MATH40005: Probability and Statistics

Almut E. D. Veraart
Room 551, Huxley Building
Department of Mathematics
Imperial College London
180 Queen's Gate, London, SW7 2AZ
E-Mail: a.veraart@imperial.ac.uk

Autumn 2019

(Last Update: Thursday 12th December, 2019)

Contents

Chapter 1

Additional examples/exercises discussed in lectures

These notes contain the additional examples, exercises and results which were discussed during the lectures in addition to our regular lecture notes.

1.1 Lecture 1

Recall the definition of a bijection:

Definition 1.1.1. *Consider a function $f : X \to Y$. f is injective, if for all $a, b \in X$, $f(a) = f(b) \Rightarrow a = b$. f is surjective, if for all $y \in Y$, there exists an $x \in X$ such that $f(x) = y$. f is bijective if it is injective and surjective.*

Definition 1.1.2. *Two sets have the same cardinality if there is a bijection between the two sets.*

Exercise 1.1.3. *Show that $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ and $\mathbb{N} \cup \{0\} = \{0, 1, \dots\}$ have the same cardinality.*

Proof. We use the convention that 0 is even. We define a function $f : \mathbb{N} \cup \{0\} \to \mathbb{Z}$ such that

$$f(x) = \begin{cases} \frac{x}{2}, & \text{if } x \text{ is even,} \\ -\frac{(x+1)}{2}, & \text{if } x \text{ is odd} \end{cases}$$

We need to show that this function is bijective.

First, we show that it is injective: For all $x, y \in \mathbb{N} \cup \{0\}$ with $f(x) = f(y)$ we have that $f(x)$ and $f(y)$ have the same sign. So either

$$f(x) = \frac{x}{2} = \frac{y}{2} = f(y) \Rightarrow x = y,$$

or

$$f(x) = -\frac{x+1}{2} = -\frac{y+1}{2} = f(y) \Rightarrow x = y.$$

Next, we show that f is surjective. For all $y \in \mathbb{Z}, y < 0$, choose $x = -2y-1$, then $f(x) = -\frac{-2y-1+1}{2} = y$. For all $y \in \mathbb{Z}, y \geq 0$, choose $x = 2y$, then $f(x) = \frac{2y}{2} = y$. □

1.2 Lecture 2

Example 1.2.1. *We flip a coin three times. We write "0" for heads and "1" for tails. Then the sample space is given by*

$$\Omega = \{(0,0,0),(0,0,1),\ldots,(1,1,0),(1,1,1)\},$$

which are all the ordered triplets of zeros and ones. By the multiplication principle we have that $\mathrm{card}(\Omega) = 2^3 = 8$*, hence* $\mathrm{P}_{\mathrm{Naive}}(\{\omega\}) = 1/8$ *for all* $\omega \in \Omega$*.*

Let us now consider the event $B :=$ *the first and the third flip are tails* $= \{(1,0,1),(1,1,1)\}$*. Then* $\mathrm{P}_{\mathrm{Naive}}(B) = 2/8 = 1/4$*.*

Remark 1.2.2. *If we deal with repetitions of experiments (coin toss, rolling a die), the corresponding sample spaces are given by* Cartesian product spaces*.*

For sets $A_1,\ldots,A_n$*, we define the Cartesian product as*

$$A_1 \times \cdots \times A_n = \{(x_1,\ldots,x_n) : x_i \in A_i \text{ for } i = 1,\ldots,n\}.$$

So in our example above, we can write $\Omega = \{0,1\} \times \{0,1\} \times \{0,1\} = \{0,1\}^3$*.*

1.3 Lecture 3

Exercise 1.3.1. *A college has 10 non-overlapping time slots for its courses and assigns courses to time slots randomly and independently. A student randomly chooses three of the courses to enroll in. What is the probability that there is a conflict in the student's schedule?*

Proof. Using the multiplication principle and the naive probability, we can compute the probability of no schedule conflict as $\frac{10\cdot 9\cdot 8}{10^3}(= 0.72)$. So the probability that there is at least one schedule conflict is given by $1 - \frac{10\cdot 9\cdot 8}{10^3}(= 0.28)$. □

Exercise 1.3.2. *A family has 6 children consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?*

Proof. Label the girls as 1, 2, 3, and the boys as 4, 5, 6. Then the birth order is a permutation of 1,2,3,4,5,6. So, 236514 means that child 2 was born first, then child 3 etc. The number of possible permutations is $6!$. For the three girls to be the eldest children, we need a permutation of 1,2,3, followed by a permutation of 4,5,6. Hence

$$\mathrm{P}(\text{the 3 girls are the 3 eldest children}) = \frac{3!3!}{6!} = \frac{1}{20} = 0.05.$$

Alternative solution: There are $\binom{6}{3}$ ways to chose where the three girls appear in the birth order (without taking ordering of the girls into account). Of these cases, there is only one where the three girls are the three eldest children. Hence

$$\mathrm{P}(\text{the 3 girls are the 3 eldest children}) = \frac{1}{\binom{6}{3}} = \frac{3!3!}{6!} = \frac{1}{20} = 0.05.$$

□

1.4 Lecture 4

Exercise 1.4.1. *Consider a group of four people.*

1. *How many ways are there to choose a two-person committee?*
2. *How many ways are there to break the people into two teams of two?*

Proof. 1. We could list all possibilities: Label the people as 1,2,3,4. Then the possibilities are $\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}$, i.e. we have 6 possibilities.

Alternatively, we can use the multiplication rule and account for overcounting: There are 4 possibilities for choosing the first person on the committee and 3 to choose the second person. Note, however, that this counts every possibility twice since picking 1 and 2 is the same as picking 2 and 1. We have overcounted by a factor of 2. So the number of possibilities is given by $4 \cdot 3/2 = 6 (= \binom{4}{2})$.

2. There are three possibilities:

Again, we could list all possibilities and count them: $\{1,2\},\{3,4\};\{1,3\},\{2,4\};\{1,4\},\{2,3\}$, which is rather tedious...

Alternatively, we could just specify the first person's teammate, then the other team is determined.

Or, we use 1) to deduce that there are 6 possibilities of choosing one team. Here we overcount by a factor of 2 since picking $\{1,2\}$ as a team is equivalent to picking $\{3,4\}$ as a team.

□

Exercise 1.4.2. *How many ways are there to permute the letters in the word STATISTICS? Note that there are 10 letter in total, "S" and "T" appear three times, "I" twice and "A" and "C" once.*

Proof. Approach 1: There are 10 positions in total, first we choose the 3 positions for the "S"s out of 10, then the three positions for the "T"s out of the remaining 7 positititions etc. Hence we get

$$\underbrace{\binom{10}{3}}_{"S"}\underbrace{\binom{7}{3}}_{"T"}\underbrace{\binom{4}{1}}_{"A"}\underbrace{\binom{3}{2}}_{"I"}\underbrace{\binom{1}{1}}_{"C"} = 50400$$

Approach 2: Start with $10!$ permutations of the the 10 letters and adjust for overcounting:

$$\frac{10!}{3!3!2!} = 50400.$$

□

Exercise 1.4.3. *Let's review the birth order example again: We have a family with 6 children, 3 girls, 3 boys. How many possibilities are there such that the 3 eldest children are the 3 girls?*

Proof. Approach 1:

$$\underbrace{\binom{6}{3}}_{girls}\underbrace{\binom{3}{3}}_{boys} = \binom{6}{3}$$

Approach 2: Start with $6!$ possible birth orders and adjust for overcounting:

$$\frac{6!}{3!3!}$$

□

1.5 Lecture 5

Exercise 1.5.1. *Let $\Omega = \mathbb{N} \cup \{0\}, \mathcal{F} = \mathcal{P}(\mathbb{N} \cup \{0\}), \mathrm{P}(A) = \sum_{x \in A} \frac{e^{-\lambda}\lambda^x}{x!}$ for $\lambda > 0$. Show that $(\Omega, \mathcal{F}, \mathrm{P})$ is a probability space.*

Proof. We know from lectures that the power set is a σ-algebra. So we only need to show that P is a probability measure. We note that $\mathrm{P} : \mathcal{F} \to \mathbb{R}$ and hence it remains to check the three axioms of the definition of a probability measure:

Axiom (ii) We use the fact that $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$, to deduce that

$$\mathrm{P}(\Omega) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{\lambda} e^{-\lambda} = 1.$$

Axiom (i) Since $\lambda > 0$, we have that for any $A \in \mathcal{F}$,

$$0 \leq \mathrm{P}(A) = \sum_{x \in A} \frac{\lambda^x}{x!} e^{-\lambda} \leq \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = 1.$$

Axiom (iii) Let $A_1, A_2, \ldots \in \mathcal{F}$ be disjoint, then

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{x \in \cup_{i=1}^{\infty} A_i} \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{i=1}^{\infty} \sum_{x \in A_i} \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{i=1}^{\infty} \mathrm{P}(A_i),$$

where we interchanged the order of summation (to be covered in Analysis!).

□

1.6 Lecture 6

Example 1.6.1. *We consider an example of an algebra, which is not a σ-algebra. Let $\Omega = \mathbb{R}$ and define $\mathcal{F}$ as the collection of finite disjoint unions of sets of the form $(a, b], (-\infty, a], (b, \infty)$ for all $a, b \in \mathbb{R}$.*
Then $\mathcal{F}$ is an algebra:

1. $\emptyset = (a, b] \in \mathcal{F}$ *for* $b < a$,
2. $\emptyset^c = \mathbb{R} = (-\infty, a] \cup (a, \infty) \in F$. *Also, for all $a < b$ we have*

$$(a, b]^c = (-\infty, a] \cup (b, \infty) \in \mathcal{F},$$
$$(-\infty, a]^c = (a, \infty) \in \mathcal{F},$$
$$(b, \infty)^c = (-\infty, b] \in \mathcal{F}.$$

The above results imply that for any $A \in \mathcal{F}$, we also have that $A^c \in \mathcal{F}$.

3. *The closedness under unions of pairs follows directly from the definition of $\mathcal{F}$.*

Hence $\mathcal{F}$ is an algebra. However, it is NOT a σ-algebra. To see this, note that $A_i = (0, 1 - \frac{1}{i}] \in \mathcal{F}$, but

$$\bigcup_{i=1}^{\infty} A_i = (0, 1) \notin \mathcal{F}.$$

Example 1.6.2. *Let $\Omega = \mathbb{R}$ and $\mathcal{F} = \{A \subset \Omega : A \text{ is finite or } A^c \text{ is finite}\}$. Then $\mathcal{F}$ is an algebra since:*

1. $\emptyset \in \mathcal{F}$ *since it is finite,*
2. *For any $A \in \mathcal{F}$, we have either*

$$A \text{ is finite} \Rightarrow (A^c)^c = A \text{ is finite}, \text{ hence } A^c \in \mathcal{F},$$

or

$$A^c \text{ is finite} \Rightarrow A^c \in \mathcal{F}.$$

3. *For any* $A_1, A_2 \in F$*, we have either*

$$A_1, A_2 \text{ are both finite} \Rightarrow A_1 \cup A_2 \text{ finite } \Rightarrow A_1 \cup A_2 \in \mathcal{F},$$

or at least one A_i^c *is finite for* $i = 1, 2$*. Without loss of generality assume that* A_2^c *is finite. Then*

$$(A_1 \cup A_2)^c = (A_1^c \cap A_2^c) \subseteq A_2^c \text{ is finite} \Rightarrow A_1 \cup A_2 \in \mathcal{F}.$$

Hence $\mathcal{F}$ *is an algebra. However,* $\mathcal{F}$ *is NOT a* σ*-algebra. To see this, note that* $A_i = \{i\} \in \mathcal{F}$ *since it is finite, but* $\cup_{i=1}^{\infty} A_i = \mathbb{N} \notin \mathcal{F}$ *since it is not finite and* $\mathbb{N}^c = \mathbb{R} \setminus \mathbb{N}$ *is not finite either!*

Note that if we work with $\Omega = \mathbb{N}$ *instead in the above example, we still get that* $\mathcal{F}$ *is an algebra, but not a* σ*-algebra. Here we could take* $A_i = \{2i\}$*, the* $\cup_{i=1}^{\infty} A_i$ *are the even natural numbers, which are infinite, and* $(\cup_{i=1}^{\infty} A_i)^c$ *are the odd natural numbers, which are infinite, too. So* $\cup_{i=1}^{\infty} A_i \notin \mathcal{F}$*.*

1.7 Lecture 7

Example 1.7.1. *We pick 2 cards at random from a well-shuffled 52-card deck. We consider the event* E*:="2nd card is red". What is* $\mathrm{P}(E)$*?*

We use the law of total probability to deduce that

$$\begin{aligned}\mathrm{P}(E) &= \mathrm{P}(E|\text{ 1st card red})\mathrm{P}(\text{1st card red}) + \mathrm{P}(E|\text{1st card black})\mathrm{P}(\text{1st card black}) \\ &= \frac{25}{51}\cdot\frac{1}{2} + \frac{26}{51}\cdot\frac{1}{2} = \frac{1}{2}.\end{aligned}$$

Example 1.7.2. *You have three bags that each contain 100 marbles.*

- *Bag 1 has 70 red and 30 green marbles.*
- *Bag 2 has 60 red and 40 green marbles.*
- *Bag 1 has 50 red and 50 green marbles.*

1. *If you choose one bag at random and then pick a marble at random from the chosen bag. What is the probability that the chosen marble is red?*

   We define the events B_i*:=bag* i *was chosen,* $\mathrm{P}(B_i) = \frac{1}{3}$*,* $i = 1, 2, 3$*.* $R :=$ *marble is red. Then*

$$\mathrm{P}(R|B_1) = \frac{7}{10}, \qquad \mathrm{P}(R|B_2) = \frac{6}{10}, \qquad \mathrm{P}(R|B_3) = \frac{5}{10},$$

   We use the law of total probability:

$$\mathrm{P}(R) = \mathrm{P}(R|B_1)\mathrm{P}(B_1) + \mathrm{P}(R|B_2)\mathrm{P}(B_2) + \mathrm{P}(R|B_3)\mathrm{P}(B_3) = \left(\frac{7}{10} + \frac{6}{10} + \frac{5}{10}\right)\frac{1}{3} = 0.6.$$

2. *Suppose that the chosen marble was red, what is the probability that bag 1 was chosen?*

   We use Bayes' rule:

$$\mathrm{P}(B_1|R) = \frac{\mathrm{P}(R|B_1)\mathrm{P}(B_1)}{\mathrm{P}(R)} = \frac{\frac{7}{10}\cdot\frac{1}{3}}{\frac{6}{10}} = \frac{7}{18} \approx 0.38 > \frac{1}{3}.$$

1.8 Lecture 8

Example 1.8.1. *You have a fair and an "unfair" coin. The unfair coin lands Heads with probability $\frac{3}{4}$. You pick one coin at random and toss it three times. It lands Heads three times. Given this information, what is the probability that you picked the fair coin?*

Let A:=the chosen coin lands Heads three times, F:= you picked the fair coin. We want to find $\mathrm{P}(F|A)$. We use the generalised Bayes rule:

$$\mathrm{P}(F|A) = \frac{\mathrm{P}(A|F)\mathrm{P}(F)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A|F)\mathrm{P}(F)}{\mathrm{P}(A|F)\mathrm{P}(F) + \mathrm{P}(A|F^c)\mathrm{P}(F^c)} = \frac{\left(\frac{1}{2}\right)^3 \cdot \frac{1}{2}}{\left(\frac{1}{2}\right)^3 \cdot \frac{1}{2} + \left(\frac{3}{4}\right)^3 \cdot \frac{1}{2}} \approx 0.23.$$

After having seen that it landed Heads three times, what is the probability that, if we toss the same coin a fourth time, it lands Heads a fourth time?

Let H := chosen coin lands Heads at 4th toss. We want to find $\mathrm{P}(H|A)$. Here we use the law of total probability with extra conditioning:

$$\mathrm{P}(H|A) = \mathrm{P}(H|A\cap F)\mathrm{P}(F|A) + \mathrm{P}(H|A\cap F^c)\mathrm{P}(F^c|A) \approx \frac{1}{2}\cdot 0.23 + \frac{3}{4}(1-0.23) = 0.69.$$

Note that H and A are conditional independent given F:

$$\mathrm{P}(H\cap A|F) = \frac{1}{2^4} = \frac{1}{2}\cdot\left(\frac{1}{2}\right)^3 = \mathrm{P}(H|F)\mathrm{P}(A|F).$$

Hence $\mathrm{P}(H|A\cap F) = \mathrm{P}(H|F) = \frac{1}{2}$. Similarly, $\mathrm{P}(H|A\cap F^c) = \mathrm{P}(H|F^c) = \frac{3}{4}$.

Example 1.8.2. *Example of an increasing sequence of sets: Let $\Omega = (0,\infty)$ and define $A_n := \left(0, 1-\frac{1}{n}\right]$. Then*

$$A_1 = (0,0] = \emptyset \subseteq A_2 = \left(0, \frac{1}{2}\right] \subseteq A_3 = \left(0, \frac{2}{3}\right] \subseteq \cdots$$

with

$$A_n \uparrow A := \bigcup_{n=1}^{\infty} A_n = (0,1).$$

We note that the sequence of complements is decreasing: $A_n^c \downarrow A^c$: We have $A_n^c = \left(1-\frac{1}{n}, \infty\right)$ and hence

$$A_1^c = (0,\infty) = \Omega \supseteq A_2^c = \left(\frac{1}{2}, \infty\right) \supseteq A_3^c = \left(\frac{2}{3}, \infty\right) \supseteq \cdots$$

with

$$A_n^c \downarrow A^c = \bigcap_{n=1}^{\infty} A_n^c = [1,\infty).$$

Example 1.8.3. *Example of an increasing sequence of sets: Let $\Omega = \mathbb{R}$ and define $A_n := (-\infty, n]$. Then*

$$A_1 = (-\infty,1] \subseteq A_2 = (-\infty,2] \subseteq A_3 = (-\infty,3] \subseteq \cdots$$

with

$$A_n \uparrow A := \bigcup_{n=1}^{\infty} A_n = \Omega = \mathbb{R}.$$

We note that the sequence of complements is decreasing: $A_n^c \downarrow A^c$: We have $A_n^c = (n,\infty)$ and hence

$$A_1^c = (1,\infty) = \Omega \supseteq A_2^c = (2,\infty) \supseteq A_3^c = (3,\infty) \supseteq \cdots$$

with

$$A_n^c \downarrow A^c = \bigcap_{n=1}^{\infty} A_n^c = \emptyset$$

1.9 Lecture 9

Example 1.9.1. *Consider the experiment where we toss a fair coin twoce. Write H for Heads and T for Tails. Then $\Omega = \{HH, HT, TH, TT\}$. Define random variables on Ω!*

- $X =$ *number of Heads:*

$$X(HH) = 2, \quad X(HT) = H(TH) = 1, \quad X(TT) = 0.$$

- $Y =$ *number of Tails:* $Y = 2 - X$.

- $I = 1$ *if first toss lands Heads and* 0 *otherwise.*

$$I(HH) = I(HT) = 1, \quad I(TH) = I(TT) = 0.$$

  This is a so-called indicator random variable *indicating whether or not the first toss lands Heads (1="yes", 0="no").*

We can write 1 for H and 0 for T, then $\Omega = \{(1,1), (1,0), (0,1), (0,0)\}$. I.e. $\omega = (\omega_1, \omega_2) \in \Omega$ if $\omega_i \in \{0,1\}, i = 1, 2$. Then we can express the three random variables defined above as follows:

$$\begin{aligned} X(\omega_1, \omega_2) &= \omega_1 + \omega_2, \\ Y(\omega_1, \omega_2)) &= 2 - \omega_1 - \omega_2, \\ I(\omega_1, \omega_2) &= \omega_1. \end{aligned}$$

Note that we can define the event $A := \{(1,1), (1,0)\}$, then $I = \mathbb{I}_A$. Also, note that

$$\mathrm{P}(\mathbb{I}_A = 1) = \mathrm{P}(\{\omega \in \Omega : \mathbb{I}_A(\omega) = 1\}) = \mathrm{P}(\{\omega \in \Omega : \omega \in A\}) = \mathrm{P}(A).$$

1.10 Lecture 10

Example 1.10.1. *Show Vandermonde's identity: For $k, n, m \in \mathbb{N} \cup \{0\}, k \leq n + m$, we have*

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

Story proof/Combinatorial proof: *Consider selecting a committee of k people from a group of people consisting of m men and n women. The left hand side describes the number of possibilities of selecting k from $m+n$ people (without replacement, order irrelevant). On the right hand side we consider all possible combinations when we choose j men out of m men, then we need to choose $k - j$ women out of n women to obtain a committee of k people. We then need to sum of all possible values of j which gives us the right hand side.*

Algebraic proof: *Using the binomial theorem, we get*

$$(1 + x)^{m+n} = \sum_{k=0}^{m+n} \binom{m+n}{k} x^k,$$

and also

$$(1 + x)^{m+n} = (1 + x)^m (1 + x)^n = \sum_{i=0}^{m} \binom{m}{i} x^i \sum_{j=0}^{n} \binom{n}{j} x^j = \sum_{i=0}^{m} \sum_{j=0}^{n} \binom{m}{i} \binom{n}{j} x^{i+j}.$$

Next, we change the summation indices: We have that $0 \leq i \leq m, 0 \leq j \leq n$. We set $k = i + j$, then $0 \leq k \leq m + n$. Hence

$$(1 + x)^{m+n} = \sum_{i=0}^{m} \sum_{j=0}^{n} \binom{m}{i} \binom{n}{j} x^{i+j} = \sum_{k=0}^{m+n} \sum_{i=0}^{k} \binom{m}{i} \binom{n}{k-i} x^k$$

So, overall, we found that

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

We note that two polynomials are identical if they have the same degree and the corresponding coefficients are identical, which implies that for all $0 \leq k \leq n+m$*, we have*

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

We note that the Vandermonde's identity is useful in showing that the probability mass function of the hypergeometric distribution is indeed a valid probability mass function.

Example 1.10.2. *We study the generalisation of the binomial coefficient: For* $\alpha \in \mathbb{C}, k \in \mathbb{N}$*, we define*

$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}.$$

The generalised binomial formula is then given by

$$(1+x)^\alpha = \sum_{k=0} \binom{\alpha}{k} x^k, \text{ for } |x| < 1.$$

We can then derive the following identity for $x \in \mathbb{N} \cup \{0\}, r \in \mathbb{N}$*:*

$$\binom{x+r-1}{r-1} = (-1)^x \binom{-r}{x}.$$

To see this, note that

$$\binom{x+r-1}{r-1} = \frac{(x+r-1)!}{x!(r-1)!} = \frac{(x+r-1)(x+r-2)\cdots r}{x!}$$
$$= (-1)^x \frac{(-r)(-r-1)\cdots(-r-x+1)}{x!} = (-1)^x \binom{-r}{x}.$$

This result (together with the generalised Binomial formula stated above) can be used to show that if $X \sim NBin(r,p)$ *for* $p \in (0,1)$*, then*

$$\sum_{x=0}^{\infty} \mathrm{P}(X=x) = \sum_{x=0}^{\infty} \binom{x+r-1}{r-1} p^r (1-p)^x = p^r \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} (1-p)^x$$
$$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (p-1)^x = p^r(1+(p-1))^{-r} = p^r p^{-r} = 1,$$

where the generalised Binomial theorem was applicable since $|1-p| = (1-p) < 1$.

1.11 Lecture 11

Remark 1.11.1. *A cumulative distribution function (c.d.f) of a random variable* X *say is right continuous, but not in general left continuous. To see the latter, consider a point* $x \in \mathbb{R}$ *and an arbitrary sequence* $(x_n)_{n\in\mathbb{N}}$ *approaching* x *from the left, i.e.* $x_n \leq x$ *and* $\lim_{n\to\infty} x_n = x$*. We can then define* $\delta_n = x - x_n$ *for* $n \in \mathbb{N}$*. Clearly,* $\delta_n \geq 0, \lim_{n\to\infty} \delta_n = 0$*. Then*

$$(-\infty, x_n] = (-\infty, x-\delta_n] \uparrow (-\infty, x).$$

Hence,

$$\{\omega \in \Omega : X(\omega) \in (-\infty, x_n]\} \uparrow \{\omega \in \Omega : X(\omega) \in (-\infty, x)\}.$$

Hence, by the continuity of the probability measure

$$\begin{aligned}
\lim_{n\to\infty} F_X(x_n) &= \lim_{n\to\infty} \mathrm{P}(\{\omega \in \Omega : X(\omega) \le x_n\}) \\
&= \lim_{n\to\infty} \mathrm{P}(\{\omega \in \Omega : X(\omega) \in (-\infty, x_n]\}) \\
&= \mathrm{P}(\lim_{n\to\infty} \{\omega \in \Omega : X(\omega) \in (-\infty, x_n]\}) \\
&= \mathrm{P}(\lim_{n\to\infty} \{\omega \in \Omega : X(\omega) \in (-\infty, x - \delta_n]\}) \\
&= \mathrm{P}(\{\omega \in \Omega : X(\omega) \in (-\infty, x)\}) \\
&= \mathrm{P}(\{\omega \in \Omega : X(\omega) < x\}) \\
&= \mathrm{P}(\{\omega \in \Omega : X(\omega) \le x\}) - \mathrm{P}(\{\omega \in \Omega : X(\omega) = x\}) \\
&\le \mathrm{P}(\{\omega \in \Omega : X(\omega) \le x\} = F_X(x).
\end{aligned}$$

So, we observe that F_X is left-continuous in x iff $\mathrm{P}(\{\omega \in \Omega : X(\omega) = x\}) = 0$.

1.12 Lecture 12

Example 1.12.1. *Let us consider an example of a random variable which is neither discrete nor continuous: We flip an unfair coin infinitely many times and assume that we obtain Heads with probability $p \in (0,1)$. We denote the outcomes by $X_1, X_2, \ldots$ with $X_i = 0$ if we obtain Tails in the ith flip and $X_i = 1$ if we obtain Heads in the ith flip. Define a random number $Y = 0.X_1X_2X_3\ldots$ (in base 2). I.e. $Y = X_1 \cdot \frac{1}{2} + X_2 \left(\frac{1}{2}\right)^2 + X_3 \cdot \left(\frac{1}{2}\right)^3 + \cdots$. Then*

- *X_1 determines whether Y is in the first half $[0, \frac{1}{2})$ (if $X_1 = 0$) or in the second half $[\frac{1}{2}, 1]$ (if $X_1 = 1$).*
- *X_2 determines whether Y is in the first half (if $X_2 = 0$) (either in $[0, \frac{1}{4})$ or in $[\frac{1}{2}, \frac{3}{4})$) or in the second half (if $X_2 = 1$) (either in $[\frac{1}{4}, \frac{1}{2})$ or in $[\frac{3}{4}, 1]$) of the previous half.*
- *etc.*

We note that for any $y = 0.x_1x_2x_3\ldots$ in base 2 representation, we have

$$\mathrm{P}(Y = y) = \mathrm{P}(X_1 = x_1)\mathrm{P}(X_2 = x_2)\cdots = 0$$

since each term in the product is either equal to p or $1-p$ which are both smaller than 1, so their infinite product will converge to 0. Hence Y cannot be a discrete random variable.

One (not we!) can show that for $p \neq 0.5$, Y does not have a density, whereas if $p = 0.5$, then Y is uniformly distributed on $[0,1]$ and hence has a density.

1.13 Lecture 13

Example 1.13.1 (Not examinable, but good to know :))**.** *Show that the standard normal density integrates to 1. We want to show that*

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1,$$

which is equivalent to showing that

$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}.$$

In fact, we will square both sides and show that

$$\left(\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx\right)^2 = 2\pi.$$

We write

$$\begin{aligned} A := \left(\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx\right)^2 &= \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2+y^2)\right) dxdy \end{aligned}$$

Now we consider the (invertible) transformation to polar coordinates:

$$x = r\cos(\theta), \qquad y = r\sin(\theta),$$

for $r > 0$ and $\theta \in [0, 2\pi]$. Then, we compute the Jacobian $J(r,\theta)$ of the transformation as follows:

$$J(r,\theta) = \left|\det\left(\frac{\partial(x,y)}{\partial(r,\theta)}\right)\right| = \left|\det\begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix}\right| = r(\cos^2(\theta) + \sin^2(\theta)) = r.$$

Then,

$$\begin{aligned} A &= \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{1}{2}(r^2\cos^2(\theta) + r^2\sin^2(\theta))\right) J(r,\theta) dr d\theta \\ &= \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{1}{2}r^2\right) r dr d\theta. \end{aligned}$$

You can now do another variable transformation (or integrate directly): We set $u = R = r^2/2$, the $du = rdr$ and

$$A = \int_0^{2\pi} \left(\int_0^{\infty} e^{-u} du\right) d\theta = \int_0^{2\pi} \left(-e^{-u}\big|_{u=0}^{\infty}\right) d\theta = \int_0^{2\pi} 1 d\theta = 2\pi.$$

Hence, we have that $\sqrt{A} = \sqrt{2\pi}$.

1.13.1 Pre-images and their properties

Consider a function with domain $\mathcal{X}$ and co-domain $\mathcal{Y}$, i.e. $f : \mathcal{X} \to \mathcal{Y}$. For any subset $B \subseteq \mathcal{Y}$, we define the *pre-image of* B as

$$f^{-1}(B) = \{x \in \mathcal{X} : f(x) \in B\}.$$

Please not that the pre-image should not be confused with the inverse function (despite the fact that we are using the same notation). The pre-image is well-defined for any function, whereas the inverse function obviously only exists when the function f is invertible.

The definition of the pre-image implies that $x \in f^{-1}(B) \Leftrightarrow f(x) \in B$.

Note that in the case when B is a singleton, i.e. $B = \{b\}$ for an element $b \in \mathcal{Y}$, then we often simplify the notation to $f^{-1}(\{b\}) = f^{-1}(b)$.

Lemma 1.13.2. *For any collection of subsets $B_i \subseteq \mathcal{Y}$, $i \in \mathcal{I}$ where $\mathcal{I}$ denotes an (arbitrary) index set, we have that*

$$f^{-1}\left(\bigcup_{i\in\mathcal{I}} B_i\right) = \bigcup_{i\in\mathcal{I}} f^{-1}(B_i).$$

Proof. We have that

$$\begin{aligned} x \in f^{-1}\left(\bigcup_{i\in\mathcal{I}} B_i\right) &\Leftrightarrow f(x) \in \bigcup_{i\in\mathcal{I}} B_i \\ &\Leftrightarrow \exists\, i \in \mathcal{I} \text{ such that } f(x) \in B_i \\ &\Leftrightarrow \exists\, i \in \mathcal{I} \text{ such that } x \in f^{-1}(B_i) \\ &\Leftrightarrow x \in \bigcup_{i\in\mathcal{I}} f^{-1}(B_i). \end{aligned}$$

□

Remark 1.13.3. *We used the above result to justify in Section 9.1 of the lecture notes that, given X is a discrete random variable, then $g(X)$ is a random variable for $g : \mathbb{R} \to \mathbb{R}$. To this end, note that for all $y \in \mathbb{R}$, we have*

$$\begin{aligned} Y^{-1}(\{y\}) = \{\omega \in \Omega : Y(\omega) = y\} &= \{\omega \in \Omega : g(X(\omega)) = y\} \\ &= \{\omega \in \Omega : X(\omega) \in \{x \in \mathrm{Im}X : g(x) = y\}\} \\ &= \{\omega \in \Omega : X(\omega) \in \bigcup_{x\in\mathrm{Im}X:g(x)=y} \{x\}\} \\ &= X^{-1}\left(\bigcup_{x\in\mathrm{Im}X:g(x)=y} \{x\}\right) \\ &\overset{Lemma 1.13.2}{=} \bigcup_{x\in\mathrm{Im}X:g(x)=y} X^{-1}(\{x\}) \\ &= \bigcup_{x\in\mathrm{Im}X:g(x)=y} \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}, \end{aligned}$$

since each event $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$ and, by the definition of a σ-algebra, a countable union of elements of $\mathcal{F}$ is in $\mathcal{F}$, too. Since X is discrete, we indeed have that $\{x \in \mathrm{Im}X : g(x) = y\} \subseteq \mathrm{Im}X$ is (at most) countably infinite.

Example 1.13.4. *Example of a probability space and one function which is a random variable and one function which is not a random variable on that probability space:*

Consider the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ and define the event space $\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$. Define the probability measure P to be the naive probability measure, i.e. $\mathrm{P}(A) = \mathrm{card}(A)/\mathrm{card}(\Omega)$ for $A \in \mathcal{F}$.

(i) Define $X : \Omega \to \mathbb{R}$ such that $X(\omega) = 1$ if ω is even and $X(\omega) = -1$ if ω is odd. Then $\mathrm{Im}X = \{-1, 1\}$ is finite and $X^{-1}(\{-1\}) = \{1, 3, 5\} \in \mathcal{F}$ and $X^{-1}(\{1\}) = \{2, 4, 6\} \in \mathcal{F}$. For all $x \notin \mathrm{Im}X$ we have that $X^{-1}(\{x\}) = \emptyset \in \mathcal{F}$. Hence X is a discrete random variable.

(ii) Define $X : \Omega \to \mathbb{R}$ such that $X(\omega) = \omega$. Then e.g. $X^{-1}(\{1\}) = \{1\} \notin \mathcal{F}$, hence X is not a random variable with respect to the given sigma-algebra $\mathcal{F}$. (It would be one if we had chosen $\mathcal{F}$ to be the power sigma-algebra of Ω!)

1.14 Lecture 14

Example 1.14.1. *Consider a random variable $X : \Omega \to \mathbb{R}$ with c.d.f. F_X.*

1. *Find the c.d.f. of $Y = \max\{X, 3\}$: Let $y \in \mathbb{R}$, then*

$$\begin{aligned} F_Y(y) = \mathrm{P}(\max\{X, 3\} \le y) = \mathrm{P}(X \le y, 3 \le y) &= \mathrm{P}(X \le y)\mathbb{I}_{[3,\infty)}(y) \\ &= \begin{cases} F_X(y), & \text{for } y \ge 3, \\ 0, & \text{for } y < 3. \end{cases} \end{aligned}$$

2. *Find the c.d.f. of $Y = |X|$. Let $y \in \mathbb{R}$, then*

$$F_Y(y) = \mathrm{P}(|X| \le y) = \begin{cases} 0, & \text{for } y < 0, \\ \mathrm{P}(-y \le |X| \le y), & \text{for } y \ge 0. \end{cases}$$
$$= \begin{cases} 0, & \text{for } y < 0, \\ F_X(y) - F_X((-y)-), & \text{for } y \ge 0. \end{cases}$$

Recall that $F_X((-y)-)$ is the left limit of F_X at the point $-y$.

Example 1.14.2. *Let X be a continuous random variable with density $f_X(x) = cx^2$ for $x \in [0, 2]$ and $f_X(x) = 0$ otherwise. Find c and $\mathrm{E}(X)$ and $\mathrm{E}(X^2)$. The probability density function needs to be nonnegative and integrate to 1, hence we set*

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = c\int_0^2 x^2 dx = c\frac{1}{3}x^3\Big|_0^2 = c\frac{8}{3},$$

which implies that $c = \frac{3}{8} (\ge 0)$.

Then

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^2 \frac{3}{8}x^3 dx = \frac{3}{8}\frac{1}{4}x^4\Big|_0^2 = \frac{3}{2}.$$

Using the LOTUS, we get

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^2 \frac{3}{8}x^4 dx = \frac{3}{8}\frac{1}{5}x^5\Big|_0^2 = \frac{12}{5}.$$

Example 1.14.3. *Let $X \sim \mathrm{Ber}(p)$. Find $\mathrm{E}(X)$ and $\mathrm{E}(X)$. From the definition of the expectation for discrete random variables we get*

$$\mathrm{E}(X) = \sum_{x \in \mathrm{Im}X} x\mathrm{P}(X = x) = 0 \cdot \mathrm{P}(X = 0) + 1 \cdot \mathrm{P}(X = 1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Using the LOTUS, we have

$$\mathrm{E}(X^2) = \sum_{x \in \mathrm{Im}X} x^2\mathrm{P}(X = x) = 0^2 \cdot \mathrm{P}(X = 0) + 1^2 \cdot \mathrm{P}(X = 1) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p.$$

Example 1.14.4. *Let $X \sim \mathrm{Poi}(\lambda)$ for $\lambda > 0$. Find $\mathrm{E}(X!)$.*

$$\mathrm{E}(X!) \overset{\text{LOTUS}}{=} \sum_{n=0}^{\infty} n!\mathrm{P}(X = n) = \sum_{n=0}^{\infty} n!\frac{\lambda^n}{n!}e^{-\lambda} = e^{-\lambda}\sum_{n=0}^{\infty}\lambda^n \overset{\text{geom. series}, |\lambda|<1}{=} e^{-\lambda}\frac{1}{1-\lambda},$$

if $|\lambda| = \lambda < 1$ and $\mathrm{E}(X!) = \infty$ for $\lambda \ge 1$.

1.15 Lecture 15

Example 1.15.1. *Let $F_{X,Y}$ denote the joint c.d.f. of (X, Y). For $x, y \in \mathbb{R}$, find an expression for $\mathrm{P}(X \le x, Y \ge y)$ in terms of $F_{X,Y}$. We note that*

$$\mathrm{P}(X \le x, Y \ge y) + \mathrm{P}(X \le x, Y < y) \overset{\text{Law of total prob.}}{=} \mathrm{P}(X \le x) = F_{X,Y}(x, \infty).$$

Also, $\mathrm{P}(X \le x, Y < y) = F_{X,Y}(x, y-)$. Hence,

$$\mathrm{P}(X \le x, Y \ge y) = F_{X,Y}(x, \infty) - F_{X,Y}(x, y-).$$

1.16 Lectures 16 and 17

Example 1.16.1. *Consider jointly continuous random variables X, Y with joint density given by*

$$f_{X,Y}(x,y) = \begin{cases} c(x^2+y^2), & \text{for } 0 < x < 2, 0 < y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

1. *Find c such that $f_{X,Y}$ is a p.d.f.: We need that $c \geq 0$ and*

$$1 = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = \int_{x=0}^{2}\left(\int_{y=0}^{2} c(x^2+y^2)dy\right)dx = c\int_{x=0}^{2}\left(x^2y+\frac{1}{3}y^3\Big|_{y=0}^{2}\right)dx$$
$$= c\int_{x=0}^{2}\left(2x^2+\frac{8}{3}\right)dx = c\left(\frac{2}{3}x^3+\frac{8}{3}x\Big|_{x=0}^{2}\right) = c\frac{32}{3} \Leftrightarrow c = \frac{3}{32}.$$

2. *Find the joint c.d.f. of (X,Y).*

Case 1: Let $x \leq 0$ or $y \leq 0$: $F_{X,Y}(x,y) = 0$.

Case 2: Let $x, y \in (0,2)$:

$$F_{X,Y}(x,y) = \int_{u=0}^{x}\left(\int_{v=0}^{y} c(u^2+v^2)dv\right)du = c\int_{u=0}^{x}\left(u^2v+\frac{1}{3}v^3\Big|_{v=0}^{y}\right)du$$
$$= c\int_{u=0}^{x}\left(u^2y+\frac{1}{3}y^3\right)du = c\left(\frac{1}{3}u^3y+\frac{1}{3}y^3u\Big|_{u=0}^{x}\right) = c\left(\frac{1}{3}x^3y+\frac{1}{3}y^3x\right)$$
$$= \frac{1}{32}(x^3y+xy^3).$$

Case 3: Let $x \in (0,2), y \geq 2$: $F_{X,Y}(x,y) = \frac{1}{32}\left(2x^3+8x\right) = \frac{1}{16}x^3+\frac{1}{4}x = F_X(x)$.

Case 4: Let $y \in (0,2), x \geq 2$: $F_{X,Y}(x,y) = \frac{1}{32}\left(2y^3+8y\right) = \frac{1}{16}y^3+\frac{1}{4}y = F_Y(y)$.

Case 5: Let $x, y \geq 2$: $F_{X,Y}(x,y) = 1$.

3. *Differentiate the c.d.f. to obtain the p.d.f.:*

Case 1: Let $x \leq 0$ or $y \leq 0$: $F_{X,Y}(x,y) = 0$. *Hence*

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = 0.$$

Case 2: Let $x, y \in (0,2)$: $F_{X,Y}(x,y) = \frac{1}{32}(x^3y+xy^3)$. *Hence*

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y}\left[\frac{\partial}{\partial x}\frac{1}{32}(x^3y+xy^3)\right] = \frac{\partial}{\partial y}\left[\frac{1}{32}(3x^2y+y^3)\right]$$
$$= \frac{1}{32}(3x^2+3y^2) = \frac{3}{32}(x^2+y^2).$$

Case 3: Let $x \in (0,2), y \geq 2$: $F_{X,Y}(x,y) = \frac{1}{16}x^3+\frac{1}{4}x$. *Hence* $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = 0$.

Case 4: Let $y \in (0,2), x \geq 2$: $F_{X,Y}(x,y) = \frac{1}{16}y^3+\frac{1}{4}y$. *Hence* $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = 0$.

Case 5: Let $x, y \geq 2$: $F_{X,Y}(x,y) = 1$. *Hence* $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = 0$.

4. *Find the marginal densities of X and Y.*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{0}^{2} c(x^2+y^2)dy = c\left(x^2y+\frac{1}{3}y^3\Big|_{y=0}^{2}\right)$$

$$= c\left(2x^2 + \frac{8}{3}\right) = \frac{3}{16}x^2 + \frac{1}{4},$$

for $x \in (0, 2)$ *and* $f_X(x) = 0$ *otherwise. By symmetry,* $f_Y(y) = \frac{3}{16}y^2 + \frac{1}{4}$, *for* $y \in (0, 2)$ *and* $f_Y(y) = 0$ *otherwise.*

5. *Show that* X *and* Y *are not independent.*
   We have

$$f_X(x)f_Y(y) = \left(\frac{3}{16}x^2 + \frac{1}{4}\right)\left(\frac{3}{16}y^2 + \frac{1}{4}\right) \neq f_{X,Y}(x, y),$$

   for $x, y \in (0, 2)$, *hence* X *and* Y *are not independent.*

6. *Find the marginal c.d.f.s of* X *and* Y.
   $F_X(x) = F_{X,Y}(x, \infty) = \frac{1}{16}x^3 + \frac{1}{4}x$ *for* $x \in (0, 2)$, $F_X(x) = 0$ *for* $x \leq 0$ *and* $F_X(x) = 1$ *for* $x \geq 2$. *Also,* $F_Y(x) = F_{X,Y}(\infty, y) = \frac{1}{16}y^3 + \frac{1}{4}y$ *for* $y \in (0, 2)$, $F_Y(y) = 0$ *for* $y \leq 0$ *and* $F_Y(y) = 1$ *for* $y \geq 2$.

7. *Find* $\mathrm{Cov}(X, Y)$.
   We note that $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$. *Then*

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^2 \left(\frac{3}{16}x^3 + \frac{1}{4}x\right) dx = \frac{3}{16} \cdot \frac{1}{4}x^4 + \frac{1}{4} \cdot \frac{1}{2}x^2\bigg|_{x=0}^{2} = \frac{3}{4} + \frac{1}{2} = \frac{5}{4}.$$

   By symmetry, we also have that $\mathrm{E}(Y) = \frac{5}{4}$. *Also,*

$$\mathrm{E}(XY) \overset{\text{LOTUS}}{=} \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} xyf_{X,Y}(x, y)dydx = c\int_{x=0}^{2}\left(\int_{y=0}^{2}(x^3y + xy^3)dy\right)dx$$
$$= c\int_{x=0}^{2}\left(\frac{1}{2}x^3y^2 + \frac{1}{4}xy^4\bigg|_{y=0}^{2}\right)dx = c\int_{x=0}^{2}\left(2x^3 + 4x\right)dx = c\left(\frac{2}{4}x^4 + \frac{4}{2}x^2\right)\bigg|_{x=0}^{2}$$
$$= \frac{3}{32}(8 + 8) = \frac{3}{2}.$$

   Hence

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = \frac{3}{2} - \frac{5^2}{4^2} = -\frac{1}{16}.$$

8. *Find* $\mathrm{Cor}(X, Y)$.
   Recall that $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$. *Here we have*

$$\mathrm{E}(X^2) \overset{\text{LOTUS}}{=} \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^2 \left(\frac{3}{16}x^4 + \frac{1}{4}x^2\right) dx = \frac{3}{16} \cdot \frac{1}{5}x^5 + \frac{1}{4} \cdot \frac{1}{3}x^3\bigg|_{x=0}^{2}$$
$$= \frac{6}{5} + \frac{2}{3} = \frac{28}{15}.$$

   Hence

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{28}{15} - \frac{5^2}{4^2} = \frac{73}{240},$$

   and by symmetry $\mathrm{Var}(Y) = \frac{73}{240}$. *Hence*

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-\frac{1}{16}}{\frac{73}{240}} = -\frac{15}{73} \approx -0.2.$$

1.17 Lecture 18

Example 1.17.1. *Compute the mean and variance of the Bernoulli, Binomial and Poisson distributions using the probability generating functions.*

- *For* $X \sim \mathrm{Ber}(p)$, *we have, for* $s \in \mathbb{R}$,

$$G_X(s) = \mathrm{E}(s^X) \stackrel{\text{LOTUS}}{=} \sum_x s^x \mathrm{P}(X = x) = s^0 \mathrm{P}(X = 0) + s^1 \mathrm{P}(X = 1) = 1 - p + sp.$$

Then

$$\frac{d}{ds} G_X(s)\bigg|_{s=1} = p, \qquad \frac{d^2}{ds^2} G_X(s)\bigg|_{s=1} = 0.$$

Hence

$$\mathrm{E}(X) = G_X'(1) = p, \qquad \mathrm{Var}(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2 = p - p^2 = p(1-p).$$

- *For* $X \sim \mathrm{Bin}(n, p)$, *we have, for* $s \in \mathbb{R}$,

$$\begin{aligned} G_X(s) = \mathrm{E}(s^X) &\stackrel{\text{LOTUS}}{=} \sum_x s^x \mathrm{P}(X = x) = \sum_{x=0}^n s^x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^n \binom{n}{x} (sp)^x (1-p)^{n-x} \\ &\stackrel{\text{Binomial theorem}}{=} (sp + 1 - p)^n. \end{aligned}$$

Then

$$\begin{aligned} \frac{d}{ds} G_X(s)\bigg|_{s=1} &= n(sp + 1 - p)^{n-1} p\big|_{s=1} = np, \\ \frac{d^2}{ds^2} G_X(s)\bigg|_{s=1} &= n(n-1)(sp + 1 - p)^{n-2} p^2\big|_{s=1} = n(n-1)p^2. \end{aligned}$$

Hence

$$\begin{aligned} \mathrm{E}(X) &= G_X'(1) = np, \\ \mathrm{Var}(X) &= G_X''(1) + G_X'(1) - (G_X'(1))^2 = n^2p^2 - np^2 + np - (np)^2 = np(1-p). \end{aligned}$$

- *For* $X \sim \mathrm{Poi}(\lambda)$, *we have, for* $s \in \mathbb{R}$,

$$G_X(s) = \mathrm{E}(s^X) \stackrel{\text{LOTUS}}{=} \sum_x s^x \mathrm{P}(X = x) = \sum_{x=0}^\infty s^x \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{x=0}^\infty \frac{(s\lambda)^x}{x!} e^{-\lambda} = e^{\lambda s} e^{-\lambda} = \exp(\lambda(s-1)).$$

Then

$$\begin{aligned} \frac{d}{ds} G_X(s)\bigg|_{s=1} &= \exp(\lambda(s-1))\lambda\big|_{s=1} = \lambda, \\ \frac{d^2}{ds^2} G_X(s)\bigg|_{s=1} &= \exp(\lambda(s-1))\lambda^2\big|_{s=1} = \lambda^2. \end{aligned}$$

Hence

$$\begin{aligned} \mathrm{E}(X) &= G_X'(1) = \lambda, \\ \mathrm{Var}(X) &= G_X''(1) + G_X'(1) - (G_X'(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

Example 1.17.2. *Let X be a discrete random variable with $\mathrm{Im}X \subseteq \mathbb{N} \cup \{0\}$. Suppose that $G_X(s) = \frac{1}{3} + \frac{1}{5}s^5 + \frac{1}{5}s^{10} + \frac{4}{15}s^{12}$. Find the p.m.f. of X.*
Recall that

$$G_X(s) = \mathrm{E}(s^X) = \sum_{x=0}^{\infty} s^x \mathrm{P}(X = x) = \mathrm{P}(X = 0) + s\mathrm{P}(X = 1) + s^2\mathrm{P}(X = 2) + \cdots .$$

Hence, in the example above, we can just read off the probabilities:

$$\mathrm{P}(X = 0) = \frac{1}{3}, \mathrm{P}(X = 5) = \frac{1}{5}, \mathrm{P}(X = 10) = \frac{1}{5}, \mathrm{P}(X = 12) = \frac{4}{15},$$

and $\mathrm{P}(X = x) = 0$ for $x \notin \{0, 5, 10, 12\}$.

Remark 1.17.3. *Relation between the p.g.f. and the m.g.f. for discrete random variables: Let X be a discrete random variable with $\mathrm{Im}X \subseteq \mathbb{N} \cup \{0\}$. Then*

$$M_X(t) = \mathrm{E}(e^{tX}) = \mathrm{E}((e^t)^X) = G_X(e^t).$$

Remark 1.17.4. *Why is M_X called the* moment generating function*?*
Let X be a random variable. Then its moment generating function *(m.g.f.) is defined as*

$$M_X(t) = \mathrm{E}(e^{tX}),$$

provided the expectation exists in some neighbourhood of zero, i.e. the expectation exists for all $|t| < \epsilon$ for some $\epsilon > 0$. Then

$$M_X(t) = \mathrm{E}(e^{tX}) = \mathrm{E}\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) \overset{(*)}{=} \sum_{n=0}^{\infty} \mathrm{E}(X^n)\frac{t^n}{n!}.$$

Note that, in general, we are not allowed to interchange an infinite sum with the expectation. However, here the equality in (*) holds, since we assume the existence of the moment generating function in a neighbourhood of zero.

Also, we can do a Taylor series expansion of $M_X(t)$ around 0, which leads to

$$M_X(t) = \sum_{n=0}^{\infty} M^n(0)\frac{t^n}{n!}.$$

Clearly, for the two infinite series to be the same, we need that $M^{(n)}(0) = \mathrm{E}(X^n)$.

Example 1.17.5. *Generating all the moments of the exponential distribution:*

- *Let $X \sim \mathrm{Exp}(1)$. Then $M_X(t) = (1-t)^{-1}$ for all $t < 1$. Using the geometric series for $|t| < 1$, we have*

$$M_X(t) = \frac{1}{1-t} = \sum_{n=0}^{\infty} t^n = \sum_{n=0}^{\infty} n!\frac{t^n}{n!} = \sum_{n=0}^{\infty} \mathrm{E}(X^n)\frac{t^n}{n!},$$

  hence $\mathrm{E}(X^n) = n!$ for all $n \in \mathbb{N}$.

- *Now consider the general case, when $Y \sim \mathrm{Exp}(\lambda)$. Then $X := \lambda Y \sim \mathrm{Exp}(1)$. To see this, note that, for $x > 0$,*

$$F_X(x) = \mathrm{P}(X \le x) = \mathrm{P}(\lambda Y \le x) = \mathrm{P}(Y \le x/\lambda) = F_Y(x/\lambda) = 1 - \exp(-x),$$

  and $F_X(x) = 0$ for $x < 0$, which is the c.d.f. of an $\mathrm{Exp}(1)$-distributed random variable. Then, we can deduce that, for all $n \in \mathbb{N}$, we have

$$\mathrm{E}(X^n) = n! = \mathrm{E}(\lambda^n Y^n) = \lambda^n \mathrm{E}(Y^n) \Leftrightarrow \mathrm{E}(Y^n) = \frac{n!}{\lambda^n}.$$

  In particular, $\mathrm{E}(X) = \lambda^{-1}, \mathrm{Var}(X) = \lambda^{-2}$.

1.18 Lecture 19

Example 1.18.1. *Generating all the moments of the standard normal distribution:*
Let $X \sim \mathrm{N}(0,1)$. *Then*

$$M_X(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} = \sum_{n=0}^{\infty} \frac{(2n)!}{2^n n!} \cdot \frac{t^{2n}}{(2n)!} = \sum_{n=0}^{\infty} \mathrm{E}(X^{2n}) \frac{t^{2n}}{(2n)!}.$$

I.e. $\mathrm{E}(X^{2n}) = \frac{(2n)!}{2^n n!}$ *and* $\mathrm{E}(X^{2n-1}) = 0$ *for all* $n \in \mathbb{N}$.
The even moments can be computed using the following identity:

Lemma 1.18.2. $\frac{(2n)!}{2^n n!} = (2n-1)(2n-3)\cdots 3 \cdot 1$, *for* $n \in \mathbb{N}$.

Proof. We can give a story proof/proof by interpretation. Both sides count how many ways there are to break a group of $2n$ people into n pairs:
Left hand side: Take $2n$ people and label them 1 to $2n$. We can line up the $2n$ people (there are $(2n)!$ possible permutations) and say that the first two are a pair, the next two are a pair etc. Here we overcount by a factor of $n!$ since the order of the paris does not matter and by a factor of 2^n since the order within each pair does not matter.
Right hand side: There are $2n-1$ ways to choose a partner for the first person, then there are $2n-3$ choices for person 2 (or 3 if 2 was already paired to person 1) etc. □

1.19 Lecture 20

Example 1.19.1. *Consider two jointly continuous random variables with joint density (for* $\lambda > 0$*):*

$$f_{X,Y}(x,y) = \begin{cases} \lambda^2 e^{-\lambda y}, & \text{for } 0 \leq x \leq y < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

- *Find* $f_{Y|X}$*:*
  Recall that $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$. *We compute the marginal density of* X *first:*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_x^{\infty} \lambda^2 e^{-\lambda y} dy = \lambda^2 \frac{(-1)}{\lambda} e^{-\lambda y}\Big|_{y=x}^{\infty} = \lambda e^{-\lambda x},$$

  for $x \geq 0$ *and* $f_X(x) = 0$ *for* $x < 0$. *Hence,*

$$f_{Y|X}(y|x) = \begin{cases} \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda y}} = \lambda e^{-\lambda(y-x)}, & \text{for } 0 \leq x \leq y < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

- *Find* $\mathrm{E}(Y|X=x)$.
  For $x \geq 0$, *we have*

$$\begin{aligned} \mathrm{E}(Y|X=x) &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy = \int_x^{\infty} y\lambda e^{-\lambda(y-x)}dy = e^{\lambda x}\int_x^{\infty} \lambda y e^{-\lambda y} \\ &\overset{u:=\lambda y}{=} e^{\lambda x}\int_{\lambda x}^{\infty} u e^{-u} du \lambda^{-1} = \frac{e^{\lambda x}}{\lambda}\left\{u(-e^{-u})|_{u=\lambda x}^{\infty} - \int_{\lambda x}(-e^{-u})du\right\} \\ &= \frac{e^{\lambda x}}{\lambda}\left\{\lambda x e^{-\lambda x} + e^{-\lambda x}\right\} = \frac{1}{\lambda}(\lambda x + 1). \end{aligned}$$

Remark 1.19.2.
- *If* (X,Y) *is bivariate normal and* $\mathrm{Cov}(X,Y) = 0 \Rightarrow X, Y$ *are independent.*
- *However, if* X *and* Y *follow a univariate normal distribution and* $\mathrm{Cov}(X,Y) = 0 \not\Rightarrow X, Y$ *are independent.*

Example 1.19.3. *Let $X \sim \mathrm{N}(0,1)$. Let Z be a discrete random variable, independent of X with $\mathrm{P}(Z = -1) = \mathrm{P}(Z = 1) = \frac{1}{2}$. Let $Y := Z \cdot X$. We want to show that 1) $Y \sim \mathrm{N}(0,1)$, 2) $\mathrm{Cov}(X,Y) = 0$, 3) X and Y are not independent:*

1. *$Y \sim \mathrm{N}(0,1)$:*
   Let $y \in \mathbb{R}$. Then, using the law of total probability and the independence of Z and X, we have

$$
\begin{aligned}
F_Y(y) &= \mathrm{P}(Y \le y) = \mathrm{P}(ZX \le y) \\
&= \mathrm{P}(ZX \le y|Z = -1)\mathrm{P}(Z = -1) + \mathrm{P}(ZX \le y|Z = 1)\mathrm{P}(Z = 1) \\
&= \frac{1}{2}\left(\mathrm{P}(-X \le y) + \mathrm{P}(X \le y)\right) = \frac{1}{2}(\mathrm{P}(X \ge -y) + \Phi(y)) \\
&= \frac{1}{2}(1 - \mathrm{P}(X \le -y) + \Phi(y)) = \frac{1}{2}(1 - \Phi(-y) + \Phi(y)) \\
&= \frac{1}{2}(1 - (1 - \Phi(y)) + \Phi(y)) = \Phi(y).
\end{aligned}
$$

   Hence $Y \sim \mathrm{N}(0,1)$.

2. *$\mathrm{Cov}(X,Y) = 0$:*

$$
\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = \mathrm{E}(X^2 Z) \overset{\text{independence of } X,Z}{=} \mathrm{E}(X^2)\mathrm{E}(Z) = 1 \cdot 0 = 0.
$$

3. *X and Y are not independent:*
   We note that $|X| = |Y|$ is always true, hence X and Y are not independent.

As a side remark, we note that the sum of X and Y is not normally distributed (and not even continuous!):

$$
\begin{aligned}
\mathrm{P}(X + Y = 0) &= \mathrm{P}(X + ZX = 0) = \mathrm{P}(X(1 + Z) = 0) \\
&= \mathrm{P}(X(1 + Z) = 0|Z = 1)\mathrm{P}(Z = 1) + \mathrm{P}(X(1 + Z) = 0|Z = -1)\mathrm{P}(Z = -1) \\
&= \mathrm{P}(2X = 0) \cdot \frac{1}{2} + \mathrm{P}(0 = 0) \cdot \frac{1}{2} = \frac{1}{2} \neq 0.
\end{aligned}
$$

Conclusions: If we only know that X and Y follow univariate normal distributions, we **cannot** *conclude that*

- *(X,Y) has bivariate normal distribution,*
- *(X,Y) are jointly continuous,*
- *$\mathrm{Cov}(X,Y) = 0 \Rightarrow X, Y$ are independent.*