until k heads are obtained, where $k \in \{1, 2, \dots\}$.

1. Find the distribution of the number of flips Y.
2. Show that Y is a discrete random variable.

Solution. For ((1)), fix $n \geq k$. The event $\{Y = n\}$ means that in the first $n - 1$ tosses there are exactly $k - 1$ heads, and toss n is a head. Therefore

$$\mathbb{P}(Y = n) = \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} \cdot p = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n = k, k+1, \dots$$

and $\mathbb{P}(Y = n) = 0$ for $n < k$. This is the **negative binomial distribution** $\mathrm{NegBin}(k, p)$.

For ((2)), let

$$C := \{k, k+1, k+2, \dots\}.$$

The set C is countable, and at least k tosses are needed to obtain k heads. To verify that Y takes a value in C with probability one, compute

$$\sum_{n \in C} \mathbb{P}(Y = n) = \sum_{n=k}^{\infty} \binom{n-1}{k-1} p^k (1-p)^{n-k} = p^k \sum_{m=0}^{\infty} \binom{m+k-1}{k-1} (1-p)^m,$$

where $m = n - k$. By a Taylor expansion,

$$(1-t)^{-k} = \sum_{m=0}^{\infty} \binom{m+k-1}{k-1} t^m, \quad |t| < 1,$$

and putting $t = 1 - p \in [0, 1)$ for $p > 0$, we get

$$\sum_{m=0}^{\infty} \binom{m+k-1}{k-1} (1-p)^m = (1 - (1-p))^{-k} = p^{-k}.$$

Therefore
$$\sum_{n\in C} \mathbb{P}(Y=n) = p^k \cdot p^{-k} = 1.$$
so this is a valid mass function on a countable set. Hence Y is a discrete random variable.

In the special case $k=1$,
$$\mathbb{P}(Y=n) = p(1-p)^{n-1}, \quad n = 1,2,3,\dots,$$
which is the **geometric distribution**. □

Example 15.9 (Problem 8.44) Fix $\lambda > 0$, and for each integer $n \geq \lceil \lambda \rceil$ let X_n have binomial distribution $\mathrm{Bin}\big(n, \frac{\lambda}{n}\big)$. Show that, for every fixed $k \in \{0,1,2,\dots\}$,
$$\mathbb{P}(X_n = k) \to e^{-\lambda}\frac{\lambda^k}{k!}$$
as $n \to \infty$.

Solution. For fixed k and all sufficiently large n (so that $n \geq k$ and $n > \lambda$),
$$\mathbb{P}(X_n = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}.$$
We can rewrite this as
$$\mathbb{P}(X_n = k) = \frac{n(n-1)\dots(n-k+1)}{k!}\cdot\frac{1}{n^k}\lambda^k\cdot\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}.$$
Therefore, taking the limit as $n\to\infty$, we obtain
$$\begin{aligned}\mathbb{P}(X_n = k) &= \frac{n(n-1)\dots(n-k+1)}{n\cdot n\dots n}\cdot\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \xrightarrow{n\to\infty} 1\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}\cdot 1\\ &= e^{-\lambda}\frac{\lambda^k}{k!}.\end{aligned}$$
This is the **Poisson distribution**, also called the law of rare events. We also have
$$\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}\cdot e^{\lambda} = 1.$$

Taylor series are pretty useful. □

Example 15.10 Imagine that there are r red candies and b blue candies in a box, and we draw n of them. What is the probability that exactly x are red?

Solution. The possible values satisfy $\max\{0, n-b\} \leq x \leq \min\{n, r\}$; outside this range the probability is 0. In the valid range, there are $\binom{r}{x}$ ways to choose the red candies and $\binom{b}{n-x}$ ways to choose the blue candies. There are $\binom{r+b}{n}$ ways to choose n candies from the entire box, so the probability is

$$\frac{\dbinom{r}{x}\dbinom{b}{n-x}}{\dbinom{r+b}{n}},$$

which is the **hypergeometric distribution**. We've seen this before! □

Absolutely Continuous Random Variables

Definition 15.11 A random variable is said to be **absolutely continuous** with **density function** $f_X : \mathbb{R} \to [0, \infty)$ if f_X is Borel measurable and

$$\mathbb{P}_X((a, b]) = \mathbb{P}(a < X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x$$

for all $a, b \in \mathbb{R}$ with $a \leq b$.

Warning Do not confuse this with continuity of the function itself, or with a mass function.

It is common to use "absolutely continuous" and "continuous" interchangeably for random variables, but the terms are not synonymous. A **continuous** random variable X has a continuous distribution function F_X; recall that a general distribution function need only be right-continuous. Continuity implies

$$\mathbb{P}(X = x) = \mathbb{P}_X(\{x\}) = 0$$

for every $x \in \mathbb{R}$. Continuous random variables that are not absolutely continuous are quite pathological and rarely appear in elementary applications.

The terminology here is not an accident. X is an absolutely continuous random variable if and only if its distribution function F_X is absolutely continuous as a function on $\mathbb{R}$ in the sense of real analysis: for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\sum_{i=1}^{n}(y_i - x_i) < \delta \Longrightarrow \sum_{i=1}^{n} |F_X(y_i) - F_X(x_i)| < \epsilon$$

for any $x_1 < y_1 < x_2 < y_2 < \cdots < x_n < y_n$ and any $n \in \mathbb{N}$. An important result in measure theory shows that this condition is equivalent to the existence of a Lebesgue integrable function f defined almost everywhere on $\mathbb{R}$ which satisfies

$$F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \quad x \in \mathbb{R}.$$

In contrast, a continuous random variable only needs to have a continuous distribution function, which is a strictly weaker condition.

Example 15.12 (Problem 8.48) Let X be absolutely continuous with density f_X. Show that

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1.$$

Solution. An *incorrect* argument would be

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = \mathbb{P}(-\infty < X \le \infty) = 1.$$

This is not valid, because ∞ and $-\infty$ do not lie in $\mathbb{R}$. The correct argument is

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = \lim_{n\to\infty}\int_{-n}^{n} f_X(x)\,\mathrm{d}x = \lim_{n\to\infty}\mathbb{P}(-n < X \le n) = \lim_{n\to\infty}\mathbb{P}_X((-n, n]) = \mathbb{P}_X(\mathbb{R}) = 1,$$

because $\bigcup_{n=1}^{\infty}(-n, n] = \mathbb{R}$ by continuity from below (Problem 6.8). □

Lecture 16 — Tuesday, November 05, 2013

Example 16.1 (Problem 8.49) Let X be absolutely continuous. Show that $\mathbb{P}_X(\{y\}) = 0$ for any $y \in \mathbb{R}$.

In other words, absolutely continuous random variables are continuous.

Solution. Using continuity from above (Problem 6.9), we have

$$\mathbb{P}_X(\{y\}) = \lim_{n\to\infty} \mathbb{P}_X\left(\left(y - \frac{1}{n}, y\right]\right) = \lim_{n\to\infty} \int_{y-\frac{1}{n}}^{y} f_X(x)\,\mathrm{d}x = \int_y^y f_X(x)\,\mathrm{d}x = 0,$$

using the fact that the function

$$x \mapsto \int_x^y f(t)\,\mathrm{d}t$$

is continuous for any integrable function f. □

Example 16.2 (Problem 8.50) Let X be absolutely continuous. Find a formula for $F_X(y)$ in terms of f_X.

Solution. Again using continuity from below (Problem 6.8) and continuity of

$$x \mapsto \int_x^y f(t)\,\mathrm{d}t,$$

$$F_X(y) = \mathbb{P}_X((-\infty, y]) = \lim_{n\to\infty} \mathbb{P}_X((-n, y]) = \lim_{n\to\infty} \int_{-n}^{y} f_X(x)\,\mathrm{d}x = \int_{-\infty}^{y} f_X(x)\,\mathrm{d}x.$$

□

Example 16.3 (Problem 8.52) Show that if the distribution function F_X is continuously differentiable, then X is an absolutely continuous random variable with density

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}.$$

Solution. We check the definition. First, F_X is nondecreasing by monotonicity of probability measures (Problem 4.12), so its derivative is nonnegative. Moreover,

$$\begin{aligned}\mathbb{P}_X((a,b]) &= \mathbb{P}_X((-\infty,b]\setminus(-\infty,a]) \\ &= F_X(b) - F_X(a) \\ &= \int_a^b \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}\,\mathrm{d}x\end{aligned}$$

by the fundamental theorem of calculus. Thus $\dfrac{\mathrm{d}F_X(x)}{\mathrm{d}x}$ satisfies the definition of a density. □

Example 16.4 (Problem 8.53) A random variable X has the **normal** (or **Gaussian**) **distribution** $\mathcal{N}(m,\sigma^2)$ where $m \in \mathbb{R}$ and $\sigma > 0$ if it is an absolutely continuous random variable with density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Verify that f_X is a density.

The standard normal density is shown in Figure 6.

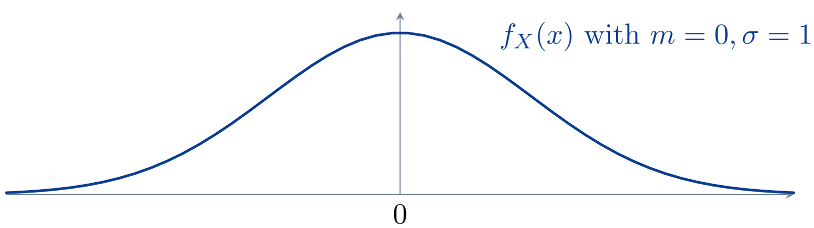

FIGURE 6

Solution. The function f_X is nonnegative, so it remains to verify that it integrates to 1. This is the famous Gaussian integral. The standard calculus derivation writes the integral as the square root of its square and then makes a clever change to polar coordinates. Treating the two improper integrals as a double integral requires Tonelli's theorem for a completely rigorous justification; that theorem is beyond the scope of this course.

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right)\mathrm{d}t \quad \text{substituting } t = \frac{x-m}{\sigma}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \sqrt{\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y} \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2+y^2}{2}\right) \mathrm{d}x\, \mathrm{d}y} \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{\int_{0}^{2\pi} \int_{0}^{\infty} r \exp\left(-\frac{r^2}{2}\right) \mathrm{d}r\, \mathrm{d}\theta} \quad \text{substituting } \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{\int_{0}^{2\pi} \int_{0}^{\infty} e^{-u}\, \mathrm{d}u\, \mathrm{d}\theta} \quad \text{substituting } u = \frac{r^2}{2} \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{\int_{0}^{2\pi} 1\, \mathrm{d}\theta} \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} \\
&= 1.
\end{aligned}
$$

$\square$

Example 16.5 (Problem 8.54) Let X be a random variable with the normal distribution $\mathcal{N}(0,1)$. Show that $m+\sigma X$, where $m \in \mathbb{R}$ and $\sigma > 0$, has the normal distribution $\mathcal{N}(m,\sigma^2)$.

Solution. We have

$$
\mathbb{P}_{m+\sigma X}((a,b]) = \mathbb{P}(a < m+\sigma X \le b) = \mathbb{P}\left(\frac{a-m}{\sigma} < X \le \frac{b-m}{\sigma}\right) = \int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} f_X(y)\, \mathrm{d}y.
$$

Substituting $x = \sigma y + m$ makes the integral equal to

$$
\frac{1}{\sigma} \int_a^b f_X\left(\frac{x-m}{\sigma}\right) \mathrm{d}x = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \mathrm{d}x.
$$

$\square$

Example 16.6 (Problem 8.55) Let X be a random variable with the normal distribution $\mathcal{N}(0,1)$. Find the density of $Y = e^X$. (The distribution of Y is called the **lognormal distribution**.)

The resulting lognormal density is shown in Figure 7.

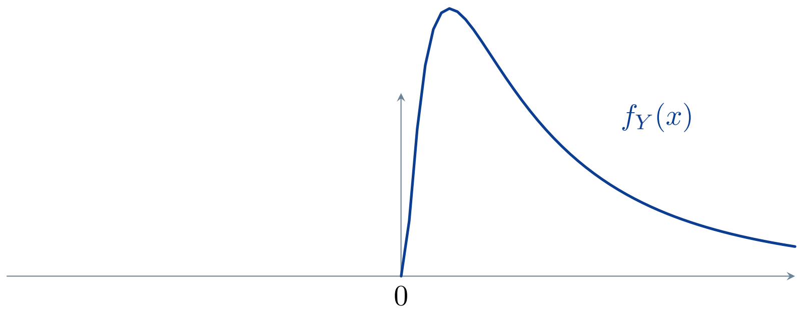

FIGURE 7

Solution. Using the chain rule, the density at $x > 0$ is

$$f_{e^X}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{e^X}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}(e^X \le x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}(X \le \log x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(\log x) = \frac{1}{x} f_X(\log x)$$

so

$$f_Y(x) = \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{(\log x)^2}{2}\right), \qquad x > 0,$$

and $f_Y(x) = 0$ for $x \le 0$. □

Example 16.7 (Problem 8.56) Let X be an absolutely continuous random variable with the uniform distribution on $[0,1]$, i.e., with $f_X(x) = 1$ for any $x \in [0,1]$ and $f_X(x) = 0$ otherwise. Find a function $g : \mathbb{R} \to \mathbb{R}$ such that $g(X)$ has the normal distribution $\mathcal{N}(0,1)$.

Solution. Let Φ be the standard normal distribution function,

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t.$$

The function Φ is a continuous, strictly increasing bijection from $\mathbb{R}$ to $(0,1)$. Define

$$g(u) = \Phi^{-1}(u), \qquad 0 < u < 1,$$

and assign arbitrary values to $g(0)$ and $g(1)$; these endpoints have probability zero under the uniform distribution. For every $x \in \mathbb{R}$,

$$\mathbb{P}(g(X) \le x) = \mathbb{P}(X \le \Phi(x)) = \Phi(x).$$

Hence $g(X)$ has distribution function Φ and therefore has the $\mathcal{N}(0,1)$ distribution. The function Φ^{-1} is the **quantile function** of the standard normal distribution. $\square$

Lecture 17 — Thursday, November 07, 2013

Joint Distributions and Independence of Random Variables

The Borel sets in $\mathbb{R}^2$ are generated by rectangles of the form $(a,b) \times (c,d)$, where $a < b$ and $c < d$. More generally, the Borel sets in $\mathbb{R}^n$ are generated by products of open intervals of the form $(a_1,b_1) \times (a_2,b_2) \times \cdots \times (a_n,b_n) \subseteq \mathbb{R}^n$.

Example 17.1 (Problem 8.57) Suppose X and Y are random variables defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Show that $\{(X,Y) \in B\} \in \mathscr{F}$ for any Borel set $B \subseteq \mathbb{R}^2$.

Solution. Define

$$\mathcal{G} := \{B \subseteq \mathbb{R}^2 : \{(X,Y) \in B\} \in \mathscr{F}\}.$$

We show first that $\mathcal{G}$ is a σ-field on $\mathbb{R}^2$.

Since $\{(X,Y) \in \mathbb{R}^2\} = \Omega \in \mathscr{F}$, we have $\mathbb{R}^2 \in \mathcal{G}$. If $B \in \mathcal{G}$, then

$$\{(X,Y) \in B^c\} = \{(X,Y) \in B\}^c \in \mathscr{F},$$

so $B^c \in \mathcal{G}$.

If $B_1, B_2, \dots \in \mathcal{G}$, then

$$\left\{(X,Y) \in \bigcup_{n=1}^{\infty} B_n\right\} = \bigcup_{n=1}^{\infty} \{(X,Y) \in B_n\} \in \mathscr{F},$$

so $\bigcup_{n=1}^{\infty} B_n \in \mathcal{G}$. Hence $\mathcal{G}$ is a σ-field. Next, let

$$\mathcal{A} := \{(a,b) \times (c,d) : a < b,\ c < d\}.$$

For any rectangle $(a, b) \times (c, d) \in \mathcal{A}$,

$$\{(X, Y) \in (a, b) \times (c, d)\} = \{X \in (a, b)\} \cap \{Y \in (c, d)\}.$$

Because X and Y are random variables, $\{X \in (a, b)\}, \{Y \in (c, d)\} \in \mathscr{F}$, so the intersection is in $\mathscr{F}$. Therefore $\mathcal{A} \subseteq \mathcal{G}$. The Borel σ-field $\mathscr{B}(\mathbb{R}^2)$ is generated by $\mathcal{A}$, i.e. $\mathscr{B}(\mathbb{R}^2) = \sigma(\mathcal{A})$. Since $\mathcal{G}$ is a σ-field containing $\mathcal{A}$, we get

$$\mathscr{B}(\mathbb{R}^2) = \sigma(\mathcal{A}) \subseteq \mathcal{G}.$$

Hence for every Borel set $B \subseteq \mathbb{R}^2$,

$$\{(X, Y) \in B\} \in \mathscr{F}.$$

$\square$

Definition 17.2 The **joint distribution** of random variables X and Y is a probability measure $\mathbb{P}_{X,Y}$ on $\mathbb{R}^2$ such that $\mathbb{P}_{X,Y}(B) = \mathbb{P}(\{(X, Y) \in B\})$ for every Borel set $B \subseteq \mathbb{R}^2$.

Example 17.3 (Problem 8.58) Let X and Y be random variables on the same probability space. Show that

$$\mathbb{P}_X(B) = \mathbb{P}_{X,Y}(B \times \mathbb{R}) \qquad \text{and} \qquad \mathbb{P}_Y(B) = \mathbb{P}_{X,Y}(\mathbb{R} \times B)$$

for any Borel set $B \subseteq \mathbb{R}$.

Proof. Let $B \in \mathscr{B}(\mathbb{R})$. By definition of the distribution of X,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

But

$$\{X \in B\} = \{(X, Y) \in B \times \mathbb{R}\},$$

since $(X(\omega), Y(\omega)) \in B \times \mathbb{R}$ if and only if $X(\omega) \in B$. Therefore, by definition of the joint distribution,

$$\mathbb{P}_X(B) = \mathbb{P}(\{(X, Y) \in B \times \mathbb{R}\}) = \mathbb{P}_{X,Y}(B \times \mathbb{R}).$$

The second identity is analogous:

$$\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(\{(X, Y) \in \mathbb{R} \times B\}) = \mathbb{P}_{X,Y}(\mathbb{R} \times B).$$

$\square$

Definition 17.4 The functions

$$\mathbb{P}_X(B) = \mathbb{P}_{X,Y}(B \times \mathbb{R}) \qquad \text{and} \qquad \mathbb{P}_Y(B) = \mathbb{P}_{X,Y}(\mathbb{R} \times B)$$

in Example 17.3 are called the marginal distributions of X and Y, respectively.

Definition 17.5 X and Y are said to be **jointly discrete** if there is some countable set $C \subseteq \mathbb{R}^2$ such that $\mathbb{P}_{X,Y}(C) = 1$.

Example 17.6 (Problem 8.62) Prove that random variables X and Y are jointly discrete if and only if they are both discrete.

Proof. *Forward implication.* If X and Y are jointly discrete, there exists a countable $C \subseteq \mathbb{R}^2$ with $\mathbb{P}_{X,Y}(C) = 1$. Let

$$A = \{x \in \mathbb{R} : \exists y,\ (x, y) \in C\} \qquad \text{and} \qquad B = \{y \in \mathbb{R} : \exists x,\ (x, y) \in C\}.$$

Then A, B are countable because they are projections of a countable set. Moreover, $C \subseteq A \times B$, so

$$\mathbb{P}(X \in A,\ Y \in B) \geq \mathbb{P}_{X,Y}(C) = 1.$$

Hence $\mathbb{P}(X \in A) = 1$ and $\mathbb{P}(Y \in B) = 1$, so X and Y are discrete.

Reverse implication. If X and Y are discrete, there exist countable $A, B \subseteq \mathbb{R}$ with $\mathbb{P}(X \in A) = 1$, $\mathbb{P}(Y \in B) = 1$. Then

$$C := A \times B$$

is countable. Since the events $\{X \in A\}$ and $\{Y \in B\}$ both have probability one,

$$\mathbb{P}_{X,Y}(C) = \mathbb{P}(X \in A,\ Y \in B) = 1.$$

So X and Y are jointly discrete. $\square$

Definition 17.7 If $C \subseteq \mathbb{R}^2$ is a countable set such that $\mathbb{P}_{X,Y}(C) = 1$, we define the **joint mass function** of X and Y to be $f_{X,Y} : C \to [0, 1]$ by

$$f_{X,Y}(x, y) = \mathbb{P}(X = x \text{ and } Y = y) = \mathbb{P}_{X,Y}(\{(x, y)\}).$$

Example 17.8 (Problem 8.63) A fair coin is tossed repeatedly. Suppose heads appears for the first time after X tosses and tails appears for the first time after Y tosses. Show that X and Y are jointly discrete and find their joint distribution. Then compute the marginal distributions of X and Y.

Solution. On each sample path, exactly one of X and Y equals 1: if the first toss is heads then $X = 1$, and if the first toss is tails then $Y = 1$. Therefore (X, Y) can only take values on

$$C = \{(1, y) : y = 2, 3, \dots\} \cup \{(x, 1) : x = 2, 3, \dots\},$$

which is countable. Hence X and Y are jointly discrete. Next we compute the joint mass function. For $y \geq 2$,

$$\{X = 1, Y = y\} = \{H \cdots HT \text{ with } y - 1 \text{ heads then one tail}\},$$

so

$$f_{X,Y}(1, y) = \mathbb{P}(X = 1, Y = y) = \left(\frac{1}{2}\right)^y.$$

For $x \geq 2$,

$$\{X = x, Y = 1\} = \{T \cdots TH \text{ with } x - 1 \text{ tails then one head}\},$$

so

$$f_{X,Y}(x, 1) = \mathbb{P}(X = x, Y = 1) = \left(\frac{1}{2}\right)^x.$$

For all other (x, y), the probability is 0. Thus

$$f_{X,Y}(x, y) = \begin{cases} 2^{-y}, & x = 1,\ y \geq 2, \\ 2^{-x}, & y = 1,\ x \geq 2, \\ 0, & \text{otherwise.} \end{cases}$$

The marginals are obtained by summing the joint mass over the other variable. For X,

$$f_X(1) = \sum_{y=2}^{\infty} f_{X,Y}(1, y) = \sum_{y=2}^{\infty} 2^{-y} = \frac{1}{2},$$

and for $x \geq 2$,

$$f_X(x) = \sum_{y \geq 1} f_{X,Y}(x, y) = f_{X,Y}(x, 1) = 2^{-x}.$$

So

$$f_X(x) = \begin{cases} \frac{1}{2}, & x = 1, \\ 2^{-x}, & x = 2, 3, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

By symmetry,

$$f_Y(y) = \begin{cases} \frac{1}{2}, & y = 1, \\ 2^{-y}, & y = 2, 3, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

□

Definition 17.9 X, Y are said to be **jointly absolutely continuous** if there exists a Borel measurable function $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ called the **joint density** such that

$$\mathbb{P}_{X,Y}((a, b] \times (c, d]) = \int_c^d \int_a^b f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y$$

for all $a < b$ and $c < d$.

Example 17.10 (Problem 8.65) Show that if X and Y are jointly absolutely continuous with joint density $f_{X,Y}$, then both are absolutely continuous with marginal densities

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}y \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}x.$$

Solution. The marginalization rule for nonnegative double integrals is a consequence of Tonelli's theorem. A complete proof of that theorem requires integration theory beyond the scope of this course. Applying the rule here gives

$$\mathbb{P}_X((a, b]) = \mathbb{P}_{X,Y}((a, b] \times \mathbb{R}) = \int_a^b \left(\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}y \right) \mathrm{d}x,$$

so X is absolutely continuous with the stated density. The argument for Y is identical after interchanging the variables. □

Example 17.11 (Problem 8.66) Find two absolutely continuous random variables X and Y that are not jointly absolutely continuous.

Solution. Take $X \sim \mathcal{N}(0,1)$ and define $Y = X$. Then X is absolutely continuous, and since Y has the same distribution as X, Y is also absolutely continuous. However,

$$\mathbb{P}\big((X,Y) \in D\big) = 1 \qquad \text{and} \qquad D := \{(x,y) \in \mathbb{R}^2 : y = x\},$$

because $(X,Y) = (X,X)$ always lies on the diagonal D. If (X,Y) had a joint density $f_{X,Y}$, then the diagonal's two-dimensional Lebesgue measure zero would give

$$\mathbb{P}((X,Y) \in D) = \int_D f_{X,Y}(x,y)\,\mathrm{d}(x,y) = 0,$$

a contradiction. Therefore (X,Y) is not jointly absolutely continuous. $\square$

Definition 17.12 Two random variables X and Y are **independent** when the events $\{X \leq a\}$ and $\{Y \leq b\}$ are independent for all $a, b \in \mathbb{R}$. In general, $X_1, X_2, \ldots, X_n$ are independent if

$$\{X_1 \leq a_1\}, \{X_2 \leq a_2\}, \ldots, \{X_n \leq a_n\}$$

are independent events for all $a_1, a_2, \ldots, a_n \in \mathbb{R}$.

Example 17.13 (Problem 8.68) Show that X and Y are independent if and only if $\{X \in A\}$ and $\{Y \in B\}$ are independent for all Borel sets $A, B \subseteq \mathbb{R}$.

Solution. *Forward implication.* If $\{X \in A\}$ and $\{Y \in B\}$ are independent for all Borel A, B, then in particular this holds for

$$A = (-\infty, a] \qquad \text{and} \qquad B = (-\infty, b],$$

so $\{X \leq a\}$ and $\{Y \leq b\}$ are independent for all $a, b \in \mathbb{R}$.

Reverse implication. Assume

$$\mathbb{P}(X \leq a,\, Y \leq b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b), \quad \text{for all } a, b \in \mathbb{R}.$$

Fix $b \in \mathbb{R}$ and define

$$\mathcal{D}_b = \{A \in \mathscr{B} : \mathbb{P}(X \in A, Y \leq b) = \mathbb{P}(X \in A)\mathbb{P}(Y \leq b)\}.$$

This is a Dynkin system: it contains $\mathbb{R}$, is closed under complements, and is closed under countable disjoint unions. By assumption it contains every interval $(-\infty, a]$. These intervals form a π-system that generates $\mathscr{B}$, so the $\boldsymbol{\pi}$–$\boldsymbol{\lambda}$ theorem gives $\mathcal{D}_b = \mathscr{B}$.

Now fix $A \in \mathscr{B}$ and define

$$\mathcal{E}_A = \{B \in \mathscr{B} : \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\}.$$

The first part shows that $\mathcal{E}_A$ contains every $(-\infty, b]$. It is again a Dynkin system, so another application of that theorem gives $\mathcal{E}_A = \mathscr{B}$. Therefore, for all Borel sets $A, B \subseteq \mathbb{R}$,

$$\mathbb{P}(X \in A,\, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B),$$

so the events $\{X \in A\}$ and $\{Y \in B\}$ are independent. □

Example 17.14 (Problem 8.70) Show that jointly absolutely continuous random variables X and Y are independent if and only if

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

for almost every $(x, y) \in \mathbb{R}^2$.

Solution. *Forward implication.* Assume X and Y are independent. Then for every $a, b \in \mathbb{R}$,

$$F_{X,Y}(a, b) = \mathbb{P}(X \leq a,\, Y \leq b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b) = F_X(a)F_Y(b).$$

Since X and Y are jointly absolutely continuous,

$$F_{X,Y}(a, b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x, y)\, \mathrm{d}y\, \mathrm{d}x.$$

Also,

$$F_X(a)F_Y(b) = \left(\int_{-\infty}^{a} f_X(x)\, \mathrm{d}x\right)\left(\int_{-\infty}^{b} f_Y(y)\, \mathrm{d}y\right) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_X(x)f_Y(y)\, \mathrm{d}y\, \mathrm{d}x.$$

Therefore the measure with density $f_{X,Y}$ and the measure with density $(x, y) \mapsto f_X(x)f_Y(y)$ agree on every lower-left rectangle $(-\infty, a] \times (-\infty, b]$. These rectangles, together with $\mathbb{R}^2$, form a π-system generating the Borel sets of $\mathbb{R}^2$. The π–λ theorem shows that the two probability measures agree on every Borel set. By uniqueness of densities,

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

for almost every (x, y).

Reverse implication. Assume now that $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ almost everywhere. Then for all $a, b \in \mathbb{R}$,

$$\begin{aligned}
F_{X,Y}(a,b) &= \mathbb{P}(X \le a,\, Y \le b) \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} f_X(x)f_Y(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \left(\int_{-\infty}^{a} f_X(x)\,\mathrm{d}x\right)\left(\int_{-\infty}^{b} f_Y(y)\,\mathrm{d}y\right) \\
&= F_X(a)F_Y(b).
\end{aligned}$$

Hence $\{X \le a\}$ and $\{Y \le b\}$ are independent for all a, b, so X and Y are independent. $\square$

Example 17.15 (Problem 8.72) Let X and Y be independent random variables with binomial distributions $\mathrm{Bin}(m, p)$ and $\mathrm{Bin}(n, p)$, respectively. What is the distribution of $X + Y$?

Solution. The marginal mass function of X is

$$f_X(x) = \binom{m}{x} p^x (1-p)^{m-x}, \quad x = 0, 1, 2, \dots, m,$$

and the marginal mass function of Y is

$$f_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, 2, \dots, n.$$

Therefore, the joint mass function is

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = \binom{m}{x}\binom{n}{y} p^{x+y}(1-p)^{(m+n)-(x+y)},$$

for $x = 0, 1, 2, \ldots, m$ and $y = 0, 1, 2, \ldots, n$. For $z = 0, 1, \ldots, m + n$,

$$\{X + Y = z\} = \bigcup_{x=0}^{z} \{X = x, Y = z - x\}.$$

Therefore

$$\begin{aligned} f_{X+Y}(z) &= \mathbb{P}(X + Y = z) \\ &= \sum_{x=0}^{z} f_{X,Y}(x, z - x) \\ &= \sum_{x=0}^{z} f_X(x) f_Y(z - x) \\ &= \sum_{x=0}^{z} \binom{m}{x}\binom{n}{z-x} p^z (1-p)^{(m+n)-z} \end{aligned}$$

and hence

$$f_{X+Y}(z) = p^z(1-p)^{(m+n)-z} \sum_{x=0}^{z} \binom{m}{x}\binom{n}{z-x} = \binom{m+n}{z} p^z (1-p)^{(m+n)-z},$$

where binomial coefficients outside their usual range are understood to be 0 as before. Thus $X + Y \sim \mathrm{Bin}(m + n, p)$. □

Example 17.16 (Problem 8.73) Let X and Y be independent random variables with Poisson distributions of parameters λ and μ, respectively. Find the distribution of $X + Y$.

Solution. Let $z \in \{0, 1, 2, \dots\}$. Since X and Y are independent,

$$\begin{aligned}
\mathbb{P}(X+Y=z) &= \sum_{x=0}^{z} \mathbb{P}(X=x,\, Y=z-x) \\
&= \sum_{x=0}^{z} \mathbb{P}(X=x)\mathbb{P}(Y=z-x) \\
&= \sum_{x=0}^{z} \frac{\lambda^x e^{-\lambda}}{x!} \cdot \frac{\mu^{z-x} e^{-\mu}}{(z-x)!} \\
&= e^{-(\lambda+\mu)} \sum_{x=0}^{z} \frac{\lambda^x \mu^{z-x}}{x!(z-x)!} \\
&= \frac{e^{-(\lambda+\mu)}}{z!} \sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x} \\
&= \frac{e^{-(\lambda+\mu)}}{z!} (\lambda+\mu)^z,
\end{aligned}$$

by the binomial theorem. Also $\mathbb{P}(X+Y=z)=0$ for $z<0$. Therefore $X+Y$ has the Poisson distribution with parameter $\lambda+\mu$. □

Lecture 18 — Tuesday, November 12, 2013

9. Expectation and Variance

Expectation of Simple Random Variables

Definition 18.1 A random variable X is called **simple** if there exists a finite set $C \subseteq \mathbb{R}$ such that $\mathbb{P}(X \in C) = 1$. The elements of C are called the values of X. For a simple random variable, define

$$\mathbb{E}[X] = \sum_{x \in C} x \cdot \mathbb{P}(X=x).$$

Equivalently, if $C = \{x_1, x_2, \ldots, x_n\}$, then

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \cdot f_X(x_i),$$

where f_X is the mass function of X. This is the **expectation** (or **expected value**).

Warning Expectation is also called the population mean. It should not be confused with an observed sample average.

Example 18.2 (Problem 9.2) What is the expectation of a random variable with the Bernoulli distribution?

Solution. Here $C = \{0, 1\}$. On C, we have $f_X(0) = 1 - p$ and $f_X(1) = p$. Therefore

$$\mathbb{E}[X] = \sum_{i=1}^{2} x_i \cdot f_X(x_i) = 0 \cdot (1-p) + 1 \cdot p = p.$$

□

Example 18.3 (Problem 9.3) A friend forgot the 8-digit password required to log into a computer. If all possible passwords are tried completely at random, discarding unsuccessful ones, what is the expected number of attempts needed to find the correct password?

Solution. We are assuming that the sampling is done without replacement because your friend is probably not that dumb. Let X be the number of attempts. Then $X \in \{1, 2, \ldots, 10^8\}$, where $10^8 = n$. The mass function is

$$f_X(1) = \mathbb{P}(X = 1) = \frac{1}{n}$$

and

$$f_X(2) = \mathbb{P}(X = 2) = \left(\frac{n-1}{n}\right) \cdot \left(\frac{1}{n-1}\right) = \frac{1}{n}.$$

In general, $f_X(k) = \frac{1}{n}$ for $1 \le k \le n$. Therefore

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \cdot f_X(x_i) = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{n+1}{2} = 50,000,000.5.$$

□

Example 18.4 (Problem 9.5) What is the expectation of a random variable with the binomial distribution?

Solution. Let $X \sim \mathrm{Bin}(n,p)$, so for $k = 0, 1, \ldots, n$ we have

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Therefore,

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

Using

$$k\binom{n}{k} = n\binom{n-1}{k-1},$$

we get

$$\mathbb{E}[X] = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}.$$

Let $j = k-1$. Then $j = 0, 1, \ldots, n-1$, so

$$\mathbb{E}[X] = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j}.$$

The sum equals $(p + (1-p))^{n-1} = 1$ by the binomial theorem. Hence $\mathbb{E}[X] = np$. □

Example 18.5 (Problem 9.6: Linearity of expectation) Show that $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ and $\mathbb{E}[cX] = c\mathbb{E}[X]$ for simple random variables X, Y and any $c \in \mathbb{R}$.

Solution. Suppose that the values of X are $x_1, \ldots, x_n$ and the values of Y are $y_1, \ldots, y_m$. Then

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \cdot f_X(x_i) \qquad \text{and} \qquad \mathbb{E}[Y] = \sum_{j=1}^{m} y_j \cdot f_Y(y_j).$$

The class of random variables forms a vector space, so

$$X + Y \in \{x_i + y_j \mid 1 \le i \le n,\ 1 \le j \le m\}.$$

Thus $X + Y$ is simple. Let $Z = X + Y$, with values $z_1, \ldots, z_k$, where $k \le m \cdot n$. Here,

$$f_Z(z_\ell) = \sum_{i,j : x_i + y_j = z_\ell} f_{X,Y}(x_i, y_j),$$

which is a fancy version of the total probability formula. Therefore

$$\begin{aligned}
\mathbb{E}[Z] &= \sum_{\ell=1}^{k} z_\ell \cdot f_Z(z_\ell) \\
&= \sum_{\ell=1}^{k} z_\ell \sum_{i,j : x_i + y_j = z_\ell} f_{X,Y}(x_i, y_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) \cdot f_{X,Y}(x_i, y_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i \cdot f_{X,Y}(x_i, y_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j \cdot f_{X,Y}(x_i, y_j) \\
&= \sum_{i=1}^{n} x_i \left[\sum_{j=1}^{m} f_{X,Y}(x_i, y_j) \right] + \sum_{j=1}^{m} y_j \left[\sum_{i=1}^{n} f_{X,Y}(x_i, y_j) \right] \\
&= \sum_{i=1}^{n} x_i \cdot f_X(x_i) + \sum_{j=1}^{m} y_j \cdot f_Y(y_j) \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

This calculation proves that $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

We now prove that $\mathbb{E}[cX] = c\,\mathbb{E}[X]$. If $c = 0$, both sides are 0. Suppose therefore that $c \neq 0$ and that X takes the distinct values $x_1, \ldots, x_n$. Then $cx_1, \ldots, cx_n$ are also distinct, and for each i we have

$$\mathbb{P}(cX = cx_i) = \mathbb{P}(X = x_i) = f_X(x_i).$$

Therefore,

$$\mathbb{E}[cX] = \sum_{i=1}^{n}(cx_i)\mathbb{P}(cX = cx_i) = \sum_{i=1}^{n}(cx_i)f_X(x_i) = c\sum_{i=1}^{n}x_i f_X(x_i) = c\,\mathbb{E}[X].$$

□

Example 18.6 (Problem 9.7) Show that if X is a simple random variable and $X \geq 0$, then $\mathbb{E}[X] \geq 0$.

Solution. Note that $X \geq 0$ means $X(\omega) \geq 0$ for all $\omega \in \Omega$. This is immediate from the definition, and then every term in the defining finite sum for $\mathbb{E}[X]$ is nonnegative. □

Example 18.7 (Problem 9.8: Monotonicity of expectation) If simple random variables X and Y satisfy $X \leq Y$, show that $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Solution. If $X \leq Y$, then $Y - X \geq 0$. Therefore, Example 18.6 gives

$$0 \leq \mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X],$$

so $\mathbb{E}[X] \leq \mathbb{E}[Y]$. □

Example 18.8 (Problem 9.9) Show that $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ for any simple random variable X with values $x_1, \ldots, x_n$.

Solution. From Example 18.7, $\mathbb{E}[X] \leq \mathbb{E}[|X|]$ and $-\mathbb{E}[X] \leq \mathbb{E}[|X|]$, so $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$. Alternatively, by the triangle inequality

$$|\mathbb{E}[X]| = \left|\sum_{i=1}^{n} x_i \cdot f_X(x_i)\right| \leq \sum_{i=1}^{n}|x_i| \cdot f_X(x_i) = \mathbb{E}[|X|].$$

□

Example 18.9 (Problem 9.11: Law of the unconscious statistician for simple random variabls) Let X be a simple random variable with values $x_1, \ldots, x_n$. Show that for any function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\left[g(X)\right] = \sum_{i=1}^{n} g(x_i) f_X(x_i).$$

Solution. Define $Y = g(X)$. Let Y have values $y_1, \ldots, y_m$, where $m \le n$. Then

$$\begin{aligned}
\mathbb{E}\left[Y\right] &= \sum_{j=1}^{m} y_j \cdot f_Y(y_j) = \sum_{j=1}^{m} y_j \left[\sum_{i:g(x_i)=y_j} f_X(x_i) \right] \\
&= \sum_{j=1}^{m} \sum_{i:g(x_i)=y_j} y_j \cdot f_X(x_i) \\
&= \sum_{j=1}^{m} \sum_{i:g(x_i)=y_j} g(x_i) \cdot f_X(x_i) \\
&= \sum_{i=1}^{n} g(x_i) \cdot f_X(x_i).
\end{aligned}$$

□

Lecture 19 — Thursday, November 14, 2013

Variance, Covariance and Jensen's Inequality

Example 19.1 (Problem 9.12: Cauchy–Schwarz–Buniakovsky inequality) Let X and Y be simple random variables. Verify

$$\mathbb{E}\left[XY\right]^2 \le \mathbb{E}\left[X^2\right] \mathbb{E}\left[Y^2\right].$$

Proof. Suppose that $\mathbb{E}\left[Y^2\right] > 0$. Otherwise the result is immediate, because $\mathbb{P}(Y = 0) = 1$ and both sides are zero. Note that $(X - tY)^2 \ge 0$ for all $t \in \mathbb{R}$, so by Problem 9.7 we have

$\mathbb{E}\left[(X-tY)^2\right] \geq 0$. Now

$$0 \leq \mathbb{E}\left[(X-tY)^2\right] = \mathbb{E}\left[X^2\right] - 2t\mathbb{E}\left[XY\right] + t^2\mathbb{E}\left[Y^2\right],$$

which is a quadratic polynomial in t. Choosing

$$t = \frac{\mathbb{E}\left[XY\right]}{\mathbb{E}\left[Y^2\right]}$$

minimizes this expression, and substitution yields

$$0 \leq \mathbb{E}\left[X^2\right] - \frac{\mathbb{E}\left[XY\right]^2}{\mathbb{E}\left[Y^2\right]}.$$

Rearranging gives

$$\mathbb{E}\left[XY\right]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right].$$

$\square$

Definition 19.2 We can rewrite the result as

$$|\mathbb{E}\left[XY\right]| \leq \sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}.$$

Replacing X by $X - m_X$, where $m_X = \mathbb{E}\left[X\right]$, and Y by $Y - m_Y$ gives

$$|\mathbb{E}\left[(X-m_X)(Y-m_Y)\right]| \leq \sqrt{\mathbb{E}\left[(X-m_X)^2\right]} \cdot \sqrt{\mathbb{E}\left[(Y-m_Y)^2\right]}.$$

This is the **covariance inequality**. We define

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-m_X)(Y-m_Y)\right]$$

to be the **covariance** of X and Y, and

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-m_X)^2\right]$$

to be the **variance** of X. If $\mathrm{Var}(X) > 0$ and $\mathrm{Var}(Y) > 0$, the **coefficient of linear correlation** is defined by

$$-1 \leq \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \leq 1,$$

where the bounds follow from the covariance inequality.

Definition 19.3 A function $h : \mathbb{R} \to \mathbb{R}$ is **convex** if

$$h(ax + (1-a)y) \leq ah(x) + (1-a)h(y)$$

for all $x, y \in \mathbb{R}$ and all $a \in [0, 1]$.

A useful example is $|x| = \max(x, -x)$, which is convex.

Example 19.4 (Problem 9.14: Jensen's inequality) Let X be a simple random variable with values in an interval (a, b), and let $h : (a, b) \to \mathbb{R}$ be convex. Prove Jensen's inequality:

$$h(\mathbb{E}[X]) \leq \mathbb{E}[h(X)].$$

Proof. Because X has finitely many values in (a, b), their convex combination $x_0 = \mathbb{E}[X]$ also lies in (a, b). We use the **supporting line theorem**: for this x_0, there are constants $\alpha, \beta \in \mathbb{R}$ such that $\alpha + \beta x \leq h(x)$ for all $x \in (a, b)$ and $\alpha + \beta x_0 = h(x_0)$. Then

$$\mathbb{E}[h(X)] \geq \mathbb{E}[\alpha + \beta X] = \alpha + \beta\mathbb{E}[X] = \alpha + \beta x_0 = h(x_0) = h(\mathbb{E}[X]).$$

We do not need to prove the supporting line theorem here. $\square$

Special cases include $\mathbb{E}[|X|] \geq |\mathbb{E}[X]|$, which is exactly Problem 9.9, and $\mathbb{E}\left[X^2\right] \geq [\mathbb{E}[X]]^2$, which implies that the variance is nonnegative. We also obtain $\mathbb{E}\left[e^X\right] \geq e^{\mathbb{E}[X]}$. For concave functions, the inequality is reversed.

Example 19.5 (Problem 9.19) Show that for any simple random variable X,

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

Solution. By expanding the definition of variance, we obtain

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[(X - m_X)^2\right] \\
&= \mathbb{E}\left[X^2 - 2m_X X + m_X^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2(\mathbb{E}[X])^2 + \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.
\end{aligned}$$

□

Example 19.6 (Problem 9.20) Compute the variance of a random variable with the Bernoulli distribution.

Solution. For a Bernoulli random variable, $X \sim \text{Bin}(1, p)$. Recall that $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$. Since $X \in \{0, 1\}$, we have $X^2 = X$, and therefore

$$\text{Var}(X) = \mathbb{E}\left[X^2\right] - [\mathbb{E}[X]]^2 = \mathbb{E}[X] - \mathbb{E}[X]^2 = \mathbb{E}[X][1 - \mathbb{E}[X]] = p(1 - p).$$

□

Example 19.7 (Problem 9.23) Show that $\text{Var}(bX + a) = b^2 \text{Var}(X)$ for any $a, b \in \mathbb{R}$.

Solution. Let $m_X = \mathbb{E}[X]$. Then

$$\mathbb{E}[bX + a] = b\mathbb{E}[X] + a = bm_X + a.$$

Hence

$$\begin{aligned}
\text{Var}(bX + a) &= \mathbb{E}\left[\left((bX + a) - \mathbb{E}[bX + a]\right)^2\right] \\
&= \mathbb{E}\left[\left((bX + a) - (bm_X + a)\right)^2\right] \\
&= \mathbb{E}\left[\left(b(X - m_X)\right)^2\right] \\
&= b^2\mathbb{E}\left[(X - m_X)^2\right] \\
&= b^2 \text{Var}(X).
\end{aligned}$$

□

Example 19.8 (Problem 9.24) Show that

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

whenever X and Y are independent simple random variables.

Proof. The key identity is

$$\operatorname{Cov}(X,Y) = \mathbb{E}\left[(X-m_X)(Y-m_Y)\right] = \mathbb{E}\left[XY\right] - m_X m_Y = \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right].$$

If X and Y are independent, then their singleton events are independent by Problem 8.68. Hence, writing their finite sets of values as $\{x_i\}$ and $\{y_j\}$,

$$\mathbb{E}[XY] = \sum_{i,j} x_i y_j \mathbb{P}(X=x_i, Y=y_j) = \sum_{i,j} x_i y_j \mathbb{P}(X=x_i)\mathbb{P}(Y=y_j) = \mathbb{E}[X]\mathbb{E}[Y].$$

Thus $\operatorname{Cov}(X,Y)=0$. Therefore

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

□

The **covariance identity** is

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y),$$

Indeed, expanding the centered square gives

$$\mathbb{E}\left[((X-m_X)+(Y-m_Y))^2\right] = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y).$$

In particular, $\operatorname{Cov}(X,Y)$ is a bilinear, symmetric, positive semidefinite form on the space of random variables with finite variances. It becomes an inner product after we identify random variables that differ by an almost-sure constant. In particular, if X and Y are independent, then $\operatorname{Cov}(X,Y)=0$, so $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$. More generally,

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i) + 2\sum_{1\le i<j\le n} \operatorname{Cov}(X_i,X_j).$$

When $X_1,\dots,X_n$ are independent, this reduces to

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i).$$

Definition 19.9 Let X be a discrete random variable with values $x_1, x_2, \ldots$. Define the expected value of X by

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \cdot f_X(x_i),$$

provided the series is **absolutely convergent**; that is, provided

$$\sum_{i=1}^{\infty} |x_i| \cdot f_X(x_i) < \infty.$$

Example 19.10 (Problem 9.27) A biased coin with probability of heads $p > 0$ and tails $1 - p$ is tossed repeatedly. Let X be the number of tosses until heads appears for the first time. Compute $\mathbb{E}[X]$.

Solution. Here $X \in \{1, 2, 3, \dots\}$, with $f_X(k) = (1-p)^{k-1} \cdot p$ for $k = 1, 2, \dots$. Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p = p \sum_{k=1}^{\infty} k(1-p)^{k-1}.$$

Using

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{\mathrm{d}}{\mathrm{d}x} \sum_{k=1}^{\infty} x^k = \frac{\mathrm{d}}{\mathrm{d}x} \frac{x}{1-x} = \frac{1}{(1-x)^2},$$

we obtain

$$\mathbb{E}[X] = p\left(\frac{1}{(1-(1-p))^2}\right) = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

Absolute convergence is automatic because every term is nonnegative and the sum is finite. □

Example 19.11 (Problem 9.29) What is the expectation of a random variable with the Poisson distribution?

Solution. Using the Poisson mass function and Taylor's theorem, we obtain

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x \cdot f_X(x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} \frac{\lambda^x \cdot e^{-\lambda}}{(x-1)!} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} = \lambda.$$

□

Lecture 20 — Tuesday, November 19, 2013

Moments and Continuous Expectation

Example 20.1 (Problem 9.28: St. Petersburg paradox) This problem was invented by Nicolas Bernoulli. A fair coin is tossed repeatedly until the first tail appears. If the first tail appears after k heads, the payoff is 2^k dollars. What is the expected payoff?

Solution. For the St. Petersburg paradox, W takes the values $1, 2, 4, 8, \ldots$, with $f_W(2^k) = 2^{-(k+1)}$ for $k = 0, 1, 2, \ldots$. Then

$$\mathbb{E}[W] = \sum_{k=0}^{\infty} 2^k \cdot 2^{-(k+1)} = \infty,$$

so the expectation is not finite. (The "paradox" is that you should be willing to pay any amount of money to play this game if your sole consideration is your expected profit.) $\square$

Example 20.2 (Problem 9.31: Law of the unconscious statistician for discrete random variables) Let X be a discrete random variable with values $x_1, x_2, \ldots$ and mass function f_X. Show that for any Borel function $g : \mathbb{R} \to \mathbb{R}$, the random variable $g(X)$ is discrete and

$$\mathbb{E}[g(X)] = \sum_{i=1}^{\infty} g(x_i) f_X(x_i),$$

provided this series is absolutely convergent.

Solution. Let $C = \{x_1, x_2, \ldots\}$. Since $\mathbb{P}(X \in C) = 1$ and $g(C)$ is countable, we have $\mathbb{P}(g(X) \in g(C)) = 1$, so $g(X)$ is discrete. Let $Y = g(X)$, and list the distinct values of Y as $y_1, y_2, \ldots$. For each j,

$$\mathbb{P}(Y = y_j) = \sum_{i:\, g(x_i) = y_j} f_X(x_i).$$

Therefore, the expectation is

$$\mathbb{E}[Y] = \sum_{j=1}^{\infty} y_j \mathbb{P}(Y = y_j) = \sum_{j=1}^{\infty} y_j \sum_{i:\, g(x_i)=y_j} f_X(x_i) = \sum_{i=1}^{\infty} g(x_i) f_X(x_i),$$

where the rearrangement is justified by absolute convergence. □

Example 20.3 (Problem 9.33) Find an example of a discrete random variable whose expectation exists but whose variance is undefined.

Solution. Let the values of X be $x_k = (3/2)^k$ for $k = 0, 1, 2, \ldots$, and let $f_X(x_k) = 2^{-(k+1)}$. These probabilities sum to one, and

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} (3/2)^k \cdot 2^{-(k+1)} = \frac{1}{2} \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k = 2.$$

However,

$$\mathbb{E}\left[X^2\right] = \sum_{k=0}^{\infty} \left(\frac{3}{2}\right)^{2k} \cdot 2^{-(k+1)} = \frac{1}{2} \sum_{k=0}^{\infty} \left(\frac{9}{8}\right)^k = \infty.$$

Hence X has a finite expectation but no finite second moment. Under our convention its variance is therefore undefined (or, in the extended nonnegative sense, equal to ∞). □

Example 20.4 (Problem 9.34) Compute the variance of a random variable with the Poisson distribution.

Solution. We found earlier that $\mathbb{E}[X] = \lambda$. What is $\mathrm{Var}(X)$? Using the simplifying formula,

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \mathbb{E}\left[X^2\right] - \lambda^2.$$

Since $f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, the expression

$$\mathbb{E}\left[X^2\right] = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x e^{-\lambda}}{x!}$$

is not especially helpful. The trick is to calculate

$$\mathbb{E}\left[X(X-1)\right] = \sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=2}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-2)!} = \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2} e^{-\lambda}}{(x-2)!} = \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = \lambda^2.$$

Returning to the original problem, we have $\mathbb{E}\left[X(X-1)\right] = \lambda^2$. Therefore

$$\mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right] = \lambda^2,$$

so

$$\mathbb{E}\left[X^2\right] = \lambda + \lambda^2 \qquad \text{and hence} \qquad \operatorname{Var}(X) = \lambda.$$

□

Definition 20.5 Different expectations are called **moments**: $\mathbb{E}\left[X^2\right]$ is the 2nd moment, $\mathbb{E}\left[X(X-1)\right]$ is the factorial moment.

Definition 20.6 If X is absolutely continuous with density $f_X(x)$, then define the expected value of X by

$$\mathbb{E}\left[X\right] = \int_{-\infty}^{\infty} x \cdot f_X(x)\,\mathrm{d}x,$$

provided the integral is absolutely convergent.

Theorem 20.7 (Law of the unconscious statistician for continuous random variables) If X is absolutely continuous with density f_X and $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\,\mathrm{d}x,$$

provided that the integral is absolutely convergent.

Remark 20.8 A complete proof of the law of the unconscious statistician uses pushforward measures and the monotone convergence theorem, both beyond the scope of this course.

Example 20.9 (Problem 9.38) Compute $\mathbb{E}\left[X\right]$ when X has the **exponential distribu-**

tion $\text{Exp}(\lambda)$, where $\lambda > 0$, with density

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Here λ is called the **intensity parameter**.

The exponential density from Example 20.9 is shown in Figure 8.

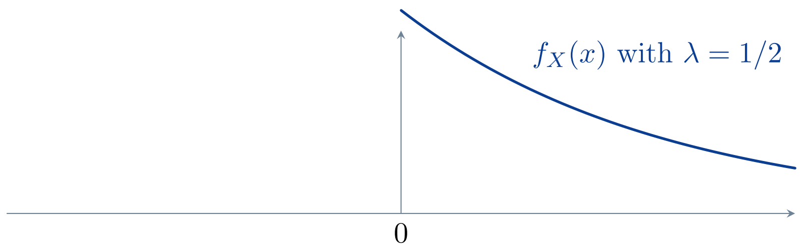

FIGURE 8

Solution. We have

$$\mathbb{E}[X] = \int_0^\infty \lambda x e^{-\lambda x}\,\mathrm{d}x = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}.$$

A useful formula to remember is

$$\int_0^\infty x^n e^{-x}\,\mathrm{d}x = n!, \qquad n = 0, 1, 2, \ldots.$$

$\square$

Example 20.10 (Problem 9.39) Compute $\mathbb{E}[X]$ when X has the **Cauchy distribution** with density

$$f_X(x) = \frac{1}{\pi(1+x^2)}.$$

The Cauchy density from Example 20.10 is shown in Figure 9.

Solution. First, we verify that

$$\int_{-\infty}^\infty f_X(x)\,\mathrm{d}x = \frac{1}{\pi}\int_{-\infty}^\infty \frac{1}{1+x^2}\,\mathrm{d}x$$

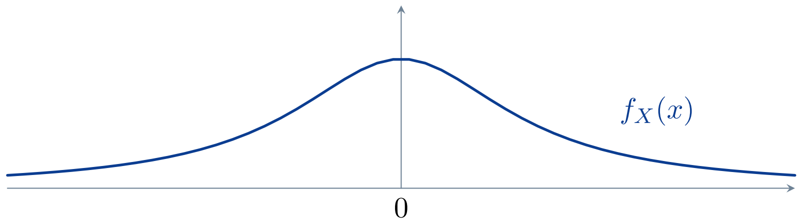

FIGURE 9

$$= \frac{1}{\pi}\left[\arctan(x)\right]_{-\infty}^{\infty}$$
$$= 1.$$

On the other hand, to define $\mathbb{E}[X]$ we must check the positive and negative parts separately. Using the substitution $y = 1 + x^2$, we compute

$$\int_0^\infty \frac{x}{\pi(1+x^2)}\,\mathrm{d}x = \frac{1}{\pi}\cdot\frac{1}{2}\int_1^\infty \frac{1}{y}\,\mathrm{d}y = \frac{1}{2\pi}\lim_{R\to\infty}\log R = \infty,$$

Similarly,

$$\int_{-\infty}^0 \frac{x}{\pi(1+x^2)}\,\mathrm{d}x = -\infty.$$

Therefore the positive and negative parts are both infinite, so the expectation is not defined. $\square$

You are shooting bullets at a wall because you are very mad at it. The wall is infinitely long and one unit in front of you. Your range of fire is 180°, so bullets are more likely to hit the region directly in front of you, as shown in Figure 10.

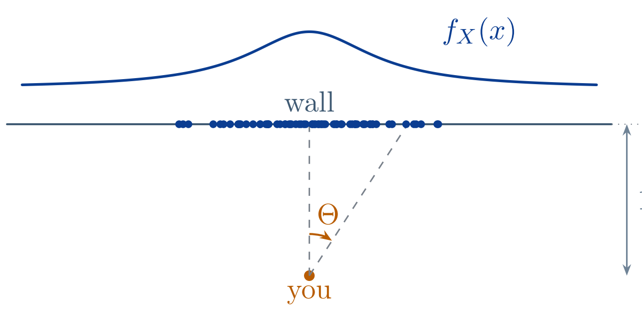

FIGURE 10

If Θ is the firing angle relative to the dashed center line and Θ is uniformly distributed on

$(-\pi/2, \pi/2)$, then the horizontal hit location X on the wall is

$$X = \tan\Theta.$$

The density of X is

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}(X \le x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}(\Theta \le \arctan x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\arctan x}{\pi} + \frac{1}{2}\right) = \frac{1}{\pi}\cdot\frac{1}{1+x^2}$$

which is exactly the Cauchy distribution density. The unscaled curve $y = \frac{1}{(1+x^2)}$ is called the **Witch of Agnesi**, and multiplying by $1/\pi$ rescales it to have total area 1.

Example 20.11 (Problem 9.40) Compute $\mathbb{E}[X]$ when X has normal distribution $\mathcal{N}(a, b^2)$, where $b > 0$.

Solution. Suppose first that $a = 0$ and $b = 1$, so the distribution is standard. Then

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x.$$

Call the integrand $g(x)$. Notice that it is an odd function, since $g(-x) = -g(x)$. Therefore $\mathbb{E}[X] = 0$. Now suppose that $X \sim \mathcal{N}(a, b^2)$. Then

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}b} e^{-\frac{(x-a)^2}{2b^2}}\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} (x-a)\frac{1}{\sqrt{2\pi}b} e^{-\frac{(x-a)^2}{2b^2}}\,\mathrm{d}x + a.\end{aligned}$$

Let $y = x - a$. Then

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{2\pi}b} e^{-\frac{y^2}{2b^2}}\,\mathrm{d}y + a = a,$$

because the integrand in the first term is odd, so that integral is 0. Thus $\mathbb{E}[X] = a$. $\square$

Lecture 21 — Thursday, November 21, 2013

Expectation and Variance for Common Distributions

Example 21.1 (Problem 9.25) Compute the variance of a random variable with the binomial distribution $\operatorname{Bin}(n, p)$.

Solution. Suppose that $X \sim \operatorname{Bin}(n, p)$. The case $n = 1$ is the Bernoulli calculation from Problem 9.20, so assume $n \geq 2$. Then $\mathbb{E}[X] = np$, and

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

We also have

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x (1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} (1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^k (1-p)^{n-2-k} \\
&= n(n-1)p^2 \cdot 1 = n(n-1)p^2.
\end{aligned}$$

Therefore $\mathbb{E}\left[X^2\right] - \mathbb{E}[X] = n(n-1)p^2$, so

$$\mathbb{E}\left[X^2\right] = n(n-1)p^2 + np = n^2p^2 - np^2 + np.$$

Hence, the variance is

$$\operatorname{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$

□

This is the natural generalization of the Bernoulli case. In fact, we can obtain it by adding n independent Bernoulli trials.

Definition 21.2 Suppose that $X \geq 0$ is a random variable. Then expectation is a function $X \mapsto \mathbb{E}[X] \in [0, \infty]$ satisfying the following properties:

1. $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$
2. If $c \geq 0$ and $Y \geq 0$, then $\mathbb{E}[cX + Y] = c\mathbb{E}[X] + \mathbb{E}[Y]$.
3. (**Monotone convergence**) If $0 \leq X_n \nearrow X$, then $0 \leq \mathbb{E}[X_n] \nearrow \mathbb{E}[X]$

Warning Note that in general $X_n \to X$ does not imply $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Definition 21.3 We define $X^+ = \max\{X, 0\}$, that is, $X^+(\omega) = \max\{X(\omega), 0\}$, the **positive part** of X. We also define $X^- = (-X)^+ = \max\{-X, 0\}$, the **negative part** of X.

Note that both X^+ and X^- are nonnegative. We have $X^+ - X^- = X$ and $X^+ + X^- = |X|$.

Definition 21.4 We say that X is **summable** (or **integrable**) if $\mathbb{E}[X^+], \mathbb{E}[X^-] < \infty$. For summable X, we define

$$\mathbb{E}[X] = \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right].$$

10. Limit Theorems

The Weak Law of Large Numbers

There are three major limit theorems in probability: the weak law of large numbers (**WLLN**), the **strong law of large numbers** (**SLLN**), and the central limit theorem (**CLT**).

Suppose $X_1, X_2, \ldots$ are independent and identically distributed, and let

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Then we expect $\overline{X}_n \to \mathbb{E}[X_1]$ as $n \to \infty$. Why is this plausible? If $\mathrm{Var}(X_i) = \sigma^2$, then

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n},$$

which tends to 0 as $n \to \infty$.

Example 21.5 (Problem 12.3: Markov's inequality) Suppose $X \geq 0$ and that $\mathbb{E}[X] < \infty$. Then for all $a > 0$, $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$.

Proof. Let

$$\mathbb{1}_{X \geq a} = \begin{cases} 1, & X \geq a, \\ 0, & X < a, \end{cases} \qquad \text{and} \qquad \mathbb{1}_{X < a} = \begin{cases} 1, & X < a, \\ 0, & X \geq a. \end{cases}$$

Then

$$\mathbb{1}_{X \geq a} + \mathbb{1}_{X < a} = 1,$$

so

$$\mathbb{E}[X] = \mathbb{E}[X \cdot 1] = \mathbb{E}[X(\mathbb{1}_{X \geq a} + \mathbb{1}_{X < a})] = \mathbb{E}[X \cdot \mathbb{1}_{X \geq a}] + \mathbb{E}[X \cdot \mathbb{1}_{X < a}].$$

Since $X \geq 0$, we have

$$\mathbb{E}[X] \geq \mathbb{E}[X \cdot \mathbb{1}_{X \geq a}] \geq \mathbb{E}[a \cdot \mathbb{1}_{X \geq a}] = a \cdot \mathbb{E}[\mathbb{1}_{X \geq a}] = a \cdot \mathbb{P}(X \geq a).$$

Therefore

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

□

Example 21.6 (Problem 12.4: Chebyshev's inequality) Let X have finite variance $\mathrm{Var}(X)$. Then for $a > 0$, $\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$.

Proof. Let $Y = (X - \mathbb{E}[X])^2 \geq 0$. Applying Markov's inequality to Y, we get

$$\mathbb{P}(Y \geq a^2) \leq \frac{\mathbb{E}[Y]}{a^2}.$$

Thus

$$\mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2) \leq \frac{\text{Var}(X)}{a^2}.$$

Therefore

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2},$$

by taking square roots. □

Example 21.7 (Problem 12.7) In the setting of repeated fair coin tosses, let S_n be the number of heads in n tosses. Show that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| < \varepsilon\right) = 1.$$

Solution. We compute

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n}\mathbb{E}[S_n] = \frac{1}{n}\left(n \cdot \frac{1}{2}\right) = \frac{1}{2}$$

and

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\,\text{Var}(S_n) = \frac{1}{n^2} \cdot \frac{n}{4} = \frac{1}{4n}.$$

Now, for any $\varepsilon > 0$,

$$0 \leq \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \frac{\text{Var}(S_n/n)}{\varepsilon^2} = \frac{1}{4n\varepsilon^2} \longrightarrow 0,$$

by Chebyshev's inequality. Therefore

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| < \varepsilon\right) = 1.$$

□

Theorem 21.8 (Weak law of large numbers) Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with finite variances. Let

$$S_n = \sum_{i=1}^{n} X_i$$

and $\overline{X}_n = \frac{S_n}{n}$. Then for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\left|\overline{X}_n - \mathbb{E}\left[X_1\right]\right| < \varepsilon) = 1.$$

Proof. Write $\mu = \mathbb{E}\left[X_1\right]$ and $\sigma^2 = \operatorname{Var}(X_1) < \infty$. Since the X_i are identically distributed, $\mathbb{E}\left[X_i\right] = \mu$ and $\operatorname{Var}(X_i) = \sigma^2$ for each i. First,

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] = \frac{1}{n}\cdot n\mu = \mu.$$

Next, using independence,

$$\operatorname{Var}(S_n) = \sum_{i=1}^{n}\operatorname{Var}(X_i) = n\sigma^2,$$

so

$$\operatorname{Var}(\overline{X}_n) = \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\operatorname{Var}(S_n) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

By Chebyshev's inequality, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\operatorname{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

Therefore

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1 - \mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \xrightarrow[n\to\infty]{} 1.$$

Since $\mu = \mathbb{E}\left[X_1\right]$, this is exactly the claimed statement. □

Why is this called the *weak* law? Compare it with the statement

$$\mathbb{P}\left(\lim_{n\to\infty}\overline{X}_n = \mathbb{E}\left[X_1\right]\right) = 1,$$

which is stronger and corresponds to the strong law of large numbers. The distinction comes from where the limit appears.

Lecture 22 — Tuesday, November 26, 2013

Central Limit Theorem

De Moivre's version of the central limit theorem uses **Stirling's approximation** $n! \sim \sqrt{2\pi n} n^n e^{-n}$ to approximate the distribution of a normalized binomial random variable as $n \to \infty$, which more generally leads to the density

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

for the limit of $\dfrac{\overline{X}_n - \mathbb{E}[X_1]}{\sqrt{\mathrm{Var}(X_1)}/\sqrt{n}}$ as $n \to \infty$ in the independent and identically distributed setting. Recall that the normal (or Gaussian) distribution has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma^2 > 0$, so f_Z is a special case of f_X; the associated distribution is called the **standard normal distribution**. Its expectation is $\mathbb{E}[X] = \mu$. The distribution function of the general normal distribution is

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t,$$

which cannot usually be evaluated in elementary terms, so we use tables or software.

To standardize, write

$$\mathbb{P}(a \le X \le b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right).$$

If $Z \sim \mathcal{N}(0,1)$, this becomes

$$F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right).$$

Theorem 22.1 (Central limit theorem) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with $\mathbb{E}[X_n] = \mu$ and $0 < \mathrm{Var}(X_n) = \sigma^2 < \infty$. Let S_n be the sum of the first n random variables. Then, as $n \to \infty$,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Longrightarrow \mathcal{N}(0,1).$$

More precisely,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t.$$

Roughly speaking, the central limit theorem says that S_n is approximately $\mathcal{N}(n\mu, n\sigma^2)$ and $\overline{X}_n$ is approximately $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ when n is large.

The proof is beyond the scope of our course.

If we toss a coin 100 times, the probability of getting exactly 50 heads is

$$\binom{100}{50}\left(\frac{1}{2}\right)^{100}.$$

It is quite small. But how small?

Example 22.2 (Problem 12.20) Estimate the probability that the number of heads in 10,000 independent tosses of a fair coin differs by less than 1% from 5000.

Solution. This is similar to the previous problem, but it is easier to use the central limit theorem. We use the basic normal approximation without a continuity correction. Let S_n be the number of heads in n tosses, and define

$$X_i = \begin{cases} 1, & \text{if toss } i \text{ is H,} \\ 0, & \text{if toss } i \text{ is T.} \end{cases}$$

Then $\mathbb{E}\left[X_i\right] = \frac{1}{2} = \mu$, $\mathrm{Var}(X_i) = \frac{1}{4} = \sigma^2$, and $S_n = X_1 + \cdots + X_n$ with $n = 10{,}000$. Therefore

$$\begin{aligned} \mathbb{P}(|S_{10{,}000} - 5000| < 50) &= \mathbb{P}(-50 < S_{10{,}000} - 5000 < 50) \\ &= \mathbb{P}\left(\frac{-50}{\sqrt{10,000}\cdot\frac{1}{2}} < \frac{S_{10{,}000} - 5000}{\sqrt{10,000}\cdot\frac{1}{2}} < \frac{50}{\sqrt{10,000}\cdot\frac{1}{2}}\right) \\ &\approx \mathbb{P}(-1 < Z < 1) \quad [\text{where } Z \text{ is } \mathcal{N}(0,1)] \\ &= F_Z(1) - F_Z(-1) \end{aligned}$$

$$= \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt \approx 0.68.$$

□

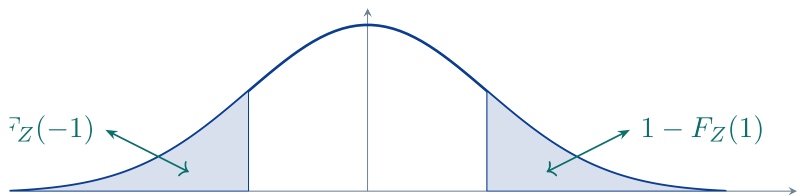

FIGURE 11

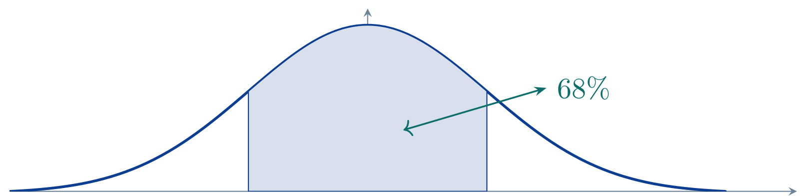

FIGURE 12

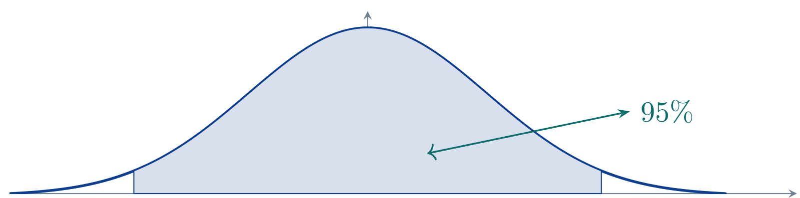

FIGURE 13

Figures 11 to 13 illustrate the area under the standard normal curve f_Z corresponding to the probabilities calculated in Problem 12.20. The first shows the tails $F_Z(-1)$ and $1 - F_Z(1)$, while the subsequent plots show the central intervals containing 68% and 95% of the total mass (corresponding to upper endpoints of approximately 1σ and 1.96σ, respectively).

Appendix: Lecture and Tutorial Index