MA 590 Notes

January–April 2016

Contents

1 Experiments, outcomes, sample space, events, probability

`Wednesday 6 January and Monday 11 January 2016`

1. An experiment is something that produces an outcome (out of a number of possible outcomes) each time it is carried out. We shall regard experiments as repeatable, which means that the likelihood of a particular outcome turning up remains the same each time the experiment is carried out. Examples of experiments are tossing a fair coin, rolling a die, tossing a fair coin 10 times in succession, tossing a biased coin 10 times in succession, etc.

2. For an experiment as defined above, the set of all possible outcomes constitutes its sample space, denoted Ω. (In other words, each time an experiment is carried out, the result is always some element or the other of the corresponding Ω.) For tossing a particular coin once, $\Omega = \{H, T\}$. For tossing a coin twice, $\Omega = \{HH, HT, TH, TT\}$.

3. Consider $\Omega = \{H, T\}$ (tossing a particular coin once). This coin will have some chance of turning up H, but by itself Ω does not tell us anything about the chance of turning up H. For that one must bring in the additional concept of a *probability* which is a mathematical notion that captures the 'chance of turning up a particular way'. Let X denote the outcome of tossing our coin. Interpret '$\mathbb{P}\{X = H\} = p$' to mean that in a large number of coin tosses, the fraction of tosses turning up H is most likely to be p. Being the most likely value for a fraction, p should lie between 0 and 1.

4. For a fair coin, the chance of turning up H should equal the chance of turning up T. In a large number of tosses, we would expect the number of H tosses to most likely be half the total number of tosses. Or, the fraction of H tosses to most likely be 1/2. Thus, for a fair coin, the 'fairness' by captured by the equation '$\mathbb{P}\{X = H\} = 1/2$'.

5. Tossing a fair coin once, and tossing a biased coin once, both experiments generate the same sample space $\Omega = \{H, T\}$. The difference between the experiments is captured by the fact that $\mathbb{P}_{\text{fair}}$ and $\mathbb{P}_{\text{biased}}$ are different.

6. When the sample space Ω is finite, any subset of Ω is also known as an event. Thus for $\Omega = \{1, 2, 3, 4, 5, 6\}$ (throwing a die), all of the following are events: $\emptyset, \{1\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}$. In general, for a finite Ω, each and every subset of Ω is an event.

7. Typically events are subsets of Ω that are describable by some natural property. For example, the event $\{HTT, THT, TTH\}$ can be described by the property "in 3 coin tosses, H turns up exactly once". The event $\{(6, 4), (4, 6), (5, 5), (5, 6), (6, 5), (6, 6)\}$ can be described as "in a throw of 2 dice you get a sum of 10 or more".

8. If A represents an event (a collection of outcomes) and we think of '$\mathbb{P}\{X \in A\}$' as 'the most likely fraction of experiments with outcome contained in A' then the following makes sense

   - For any event A its probability $\mathbb{P}\{A\}$ satisfies $\boxed{\mathbb{P}(A) \in [0, 1]}$.
   - $\boxed{\mathbb{P}(\Omega) = 1 \text{ always!}}$
   - Probabilities add up for disjoint events: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for $A \cap B = \emptyset$. In particular, for finite sample spaces $\mathbb{P}(A) = \sum_{a \in A} \mathbb{P}\{a\}$. Actually more can be said. Probabilities are countably additive:
     $\boxed{\text{If } A_1, A_2, \ldots \text{ are pairwise disjoint events, then } \mathbb{P}(\cup_1^\infty A_i) = \sum_1^\infty \mathbb{P}(A_i).}$

   Any $P(\cdot)$ defined on the collection of all events and satisfying the 3 $\boxed{\text{boxed}}$ properties is called a probability.

9. When Ω is uncountable, labeling every subset of Ω as an event will make it impossible to define a probability on all events in a consistent way. (A function $\mathbb{P}(\cdot)$ defined on all subsets of Ω will fail to always satisfy the 3 $\boxed{\text{boxed}}$ properties.) It turns out, however, that if we regard as events only those subsets of Ω that are describable by natural properties then it becomes possible to define a probability for all events in a consistent way – ie, in a way which never violates any of the 3 $\boxed{\text{boxed}}$ properties.

10. To summarize: for practical purposes for this course think of any subset of Ω that comes up while doing calculations as an event – ie, something for which a probability is defined. (If curious, you can return to this point later – after studying measure theory – to see why defining a probability for every subset of Ω may not be possible.)

11. Let E and F be two events (subsets of Ω) for an experiment. Set theoretic operations lead naturally to new events $E \cup F$ and $E \cap F$ (also written as EF) derived from E and F . In words, $E \cup F$ is the collection of outcomes that lie in E or F (or both). Similarly, $E \cap F$ is the collection of outcomes that lie in both E and F.

12. Two events E and F are said to be *mutually exclusive* if $E \cap F = \emptyset$. In words, E and F are mutually exclusive if no possible outcome lies simultaneously in both E and F – if it is in one then it cannot be in the other.

13. For an event E, its set theoretic complement $\Omega \backslash E$ represents a new event, the *complement* of E (written E^{c}). (i) Are E and E^{c} always mutually exclusive? (ii) What is $\mathbb{P}(E^{\mathrm{c}})$? Justify using properties of a probability that it is $1 - \mathbb{P}(E)$.

14. Tossing a fair die. (i) Write down Ω for tossing a die (*any* die, not necessarily a fair one). (ii) Give a $\mathbb{P}(\cdot)$ for a die which always comes up odd. (iii) Give a $\mathbb{P}(\cdot)$ for a fair die. (iv) For a fair die, what is the probability of getting an even number?

2 Venn diagrams, conditional probability, independent events, Baye's formula

`Tuesday 12 January and Wednesday 13 January 2016`

1. Given two events E and F there will be outcomes in Ω that lie in both events, in exactly one, and in neither. This state of affairs is captured by a Venn diagram. In the diagram identify the largest disjoint sets that make up all the sets appearing in the diagram. Thus, for general E and F the three disjoint sets $A = E \backslash F, B = F \backslash E$ and $C = EF$ are the building blocks. An identity like $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(EF)$ can be verified by writing each set in terms of the disjoint building blocks. Thus

$$\begin{aligned}
\mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(EF) &= \mathbb{P}(A \cup C) + \mathbb{P}(B \cup C) - \mathbb{P}(C) \\
&= \{\mathbb{P}(A) + \mathbb{P}(C)\} + \{\mathbb{P}(B) + \mathbb{P}(C)\} - \mathbb{P}(C) \\
&= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\
&= \mathbb{P}(E \cup F).
\end{aligned}$$

2. Try writing $\mathbb{P}(E \cup F \cup G)$ in terms of $\mathbb{P}(E), \mathbb{P}(F), \mathbb{P}(G)$ and the probabilities of various intersections of these events – ie, the events $E \cap F, F \cap G, E \cap G$ and $E \cap F \cap G$.

3. Throw a fair die. Let D denote the outcome. What is $\mathbb{P}\{D = 4\}$? Next, let E be the event that the outcome is even and ask a friend to tell you whether E occurs or E^c occurs. Will the friend's answer change the likelihood that $D = 4$? *Conditional probability* is used to describe the chance of some event based on the information that some other event has occurred. Here, we ask for the chance of $\{D = 4\}$ based on the knowledge that event E has occurred. Knowing E has occurred, we would expect equal chances for D being $2, 4$ or 6, assuming a fair die, and no chance for D being $1, 3$ or 5. We write this as

$$\mathbb{P}(\{D = 4\}|E) = 1/3.$$

4. `Question.` Throw a biased die. Let $p_1, p_2, \ldots, p_6$ be the six probabilities (interpreted naturally). Let D denote the random outcome of this experiment. Let $E = \{2, 4, 6\}$ (die comes up even) and $F = \{1, 2\}$. The probabilities for the two events are $\mathbb{P}(E) = p_2 + p_4 + p_6$ and $\mathbb{P}(F) = p_1 + p_2$. Suppose we know that event E has occurred but don't know anything more about the outcome. What should be our best estimate for the (conditional) likelihood of event F?

5. `Answer.` Imagine repeating the experiment a large number of times, say N times. Let N_E be the number of occurrences of event E. From among the N_E occurrences of E, further count the number of occurrences of F. These are outcomes where we first filter for E, and next filter for F. Denote this count as $N_{E\cap F}$. Our best estimate for the (conditional) likelihood of event F given that event E has occurred should be $N_{E\cap F}/N_E$. Recall that this is an imagined ratio, so we should talk about our best belief of its most likely value. Since $E = \{2, 4, 6\}$ and $E \cap F = \{2\}$, we should estimate the most likely value for N_E to be $N(p_2 + p_4 + p_6)$, and for $N_{E\cap F}$ to be Np_2. For the ratio $N_{E\cap F}/N_E$ we would estimate the most likely value to be $p_2/(p_2 + p_4 + p_6)$.

6. Guided by this argument we can give a general definition for the *conditional probability* of event F given the occurrence of event E: $\boxed{\mathbb{P}(F|E) = \frac{\mathbb{P}(E\cap F)}{\mathbb{P}(E)}}$.

7. In thinking probabilistically it often helps to think of a Venn diagram where the various sets (ie, the various events) have a size equal to their probability. Imagine throwing a dart uniformly at random at such a Venn diagram. Such a dart never falls outside the diagram and inside it, it falls uniformly (ie, without any bias or preference for any part of the diagram). The probability for such a random dart to hit a set E will be exactly $\mathbb{P}(E)$. Now imagine throwing darts till one hits E. What is the probability that this dart hits F? We know only that the dart has hit E, so we should believe that the likelihood of its exact location is evenly spread over E in the diagram. The chance of also being in the F part of the Venn diagram should just be the ratio of the area of $E \cap F$ (as in the diagram) to the area of E. By properties of our dart and diagram, this is just $\mathbb{P}(E \cap F)/\mathbb{P}(E)$, exactly matching the definition of $\mathbb{P}(F|E)$.

8. Let Ω be a sample space. Let $F_1, \ldots, F_n$ be mutually exclusive with $F_1 \cup \ldots \cup F_n = \Omega$ (illustrate this on a Venn diagram), and let E be arbitrary. *Baye's formula* expresses $\mathbb{P}(F_j|E)$ in terms of the various $\mathbb{P}(E|F_i)$, the conditional probabilities conditioned the opposite way.

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i)}.$$

   It is helpful to draw a Venn diagram and locate each of the above terms in the diagram.

9. Two events E and F are said to be *independent* if

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

Independence can be defined for more than 2 events also.

10. A collection of finitely many events $E_1, E_2, \ldots, E_n$ is said to be (jointly) independent if for every sub–collection of the original collection (denoted $E_{1'}, E_{2'}, \ldots, E_{r'}$ with $r \leq n$) the following holds

$$\mathbb{P}(E_{1'}E_{2'} \cdots E_{r'}) = \mathbb{P}(E_{1'})\mathbb{P}(E_{2'}) \cdots \mathbb{P}(E_{r'}).$$

11. On $\Omega = \{1, 2, 3, 4\}$ give a probability, and events A, B and C, such that the events are pairwise independent, but not jointly independent. (Give an example different from the solved example in the book.)

3 Examples

`Monday 18 January to Friday 22 January 2016`

1. Example 1.4 of 'Introduction to Probability Models, 10^{th} edition (Sheldom M. Ross). Draw a Venn diagram. One of ten cards, each with a number from 1 to 10, is pulled out at random, each number equally likely with probability 1/10. If the pulled out card has a number from 5 to 10, what is the (conditional) probability that the number is 10?

   Draw a Venn diagram. Verify that the answer is $\frac{1/10}{6 \times 1/10} = 1/6$.

2. Example 1.5 A couple has two children. Given that at least one is boy, what is the probability that both are boys?

   Imagine doing the experiment a large number of times: a couple has two children and the whole process is repeated N times (say, a million). Each time, the possible outcomes are $(g, g), (g, b), (b, g)$, and (b, b). What are the most likely counts for each? What is the sum of the four counts? (This shows that we are splitting a count into four sub–counts.) Draw a Venn diagram with these four mutually exclusive outcomes. Give each a probability equal to its most likely count divided by N. The required conditional probability is the ratio of the probabilities of two events. Identify the corresponding events on the Venn diagram. Check that the conditional probability is 1/3.

3. Example 1.6. A student tosses a fair coin and takes either computer science (with 50% chance of grade A) or takes chemistry (with a 1/3 chance of grade A). What is the probability that the student ends up with grade A in chemistry?

   Imagine repeating the experiment a large number of times (say N times).

   - Split N into two counts, one for {student takes computer science}, and one for {student takes chemistry}.
   - Split the count for {student takes computer science} further into two sub–counts, one for {student takes computer science and gets A} and one for {student takes computer science and doesn't get A}. Repeat for chemistry.

- Draw a Venn diagram with these four mutually exclusive events and give each a probability proportional to its count.
- Probability that the student gets an A in chemistry is $\frac{1}{2} \times \frac{1}{3} = 1/6$.

4. Example 1.8. Three men with a hat each. After mixing the hats, what is the chance that no man gets his own hat?

   Again imagine repeating the experiment N times, and estimate the most likely value for each possible outcome. So {M1 get H1} occurs most likely in $N/3$ cases. These $N/3$ cases are the disjoint union of the two cases {M1 gets H1; M2 gets H2} and {M1 gets H1; M2 gets H3}, each with a most likely count of $N/6$. Similarly, {M1 get H2} leads to two more disjoint events, and {M1 get H3} leads to two more still. We finally get 6 disjoint events each with a most likely count of $N/6$. So we can give each a probability of 1/6. From the Venn diagram, the probability that no man gets his own hat is now 2/6.

5. Note the process of splitting a count into sub–counts. Thus, in the above example, we split N experiments into 3 events {M1 get H1}, {M1 get H2}, {M1 get H3}) each having a most likely count of $N/3$. (Note the conservation of counting: the sub–counts add up to the original count.) Again, the count for {M1 get H1} is further split into two sub–counts, {M1 gets H1 and M2 gets H2} and {M1 gets H1 and M2 gets H3}, each with a most likely count of $N/6$ since the splitting should show no preference either way. Note that $N/6 + N/6 = N/3$, so the sub–counts adds up to the original count.

6. Example 1.12. There is an urn with 2 white balls and 7 black ones. A second urn has 5 white balls and 6 black ones. 'H' on a fair coin leads to the first urn, and 'T' to the second urn. The experimenter pulls out a ball uniformly at random from the chosen urn.

7. Example 1.15. A letter is in one of three folders, each equally likely. The man chooses a folder at random and checks it for the letter. His success probability is α_i if he looks up folder i and it holds the letter ($i = 1, 2, 3$). He decides to look up folder 1 but fails to find the letter. What is the probability that the letter is actually in folder 1?

   Again, imagine repeating the experiment N times and try estimating the most likely counts for various outcomes. So a most likely count of $N/3$ for {letter in folder i} ($i = 1, 2, 3$). The count for {letter in folder i} further splits into two sub–counts, depending on whether the man finds the letter or not. So the count for finding the letter should most likely be $\alpha_1 \times N/3$ and for not finding the letter should most likely be $(1 - \alpha_1) \times N/3$. With these counts, we can associate probabilities for the various events on the Venn diagram, and, with that done, calculate the conditional probability.

8. Pairwise independent, but not jointly independent.

   Let $\Omega = \{1, 2, 3, 4\}$ with $\mathbb{P}\{1\} = \mathbb{P}\{2\} = \mathbb{P}\{3\} = \mathbb{P}\{4\}$. Let $R = \{1, 2\}, C = \{1, 4\}$ and $D = \{1, 3\}$. Then the three are not jointly independent. (Can choose $D' = \{2, 4\}$, and work with R, C and D' also.)

4 Random variables, cumulative distribution function

`Monday 25 January and Wednesday 27 January 2016`

1. Let Ω be the sample space for an experiment. A random variable (often written X, Y, Z etc) is a function which assigns a real number to each outcome of the experiment ($X : \Omega \to \mathbb{R}$).

   Consider tossing a coin till it first shows up an H. Let N denote the (random) number of tosses, counting up till the first H. Then N is a random variable. Assume each toss is independent of all other tosses and H comes up with probability p. Then

   $$\mathbb{P}\{N = n\} = (1-p)^{n-1}p \text{ for } n = 1, 2, \ldots.$$

   All possible outcomes for this experiment can be written as a union of mutually exclusive events, $\cup_1^\infty \{N = n\}$. Verify

   $$\mathbb{P}(\text{ all possible outcomes }) = 1.$$

2. Given a random variable X (on an underlying sample space), the cumulative distribution function for the random variable X is the function $F(\cdot) \mapsto [0, 1]$ given by

   $$F(b) = \mathbb{P}\{X \leq b\} \ \forall \ b \in \mathbb{R}.$$

3. If $b_1 \leq b_2$ then $F(b_1) \leq F(b_2)$.

   Since $\{X \leq b_2\} = \{X \leq b_1\} \cup \{b_1 < X \leq b_2\}$, a union of disjoint sets, we must have $\mathbb{P}\{X \leq b_2\} = \mathbb{P}\{X \leq b_1\} + \mathbb{P}\{b_1 < X \leq b_2\}$. Since probabilities are always nonnegative, the result follows.

4. $\lim_{b\to\infty} F(b) = F(\infty) = 1$.

   Note that $\{X \in \mathbb{R}\} = \{X \leq 0\} \cup \{0 < X \leq 1\} \cup \{1 < X \leq 2\} \cup \ldots$, a countable union of disjoint sets. Properties of a probability give

   $$\mathbb{P}\{X \in \mathbb{R}\} = F(0) + \big(F(1) - F(0)\big) + \big(F(2) - F(1)\big) + \ldots.$$

   Recall the definition of an infinite sum: $\sum_1^\infty a_i = \lim_{m\to\infty} \Big(\sum_1^m a_i\Big)$. Next, note that for a function $X : \Omega \to \mathbb{R}$ we must have $X^{-1}(\mathbb{R}) = \Omega$, or, $\{X \in \mathbb{R}\} = \Omega$. Consequently, $\mathbb{P}\{X \in \mathbb{R}\} = 1$. The two observations give

   $$\begin{aligned} 1 &= F(0) + \sum_1^\infty \Big(F(i) - F(i-1)\Big) \\ &= F(0) + \lim_{m\to\infty} \sum_1^m \Big(F(i) - F(i-1)\Big) \\ &= \lim_{m\to\infty} F(m). \end{aligned}$$

   Note that we have let $m \to \infty$ along integer values. Extending to the case where m is real valued is a simple exercise in using the definition of convergence and the fact that $F(m) \geq F(\lfloor m \rfloor)$ for all $m \in \mathbb{R}$. (Write it formally.)

5. $\lim_{b\to-\infty} F(b) = F(-\infty) = 0$.

   Let $m \in \mathbb{Z}$. Write $\{X \in \mathbb{R}\} = \{0 < X\} \cup \{-1 < X \leq 0\} \cup \{-2 < X \leq -1\} \cup \ldots$, a countable union of disjoint sets. We get

   $$1 = \mathbb{P}\{0 < X\} + \mathbb{P}\{-1 < X \leq 0\} + \mathbb{P}\{-2 < X \leq -1\} + \ldots,$$

which gives

$$1 = \Big(1 - F(0)\Big) + \lim_{m\to\infty} \sum_{1}^{m} \Big(F(1-i) - F(-i)\Big) = 1 - \lim_{m\to\infty} F(-m),$$

giving $\lim_{m\to\infty} F(-m) = 0$. The proof for $m \in \mathbb{Z}$ extends as before to $m \in \mathbb{R}$.

6. Here's another proof. Let $F(\cdot)$ be the distribution function for X. Define a new random variable, $Y = -X$. This means that X and Y are both defined on the same sample space Ω, and $Y(\omega) = -X(\omega)$ for all $\omega \in \Omega$. Let $G(\cdot)$ be the c.d.f. for Y. Recall that $A \subset B \Rightarrow \mathbb{P}(A) \le \mathbb{P}(B)$. We get

$$0 \le F(b) = \mathbb{P}\{X \le b\} = \mathbb{P}\{Y \ge -b\} \le \mathbb{P}\{Y > -b-1\} = 1 - G(-b-1).$$

Since $G(-b-1) \to 1$ as $b \to -\infty$ (property of a c.d.f.), it follows that $\lim_{b\to-\infty} F(b) = 0$.

7. Show that $\mathbb{P}\{X \in [a,b]\} = F(b) - \lim_{h\downarrow 0} F(a-h)$.

5 Two problems

`Monday 1 February and Tuesday 2 February 2016`

1. Toss a fair coin repeatedly, each toss independent of all preceding tosses. What is the probability that:

   (i) The first occurrence of HH or TT happens in an even number of tosses?

   (ii) The first occurrence of HH happens in an even number of tosses?

2. (i) Define $P_{HH} = \{HT, HTHT, HTHTHT, HTHTHTHT, \ldots\}$. We calculate

$$\mathbb{P}(P_{HH}) = \frac{1}{4}\Big(1 + \big(1/4\big) + \big(1/4\big)^2 + \ldots\Big) = \frac{1}{3}.$$

Next, define

$$F_{HH} = \{\text{first occurrence of HH or TT is an HH on an even toss}\}.$$

Think of P_{HH} as a collection of prefixes for outcomes in F_{HH}. We can write

$$F_{HH} = \{HH\} \cup \{xHH : x \in P_{HH}\}.$$

We get

$$\mathbb{P}(F_{HH}) = \frac{1}{4} + \frac{1}{4} \times \mathbb{P}(P_{HH}) = \frac{1}{3}.$$

Similarly, $\mathbb{P}(F_{TT}) = \frac{1}{3}$. (This follows by symmetry, or else, by repeating the calculation above with the analogous events P_{TT} and F_{TT}.)

Now, the event that the first occurrence of HH or TT happens in an even number of tosses is just $F_{HH} \cup F_{TT}$. As F_{HH} and F_{TT} are disjoint events, we get

$$\mathbb{P}(\{\text{first occurrence of HH or TT happens on an even toss}\}) = \frac{2}{3}.$$

(ii) Define the sample space Ω and event E as

$$\Omega = \{\text{finite sequence of H and T ending with first occurrence of HH}\},$$

$$E = \{\text{first occurrence of HH happens after an even number of tosses}\}.$$

Write $p = \mathbb{P}(E)$. We give two different ways to calculate p.

(a) Define the events

$$E_T = \{\text{first toss is a T}\},$$
$$E_{HT} = \{\text{first two tosses are HT}\},$$
$$E_{HH} = \{\text{first two tosses are HH}\}.$$

We can write Ω as the disjoint union of E_T, E_{HT} and E_{HH}. Let . We get

$$p = \mathbb{P}(E \cap E_T) + \mathbb{P}(E \cap E_{HT}) + \mathbb{P}(E \cap E_{HH}).$$

Look at sequences in $E \cap E_T$. If we chop off the first toss from every sequence in $E \cap E_T$, the resulting collection just looks like $\Omega \backslash E$, the event that the first occurrence of HH happens after an odd number of tosses. In other words, if we prefix every element of $\Omega \backslash E$ with a T, we get precisely the event $E \cap E_T$. So $\mathbb{P}(E \cap E_T) = \frac{1}{2} \times (1 - p)$.

Next, look at sequences in $E \cap E_{HT}$. If we chop off the first two tosses from every sequence in $E \cap E_{HT}$, the resulting collection just looks like E. So $\mathbb{P}(E \cap E_{HT}) = \frac{1}{4} \times p$. We get

$$p = \frac{1}{2} \times (1 - p) + \frac{1}{4} \times p + \frac{1}{4}.$$

Rearranging gives $\frac{5p}{4} = \frac{3}{4}$, or, $p = \frac{3}{5}$.

(b) Let $y \in E$. Since y ends on the first occurrence of the pattern HH, we can write $y = xHH$, where every H appearing in x is immediately followed by an occurrence of T. Therefore, we can think of x as a finite sequence of the two patterns, HT and T, with T occurring an even number of times. Let $E_{(2i,k-2i)}$ be all $y \in E$ such that the corresponding x has $2i$ occurrences (even number) of the pattern T, and $k - 2i$ occurrences of the pattern HT. The number of ways of choosing such an x is $\binom{k}{2i}$, and for each such x, $\mathbb{P}(x) = \left(\frac{1}{2}\right)^{2i}\left(\frac{1}{4}\right)^{k-2i}$. Therefore

$$\mathbb{P}(E_{(2i,k-2i)}) = \frac{1}{4}\binom{k}{2i}(1/2)^{2i}(1/4)^{k-2i}.$$

We get

$$\begin{aligned}
\mathbb{P}(E) &= \frac{1}{4}\sum_k \sum_i \binom{k}{2i}(1/2)^{2i}(1/4)^{k-2i} \\
&= \frac{1}{4}\sum_k \sum_j \binom{k}{j}\frac{(1/2)^j(1/4)^{k-j} + (-1/2)^j(1/4)^{k-j}}{2} \\
&= \frac{1}{8}\Big(\sum_k (1/2 + 1/4)^k + \sum_k (-1/2 + 1/4)^k\Big) \\
&= \frac{1}{8}\Big(\frac{1}{1 - 3/4} + \frac{1}{1 + 1/4}\Big) \\
&= \frac{3}{5}.
\end{aligned}$$

6 Exercises

`Tuesday 2 February and Wednesday 3 February 2016`

1. Exercise 1.5. A man bets a dollar quitting if he wins. If he loses, he bets 2 dollars and quits whatever the outcome.

   Draw a Venn diagram, with three exclusive events W, LW and LL. Calculate probabilities for each.

2. Exercise 1.13. Game of craps. A man throws a pair of dice. He wins if he gets one of $\{7, 11\}$, loses if he gets one of $\{2, 3, 12\}$ and goes to a second stage otherwise - ie, if he gets one of $\{4, 5, 6, 8, 9, 10\} = F$. In the second stage he keeps throwing the dice till he gets a 7 (and he loses) or a repeat of the number he got in the first stage (when he wins). What is his winning probability?

$$\begin{aligned}
\mathbb{P}(\text{win}) &= \mathbb{P}\{7\} + \mathbb{P}\{11\} + \sum_{f \in F} \mathbb{P}\{f\}\mathbb{P}(f|\{f, 7\}) \\
&= \frac{3}{36} + \frac{2}{36} + 2 \times \left[\frac{3}{36} \cdot \frac{3}{9} + \frac{4}{36} \cdot \frac{4}{10} + \frac{5}{36} \cdot \frac{5}{11}\right] \\
&= \frac{976}{1980}
\end{aligned}$$

3. A and B play a game with A likely to win with probability p. The first to win 2 more games than the other, wins overall. What is the chance of A winning? What is the chance that a player wins in exactly $2m$ rounds?

   Look at the result of two games combined with the players balanced when they start the two games. Either A wins with probability p^2, or B wins with probability q^2, or they are still balanced (with probability $2pq$). The probability that some player wins in exactly $2m$ rounds is thus $(2pq)^{n-1}(1 - 2pq)$. And, $\mathbb{P}(\text{ A wins }) = \frac{p^2}{p^2+q^2}$.

4. Exercise 1.24. Conjecture is that $\mathbb{P}_{n,m} = \frac{n-m}{n+m}$.

5. Exercise 1.28. If B makes A more likely, does A also make B more likely?

   Both conditions are equivalent to $\mathbb{P}(AB) > \mathbb{P}(A)\mathbb{P}(B)$. Hence one is true if and only if the other is true.

6. Exercise 1.30. Both George and Bill hit at a target independently. George hits it with probability g and Bill succeeds with probability b. Given that only one has hit the target, what is the probability that George has hit it? Given that the target has been hit at least once, what is the probability that George has hit it?

   Let B and G respectively denote the events that B alone hits the target and George alone hits the target. Let T denote the event that there are two hits - ie, both George and Bill have hit the target. Then $\mathbb{P}(B) = b(1 - g)$, $\mathbb{P}(G) = g(1 - b)$ and $\mathbb{P}(T) = bg$. We get

$$\mathbb{P}(G|B \cup G) = \frac{g(1-b)}{g(1-b) + b(1-g)},$$

$$\mathbb{P}(\text{ George hits } | \text{ at least one hit }) = \mathbb{P}(\text{ George hits } | B \cup G \cup T),$$

   which gives $\mathbb{P}(\text{ George hits } | \text{ at least one hit }) = \frac{g}{g+b-gb}$.

7 Bernoulli, binomial, geometric, & Poisson random variables

`Monday 8 February to Wednesday 10 February 2016`

1. A `Bernoulli random variable` takes two values. One can think of these two values as a $1/0$ or H/T or success/failure. Such a random variable is completely described by its success probability p. The failure probability is often written q - ie, $q = 1 - p$.

2. A `binomial random variable` is obtained by performing a given number of independent and identical Bernoulli trials and counting the number of successes therein. Thus, X is a binomial random variable if for suitable n and p we have

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k q^{n-k}.$$

3. A `geometric random variable` counts the position of the first success in a sequence of i.i.d. Bernoulli trials. Thus, X is a geometric random variable if
$\boxed{\mathbb{P}\{X = k\} = q^{k-1}p \text{ with } p + q = 1.}$
To see this, imagine a sequence of i.i.d. Bernoulli trials, each with success probability p. X will equal k if the first $k-1$ trials are failures and the k^{th} trial is a success. The corresponding probability is exactly as given.

4. A `Poisson random variable` is a limiting binomial random variable obtained as follows.

   - Let B_n be a binomial random variable with n i.i.d. Bernoulli trials, each with success probability $p = \lambda/n$. Then

   $$\mathbb{P}\{B_n = k\} = \left(\begin{smallmatrix} n \\ k \end{smallmatrix}\right) (\lambda/n)^k (1 - \lambda/n)^{n-k}.$$

   - Verify that $\lim_{n\to\infty} \mathbb{P}\{B_n = k\} = \exp(-\lambda)\frac{\lambda^k}{k!}$. The limit is obtained when the number n of Bernoulli trials goes to infinity, and the success probability p of each trial goes to zero, with the product np held constant at λ.
   - X is said to be a Poisson random variable of intensity λ if
   $\boxed{\mathbb{P}\{X = k\} = \exp(-\lambda)\frac{\lambda^k}{k!} = \lim_{n\to\infty} \mathbb{P}\{B_n = k\}.}$

5. An example to clarify how the Poisson random variable comes about.

6. Example 2.8. An engine can fail with probability q. A plane crashes if more than half of its engines fail. When is a four engine plane safer than a two engine one?

   The book works with survival probability. Working with crash probability reduces a term each. The four engine plane is safer if its crash probability is lower - ie, if $q^4 + \binom{4}{3}q^3p < q^2$. This gives

   $$0 < 1 - 4q + 3q^2 = (1 - 3q)(1 - q).$$

   This gives $q < 1/3$, which is the same as saying $p > 2/3$ (as given in the book).

7. A continuous random variable is one for which probabilities can be calculated by integration. Thus, X is said to be a continuous R. V. if
$$\mathbb{P}\{X \in B\} = \int_B f(x)\mathrm{d}x, \text{ valid for all } B \subset \mathbb{R}.$$
The function $f(\cdot)$ is called the probability distribution function, and since $\mathbb{P}\{X \in B\} \geq 0$ for all B, we assume that $f(x) \geq 0$ for all x.

8. $\mathbb{P}\{X = a\} = 0$ for all $a \in \mathbb{R}$, and $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$.

9. The cumulative distribution function for a random variable which has a p.d.f. (ie, a continuous random variable) is defined as
$$F(a) = \int_{-\infty}^{a} f(x)\mathrm{d}x = \mathbb{P}\{X \in (-\infty, a]\}.$$
Note that this is the same as $\mathbb{P}\{X \in (-\infty, a)$.

10. A uniform random variable on $(0, 1)$ has p.d.f. given by
$$f(x) = \begin{cases} 1 & \text{if } x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$
How will the c.d.f. of this random variable look like?

11. For a random variable uniform on (α, β) the p.d.f. will be
$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } x \in (\alpha,\beta), \\ 0 & \text{otherwise.} \end{cases}$$

12. An exponential random variable is a random variable with p.d.f.
$$f(x) = \begin{cases} \lambda \exp^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$
Its c.d.f. is
$$F(a) = \int_0^a f(x) = 1 - \exp^{-\lambda a} \text{ for } a \geq 0.$$

13. Imagine a very small number ϵ. Given $\lambda > 0$ consider a (sufficiently long) line segment with contiguous ϵ–sized pieces marked on it. Each piece "lights up" with probability $\lambda\epsilon$. The probability that the ϵ–piece at x is the first to light up is $(1-\lambda\epsilon)^{x/\epsilon}\lambda\epsilon$. If we think of this as $f(x)\epsilon$, and work out $f(x)$ on the assumptions that ϵ is very small, we get
$$f(x) = \lim_{\epsilon\to 0}(1-\lambda\epsilon)^{x/\epsilon}\lambda = \lambda \cdot \exp^{-\lambda x}.$$
This heuristic should give some idea of where the exponential distribution comes from.

14. If we think of pieces of size Δ (not necessarily small), then the number of pieces that light up in $[0, x]$ will have binomial distribution with $n = x/\Delta$ trials (assume $n \in \mathbb{Z}$), and success probability $p = \lambda\Delta$. So probability of k pieces lighting up will be $\binom{n}{k}p^k q^{n-k}$.

    If each piece lights up probability $\lambda\Delta$ (where Δ is the piece size), then in the limit as the piece size goes to zero (taking with it to zero the probability of each piece lighting up) then the probability of k pieces lighting up will be $\exp^{-\lambda x}\frac{(\lambda x)^k}{k!}$, which is the Poisson distribution.

15. Note the similarity in the heuristics used to understand the exponential distribution and the Poisson distribution. In both we think of very small contiguous pieces which light up independently, each with a very small probability. (The smaller you think each piece, the smaller its probability of lighting up).

8 Expected value for discrete random variables

`Monday 15 February 2016`

1. Let X be a random variable taking countably many values $x_1, x_2, \ldots$ (this could be finite too) with probabilities $p_1, p_2, \ldots$. If $\sum |x_i| p_i < \infty$ then the sum $\sum x_i p_i$ is well defined and remains the same no matter in what order we add up the various terms. (Given an absolutely convergent series, the sum remains the same for every rearrangement of terms.) If $\sum |x_i| p_i < \infty$, we define the *expected value* of the random variable X to be

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i p_i.$$

2. Bernoulli
3. Binomial
4. Geometric
5. Poisson

9 Exercises

`Friday 19 February 2016`

1. Exercise 2.19.
2. Exercise 2.20.
3. Exercise 2.21.
4. Exercise 2.26.
5. Exercise 2.27.
6. Exercise 2.28.
7. Exercise 2.29.

10 Background on rearranging terms of a series

`Tuesday 16, Wednesday 17 & Monday 21 February 2016`

10.1 Rearranging terms of an absolutely convergent series

1. Knowing how to add two numbers tells us how to add finitely many numbers, but not how to add infinitely many numbers. To begin making sense of infinite sums, we first calculate finite sums (which we know how to do): let $s_n = a_1 + \ldots + a_n$ be the n^{th} partial sum. Next, consider the sequence (s_n). Each term of this sequence makes sense since it is a just a sum of finitely many numbers.

   Armed with this sequence, we can define infinite sums. If the sequence (s_n) converges to a limit, say l, define the infinite sum $\sum_1^\infty a_i$ to be the number l. Thus, an infinite sum is the limit of finite sums whenever the limit exists.

2. Recall that a sum of finitely many finite terms is always finite (follows from induction). Also, finite sums remain the same even if the terms are rearranged. So, $1+2-3 = 1-3+2 = 2+1-3 = -3+2+1$. Remember that this works only for a finite sum; for an infinite sum if infinitely many terms are rearranged the sum can change (as happens with conditional convergence).

   Now, $\sum_1^\infty a_i = \lim_{n\to\infty} \left(\sum_1^n a_i\right)$ if the limit exists. So, an infinite sum is calculated by first calculating finite sums and then taking the limit. Before we take the limit we still have only a finite sum, and so we can rearrange the (finitely many) terms. This shows

$$
\begin{aligned}
\sum_1^\infty (a_n + \alpha b_n) &= \lim_{N\to\infty} \Big((a_1 + \alpha b_1) + \cdots + (a_N + \alpha b_N)\Big) \\
&= \lim_{N\to\infty} \Big((a_1 + \ldots + a_N) + \alpha(c_1 + \ldots + c_N)\Big) \\
&= \lim_{N\to\infty} \Big(\sum_1^N a_n + \alpha \sum_1^N b_n\Big) \\
&= \sum_1^\infty a_n + \alpha \sum_1^\infty b_n
\end{aligned}
$$

   provided the two sums, $\sum_1^\infty a_n$ and $\sum_1^\infty b_n$, both exist.

3. If there is a $1:1$, onto map $\sigma(\cdot) : \mathbb{N} \to \mathbb{N}$ such that $a_i = b_{\sigma(i)}$ for all $i \in \mathbb{N}$ then the sequence (b_n) is said to be a rearrangement of the sequence (a_n). Let $\sum_1^\infty a_n$ be convergent with all $a_n \geq 0$. Let (b_n) be a rearrangement of (a_n). Then $\sum_1^\infty b_n = \sum_1^\infty a_n$.

   Define $s_n = a_1 + \ldots + a_n$. Since $a_n \geq 0$ for all n, it follows that $s_n \leq s_{n+1}$ for all n. It is a property of real numbers that monotonic sequences have limits. It follows that $\lim s_n = s$ exists in $\mathbb{R} \cup \{\infty\}$. By definition of infinite sum, $\sum a_i = s$.

   Define $t_k = \sum_{i=1}^k b_i$. Since (t_k) too is a monotonic increasing sequence, we must have $t_n \uparrow t$ for some $t \in \mathbb{R} \cup \{\infty\}$. Let

$$m = \max\{\sigma(1), \ldots, \sigma(n)\}.$$

   It follows that $s_n \leq t_m$ and therefore $s_n \leq t$ for all n. Since $s_n \uparrow s$, it follows that $s \leq t$. A similar argument proves that $t \leq s$. In other words, $s = t$ or $\sum a_i = \sum b_i$.

4. The above result assumes that all terms a_n are nonnegative. What happens if both positive and negative terms are allowed? Answer: If a series is absolutely convergent then rearranging

the terms doesn't change the sum. In other words, if $\sum_1^\infty a_n$ is absolutely convergent, and (b_n) is a rearrangement of (a_n), then $\sum_1^\infty b_n = \sum_1^\infty a_n$.

By assumption $\sum_1^\infty |a_n|$ converges. Define $A_n = |a_n| + a_n$ and $B_n = |b_n| + b_n$. Then $A_n \geq 0$ for all n, and (B_n) is a rearrangement of (A_n). Let us look at the partial sums of A_n. The partial sums of A_n are monotonically increasing. They are also bounded by a finite constant,

$$\sum_1^N A_n = \sum_1^N (|a_n| + a_n) \leq \sum_1^N (|a_n| + |a_n|) = 2\sum_1^N |a_n| \leq 2\sum_1^\infty |a_n|.$$

Being bounded and monotone increasing the partial sums must converge, and so $\sum_1^\infty A_n$ converges. The previous result tells us that

$$\sum_1^\infty A_n = \sum_1^\infty B_n, \text{ and } \sum_1^\infty |a_n| = \sum_1^\infty |b_n|.$$

Subtracting gives us the desired result.

10.2 Three facts

The series $\sum_1^\infty x_i$ is said to be absolutely convergent if $\sum_1^\infty |x_i| < \infty$. It is said to be conditionally convergent if $\sum_1^\infty x_i$ converges, but $\sum_1^\infty |x_i| = \infty$. Define $x_i^+ = \max\{x_i, 0\}$ and $x_i^- = \max\{-x_i, 0\}$. Thus $x_i = x_i^+ - x_i^-$ with x_i^+ and x_i^- both nonnegative.

1. Conditionally convergent implies $\sum |x_i|$ diverges.
2. Conditionally convergent implies $\sum |x_i^+|$ and $\sum |x_i^-|$ diverge.
3. Conditional convergence implies $|x_i| \to 0$.

10.3 Riemann's rearrangement theorem

Let $\sum_1^\infty x_i$ be conditionally convergent. For $s \in \mathbb{R}$, there exists a rearrangement (y_n) of (x_n) such that $\sum_1^n y_i \to s$ as $n \to \infty$.

We construct two sequence (p_j) and (m_j) using the terms in (x_n) as follows. With $p_1, \ldots, p_j$ defined, define $p_{j+1} = x_n$ where n is the smallest index such that $x_n \geq 0$ and x_n has not yet been assigned to $p_1, \ldots, p_j$. Similarly, with $m_1, \ldots, m_j$, define $m_{j+1} = x_n$ where n is the smallest index such that $x_n < 0$ and x_n has not yet been assigned to $m_1, \ldots, m_j$. The following facts are easily verified.

1. Every nonnegative term x_n is included exactly once in the sequence (p_j) and never in the sequence (m_j).
2. Every negative term x_n is included exactly once in the sequence (m_j) and never in the sequence (p_j).
3. All terms used in (p_j) or (m_j) come from the terms in (x_n).
4. $\sum_1^\infty p_j = \infty$ and $\sum_1^\infty m_j = -\infty$.

5. $p_j \to 0$ and $m_j \to 0$ (since $|x_i| \to 0$ due to conditional convergence).

Define $s_0 = 0$. Once $y_1, \ldots, y_n$ have been defined, let $s_n = \sum_1^n y_n$ and define y_{n+1} as follows. If $s_n < s$, let $y_{n+1} = p_j$ where j is the smallest index such that p_j has not yet been assigned to $y_1, \ldots, y_n$. On the other hand, if $s_n \geq s$, let $y_{n+1} = m_j$ where j is the smallest index such that m_j has not yet been assigned to $y_1, \ldots, y_n$. It is easy to verify that (y_n) is a rearrangement of (x_n) such that $y_n \to s$.

10.4 Rearranging to obtain divergence

We need a rearrangement (y_n) such that $\sum_1^\infty y_n$ is divergent. Let (p_j) and (m_j) be as above. We illustrate divergence with two examples.

`Bounded oscillation.` We construct a rearrangement that oscillates between 1 and -1. We'll use an auxiliary sequence (u_n) of 0s and 1s with $u_n = 1$ indicating that y_{n+1} should be chosen nonnegative. Define $s_0 = 0$ and $u_0 = 1$. Once $u_1, \ldots, u_n$ and $y_1, \ldots, y_n$ have been defined, let $s_n = \sum_1^n y_n$ and define y_{n+1} as follows.

1. If $u_n = 1$, set $y_{n+1} = p_j$ where j is the smallest index such that p_j has not yet been assigned to $y_1, \ldots, y_n$. Set
$$u_{n+1} = \begin{cases} 1 & \text{if } s_{n+1} < 1, \\ 0 & \text{otherwise.} \end{cases}$$

2. If $u_n = 0$, set $y_{n+1} = m_j$ where j is the smallest index such that m_j has not yet been assigned to $y_1, \ldots, y_n$. Set
$$u_{n+1} = \begin{cases} 1 & \text{if } s_{n+1} < -1, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify that (y_n) is a rearrangement of (x_n) such that s_n has two limit points, -1 and 1.

`Unbounded oscillation.` Once $y_1, \ldots, y_n$ have been defined, let $s_n = \sum_1^n y_n$ and define y_{n+1} as follows. If $\max\{\lfloor |s_1| \rfloor, \lfloor |s_2| \rfloor, \ldots, \lfloor |s_n| \rfloor\}$ is even, let $y_{n+1} = p_j$ where j is the smallest index such that p_j has not yet been assigned to $y_1, \ldots, y_n$. While if $\max\{\lfloor |s_1| \rfloor, \lfloor |s_2| \rfloor, \ldots, \lfloor |s_n| \rfloor\}$ is odd, let $y_{n+1} = m_j$ where j is the smallest index such that m_j has not yet been assigned to $y_1, \ldots, y_n$. It is easy to verify that (y_n) is a rearrangement of (x_n) such that s_n will keep oscillating and the oscillations are unbounded.

11 Expected value, variance, $N(0,1)$ RV, jointly distributed RVs, independence

`Tuesday 23 & Wednesday 24 February 2016`

1.

2.

12 Exercises

Friday 26 February 2016

1. Exercise 2.33. A probability density function is given by

$$f(x) = \begin{cases} c(1-x^2) & \text{for } x \in (-1,1), \\ 0 & \text{for } x \notin (-1,1). \end{cases}$$

To calculate c, solve the equation $\int_{-1}^{1} c(1-x^2)\mathrm{d}x = 1$ to get $c = 3/4$. The cumulative distribution function will be

$$F(p) = \begin{cases} 0 & \text{for } p < -1 \\ \frac{1}{2} + \frac{3p}{4} - \frac{p^3}{4} & \text{for } p \in [-1,1] \\ 1 & \text{for } p > 1 \end{cases}$$

2. Exercise 2.42. Coupons are drawn independently, each coupon equally likely to be one of m types. Let X be the number of coupons drawn till one of each type is obtained. What is $\mathbb{E}[X]$?

Let $X_1 + \ldots + X_i$ denote the number of coupons drawn till the number of different types obtained first becomes i. We want $\mathbb{E}[X_1 + \ldots + X_m]$. Now, $\mathbb{E}[X_i] = p(1 + 2q + 3q^2 + \ldots) = \frac{1}{p}$ where $p = \frac{m-(i-1)}{m}$. Therefore,

$$\mathbb{E}[X_1 + \ldots + X_m] = m\Big(\frac{1}{m} + \frac{1}{m-1} + \frac{1}{m-2} + \ldots + 1\Big).$$

3. Exercise 2.43. An urn has r red balls and b black balls, to be drawn out one at a time, without replacement. What is the expected number of red balls pulled out before the first black ball is pulled out?

On each red ball, put a number. This should not change the answer. Let $R = R_1 + \ldots + R_r$ where R_i is an indicator variable that is 1 if the i^{th} red ball is pulled out before the first black ball. We want to calculate $\mathbb{E}[R]$. By symmetry and linearity of expectation, $\mathbb{E}[R] = r\mathbb{E}[R_1]$. Now,

$$\mathbb{E}[R_1] = \mathbb{P}\{R_1 = 1\} = \frac{1}{b+1}.$$

Therefore $\mathbb{E}[R] = \frac{r}{b+1}$.

4. Exercise 2.44. What is the expected number of red balls pulled out after the first black ball is pulled out, but before the second black one is pulled out?

Write again $R = R_1 + \ldots + R_r$ where R_i is an indicator variable that is 1 if the i^{th} red ball is pulled out after the first black ball is pulled out, but before the second black one is pulled out. We want to calculate $\mathbb{E}[R]$. By symmetry and linearity of expectation, $\mathbb{E}[R] = r\mathbb{E}[R_1]$. Now,

$$\mathbb{E}[R_1] = \mathbb{P}\{R_1 = 1\} = \frac{b}{b+1} \times \frac{1}{b} = \frac{1}{b+1}.$$

Therefore $\mathbb{E}[R] = \frac{r}{b+1}$.

5. Exercise 2.45. Let $p_1 + \dots p_k = 1$. A total of k keys go into k boxes, each key independently going to box i with probability p_i. What is the expected number of collisions?

   For $i < j$ let C_{ij} be an indicator variable that is 1 if keys i and j collide. We are interested in $\mathbb{E}[\sum_{i<j} C_{ij}]$. By linearity this is

   $$\sum_{i<j} \mathbb{E}[C_{ij}] = \frac{k(k-1)}{2}\mathbb{E}[C_{12}] = \frac{k(k-1)}{2}\mathbb{P}\{\text{keys 1 and 2 collide}\}.$$

   Now, $\mathbb{P}\{\text{keys 1 and 2 collide}\} = p_1^2 + \dots + p_k^2$, giving

   $$\mathbb{E}\Big[\sum_{i<j} C_{ij}\Big] = \frac{k(k-1)}{2} \times (p_1^2 + \dots + p_k^2).$$

6. Exercise 2.47. Three trials are conducted each of which can be a success or a failure. $\mathbb{E}[X] = 1.8$ where X is the number of successes. What are the largest and smallest values possible for $\mathbb{P}\{X = 3\}$?

   Think of this as an experiment which produces one of four outcomes, $\{X = 0\}, \{X = 1\}, \{X = 2\}$ and $\{X = 3\}$ with respective probabilities p_0, p_1, p_2 and p_3. Then

   $$\mathbb{E}[X] = 1.8 = p_1 + 2p_2 + 3p_3.$$

   Clearly $p_3 \leq 0.6$. Also, $p_0 = 0.4$ and $p_3 = 0.6$ shows that p_3 can actually attain the value 0.6. (Here, the three trials are clearly not independent: if one fails, all fail; if one succeeds, all succeed.)

   Next, $p_1 = 0.2$ and $p_2 = 0.8$ shows that one can have $p_3 = 0$.

   How to obtain such a joint distribution? Use a single uniform $(0, 1)$ random variable, and based on its value give suitable values to the outcome of each of the three trials. For example, let $X = X_1 + X_2 + X_3$ with $X_1 = X_2 = X_3 = 0$ if $U < 0.4$ and $X_1 = X_2 = X_3 = 1$ if $U \geq 0.4$.

7. Exercise 2.48. Let X be a nonnegative continuous random variable, and $g(\cdot)$ a differentiable function with $g(0) = 0$. Show under suitable assumptions that

   $$\mathbb{E}[g(X)] = \int_0^\infty \mathbb{P}\{X > t\}g'(t)\mathrm{d}t.$$

   Let $f(\cdot)$ be the probability density function for X. Under suitable assumptions (existence of integrals, $g(t)(1 - F(t)) \to 0$ as $t \to \infty$)

   $$\begin{aligned}
   \mathbb{E}[g(X)] &= \int_0^\infty g(t)f(t)\mathrm{d}t \\
   &= \int_0^\infty g(t)\mathrm{d}(F(t) - 1) \\
   &= g(t)(1 - F(t))|_0^\infty + \int_0^\infty (1 - F(t))\mathrm{d}g(t) \\
   &= \int_0^\infty \mathbb{P}\{X > t\}g'(t)\mathrm{d}t.
   \end{aligned}$$

13 Exercises

Monday 21 March 2016

1. Exercise 2.56. Coupons are of n types. When collected, with probability p_i a coupon is of type i. Collect k coupons with N the (random) number of different types of coupons obtained. What is $\mathbb{E}[N]$ and $Var(N)$?

   We work with indicator random variables.

   $$I_i = \begin{cases} 1 & \text{if type } i \text{ is obtained} \\ 0 & \text{otherwise} \end{cases}$$

   Write $N = I_1 + \ldots + I_n$. Since $\mathbb{P}\{I_i = 1\} = 1 - \mathbb{P}\{I_i = 0\}$, we get

   $$\mathbb{E}[N] = \sum_{i=1}^{n} \mathbb{E}[I_i] = \sum_{i=1}^{n} \Big(1 - \mathbb{P}\{I_i = 0\}\Big) = \sum_{i=1}^{n} \Big(1 - (1 - p_i)^k\Big).$$

   Next

   $$\sigma^2(N) = \sum_{i,j} \Big(\mathbb{E}[I_i I_j] - \mathbb{E}[I_i]\mathbb{E}[I_j]\Big).$$

   For $i = j$ we have $I_i I_j = I_i$, giving a contribution of

   $$\mathbb{E}[I_i](1 - \mathbb{E}[I_i]) = \Big(1 - (1 - p_i)^k\Big)(1 - p_i)^k.$$

   For $i \neq j$, we have

   $$\mathbb{E}[I_i I_j] = \mathbb{P}\{I_i = 1 = I_j\} = 1 - \mathbb{P}\{I_i = 0\} - \mathbb{P}\{I_j = 0\} + \mathbb{P}\{I_i = 0 = I_j\},$$

   giving

   $$\mathbb{E}[I_i I_j] = 1 - (1 - p_i)^k - (1 - p_j)^k + (1 - p_i - p_j)^k.$$

   Since

   $$\mathbb{E}[I_i]\mathbb{E}[I_j] = \Big(1 - (1 - p_i)^k\Big)\Big(1 - (1 - p_j)^k\Big),$$

   we get

   $$\mathbb{E}[I_i I_j] - \mathbb{E}[I_i]\mathbb{E}[I_j] = (1 - p_i - p_j)^k - (1 - p_i)^k(1 - p_j)^k.$$

   Combining gives

   $$\sigma^2(N) = \sum_{i} (1 - p_i)^k + \sum_{i \neq j} (1 - p_i - p_j)^k - \sum_{i,j} (1 - p_i)^k (1 - p_j)^k.$$

2. Exercise 2.59. Consider i.i.d. random variables X_1, X_2, X_3 and X_4 with continuous distribution function $F(\cdot)$. Calculate $p = \mathbb{P}\{X_1 < X_2 > X_3 < X_4\}$.

   Let $f(\cdot)$ be the common probability density function.

   $$p = \int_{\mathbb{R}} f(b) \Big(\int_{-\infty}^{b} f(a) \mathrm{d}a \int_{-\infty}^{b} f(c) \Big(\int_{c}^{\infty} f(d) \mathrm{d}d \Big) \mathrm{d}c \Big) \mathrm{d}b.$$

Since $f(x)\mathrm{d}x = \mathrm{d}F(x)$ we get

$$p = \int_0^1 \Big(F(b) \int_0^{F(b)} (1 - F(c))\mathrm{d}F(c) \Big) \mathrm{d}F(b).$$

Or

$$p = \int_0^1 F(b)\big(F(b) - F(b)^2/2\big)\mathrm{d}F(b) = \frac{1}{3} - \frac{1}{8} = \frac{5}{24}.$$

Since all 4! possible orderings are equally likely, and only 5 of them satisfy the required conditions, we again get

$$p = \frac{5}{4!} = \frac{5}{24}.$$

3. Exercise 2.61. Computational. Should be easy.

4. Exercise 2.69. What is $\mathbb{P}\{2 < X < 3\}$ where $X = N(\mu, \sigma^2)$ with $\mu = 1$ and $\sigma^2 = 4$?

   Write $Y = \frac{X-1}{2}$. Then $Y = N(0,1)$ and

$$\begin{aligned} \mathbb{P}\{2 < X < 3\} &= \mathbb{P}\{0.5 < Y < 1\} \\ &= \mathbb{P}\{0.5 \le Y < 1\} \\ &= \mathbb{P}\{Y < 1\} - \mathbb{P}\{Y < 0.5\}. \end{aligned}$$

   The last two terms can be looked up in tables for $N(0,1)$.

5. Exercise 2.70. Let Y be a Poisson random variable with intensity 1. (So $\mathbb{E}[Y] = 1$ and $\sigma^2(Y) = 1$). Let $Y_1, Y_2, \dots$ be i.i.d. random variables, each distributed like Y. If X_n is Poisson with intensity n, then X_n has the same distribution as $Y_1 + Y_2 + \cdots + Y_n$. It follows that $\frac{X_n - n}{\sigma(X_n)}$ is distributed like $\frac{\sum_1^n Y_i - n\mathbb{E}[Y]}{\sqrt{n}\sigma(Y)}$ and so approaches $N(0,1)$ as $n \to \infty$. Therefore

$$\exp(-n) \sum_{k=0}^{n} \frac{n^k}{k!} = \mathbb{P}\{X_n \le n\} = \mathbb{P}\Big\{ \frac{(X_n - n)}{\sigma(X_n)} \le 0 \Big\} \to \frac{1}{2}.$$

6. Exercise 2.75. Pick a permutation of $\{1, 2, \dots, n\}$ uniformly from among the $n!$ possible permutations, and let N be the number of inversions. Write $N = N_1 + \ldots + N_n$ where N_i is the number of inversions of the element i. (If i appears before $i-2$, say, then that is an inversion, and we'll associate it to i.) The N_is are independent. Therefore $\mathbb{E}[N] = \sum_1^n \mathbb{E}[N_i]$ and $\sigma^2(N) = \sum_1^n \big(\mathbb{E}[N_i^2] - (\mathbb{E}[N_i])^2\big)$.

$$\mathbb{E}[N_i] = \frac{1}{i}\big(0 + 1 + \ldots (i-1)\big) = \frac{i-1}{2}.$$

$$\mathbb{E}[N_i^2] = \frac{1}{i}\big(0^2 + 1^2 + \ldots (i-1)^2\big) = \frac{(i-1)i(2i-1)}{6i} = \frac{(i-1)(2i-1)}{6}.$$

   We get $\mathbb{E}[N] = \frac{n(n-1)}{4}$ and $\sigma^2(N) = \sum_1^n \frac{i^2-1}{12}$.