A

Overview of appendices

These appendices were written in response to questions about rigour in the Oxford Probability courses. The obvious place to point to is the Part B course on Probability, Measure and Martingales, where a general measure-theoretic approach is presented, which allows to streamline the main developments at an abstract level and to specialise to the setting of Prelims and Part A Probability. At this stage, however, this is a rather unsatisfactory answer to the sometimes rather specific questions on technical detail and fails to acknowledge that most of the developments in Prelims and Part A Probability are already rigorous with only very occasional use of tools that have not yet been rigorously established in the Analysis courses.

Indeed, Probability meets Analysis almost immediately. As Prelims Probability is taught before or in parallel to the Analysis courses, we are slightly pragmatic about series and integrals, but (discrete and continuous) random variables are introduced as functions on a probability space, and the Prelims Analysis courses then give these developments the same rigorous meaning (in the generality presented) that they have (after considerably more effort) in a general measure-theoretic approach. There are just a few technical gaps on the Analysis side that are best identified and can then be addressed in the context of Part A Integration.

The aim of these (non-examinable) appendices is to fill any Analysis gaps left in the Prelims and Part A Probability courses, by pointing to relevant developments in the Analysis courses, where available, by revisiting arguments that were not fully justified at the time and by adding a few arguments to fill gaps that may benefit from a bit more than a pointer.

As far as Prelims material is concerned, Appendix B collects mainly statements of relevant definitions and theorems taken often directly from

- Analysis I lecture notes by Vicky Neale (Michaelmas Term 2020), Hilary Priestley (2016),
- Analysis II lecture notes by Zhongmin Qian (Hilary Term 2021),
- Analysis III lecture notes by Marc Lackenby (Trinity Term 2021).

Appendix C records some consequences, notably about the notion of countable sums, but also discussing some relevant integrals. Appendix D reviews, in the context of Appendices B–C where relevant, all definitions and essentially all theorems of the

- Prelims Probability lecture notes by James Martin (Michaelmas Term 2020).

Appendix D is also useful reading before Part A Probability when not chasing gaps.

Before turning to Part A material, we include in Appendix E some further developments that allow to address at the level of Prelims Analysis and Probability some more points that have arisen in the discussion of Prelims Probability in Appendix D.

The Part A Integration course is the key course that fills or allows to fill almost all remaining gaps in the Prelims and Part A Probability courses. In Part A Probability, we also state some theorems without proof (analyic characterisations of probability distributions and their convergence), which can be proved using material from the Part A Metric Spaces and Complex Analysis course. Our preview in Appendix F contains relevant material from

- Part A Integration lecture notes by Charles Batty (Hilary Term 2021),
- Part A Metric Spaces lecture notes by Ben Green (Michaelmas Term 2021),
- Part A Complex Analysis lecture notes by Ben Green and Panos Papazoglou (Michaelmas Term 2021).

Appendix G records some consequences notably about the existence of suitable probability spaces, expectations of functions of continuous random variables and some further integrals, which were already justified less formally in Prelims Probability. We also include sections on (probability and other) measure spaces and the measure-theoretic approach to define expectations as integrals against probability measures. These latter developments are strictly parallel to the material quoted from the Part A Integration course in Appendix F. These parallels will be evident in their statements, and while we do not provide proofs, these parallels do extend to their proofs, too. The purpose here is to provide clear statements of some theorems that we later apply to provide proofs of the results mentioned above. A slightly different approach to these theorems is given in Part B Probability, Measure and Martingales.

Appendix H turns to Part A Probability and includes a chapter-by-chapter discussion, revisiting informal proofs and results stated without proof, as appropriate. This perspective of Part A Probability is non-examinable.

Markov chain theory in Part A Probability is formulated in terms of matrices and vectors, finite-dimensional, as well as (infinite-dimensional) vector spaces of sequences. Appendix I contains a small amount of material from the algebraic Prelims courses also including material on linear recurrence relations, which was included in the Prelims Probability course:

- Linear Algebra I lecture notes by Richard Earl (Michaelmas Term 2020),
- Linear Algebra II lecture notes by James Maynard (Hilary Term 2021),
- Groups and Group Actions lecture notes by Vicky Neale (Hilary Terms 2020),
- Prelims Probability lecture notes by James Martin (Michaelmas Term 2020).

Any proofs included in any of the appendices are either extrapolating from material from the various lecture notes or from textbooks on the reading list or from

- R. Durrett: *Probability: Theory and Examples.* CUP 2019 (or earlier editions).

These appendices are all new in Michaelmas Term 2021. If you spot any typos or mistakes or have other suggestions or comments, please email me at winkel@stats.ox.ac.uk.

Matthias Winkel, October 2021

B

Prelims Analysis: Relevant material

This appendix contains a selection of statements of definitions and theorems from the Prelims Analysis courses that are relevant for Prelims and Part A Probability.

B.1 Sequences and series

We denote by $\mathbb{N}$ the set of positive integers (not including 0) and by $\mathbb{R}$ the set of real numbers.

Definition B.1. A *sequence of real numbers* is a function $b\colon \mathbb{N} \to \mathbb{R}$. We usually write b_n for the real number $b(n)$ and $(b_n)_{n\geq 1}$ or (b_n) for the sequence.

Definition B.2. Let (b_n) be a sequence of real numbers and let $L \in \mathbb{R}$. We say that (b_n) *converges to L as $n \to \infty$* if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N} \text{ such that } \forall n \geq N, \text{ we have } |b_n - L| < \varepsilon.$$

In this case, we write $\lim_{n\to\infty} b_n = L$ or $b_n \to L$ as $n \to \infty$, and we say L is the *limit of* (b_n).

Definition B.3. Let (b_n) be a sequence of real numbers. We say that (b_n) *is convergent* if there is $L \in \mathbb{R}$ such that $b_n \to L$. If (b_n) is not convergent, then we say it is *divergent*.

Theorem B.4 (Uniqueness of limits)**.** *Let (b_n) be a convergent sequence of real numbers. Then the limit is unique.*

Theorem B.5 (Algebra of Limits)**.** *Let (a_n) and (b_n) be sequences with $a_n \to K$ and $b_n \to L$. Then $a_n + b_n \to K + L$ and $a_n b_n \to KL$.*

Definition B.6. Let (a_n) be a sequence. We say that (a_n) is *bounded* if

$$\exists M \in \mathbb{R} \text{ such that } \forall n \geq 1, \text{ we have } |a_n| \leq M.$$

Proposition B.7. *Let (a_n) be convergent. Then (a_n) is bounded.*

Definition B.8. Let (a_n) be a real sequence. We say that (a_n) is *increasing* if $a_n \leq a_{n+1}$ for all $n \geq 1$. We say that (a_n) is *strictly increasing* if $a_n < a_{n+1}$ for all $n \geq 1$.

Theorem B.9 (Monotone Sequences Theorem)**.** *Let (a_n) be a real sequence. If (a_n) is increasing and bounded, then (a_n) converges.*

Definition B.10. Let $(a_n)_{n\geq 1}$ be sequence. A *subsequence* $(b_r)_{r\geq 1}$ of $(a_n)_{r\geq 1}$ is defined by a function $f\colon \mathbb{N} \to \mathbb{N}$ such that f is strictly increasing (if $p < q$ then $f(p) < f(q)$), and $b_r = a_{f(r)}$ for $r \geq 1$. We often write $f(r)$ as n_r. Then $(n_r)_{r\geq 1}$ is a strictly increasing sequence of natural numbers, and $b_r = a_{n_r}$.

Proposition B.11. *Let (a_n) be a sequence. If (a_n) converges, then every subsequence (a_{n_r}) of (a_n) converges. Moreover, if $a_n \to L$ as $n \to \infty$, every subsequence also converges to L.*

Theorem B.12 (Bolzano–Weierstrass Theorem)**.** *Let (a_n) be a bounded sequence. Then (a_n) has a convergent subsequence.*

Definition B.13. Let (b_n) be a sequence of real numbers. We say that (b_n) *diverges to infinity* if

$$\forall M \in \mathbb{R}\ \exists N \in \mathbb{N} \text{ such that } \forall n \geq N, \text{ we have } a_n > M.$$

In this case, we write $\lim_{n\to\infty} b_n = \infty$ or $b_n \to \infty$.

Remark B.14. Uniqueness of limits extends to include the limit infinity. "Algebra of Limits" with infinite limits requires appropriate definitions, constraints and separate proofs. Although this was not presented in this way in Prelims Analysis, it is sometimes useful to define $\infty + \infty = a + \infty = \infty + a = \infty$ and $b\infty = \infty b = \infty\infty = \infty$ for $a \in \mathbb{R}$ and $b > 0$. If one or both of the limits K and L of (a_n) and (b_n) are infinite in the sense of Definition B.13 and their sum or product are defined in this way, Algebra of Limits extends to these cases. We leave the proofs as an exercise to the reader. We also stress that we leave the product 0∞ undefined in general, because Algebra of Limits does not extend to this case, in general.

Definition B.15. Let (b_n) be a sequence of real numbers. Its sequence (s_n) of *partial sums* is given by $s_1 = b_1$ and $s_{n+1} = s_n + b_n$, $n \in \mathbb{N}$. We usually write this as

$$s_n = \sum_{j=1}^{n} b_j, \quad n \in \mathbb{N}.$$

We say the *series of* (b_n) is *convergent* if its sequence of partial sums converges. In this case, we write the limit as

$$\lim_{n\to\infty} s_n = \sum_{j=1}^{\infty} b_j,$$

and we also say *the series* $\sum b_j$ *converges.* A series is *divergent* if its sequence of partial sums diverges, and we also say *the series* $\sum b_j$ *diverges.* If the sequence of partial sums diverges to infinity, we say *the series* $\sum b_j$ *diverges to infinity* and write $\sum b_j = \infty$.

A series $\sum b_j$ is *absolutely convergent* if $\sum |b_j|$ is convergent.

Theorem B.16 (Comparison test)**.** *Let (b_j) and (c_j) be real sequences that satisfy $0 \leq |b_j| \leq c_j$ for all $j \geq 1$. Then the convergence of $\sum c_j$ implies the convergence of $\sum b_j$.*

Theorem B.17 (Cauchy convergence criterion for series)**.** *Let (b_j) be a sequence. Then $\sum b_j$ converges if and only if*

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N} \textit{ such that } \forall n > m \geq N, \textit{ we have } \left| \sum_{j=m+1}^{n} b_j \right| < \varepsilon.$$

Theorem B.18 (Absolute convergence implies convergence)**.** *Let (b_j) be a sequence. If the series $\sum b_j$ is absolutely convergent, the series is also convergent.*

The notions of sequence, series and their convergence remain meaningful if we replace the set $\mathbb{R}$ of real numbers by the set $\mathbb{C}$ of complex numbers, giving rise to *sequences and series of complex numbers.* All theorems stated above extend to this setting. Specifically, in Theorem B.16, (c_n) will still be a real sequence and $0 \leq |b_j| \leq c_j$ are inequalities between real numbers.

Definition B.19. A *real power series* is a series of the form $\sum_{k=0}^{\infty} c_k x^k$, where $c_k \in \mathbb{R}$ for all $k \geq 0$ and x is a real variable.

A *complex power series* is a series of the form $\sum_{k=0}^{\infty} c_k z^k$ where $c_k \in \mathbb{C}$ for all $k \geq 0$ and z is a complex variable.

Definition B.20. Let $\sum c_k z^k$ be a power series. We define its *radius of convergence* to be

$$R := \begin{cases} \sup\{|z| \colon z \in \mathbb{C} \text{ and } \sum |c_k z^k| \text{ converges}\} & \text{if the sup exists} \\ \infty & \text{otherwise.} \end{cases}$$

Proposition B.21. *Let $\sum c_k z^k$ be a power series with radius of convergence R.*

(i) If $R > 0$ and $|z| < R$, then $\sum c_k z^k$ converges absolutely and hence converges.

(ii) If $R < \infty$ and $|z| > R$, then $\sum c_k z^k$ diverges.

Example B.22. The exponential series $\sum_{k=0}^{\infty} (1/k!) z^k$ has radius of convergence $R = \infty$.

B.2 Function limits, continuity and differentiability

Definition B.23. Let $E \subseteq \mathbb{C}$. Then $p \in \mathbb{C}$ is called a *limit point* of E if

$$\forall \varepsilon > 0 \; \exists z \in E \setminus \{p\} \text{ such that } |z - p| < \varepsilon.$$

Proposition B.24. *Let $a, b \in \mathbb{R}$ with $a < b$. Then $p \in \mathbb{R}$ is a limit point of the open interval (a, b) if and only if $p \in [a, b]$. The closed and half-open intervals $[a, b]$, $[a, b)$ and $(a, b]$ have the same limit points as the open interval (a, b).*

Definition B.25 (Function limits)**.** Let $E \subseteq \mathbb{C}$ and $f \colon E \to \mathbb{C}$ a function. Let p be a limit point of E, and L a number. We say f converges to L as x tends to p if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \text{ such that } \forall x \in E \colon 0 < |x - p| < \delta, \text{ we have } |f(x) - L| < \varepsilon.$$

In this case, we write $\lim_{x \to p} f(x) = L$ or $f(x) \to L$ as $x \to p$.

Proposition B.26 (Uniqueness of limits)**.** *Let $f \colon E \to \mathbb{C}$ be a function and p a limit point of E. If f has a limit as $x \to p$, then the limit is unique.*

Proposition B.27 (Algebra of Limits)**.** *Let p be a limit point of $E \subseteq \mathbb{C}$, and $f, g \colon E \to \mathbb{C}$ two functions. Suppose that $\lim_{x \to p} f(x) = A$ and $\lim_{x \to p} g(x) = B$. Then*

(1) $\lim_{x \to p} (f(x) \pm g(x)) = A \pm B$;

(2) $\lim_{x \to p} f(x) g(x) = AB$.

Definition B.28. Let $f\colon (a,b] \to \mathbb{C}$ be a function and let $p \in (a,b]$. Then we say the left-hand limit of f at p exists and equals L if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \text{ such that } \forall x \in (a,b]\colon 0 < p - x < \delta, \text{ we have } |f(x) - L| < \varepsilon.$$

In this case, we write $\lim_{x\uparrow p} f(x) = L$ or $\lim_{x\to p-} f(x) = L$. Right-hand limits denoted by $\lim_{x\downarrow p} f(x)$ or $\lim_{x\to p-} f(x)$ are defined analogously.

Definition B.29. Let $E \subseteq \mathbb{C}$. Consider a function $f\colon E \to \mathbb{C}$ and $p \in E$. We say that f is continuous at p if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \text{ such that } \forall x \in E\colon |x - p| < \delta, \text{ we have } |f(x) - f(p)| < \varepsilon.$$

We say that f is right-continuous at p if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \text{ such that } \forall x \in E\colon 0 \le x - p < \delta, \text{ we have } |f(x) - f(p)| < \varepsilon.$$

We say that f is (right-)continuous on E if f is (right-)continuous at p for all $p \in E$.

Proposition B.30. *If $f, g\colon E \to \mathbb{C}$ are continuous at $p \in E$, then so are $f \pm g$ and fg.*

Example B.31. Let $f\colon \mathbb{C} \to \mathbb{C}$ be a polynomial. Then f is continuous on $\mathbb{C}$.

Definition B.32. Let f and f_n, $n \in \mathbb{N}$, be complex-valued functions on E.

(1) We say that f_n converges to f uniformly on E if

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \text{ such that } \forall x \in E \; \forall n > N, \text{ we have } |f_n(x) - f(x)| < \varepsilon.$$

(2) We say that the series $\sum_{n=1}^{\infty} f_n$ converges uniformly on E if the sequence $s_n = \sum_{k=1}^{n} f_k$ of partial sums converges uniformly on E.

Theorem B.33 (The uniform limit of continuous functions is continuous)**.** *Let $f_n, f\colon E \to \mathbb{C}$ be functions with $f_n \to f$ uniformly in E. Suppose that all f_n are continuous at $x_0 \in E$, then the limit function f is also continuous at x_0.*

Corollary B.34. *Suppose $\sum_{n=0}^{\infty} a_n z^n$ has radius of convergence R, where $0 < R \le \infty$. Then for every $0 \le r < R$, the power series $\sum_{n=0}^{\infty} a_n z^n$ converges uniformly on the closed disk $\{z \in \mathbb{C}\colon |z| \le r\}$, and $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is continuous on the open ball $\{z \in \mathbb{C}\colon |z| < R\}$.*

Example B.35. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by the convergent power series

$$f(x) = \sum_{n=0}^{\infty} (1/n!) x^n$$

is continuous and equal to e^x, for all $x \in \mathbb{R}$, where $e := f(1)$.

Theorem B.36 (Abel's theorem)**.** *If the series $\sum_{n=0}^{\infty} a_n$ converges, then $\sum_{n=0}^{\infty} a_n x^n$ converges uniformly on $[0,1]$. In particular, $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is continuous on $[0,1]$ with*

$$\lim_{x\uparrow 1} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n.$$

Definition B.37. (1) Let $(a, b) \subseteq \mathbb{R}$ be an open interval, f a real- or complex-valued function defined on (a, b), and $x_0 \in (a, b)$. If

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists (a real or complex number), then the limit is called the *derivative* of f at x_0 and is denoted by $f'(x_0)$ or $\frac{df}{dx}(x_0)$.

(2) If $f\colon (a, b] \to \mathbb{C}$ and $x_0 \in (a, b]$, then the left-derivative of f at x_0 is defined by

$$f'(x_0-) = \lim_{x \uparrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

provided the limit exists.

(3) Let $D \subseteq \mathbb{C}$ and $z_0 \in D$ such that $B(z_0, \delta) := \{z \in \mathbb{C}\colon |z - z_0| < \delta\} \subseteq D$ for some $\delta > 0$. We define the complex derivative of $f\colon D \to \mathbb{C}$ at z_0 to be

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

provided the limit exists.

Theorem B.38. *If $f, g\colon (a, b) \to \mathbb{C}$ are differentiable at $x_0 \in (a, b)$, then*

(1) $(f \pm g)'(x_0) = f'(x_0) \pm g'(x_0)$,

(2) (Product rule) $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$.

Theorem B.39 (Chain rule)**.** *Suppose $f\colon (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ and $g\colon (c, d) \to \mathbb{R}$ is differentiable at $y_0 = f(x_0) \in (c, d)$, and $f((a, b)) \subseteq (c, d)$, then $h = g \circ f$ is differentiable at x_0 and*

$$h'(x_0) = g'(f(x_0))f'(x_0).$$

Theorem B.40. *Consider the power series*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Let R be its radius of convergence, and assume that $0 < R \leq \infty$. Then

(1) the power series obtained by differentiating f term by term

$$g(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}$$

has the same radius of convergence R. In particular, for any $0 \leq r < R$

$$\sum_{n=1}^{\infty} n|a_n|r^{n-1} < \infty,$$

(2) the derivative

$$f'(z) = \lim_{w \to z} \frac{f(w) - f(z)}{w - z}$$

exists for every $z \in \mathbb{C}$ satisfying that $|z| < R$, and $f'(z) = g(z)$. That is

$$\frac{d}{dz} \sum_{n=0}^{\infty} a_n z^n = \sum_{n=1}^{\infty} n a_n z^{n-1} \qquad \forall z \in \mathbb{C}\colon |z| < R.$$

Theorem B.41 (L'Hôpital's rule). *Suppose that $f, g \to (a-\delta, a) \to \mathbb{R}$ are differentiable for some $a \in \mathbb{R}$, $\delta > 0$, and that $\lim_{x\uparrow a} f(x) = \lim_{x\uparrow a} g(x) = 0$, then*

$$\lim_{x\uparrow a} \frac{f(x)}{g(x)} = \lim_{x\uparrow a} \frac{f'(x)}{g'(x)},$$

provided that the limit on the right-hand side exists.

B.3 Riemann Integration

Definition B.42. Let $[a, b]$ be an interval. A *partition* $\mathcal{P}$ of $[a, b]$ is a finite sequence $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$. A function $\phi\colon [a, b] \to \mathbb{R}$ is called a *step function* if there is a partition $\mathcal{P}$ given by $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$ such that ϕ is constant on each open interval (x_{i-1}, x_i), $1 \leq i \leq n$. We then say that ϕ is *adapted* to the partition $\mathcal{P}$.

Lemma B.43. *The set $\mathcal{L}_{\text{step}}[a, b]$ of all step functions on $[a, b]$ is a vector space.*

Definition B.44. Let ϕ be a step function adapted to some partition $\mathcal{P}$ given by $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$, and suppose that $\phi(x) = c_i$ on the interval (x_{i-1}, x_i), $1 \leq i \leq n$. Then we define

$$I(\phi) = \sum_{i=1}^{n} c_i(x_i - x_{i-1}).$$

Lemma B.45. *The map $I\colon \mathcal{L}_{\text{step}}[a, b] \to \mathbb{R}$ is well-defined in the sense that $I(\phi)$ does not depend on the choice of partition $\mathcal{P}$ to which ϕ is adapted. Furthermore, the map I is linear: $I(\lambda\phi_1 + \mu\phi_2) = \lambda I(\phi_1) + \mu I(\phi_2)$.*

Definition B.46. A bounded function $f\colon [a, b] \to \mathbb{R}$ is *Riemann integrable* if

$$\sup_{\phi_-} I(\phi_-) = \inf_{\phi_+} I(\phi_+)$$

where the sup is over all *minorant* step functions $\phi_- \leq f$, and the inf is over all *majorant* step functions $\phi_+ \geq f$. In this case, we define the Riemann integral $\int_a^b f$ to be the common value of the two quantities in the display.

Proposition B.47. *Suppose that f is Riemann integrable on $[a, b]$. Then, for any c with $a < c < b$, f is Riemann integrable on $[a, c]$ and on $[c, b]$. Moreover $\int_a^b f = \int_a^c f + \int_c^b f$.*

Proposition B.48. *The Riemann integrable functions on $[a, b]$ form a vector space and the Riemann integral is a linear functional on it: $\int_a^b (\lambda f + \mu g) = \lambda \int_a^b f + \mu \int_a^b g$.*

Corollary B.49. *If f is Riemann integrable on $[a, b]$, and if $\tilde{f}$ differs from f in finitely many points, then $\tilde{f}$ is also Riemann integrable, and $\int_a^b f = \int_a^b \tilde{f}$.*

Proposition B.50. *Suppose that f is Riemann integrable on $[a, b]$. If g is another Riemann integrable function on $[a, b]$, and if $f \leq g$ pointwise, then $\int_a^b f \leq \int_a^b g$. In particular, $|\int_a^b f| \leq \int_a^b |f|$.*

Theorem B.51. *Continuous functions $f\colon [a, b] \to \mathbb{R}$ are Riemann integrable.*

Theorem B.52. *Bounded continuous functions $f\colon (a,b) \to \mathbb{R}$, arbitrarily extended to $[a,b]$, are Riemann integrable.*

Theorem B.53. *Increasing functions and decreasing functions $f\colon [a,b] \to \mathbb{R}$ are Riemann integrable.*

Theorem B.54 (First fundamental theorem of calculus)**.** *Suppose that f is Riemann integrable on (a,b). Define a new function $F\colon [a,b] \to \mathbb{R}$ by*

$$F(x) := \int_a^x f.$$

Then F is continuous. Moreover, if f is continuous at $c \in (a,b)$, then F is differentiable at c and $F'(c) = f(c)$.

Theorem B.55 (Second fundamental theorem of calculus)**.** *Suppose that $F\colon [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and differentiable on (a,b). Suppose furthermore that its derivative F' is Riemann integrable on (a,b). Then*

$$\int_a^b F' = F(b) - F(a).$$

Proposition B.56 (Integration by parts)**.** *Suppose that $f, g\colon [a,b] \to \mathbb{R}$ are continuous functions, differentiable on (a,b). Suppose that the derivatives f', g' are Riemann integrable on (a,b). Then fg' and $f'g$ are Riemann integrable on (a,b), and*

$$\int_a^b fg' = f(b)g(b) - f(a)g(a) - \int_a^b f'g.$$

Proposition B.57 (Substitution)**.** *Suppose $f\colon [a,b] \to \mathbb{R}$ is continuous and $g\colon [c,d] \to [a,b]$ is continuous on $[c,d]$, has $g(c) = a$ and $g(d) = b$, and maps (c,d) to (a,b). Suppose moreover that g is differentiable on (c,d) and that its derivative g' is Riemann integrable on this interval. Then*

$$\int_a^b f = \int_c^d (f \circ g)g'.$$

Definition B.58. Let $-\infty \le a < d \le \infty$ and $f\colon (a,b) \to \mathbb{R}$ such that for all $[b,c] \subset (a,d)$, the function f is Riemann integrable on $[b,c]$. If the limit $\lim_{b\downarrow a} \int_b^c f$ exists, we denote this limit by $\int_a^c f$. Similarly, if the limits $\lim_{c\uparrow d} \int_b^c f$ or $\lim_{c\uparrow d} \int_a^c f$ exist, we denote them by $\int_b^d f$ or $\int_a^d f$. Such limits of Riemann integrals are referred to as *improper Riemann integrals*.

Proposition B.59. *If $\int_a^d f$ exists as an improper Riemann integral and $a < b < c < d$, then $\int_a^d f = \int_a^b f + \int_b^c f + \int_c^d$*

Proof. This follows from Proposition B.47 before taking limits, and from Algebra of Limits (Proposition B.27). □

C

Prelims Analysis: Relevant consequences

This appendix contains further developments that build on the Prelims Analysis courses and are relevant for rigorously following the general theory developed in Prelims and Part A Probability. This material is non-examinable.

C.1 Rearrangements and countable sums

We now formalise the theory of countable sums, which we implicitly use in Probability. The starting point is the definition of rearrangements of a series given in Analysis I. The purpose there was the following theorem, but it was given as a remark without proof.

Definition C.1. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a bijection. Given a series $\sum b_j$, write $b'_j = b_{g(j)}$, $j \in \mathbb{N}$. Then $\sum b'_j$ is a *rearrangement* of $\sum b_j$.

Theorem C.2. *Let $\sum b_j$ be a series that converges to L. If $\sum b_j$ is absolutely convergent, then any rearrangement of $\sum b_j$ also converges to L.*

Proof. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a bijection. We denote its inverse by $g^{-1}\colon \mathbb{N} \to \mathbb{N}$. Let $b'_j = b_{g(j)}$, $j \in \mathbb{N}$ and consider the sequences (s_k) and (s'_k) of partial sums of (b_j) and (b'_j). Then we have $s_k \to L$ by hypothesis and need to show that $s'_k \to L$. By Algebra of Limits, cf. Theorem B.5, it suffices to show that $(s'_k - s_k)$ converges to 0.

Fix $\varepsilon > 0$. Since $\sum |b_j|$ is convergent, the Cauchy criterion of Theorem B.17 ensures that there is $N \in \mathbb{N}$ such that

$$\forall n > m \geq N, \qquad \sum_{j=m+1}^{n} |b_j| = \left| \sum_{j=m+1}^{n} |b_j| \right| < \varepsilon/2.$$

Let $K = \max\{g^{-1}(j)\colon 1 \leq j \leq N\}$. Then $K \geq N$ and for all $k \geq K$, the finite sums s_k and s'_k have common terms $b_j = b'_{g^{-1}(j)}$, $1 \leq j \leq N$, which cancel in the difference $s'_k - s_k$ of finite sums. The other terms b_j with $j \geq N+1$ and $b'_j = b_{g(j)}$ with $g(j) \geq N+1$ are small. Specifically, for $m = N$ and $n = \max\{g(j)\colon 1 \leq j \leq k\}$, we have

$$|s'_k - s_k| \leq 2 \sum_{j=m+1}^{n} |b_j| < \varepsilon,$$

by the triangle inequality. We conclude that $(s'_k - s_k)$ converges to 0, as required. □

Remark C.3. Let $\sum b_j$ be a series that converges to L, but does not converge absolutely. Then it can shown that there are rearrangements of $\sum b_j$ that do not converge. In fact, for every $L' \in \mathbb{R}$, there are also rearrangements of $\sum b_j$ that converge to L'. We do not need this result, but mention it as a warning that absolute convergence is important in Theorem C.2.

Theorem C.4. *Let $\sum b_j$ be a series that diverges to infinity. If $b_j \geq 0$ for all $j \in \mathbb{N}$, then any rearrangement of $\sum b_j$ also diverges to infinity.*

Proof. With g, g^{-1}, b'_j, (s_k) and (s'_k) as in the proof of Theorem C.2, we have $s_k \to \infty$ by hypothesis and need to show that $s'_k \to \infty$.

Fix $M \in \mathbb{R}$. Since $s_k \to \infty$, there is $N \in \mathbb{N}$ such that $s_k \geq s_N \geq M$ for all $k \geq N$. Let $K = \max\{g^{-1}(j)\colon 1 \leq j \leq N\}$. Then for all $k \geq K$, the finite sum s'_k has non-negative terms $b'_j = b_{g(j)} \geq 0$ including all $b_j = b'_{g^{-1}(j)}$, $1 \leq j \leq N$. Hence $s'_k \geq s_N \geq M$, as required. □

Definition C.5. Let I be a finite set with $n \in \mathbb{N}$ elements and $f\colon I \to \mathbb{R}$ a function. Let $h\colon \{j \in \mathbb{N}\colon 1 \leq j \leq n\} \to I$ be a bijection and $b_j = f(h(j))$. We define the *finite sum*

$$\sum_{i \in I} f(i) = \sum_{j=1}^{n} b_j.$$

Let I be a countably infinite set and $f\colon I \to \mathbb{R}$ a function. Let $h\colon \mathbb{N} \to I$ be a bijection and $b_j = f(h(j))$. If $\sum b_j$ either converges absolutely or if $b_j \geq 0$ for all $j \in \mathbb{N}$, we define the *countably infinite sum*

$$\sum_{i \in I} f(i) = \sum_{j=1}^{\infty} b_j.$$

If both absolute convergence and non-negativity fail, we leave $\sum_{i \in I} f(i)$ *undefined*, in general. We refer to both finite sums and countably infinite sums as *countable sums*.

Proposition C.6. *In the setting of Definition C.5, the countable sum $\sum_{i \in I} f(i)$ is well-defined, i.e. the definition of $\sum_{i \in I} f(i)$ does not depend on the choice of the bijection h. More precisely, if I is countably infinite and $\sum f(h(j))$ converges absolutely for one bijection $h\colon \mathbb{N} \to I$, it converges absolutely for all bijections $h\colon \mathbb{N} \to I$ and the limits coincide.*

Proof. In the countably infinite case, consider two bijections $h, h'\colon \mathbb{N} \to I$, set $b_j = f(h(j))$ and $b'_j = f(h'(j))$, and consider the two series $\sum b_j$ and $\sum b'_j$. Suppose that $\sum b_j$ is absolutely convergent with limit L. We want to show $\sum b'_j$ is also absolutely convergent with limit L.

Let $g = h^{-1} \circ h'\colon \mathbb{N} \to \mathbb{N}$. Then g is a bijection and $b'_j = f(h'(j)) = f(h(g(j))) = b_{g(j)}$, so that $\sum b'_j$ is a rearrangement of $\sum b_j$, and also $\sum |b'_j|$ is a rearrangement of $\sum |b_j|$. Applying Theorem C.2 first to $\sum |b_j|$, we find that $\sum |b'_j|$ also converges, i.e. $\sum b'_j$ converges absolutely. Applying Theorem C.2 to $\sum b_j$, we conclude that $\sum b_j$ and $\sum b'_j$ have the same limit.

The case of finite I with $n \in \mathbb{N}$ elements is more elementary, but also follows as above if we start from $h, h'\colon \{j \in \mathbb{N}\colon 1 \leq j \leq n\} \to I$, and set $b_j = b'_j = 0$ and $g(j) = j$ for $j \geq n+1$.

The case $f \geq 0$ with infinite countable sum follows similarly from Theorem C.4. □

Proposition C.6 provides flexibility to add the terms in a countable sum in any order. However, we need to take a further step to handle double and higher order multiple countable sums, which by definition, may involve multiple limits, but are covered by natural extensions of the Algebra of Limits, as in the following two results.

Proposition C.7. *Consider $U \times V$ for countable sets U and V, and let $f: U \times V \to \mathbb{R}$ be a function. Then*

$$\sum_{u\in U}\sum_{v\in V}|f(u,v)| = \sum_{(u,v)\in U\times V}|f(u,v)| = \sum_{v\in V}\sum_{u\in U}|f(u,v)|,$$

with the convention that a double countable sum diverges to infinity whenever any of its constituent countable sums (such as $\sum_{v\in V}|f(u,v)|$ for some $u \in U$) diverges to infinity. Furthermore, if any of the three terms is finite (and hence all are finite), then we also have

$$\sum_{u\in U}\sum_{v\in V}f(u,v) = \sum_{(u,v)\in U\times V}f(u,v) = \sum_{v\in V}\sum_{u\in U}f(u,v).$$

Proof. This is elementary when U and V are finite. More generally, this is an instance of the theorems of Fubini and Tonelli in measure theory. Other instances are in Part A Integration and in Part B Probability, Measure and Martingales (under further technical assumptions).

In our context, let us first suppose that $\sum_{(u,v)\in U\times V}|f(u,v)|$ is infinite and $M \in \mathbb{R}$. Let $I = U \times V$, and fix a bijection $h: \mathbb{N} \to I$ as in Definition C.5. Let $b_j = |f(h(j))|$ and consider the partitial sums s_k. Then there is $N \in \mathbb{N}$ such that $s_k \geq s_N \geq M$ for all $k \geq N$. Moreover, there are finite $U_0 \subseteq U$ and $V_0 \subseteq V$ such that $h(\{j \in \mathbb{N}: 1 \leq j \leq N\}) \subseteq U_0 \times V_0$. But then we have

$$\sum_{u\in U}\sum_{v\in V}|f(u,v)| \geq \sum_{u\in U_0}\sum_{v\in V_0}|f(u,v)| \geq s_N \geq M,$$

and similarly $\sum_{v\in V}\sum_{u\in U}|f(u,v)| \geq M$, as required.

Now suppose $\overline{L} = \sum_{u\in U}\sum_{v\in V}|f(u,v)| < \infty$. This means $\overline{L}_u = \sum_{v\in V}|f(u,v)| < \infty$ for all $u \in U$ and $\overline{L} = \sum_{u\in U}\overline{L}_u < \infty$. Let $\varepsilon > 0$. Then there are finite $U_1 \subseteq U$ and $V_1 \subseteq V$ with

$$\sum_{u\in U\setminus U_1}\overline{L}_u = \overline{L} - \sum_{u\in U_1}\overline{L}_u < \frac{\varepsilon}{4} \quad\text{and}\quad \sum_{v\in V\setminus V_1}|f(u,v)| = \overline{L}_u - \sum_{v\in V_1}|f(u,v)| < \frac{\varepsilon}{4|U_1|},\ u \in U_1.$$

By the equality $\sum_{u\in U_1}\sum_{v\in V_1}|f(u,v)| = \sum_{v\in V_1}\sum_{u\in U_1}|f(u,v)| = \sum_{(u,v)\in U_1\times V_1}|f(u,v)|$ of finite sums, we have for all finite $I_1 \subset U \times V$ with $U_1 \times V_1 \subseteq I_1$,

$$0 \leq \overline{L} - \sum_{(u,v)\in I_1}|f(u,v)| \leq \left(\overline{L} - \sum_{u\in U_1}\overline{L}_u\right) + \sum_{u\in U_1}\left(\overline{L}_u - \sum_{v\in V_1}|f(u,v)|\right) < \frac{\varepsilon}{2}.$$

For $L_u = \sum_{v\in V}f(u,v)$, $u \in U$, and $L = \sum_{u\in U}L_u = \sum_{u\in U}\sum_{v\in V}f(u,v)$, we similarly get

$$\begin{aligned}\left|L - \sum_{(u,v)\in I_1}f(u,v)\right| &\leq \left|L - \sum_{u\in U_1}L_u\right| + \sum_{u\in U_1}\left|L_u - \sum_{v\in V_1}f(u,v)\right| + \sum_{(u,v)\in I_1\setminus(U_1\times V_1)}|f(u,v)| \\ &\leq \frac{\varepsilon}{4} + \sum_{u\in U_1}\frac{\varepsilon}{4|U_1|} + \sum_{u\in U\setminus U_1}\overline{L}_u + \sum_{u\in U_1}\sum_{v\in V\setminus V_1}|f(u,v)| < \varepsilon.\end{aligned}$$

For any bijection $h: \mathbb{N} \to U \times V$, there is $N \in \mathbb{N}$ such that $U_1 \times V_1 \subseteq h(\{j \in \mathbb{N}: 1 \leq j \leq N\})$. Then for all $n \geq N$ and $I_1 = h(\{j \in \mathbb{N}: 1 \leq j \leq n\})$, we have

$$0 \leq \overline{L} - \sum_{j=1}^{n}|f(h(j))| = \overline{L} - \sum_{(u,v)\in I_1}|f(u,v)| < \frac{\varepsilon}{2} \quad\text{and}\quad \left|L - \sum_{j=1}^{n}f(h(j))\right| < \varepsilon,$$

as required for the left-hand equalities. The others follow similarly, with roles of u and v swapped. □

Corollary C.8. *Let Q and I_q, $q \in Q$, be countable sets that are pairwise disjoint, and let $I = \bigcup_{q\in Q} I_q$. Consider a function $g\colon I \to \mathbb{R}$ such that the countable sum $\sum_{i\in I} g(i)$ is defined in the sense of Definition C.5. Then the countable sums $\sum_{i\in I_q} g(i)$ and $\sum_{q\in Q}\sum_{i\in I_q} g(i)$ are also defined. Furthermore,*

$$\sum_{i\in I} g(i) = \sum_{q\in Q}\sum_{i\in I_q} g(i).$$

Proof. This follows from Proposition C.7 if we set $U = Q$ and $V = I$, setting $f(q,i) = g(i)$ if $i \in I_q$ and $f(q,i) = 0$ if $i \in I \setminus I_q$. Specifically, we observe that

$$\sum_{i\in I} g(i) = \sum_{(q,i)\in Q\times I} f(q,i) = \sum_{q\in Q}\sum_{i\in I} f(q,i) = \sum_{q\in Q}\sum_{i\in I_q} g(i).$$

□

C.2 Some integrals

Unfortunately, the theory of Riemann integration as covered in the Prelims Analysis course is not sufficient to justify in any reasonable generality the change of the order of integration in a way analogous to the change of the order of summation in Proposition C.7. This is one of the two main obstacles to making the Prelims Probability course rigorous, which we will postpone until tools from Part A Integration are available. In this section we will merely establish some integrals that are relevant for Prelims Probability.

Proposition C.9. *Let $\alpha > 0$. Consider the function $f_\alpha\colon (0,\infty) \to \mathbb{R}$ given by $f_\alpha(u) = u^{\alpha-1}e^{-u}$. Then the improper Riemann integral $\Gamma(\alpha) := \int_0^\infty f_\alpha(u)du$ exists.*

Proof. As a product of continuous functions, f_α is continuous, hence Riemann integrable on $[\varepsilon, R]$ for all $0 < \varepsilon < R < \infty$. Both limits, as $\varepsilon \to 0$ and as $R \to \infty$ are increasing limits and hence either exist or diverge to infinity. First consider $R = 1$ and $\varepsilon \to 0$. Then $f_\alpha(u) \le u^{\alpha-1} =: g_\alpha(u)$ and since $G_\alpha(u) = u^\alpha/\alpha$ is continuously differentiable on $[\varepsilon, 1]$ with derivative g_α, order properties of integrals and the fundamental theorem of calculus yield

$$\int_\varepsilon^1 f_\alpha(u)du \le \int_\varepsilon^1 g_\alpha(u)du = G_\alpha(1) - G_\alpha(\varepsilon) = \frac{1}{\alpha} - \frac{1}{\alpha}\varepsilon^\alpha \to \frac{1}{\alpha}$$

as $\varepsilon \to 0$. Now consider $\varepsilon = 1$ and $R \to \infty$. Then let $m = \lceil \alpha - 1 \rceil$. Then dropping all but one (positive) terms from a power series yields

$$2^m m! e^{u/2} = 2^m m! \sum_{k=0}^\infty \frac{(u/2)^k}{k!} \ge u^m$$

and hence, for all $u \ge 1$,

$$f_\alpha(u) \le u^m e^{-u} \le 2^m m! e^{u/2} e^{-u} = 2^m m! e^{-u/2} =: h_\alpha(u).$$

Since $H_\alpha(u) = -2^{m+1} m! e^{-u/2}$ has derivative h_α, we similarly integrate

$$\int_1^R f_\alpha(u)du \le \int_1^R h_\alpha(u)du = H_\alpha(R) - H_\alpha(1) = 2^{m+1}m!\left(e^{-1/2} - e^{-R/2}\right) \to 2^{m+1}m!e^{-1/2}.$$

Hence, the increasing limits defining the improper Riemann integrals are bounded and hence remain finite. □

Corollary C.10. *The improper Riemann integral* $\int_{-\infty}^{\infty} e^{-x^2/2} dx$ *converges to* $\sqrt{2}\Gamma\left(\frac{1}{2}\right)$.

Proof. First note that for all $0 < \varepsilon < R < \infty$, the Substitution rule (Proposition B.57) applies with the continuous function $f(x) = e^{-x^2/2}$ on $[\varepsilon, R]$ and the differentiable function $g(u) = \sqrt{2u}$ on $[\varepsilon^2/2, R^2/2]$, giving

$$\int_{\varepsilon}^{R} e^{-x^2/2} dx = \int_{\varepsilon^2/2}^{R^2/2} \frac{1}{\sqrt{2u}} e^{-u} du.$$

By the preceding proposition, the limits as $\varepsilon \to 0$ and $R \to \infty$ exist and give the improper Riemann integral

$$\int_{0}^{\infty} e^{-x^2/2} dx = \frac{1}{\sqrt{2}} \Gamma\left(\frac{1}{2}\right).$$

By symmetry, we also have $\int_{-\infty}^{0} e^{-x^2/2} dx = \Gamma(\frac{1}{2})/\sqrt{2}$ and their sum is as claimed. □

Remark C.11. Of course, the integral $\int_{-\infty}^{\infty} e^{-x^2/2} dx$ can also be shown to converge to $\sqrt{2\pi}$, so that the corollary yields $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. However, in Prelims Probability, a proof based on changing to polar coordinates in a two-dimensional integral was given. Since this (as well as various similar proofs based on two-dimensional changes of variables) is beyond Prelims Analysis, we postpone a rigorous proof until the tools of Part A Integration are available, and then we first show that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ to deduce that $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$.

D

Prelims Probability: Definitions and theorems

This appendix contains all definitions and essentially all theorems of the Prelims Probability course, as a reminder of where we got to regarding theory. Most of the course was taught before relevant concepts from Analysis were available. This included a few informal proofs and pragmatic use of series and integrals, which we reassess now by revisiting those points, in remarks and proofs. We include references to Appendices B and C, as appropriate, and we explicitly postpone to Part A Integration any points that still remain unresolved.

D.1 Events and probability

Definition D.1 (The axioms of probability)**.** A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where

1. Ω is a set, called the *sample space*,
2. $\mathcal{F}$ is a collection of subsets of Ω, called *events*, satisfying the *σ-algebra* axioms

   $\mathbf{F}_1$: $\Omega \in \mathcal{F}$.

   $\mathbf{F}_2$: If $A \in \mathcal{F}$, then also $A^c \in \mathcal{F}$.

   $\mathbf{F}_3$: If $\{A_i, i \in I\}$ is a countable collection of members of $\mathcal{F}$, then $\bigcup_{i \in I} A_i \in \mathcal{F}$.
3. $\mathbb{P}$ is a probability measure, which is a function $\mathbb{P} \colon \mathcal{F} \to \mathbb{R}$ satisfying axioms

   $\mathbf{P}_1$: For all $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$.

   $\mathbf{P}_2$: $\mathbb{P}(\Omega) = 1$.

   $\mathbf{P}_3$: If $\{A_i, i \in I\}$ is a countable collection of members of $\mathcal{F}$, and $A_i \cap A_j = \emptyset$ for $i \neq j$, then $\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i)$.

Remark D.2. In $\mathbf{P}_3$, the index set I can be finite and the sum $\sum_{i \in I} \mathbb{P}(A_i)$ is a finite sum, or $I = \mathbb{N}$ and the sum $\sum_{i \in I} \mathbb{P}(A_i)$ is a convergent series of non-negative real numbers, or I is a more general countable set and $\sum_{i \in I} \mathbb{P}(A_i)$ is a countable sum in the sense of Definition C.5.

Remark D.3. If Ω is finite or countable, we often take $\mathcal{F}$ to be the set of all subsets of Ω (the *power set* of Ω). If Ω is uncountable, however, the set of all subsets typically turns out to be *too large*: it ends up containing sets to which we cannot consistently assign probabilities. This issue is discussed properly in Part A Integration.

Theorem D.4. *Suppose that* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and that* $A, B \in \mathcal{F}$. *Then*

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
2. *If* $A \subseteq B$ *then* $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Definition D.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $A, B \in \mathcal{F}$ and $\mathbb{P}(B) > 0$ then the *conditional probability of* A *given* B *is*

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

(If $\mathbb{P}(B) = 0$, *then* $\mathbb{P}(A|B)$ *is not defined.)*

Theorem D.6. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $B \in \mathcal{F}$ *satisfy* $\mathbb{P}(B) > 0$. *Define a new function* $\mathbb{Q}\colon \mathcal{F} \to \mathbb{R}$ *by* $\mathbb{Q}(A) = \mathbb{P}(A|B)$. *Then* $(\Omega, \mathcal{F}, \mathbb{Q})$ *is also a probability space.*

From now on, we always suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition D.7. A countable family of events $\{B_i, i \in I\}$ is a *partition of* Ω if

1. $\Omega = \bigcup_{i \in I} B_i$, and
2. $B_i \cap B_j = \emptyset$ whenever $i \neq j$.

Theorem D.8 (The law of total probability or partition theorem)**.** *Suppose* $\{B_i, i \in I\}$ *is a partition of* Ω *with* $B_i \in \mathcal{F}$ *and* $\mathbb{P}(B_i) > 0$ *for all* $i \in I$. *Then for any* $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Remark D.9. This slightly extends the cases $I = \mathbb{N}$ and finite I from Prelims Probability.

Definition D.10. 1. Two events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

2. More generally, a family of events $\{A_i, i \in I\}$ is *independent* if

$$\mathbb{P}\left(\bigcup_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \qquad \text{for all finite subsets } J \text{ of } I.$$

3. The family $\{A_i, i \in I\}$ is *pairwise independent* if $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ whenever $i \neq j$.

Theorem D.11. *Suppose that* A *and* B *are independent events. Then* A *and* B^c *are independent.*

D.2 Discrete random variables

Definition D.12. A *discrete random variable* X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X\colon \Omega \to \mathbb{R}$ such that

(a) $\{X = x\} := X^{-1}(\{x\}) = \{\omega \in \Omega\colon X(\omega) = x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$,

(b) $\mathrm{Im}X := \{X(\omega)\colon \omega \in \Omega\}$ is a countable subset of $\mathbb{R}$.

Definition D.13. The *probability mass function* (p.m.f.) of X is the function $p_X : \mathbb{R} \to [0,1]$ defined by

$$p_X(x) = \mathbb{P}(X = x) := \mathbb{P}(\{X = x\}).$$

Remark D.14. In fact, given any function $p\colon \mathbb{R} \to \mathbb{R}$ that satisfies $p(x) \geq 0$ for all $x \in \mathbb{R}$ and $\sum_{x\in I} p(x) = 1$ where $I = \{x \in \mathbb{R}\colon p(x) > 0\}$, we can write down a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X defined on it whose probability mass function is p. Most directly, we could take $\Omega = I$, take $\mathcal{F}$ to consist of all subsets of Ω, define

$$\mathbb{P}(\{\omega\}) = p(\omega) \quad \text{for each } \omega \in \Omega$$

and more generally

$$\mathbb{P}(S) = \sum_{\omega\in S} p(\omega) \quad \text{for each } S \subseteq \Omega,$$

and then take X to be the identity function, i.e. $X(\omega) = \omega$.

Remark D.15. We have been quite explicit in describing our sample space Ω. This can quickly become impractical. Although the concept of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ underlies everything, in practice it will be rare that we think about Ω itself – instead we will talk directly about *events* and their probabilities, and *random variables* and their distributions (and we can do that without assuming any particular structure for Ω).

From a theoretical perspective, two points are important. First we need to ensure that suitable probability spaces exist for all our purposes. Second we need to ensure that the subsets of Ω that we consider are indeed events in $\mathcal{F}$. For a single discrete random variable, the first is covered by the preceding remark. The second will hold on any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where a discrete random variable X with given p.m.f. can be defined, since the only events relevant for X are countable unions (in $\mathcal{F}$ by $\mathbf{F}_3$!) of events of the form $\{X = x\} \in \mathcal{F}$.

Definition D.16. The *expectation* (or *expected value* or *mean*) of X is

$$\mathbb{E}[X] = \sum_{x\in \mathrm{Im}X} x\mathbb{P}(X = x)$$

provided that $\sum_{x\in \mathrm{Im}X} |x|\mathbb{P}(X = x)$ converges. If $\sum_{x\in \mathrm{Im}X} |x|\mathbb{P}(X = x)$ diverges, we say that the *expectation does not exist* (or if only the series of positive terms diverges, we may say that $\mathbb{E}[X] = \infty$). Often, $\mathrm{Im}X \subseteq \mathbb{N} \cup \{0\}$ and the sums are finite sums or (convergent) series. In general, these sums are countable sums in the sense of Definition C.5.

Theorem D.17. *If* $h\colon \mathbb{R} \to \mathbb{R}$*, then*

$$\mathbb{E}[h(X)] = \sum_{x\in \mathrm{Im}X} h(x)\mathbb{P}(X = x)$$

provided that $\mathbb{E}[h(X)] = \sum_{x\in \mathrm{Im}X} |h(x)|\mathbb{P}(X = x)$ *converges.*

Proof. We split $I = \mathrm{Im}X$ into subsets $I_y = \{x \in \mathrm{Im}X\colon h(x) = y\}$ according to the values $y \in Q = h(\mathrm{Im}X) = \mathrm{Im}Y$ that h and hence $Y = h(X)$ can take. Then Corollary C.8 yields

$$\begin{aligned}\sum_{x\in \mathrm{Im}X} h(x)\mathbb{P}(X = x) &= \sum_{y\in Q}\sum_{x\in I_y} h(x)\mathbb{P}(X = x) = \sum_{y\in Q}\sum_{x\in \mathrm{Im}X\colon h(x)=y} y\mathbb{P}(X = x)\\ &= \sum_{y\in Q} y \sum_{x\in \mathrm{Im}X\colon h(x)=y} \mathbb{P}(X = x) = \sum_{y\in \mathrm{Im}Y} y\mathbb{P}(h(X) = y) = \mathbb{E}[h(X)]. \qquad \square\end{aligned}$$

Example D.18. Take $h(x) = x^k$. Then $\mathbb{E}[X^k]$ is called the *kth moment of X*, when it exists.

Theorem D.19. *Let X be a discrete random variable such that $\mathbb{E}[X]$ exists.*

(a) If X is non-negative then $\mathbb{E}[X] \geq 0$.

(b) If $a, b \in \mathbb{R}$ then $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

Definition D.20. For a discrete random variable X, the *variance* of X is defined by

$$\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

provided that this quantity exists in the sense of Definition D.16 applied to random variables X and $(X - \mathbb{E}[X])^2$, or, for the latter, equivalently in the sense of Theorem D.17 applied with $h(x) = (x - \mathbb{E}[X])^2$.

Theorem D.21. *For a discrete random variable X whose variance exists,*

$$\mathrm{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Theorem D.22. *Suppose that X is a discrete random variable whose variance exists. Then if a and b are (finite) fixed real numbers, then the variance of the discrete random variable $Y = aX + b$ is given by*

$$\mathrm{var}(Y) = \mathrm{var}(aX + b) = a^2 \mathrm{var}(X).$$

Definition D.23. Suppose that B is an event such that $\mathbb{P}(B) > 0$. Then the *conditional p.m.f. of X given B* is

$$p_{X|B}(x) = \mathbb{P}(X = x|B) = \frac{\mathbb{P}(\{X = x\} \cap B)}{\mathbb{P}(B)}, \qquad x \in \mathbb{R}.$$

The *conditional expectation of X given B* is

$$\mathbb{E}[X|B] = \sum_{x \in \mathrm{Im}X} x\mathbb{P}(X = x|B),$$

whenever the sum converges absolutely.

Theorem D.24 (The law of total probability or partition theorem for expectations)**.** *If $\{B_i, i \in I\}$ is a partition of Ω with $B_i \in \mathcal{F}$ and $\mathbb{P}(B_i) > 0$ for all $i \in I$ then*

$$\mathbb{E}[X] = \sum_{i \in I} \mathbb{E}[X|B_i]\mathbb{P}(B_i)$$

whenever $\mathbb{E}[X]$ exists.

Proof. By the law of total probability (Theorem D.8) and interchanging sums (Proposition C.7),

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in \mathrm{Im}X} x\mathbb{P}(X = x) = \sum_{x \in \mathrm{Im}X} x \left(\sum_{i \in I} \mathbb{P}(X = x|B_i)\mathbb{P}(B_i) \right) \\
&= \sum_{x \in \mathrm{Im}X} \sum_{i \in I} x\mathbb{P}(X = x|B_i)\mathbb{P}(B_i) = \sum_{i \in I} \sum_{x \in \mathrm{Im}X} x\mathbb{P}(X = x|B_i)\mathbb{P}(B_i) \\
&= \sum_{i \in I} \mathbb{P}(B_i) \left(\sum_{x \in \mathrm{Im}X} x\mathbb{P}(X = x|B_i) \right) = \sum_{i \in I} \mathbb{E}[X|B_i]\mathbb{P}(B_i).
\end{aligned}$$

□

Definition D.25. Given two discrete random variables X and Y their *joint p.m.f.* is

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) := \mathbb{P}(\{X = x\} \cap \{Y = y\}), \qquad x, y \in \mathbb{R}.$$

Theorem D.26. *Given two discrete random variables X and Y with joint p.m.f. $p_{X,Y}$,*

$$p_X(x) = \sum_{y \in \mathrm{Im} Y} p_{X,Y}(x,y), \quad x \in \mathbb{R}, \quad \text{and} \quad p_Y(y) = \sum_{x \in \mathrm{Im} X} p_{X,Y}(x,y), \quad y \in \mathbb{R}.$$

Definition D.27. Discrete random variables X and Y are *independent* if

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \qquad \text{for all } x, y \in \mathbb{R}.$$

Theorem D.28. *Suppose X and Y are discrete random variables and $h\colon \mathbb{R}^2 \to \mathbb{R}$. Then $h(X,Y)$ is itself a discrete random variable, and*

$$\mathbb{E}[h(X,Y)] = \sum_{x \in \mathrm{Im} X} \sum_{y \in \mathrm{Im} Y} h(x,y)\mathbb{P}(X = x, Y = y),$$

provided that these sums converge absolutely.

Proof. We adapt the proof of Theorem D.17. Specifically, we use Proposition C.7 to combine the double sum into a single countable sum and then split the countable set $I = \mathrm{Im} X \times \mathrm{Im} Y$ into subsets $I_z = \{(x,y) \in I\colon h(x,y) = z\}$, $z \in \mathrm{Im} Z := h(X,Y)$ to apply Corollary C.8 and proceed as in the proof of Theorem D.17. □

Theorem D.29 (Linearity of expectation)**.** *Suppose X and Y are discrete random variables and $a, b \in \mathbb{R}$ are constants. Then*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

provided that both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ exist.

Proof. Setting $h(x,y) = ax+by$ and applying Algebra of Limits, Proposition C.7 and Theorem D.26, we have

$$\begin{aligned}
\mathbb{E}[aX + bY] = \mathbb{E}[h(X,Y)] &= \sum_{x \in \mathrm{Im} X} \sum_{y \in \mathrm{Im} Y} (ax + by) p_{X,Y}(x,y) \\
&= a \sum_{x \in \mathrm{Im} X} \sum_{y \in \mathrm{Im} Y} x p_{X,Y}(x,y) + b \sum_{x \in \mathrm{Im} X} \sum_{y \in \mathrm{Im} Y} y p_{X,Y}(x,y) \\
&= a \sum_{x \in \mathrm{Im} X} x \left(\sum_{y \in \mathrm{Im} Y} x p_{X,Y}(x,y) \right) + b \sum_{y \in \mathrm{Im} Y} y \left(\sum_{x \in \mathrm{Im} X} p_{X,Y}(x,y) \right) \\
&= a \sum_{x \in \mathrm{Im} X} x p_X(x) + b \sum_{y \in \mathrm{Im} Y} y p_Y(y) = a\mathbb{E}[X] + b\mathbb{E}[Y].
\end{aligned}$$

□

Theorem D.30. *If X and Y are independent discrete random variables whose expectations exist, then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Definition D.31. For discrete random variables X and Y, their *covariance* is defined by

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

provided that this quantity exists.

Theorem D.32. *For discrete random variables X and Y we have*

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

provided that these expectations exist. If X and Y have variances $\text{var}(X)$ *and* $\text{var}(Y)$ *then*

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y).$$

Furthermore, $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$ *if X and Y are independent.*

Remark D.33. Independence implies that $\text{cov}(X, Y) = 0$, but the reverse implication fails in general.

Definition D.34. We can define the *multivariate p.m.f.* of discrete random variables X_i, $1 \leq i \leq n$, as

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n), \qquad x_1, \ldots, x_n \in \mathbb{R}.$$

Definition D.35. A family $\{X_i, i \in I\}$ of discrete random variables is independent if for all *finite* sets $J \subseteq I$ and all collections $\{A_i, i \in J\}$ of subsets of $\mathbb{R}$,

$$\mathbb{P}\left(\bigcap_{i \in J} \{X_i \in A_i\}\right) = \prod_{i \in J} \mathbb{P}(X_i \in A_i).$$

Remark D.36. Remark D.14 generalises to the setting of multivariate p.m.f.s of any *finite* number of discrete random variables. Specifically, for any given $p \to \mathbb{R}^n \to \mathbb{R}$ that satisfies $p(x_1, \ldots, x_n) \geq 0$ for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\sum_{(x_1,\ldots,x_n) \in I} p(x_1, \ldots, x_n) = 1$ where $I = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : p(x_1, \ldots, x_n) > 0\}$, we take $(\Omega, \mathcal{F}, \mathbb{P})$ as in Remark D.14, just using the new p and I. Then each $\omega \in \Omega \subset \mathbb{R}^n$ is of the form $\omega = (\omega_1, \ldots, \omega_n)$ and we take X_i to be the coordinate maps $X_i(\omega) = \omega_i$, $1 \leq i \leq n$. Then $X_1, \ldots, X_n$ are random variables with $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(\{\omega \in \Omega : \omega = (x_1, \ldots, x_n)) = p(x_1, \ldots, x_n)$.

This construction does *not* extend to infinite numbers of discrete random variables, in general, since a countable set Ω may not suffice. Even in one of the simplest examples of a sequence of independent ($\{0, 1\}$-valued) Bernoulli variables X_i, $i \geq 1$, with $p_{X_i}(0) = p_{X_i}(1) = 1/2$ for all $i \geq 1$, there are uncountably many $\{0, 1\}$-valued sequences (x_i), none of which has positive probability in the sense that we must require (due to part 2. of Theorem D.4) that for all $n \geq 1$

$$\mathbb{P}(X_i = x_i \text{ for all } i \geq 1) \leq \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) = \left(\frac{1}{2}\right)^n.$$

Indeed, the existence of a suitable probability space is beyond the scope of Prelims. We postpone it until the tools of Part A Integration are available. We will, however, assume the existence of a suitable probability space that allows us to consider a sequence of independent discrete random variables with prescribed distributions. The axioms and the theory we have built give us sufficient tools to establish results about sequences of independent random variables (subject only to the existence of a suitable probability space).

D.3 Probability generating functions

Definition D.37. Let X be a non-negative integer-valued random variable. Let

$$\mathcal{S} := \left\{ s \in \mathbb{R} \colon \sum_{k=0}^{\infty} |s|^k \mathbb{P}(X = k) < \infty \right\}.$$

Then the *probability generating function* (p.g.f.) of X is $G_X \colon \mathcal{S} \to \mathbb{R}$ defined by

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k).$$

Theorem D.38. *The distribution of X is uniquely determined by its probability generating function G_X.*

Proof. First note that $G_X(0) = \mathbb{P}(X = 0)$. As $G_X(1) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$, Proposition B.21 entails that the power series $G_X(s)$ has radius of convergence $R \geq 1$. For $|s| < 1$, Theorem B.40 allows us to differentiate $G_X(s)$ term-by-term to find that the derivative is a power series – with the same radius of convergence $R \geq 1$ – given by

$$G'_X(s) = \sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}(X = k).$$

Inductively, we see that the nth derivative exists also for all $n \geq 2$ and is given by

$$G_X^{(n)}(s) = \sum_{k=n}^{\infty} k(k-1) \cdots (k-n+1) s^{k-n} \mathbb{P}(X = k).$$

In particular, $G_X^{(n)}(0) = n!\mathbb{P}(X = n)$ for all $n \geq 0$. So, we can recover the p.m.f. and hence the distribution of X from G_X. □

Theorem D.39. *If X and Y are independent non-negative integer-valued random variables, then $G_{X+Y}(s) = G_X(s)G_Y(s)$.*

Theorem D.40. *If X is a non-negative integer-valued random variable for which $\mathbb{E}[X]$ exists, then $G'_X(1) = \mathbb{E}[X]$.*

More generally, if $\mathbb{E}[X^n] < \infty$ for some $n \geq 1$, then $G_X^{(n)}(1) = \mathbb{E}[X(X-1) \cdots (X-n+1)]$.

In particular, for $n = 2$, we have $\operatorname{var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$.

Proof. If $G_X(s)$ has a radius of convergence $R > 1$, this follows from the derivatives calculated in the proof of Theorem D.38, since

$$\begin{aligned} \sum_{k=n}^{\infty} k(k-1) \cdots (k-n+1) s^{k-n} \mathbb{P}(X = k) \bigg|_{s=1} &= \sum_{k=n}^{\infty} k(k-1) \cdots (k-n+1) \mathbb{P}(X = k) \\ &= \mathbb{E}[X(X-1) \cdots (X-n+1)] \end{aligned}$$

is finite as a derivative of a power series within its radius of convergence as in Theorem B.40. In particular, we easily obtain $G'_X(1) = \mathbb{E}[X]$ and $G''_X(1) = \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X] = \mathbb{E}[X^2] - G'_X(1)$ by linearity of expectation (Theorem D.29). Hence

$$\operatorname{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

In the case $R = 1$, the derivative $G'_X(1)$ cannot exist in the strict sense of Definition B.37(1), but may exist as a left-derivative $G'_X(1-)$ in the weaker sense of Definition B.37(2). However, if $\mathbb{E}[X] = \sum_{k=1}^{\infty} k\mathbb{P}(X = k)$ exists, Abel's theorem (Theorem B.36) yields that this still equals the left limit $\lim_{s\uparrow 1} G'_X(s)$, and applying l'Hôpital's rule (Theorem B.41) with $f(x) = G_X(x) - G_X(1)$ and $g(x) = x$ as $x \uparrow a = 1$, this also equals the left-derivative $G'_X(1-)$.

The formulas for higher derivatives in the case $R = 1$ follow similarly, noting that the existence of $\mathbb{E}[X^n] = \sum_{k=1}^{\infty} k^n\mathbb{P}(X = k)$ entails the existence of all lower moments $\mathbb{E}[X^m]$ for $m \leq n$ by the comparison test (Theorem B.16) since $0 \leq b_k := k^m \leq k^n =: c_k$ for all $k \geq 1$. Hence $\mathbb{E}[X(X-1)\cdots(X-n+1)]$ exists and the previous reasoning for $n = 1$, based on Abel's theorem and l'Hôpital's rule, applies inductively. □

Theorem D.41. *Let N and X_n, $n \geq 1$, be jointly independent non-negative integer-valued random variables. Denote by G_N the p.g.f. of N and suppose that X_n, $n \geq 1$, are identically distributed with p.g.f. G_X. Then the p.g.f. of $\sum_{i=1}^{N} X_i$ is given by $G_N \circ G_X$.*

D.4 Continuous random variables

Definition D.42. A *random variable* X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$.

Definition D.43. The *cumulative distribution function (c.d.f.)* of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by $F_X(x) = \mathbb{P}(X \leq x)$.

Theorem D.44. *1. F_X is non-decreasing.*

2. $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$ for $a < b$.

3. As $x \to -\infty$, $F_X(x) \to 0$.

4. As $x \to \infty$, $F_X(x) \to 1$.

Remark D.45. Conversely, any function F satisfying conditions 1, 3 and 4 of Theorem D.44 plus *right-continuity* (Definition B.29) is the c.d.f. of *some* random variable defined on *some* probability space, although a full proof of this fact is beyond the tools available from Prelims.

If F is a step function, we can show that there is a discrete random variable with c.d.f. F. More generally (cf. Remark D.14), if F has a countable set x_q, $q \in Q$, of jumps of sizes $p(x_q) = F(x_q) - F(x_q-)$ that sum to $\sum_{q\in Q} p(x_q) = 1$, then as F is non-decreasing,

$$F(x) \geq \sup_{J\subseteq Q \text{ finite}} \sum_{q\in J:\, x_q\leq x} (F(x_q) - F(x_q-)) = \sum_{q\in Q:\, x_q\leq x} p(x_q) \text{ and } 1 - F(x) \geq \sum_{q\in Q:\, x_q>x} p(x_q).$$

As both sides sum to 1, equalities hold. Let $\Omega = \{x_q, q \in Q\} \subset \mathbb{R}$ with the power set $\mathcal{F}$ and probability measure $\mathbb{P}(S) = \sum_{x\in S} p(x)$, $S \subseteq \Omega$, and define $X : \Omega \to \mathbb{R}$ by $X(\omega) = \omega$. Then

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\bigcup_{q\in Q:\, x_q\leq x} \{x_q\}\right) = \sum_{q\in Q:\, x_q\leq x} \mathbb{P}(\{x_q\}) = F(x), \quad x \in \mathbb{R}.$$

This is, of course, the same $(\Omega, \mathcal{F}, \mathbb{P})$ in the same generality as in Remark D.14, but we did start from F rather than p. In any case, an explicit probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is usually neither necessary nor helpful once its existence has been proved (or accepted as a fact).

Theorem D.46. *If A_n, $n \geq 1$, is an increasing sequence of events in the sense that $A_n \subseteq A_{n+1}$ for all $n \geq 1$, then*

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Proof of Theorem D.44. Part 4. Let (x_n) be any increasing sequence that diverges to infinity. Let $A_n = \{X \leq x_n\}$. Then $\bigcup_{n\in\mathbb{N}} A_n = \Omega$ since X is $\mathbb{R}$-valued, so for all $\omega \in \Omega$, there is $n \in \mathbb{N}$ such that $x_n \geq X(\omega)$ and hence $\omega \in A_n$. By Theorem D.46,

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

Part 3. can be proved in the same way by showing that $1 - F_X(x) \to 1$ as $x \to -\infty$. □

Definition D.47. A *continuous random variable* X is a random variable whose c.d.f. satisfies

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(u)du,$$

where $f_X : \mathbb{R} \to \mathbb{R}$ is a function such that

(a) $f_X(u) \geq 0$ for all $u \in \mathbb{R}$,

(b) the (improper Riemann) integral $\int_{-\infty}^{\infty} f_X(u)du$ exists and equals 1.

Then f_X is called *probability density function (p.d.f.)* of X or, sometimes, just its *density.*

Remark D.48. In the setting of Definition D.47, the fundamental theorem of calculus (Theorem B.55 in conjuction with Proposition B.59 to reduce the problem to proper Riemann integrals) yields that F_X is differentiable with

$$F_X'(x) = f_X(x)$$

at any point $x \in \mathbb{R}$ where f_X is continuous.

The name "continuous random variable" is not a good name since $X : \Omega \to \mathbb{R}$ need not be continuous (indeed Ω might not be a set on which continuity has a meaning), and neither does $f_X : \mathbb{R} \to \mathbb{R}$, while the continuity of F_X is not sufficient for X to be a continuous random variable. Rather, "continuous random variable" can be seen as an abbreviation of "random variable with an absolutely continuous distribution." Indeed, the notion of a continuous random variable can be generalised by replacing the Riemann integrals by Lebesgue integrals as developed in Part A Integration, and "absolute continuity" refers to a certain relationship between a suitable "probability measure" $A \mapsto \mathbb{P}(X \in A)$ and Lebesgue measure on $\mathbb{R}$.

Theorem D.49. *If X is a continuous random variable with p.d.f. f_X then*

$$\mathbb{P}(X = x) = 0 \quad \textit{for all } x \in \mathbb{R}$$

and

$$\mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f_X(x)dx.$$

Definition D.50. Let X be a continuous random variable with p.d.f. f_X. The *expectation* or *mean* of X is defined to be

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

whenever $\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$. Otherwise, we say that the mean is undefined (or as in the discrete case, if only the positive part diverges, we might say that $\mathbb{E}[X] = \infty$).

Theorem D.51. *Let X be a non-negative continuous random variable with finite expectation $\mathbb{E}[X]$ and with p.d.f. f_X. Then*

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > x)dx.$$

Remark D.52. An outline proof was given in Prelims Probability, based on changing the order of integration. This tool is developed rigorously in Part A Integration.

Remark D.53. The conclusions of Theorem E.1 in the non-negative case and of Corollary E.2 in the general case also hold for discrete random variables. Furthermore, the formulas could be used to define $\mathbb{E}[X]$ of any random variable $X : \Omega \to \mathbb{R}$, provided that the two improper Riemann integrals exist. By Theorem B.53, this includes all bounded random variables, and in the unbounded case, the limits defining the improper Riemann integrals are increasing limits (and if only the first integral diverges, we might say that $\mathbb{E}[X] = \infty$).

Theorem D.54. *Let X be a continuous random variable with p.d.f. f_X, and let $h : \mathbb{R} \to \mathbb{R}$ be a function for which $h(X)$ is a random variable. Then*

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)dx$$

provided that the integral $\int_{-\infty}^{\infty} |h(x)| f_X(x)dx$ exists and is finite.

Remark D.55. An outline proof was given in Prelims Probability, based on changing the order of integration and handling sets $\{x \in \mathbb{R} : h(x) > y\}$ as range of integration. This is developed rigorously in Part A Integration.

On the back of this theorem, we define $\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ and show that $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ and note that these expectations can be worked out without first calculating the p.d.f.s of the random variables $(X - \mathbb{E}[X])^2$ and X^2.

Theorem D.56. *Suppose X is a continuous random variable with p.d.f. f_X. Then if $a, b \in \mathbb{R}$, then $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ and $\text{var}(aX + b) = a^2 \text{var}(X)$.*

Theorem D.57. *Suppose that X is a continuous random variable with density f_X and that $h : \mathbb{R} \to \mathbb{R}$ is differentiable with positive derivative. Then $Y = h(X)$ is a continuous random variable with p.d.f.*

$$f_Y(y) = f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y),$$

where h^{-1} is the inverse function of h.

Definition D.58. Let X and Y be random variables such that

$$F_{X,Y}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)dudv = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)dvdu$$

for some function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ such that

(a) $f_{X,Y}(u,v) \geq 0$ for all $u, v \in \mathbb{R}$,

(b) the (repeated improper Riemann) integrals $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u,v)dudv$ exist and yield 1.

Then X and Y are *jointly continuous* and $f_{X,Y}$ is their *joint density function.*

Theorem D.59. *For a pair of jointly continuous random variables X and Y, we have*

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy = \int_a^b \int_c^d f_{X,Y}(x,y)dydx,$$

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

Theorem D.60. *Suppose X and Y are jointly continuous with joint density $f_{X,Y}$. Then X is a continuous random variable with density*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,$$

and similarly Y is a continuous random variable with density

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

In this context, the one-dimensional densities f_X and f_Y are called the *marginal densities* of the joint distribution with density $f_{X,Y}$.

Definition D.61. Jointly continuous random variables X an Y with joint density $f_{X,Y}$ are *independent* if

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

for all $x, y \in \mathbb{R}$.

Theorem D.62. *For a pair of jointly continuous random variables X and Y and a function $h\colon \mathbb{R}^2 \to \mathbb{R}$ such that $h(X,Y)$ is a random variable, we have*

$$\mathbb{E}[h(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f_{X,Y}(x,y)dydx,$$

provided that the integrals exist.

Remark D.63. The proof of Theorem D.62 is postponed until tools from Part A Integration are available, but the parallels to the discrete case are pointed out.

Proposition D.64. *For a pair of jointly continuous random variables X and Y and $a, b \in \mathbb{R}$, we have $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ and $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)$.*

A proof of Proposition D.64 based on Theorem D.62 is straightforward.

The definitions and results above generalise straightforwardly to the case of n random variables X_i, $1 \leq i \leq n$.

Remark D.65. Rules for calculating expectations and variances are the same for discrete and continuous random variables. This is not a coincidence. We can make a more general definition of expectation which covers both cases (and more besides) but in order to do so in a satisfactory way, we need a more general theory of integration, which is developed in Part A Integration and Part B Probability, Measure and Martingales. There are other (less satisfactory) approaches to expectations for general random variables, by approximation by discrete random variables, but at this stage, this notion is left as an unproven fact.

D.5 Random samples and the weak law of large numbers

Definition D.66. Let $X_1, \ldots, X_n$ denote independent identically distributed random variables. Then these random variables are said to constitute a *random sample* of size n from the distribution.

Remark D.67. In this definition, "random variable" may be understood as "either discrete or continuous random variable," and "independence" has been defined in Definitions D.35 and D.61, respectively. However, the arguments of this section make sense in higher generality subject to defining notions of independence and expectation for general random variables including those that are neither discrete nor continuous.

Definition D.68. The *sample mean* is defined to be $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Theorem D.69. *Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution with mean μ and variance σ^2. Then the expectation and variance of the sample mean are*

$$\mathbb{E}\big[\overline{X}_n\big] = \mu \qquad \text{and} \qquad \mathrm{var}\big(\overline{X}_n\big) = \frac{1}{n}\sigma^2.$$

Remark D.70. Inductive proofs of Theorem D.69 based on Proposition D.64 in the continuous case and on Theorems D.29 and D.32 in the discrete case are straightforward, also for general random variables, subject only to a notion of expectation that satisfies linearity of expectation.

Theorem D.71 (Weak law of large numbers)**.** *Suppose that X_i, $i \geq 1$, are independent and identically distributed random variables with mean μ. Then for any fixed $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right) \to 0.$$

Remark D.72. The proof provided in Prelims Probability covers only the case of finite variance. A proof of the general result is given in Part B Probability, Measure and Martingales, where the stronger "strong law of large numbers" is proved in the setting stated here. Different modes of convergence including the ones for the weak and strong laws of large numbers are introduced in Part A Probability. Generating function proofs of the weak law of large numbers are also presented, subject to proving convergence theorems for generating functions.

Theorem D.73 (Markov's inequality)**.** *Suppose that Y is a non-negative random variable whose expectation exists. Then for all $t > 0$*

$$\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t}.$$

Remark D.74. The proof provided in Prelims Probability uses the law of total probability for expectations (for finite partitions) and order properties of expectation, which have been established for discrete and continuous random variables. This proof is general, subject only to a theory of expectation that includes these two results for general random variables.

Theorem D.75 (Chebyshev's inequality)**.** *Suppose that Z is a random variable with a finite variance. Then for any $t > 0$,*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq \frac{\mathrm{var}(Z)}{t^2}.$$

E

Prelims Probability: Further developments

This appendix explores some of the less immediate points raised in remarks in Appendix D, notably the notions of expectation and independence for general random variables including order and linearity properties, a law of total probability for expecations and product form of expectations of powers of independent random variables such as they appear in the general definition of covariance, as well as in variance calculations for independent random variables.

E.1 Expectations of continuous random variables

The developments of expectations of continuous random variables in Prelims Probability were informal, using rules about interchanging the order of integration that are only available rigorously after Part A Integration. We demonstrate here that some of those developments can also be carried out using only the methods of Prelims Analysis, subject to some constraints.

Theorem E.1. *Let X be a non-negative continuous random variable with finite expectation $\mathbb{E}[X]$ and with p.d.f. f_X that is piecewise continuous with left and right limits at its discontinuities. Then*

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx.$$

Proof. First note from Remark D.48 that $\mathbb{P}(X > x) = 1 - F_X(x)$ has derivative $-f_X(x)$ provided that f_X is continuous at x. If $0 \leq c < d$ and f_X is continuous on $[c, d]$, we may integrate by parts (Theorem B.56) to find

$$\int_c^d x f_X(x)dx = -\Big[x\mathbb{P}(X > x)\Big]_c^d + \int_c^d \mathbb{P}(X > x)dx.$$

If f_X vanishes outside $[c, d]$, this completes the proof since then $\mathbb{P}(X > x) = 0$ for all $x \geq d$ and, in this case, $c\mathbb{P}(X > c) = \int_0^c \mathbb{P}(X > x)dx$.

In the general case, if f_X is continuous on (c, d) but not at c or d, we consider the modification $\widetilde{f}_X$ of f_X with $\widetilde{f}_X(c) = f_X(c+)$ and $\widetilde{f}_X(d) = f_X(d-)$ that is continuous on $[c, d]$. Then $\int_c^d x f_X(x)dx = \int_c^d x\widetilde{f}_X(x)dx$ and the display equation still holds. Since f_X has at most finitely many discontinuities on $[0, c]$, linearity and cancellations yield

$$\int_0^d x f_X(x)dx = -d\mathbb{P}(X > d) + \int_0^d \mathbb{P}(X > x)dx.$$

Before we let $d \to \infty$, note that the existence of the improper Riemann integral $\mathbb{E}[X] = \int_0^\infty x f_X(x)dx$ entails

$$d\mathbb{P}(X > d) = \int_d^\infty d f_X(x)dx \leq \int_d^\infty x f_X(x)dx \to 0 \quad \text{as } d \to \infty.$$

Now letting $d \to \infty$ in the previous display completes the proof. □

Corollary E.2. *Let X be a continuous random variable whose expectation $\mathbb{E}[X]$ exists, and with p.d.f. f_X that is piecewise continuous with left and right limits at its discontinuities. Then*

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx - \int_0^\infty \mathbb{P}(X < -x)dx.$$

Proof. As in the proof of Theorem E.1, we obtain

$$\int_0^\infty x f_X(x)dx = \int_0^\infty \mathbb{P}(X > x)dx.$$

Also, $Z = -X$ has c.d.f. $\mathbb{P}(-X \leq z) = \mathbb{P}(X \geq -z) = 1 - \mathbb{P}(X \leq -z)$, using $\mathbb{P}(X = -z) = 0$. By differentiation, p.d.f. $f_{-X}(z) = f_X(-z)$. Substituting $z = g(x) = -x$ and taking care with the finitely many discontinuities by linearity of the integral, and taking care with the improper nature of integrals by letting finite bounds tend to $\pm\infty$, the above also entails

$$\int_{-\infty}^0 x f_X(x)dx = -\int_0^\infty z f_{-X}(z)dz = -\int_0^\infty \mathbb{P}(-X > z)dz = -\int_0^\infty \mathbb{P}(X < -x)dx. \quad \square$$

Theorem E.3. *Let X be a continuous random variable with p.d.f. f_X as in Corollary E.2, and let $h\colon \mathbb{R} \to \mathbb{R}$ be the function $h(x) = x^m$ for some $m \geq 1$. Then*

$$\mathbb{E}[h(X)] = \int_{-\infty}^\infty h(x) f_X(x)dx$$

provided that the integral $\int_{-\infty}^\infty |h(x)| f_X(x)dx$ exists and is finite.

Proof. We adapt the proofs of Theorem E.1 and Corollary E.2. Specifically, if f_X is continuous on $[c, d]$ for some $0 \leq c < d$, then integration by parts and substituting $z = x^m$ yield

$$\begin{aligned}
\int_c^d x^m f_X(x)dx &= -\Big[x^m \mathbb{P}(X > x)\Big]_c^d + \int_c^d m x^{m-1} \mathbb{P}(X > x)dx \\
&= -\Big[x^m \mathbb{P}(X > x)\Big]_c^d + \int_c^d m x^{m-1} \mathbb{P}(X > 0, X^m > x^m)dx \\
&= -\Big[x^m \mathbb{P}(X > x)\Big]_c^d + \int_{c^m}^{d^m} \mathbb{P}(X > 0, X^m > z)dz.
\end{aligned}$$

If f_X vanishes outside $[c, d]$, this completes the proof. Otherwise, linearity, cancellations and $d^m \mathbb{P}(X > d) \leq \int_d^\infty x^m f_X(x)dx \to 0$ as $d \to \infty$ yield

$$\int_0^\infty x^m f_X(x)dx = \int_0^\infty \mathbb{P}(X > 0, X^m > x)dx.$$

If m is odd, then $X^m > x \Rightarrow X > 0$ and $\int_{-\infty}^0 x^m f_X(x)dx = \int_0^\infty \mathbb{P}(X^m < -x)dx$.
If m is even, then $\int_{-\infty}^0 x^m f_X(x)dx = \int_0^\infty \mathbb{P}(X < 0, X^m > x)dx$.

In either case, summing the left-hand sides gives $\int_{-\infty}^\infty x^m f_X(x)dx$ and summing the right-hand sides gives $\mathbb{E}[X^m]$, by an application of Corollary E.2 with X replaced by X^m. □

A similar argument establishes further cases $h(x) = ax + b$ and $h(x) = (x - \mu)^2$. Together with the special case $h(x) = x^2$, the definition $\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ makes sense and the consequence $\mathrm{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ follows, both without reference to the p.d.f.s of the random variables $(X - \mathbb{E}[X])^2$ and X^2, which are not usually helpful in variance calculations.

E.2 Expectations of general random variables

Perhaps the most accessible way to define general expectations at this stage is to use discrete approximation. Specifically, we can approximate any random variable $X : \Omega \to \mathbb{R}$, by $X_n = 2^{-n}\lfloor 2^n X \rfloor$, i.e. by the next lower multiple of 2^{-n}. Then, clearly $X_n(\omega) \to X(\omega)$ as $n \to \infty$, as an increasing limit. For non-negative $X : \Omega \to [0, \infty)$, the expected value of the discrete random variable X_n is either finite or infinite, and we can note that $(\mathbb{E}[X_n])$ is an increasing sequence. This leads to the following definition.

Definition E.4. Let $X : \Omega \to [0, \infty)$ be any non-negative random variable. Consider the *discretisations* $X_n = 2^{-n}\lfloor 2^n X \rfloor$, $n \geq 0$. Then either $\mathbb{E}[X_n] = \infty$ for some $n \geq 0$, and we define $\mathbb{E}[X] = \infty$, or $(\mathbb{E}[X_n])$ is a finite sequence, and we define $\mathbb{E}[X]$ as the increasing limit

$$\mathbb{E}[X] := \lim_{n\to\infty} \sum_{k=0}^{\infty} k2^{-n}\mathbb{P}(X_n = k2^{-n}) \in [0, \infty].$$

If $X : \Omega \to \mathbb{R}$ is any (real-valued) random variable, we consider $X^+ = \max\{X, 0\}$ and $X^- = \max\{-X, 0\}$ so that $X = X^+ - X^-$. If both of $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are finite, we define $\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-]$. (If just one is finite, we assign the value $-\infty$ or ∞, respectively.)

These limits of series are not very elegant compared to the more direct definition of Part B Probability, Measure and Martingales, where we will make sense of $\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega)$ once the notion of integration against a probability measure $\mathbb{P}$ is available (as a generalisation of integration against Lebesgue measure, which is introduced in Part A Integration). On the back of that general theory, some of the following results, in this section and the next will have shorter proofs. However, the following arguments only use Prelims Analysis.

Theorem E.5 (Order property of expectation)**.** *Let X and Y be two random variables whose expectations exist in the sense of Definition E.4. Then $X \leq Y$ implies $\mathbb{E}[X] \leq \mathbb{E}[Y]$.*

This is left as an optional exercise in Part A Probability.

Corollary E.6. *Let X be any random variable whose expectation exists in the sense of Definition E.4. Then $\mathbb{E}[X_n] \leq \mathbb{E}[X] \leq \mathbb{E}[X_n] + 2^{-n}$ for all $n \geq 0$.*

Proof. If $\mathbb{E}[X]$ is finite, then so are $\mathbb{E}[X^\pm]$. By definition, $X_n^\pm \leq X^\pm \leq X_n^\pm + 2^{-n}$. Since $X_n \leq X_n^+ - (X_n^- + 2^{-n}) \leq X^+ - X^- = X \leq (X_n^+ + 2^{-n}) - X_n^- \leq X_n + 2^{-n}$, this follows from the proposition, and from the linearity of expections in the sense of the previous definition for discrete random variables. □

Proposition E.7. *If X is a discrete or continuous random variable, the new definition of $\mathbb{E}[X]$ is consistent with the previous definitions.*

This is left as an optional exercise in Part A Probability.

The law of total probability for expectations for finite partitions applies to X_n and extends to X by a sandwich argument noting that $\mathbb{E}[X|B_i]$ is defined to be $\lim_{n\to\infty}\mathbb{E}[X_n|B_i]$ and sandwiched between $\mathbb{E}[X_n|B_i]$ and $\mathbb{E}[X_n|B_i]+2^{-n}$.

Theorem E.8 (Linearity of expectation)**.** *Suppose X and Y are any random variables and $a, b \in \mathbb{R}$. Then $\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y]$, provided that both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ exist.*

This is left as an optional exercise in Part A Probability.

Moments $\mathbb{E}[X^k]$ and $\mathbb{E}[X^kY^m]$ are just expectations of the random variables X^k and X^kY^m, similarly $\mathrm{var}(X)=\mathbb{E}[(X-\mathbb{E}[X])^2]$ and $\mathrm{cov}(X,Y)=\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$. Results $\mathrm{var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2$, $\mathrm{var}(aX+b)=a^2\mathrm{var}(X)$, $\mathrm{cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$ and $\mathrm{var}(X+Y)=\mathrm{var}(X)+\mathrm{var}(Y)+2\mathrm{cov}(X,Y)$ follow by linearity of expectation.

Proposition E.9. *$\mathbb{E}[X]=\int_0^\infty \mathbb{P}(X>x)dx$ holds for any random variable $X\colon\Omega\to[0,\infty)$.*

This is left as an optional exercise in Part A Probability.

E.3 Independence of general random variables

Recall the notion of independence of infinitely many discrete random variables in Definition D.35. We separately defined independence of two continuous random variables in Definition D.61, but for general random variables, it is easiest here to define independence in terms of cumulative distribution functions.

Definition E.10. A family $\{X_i, i\in I\}$ of random variables is independent if for all *finite* $J\subseteq I$

$$\mathbb{P}\left(\bigcap_{i\in J}\{X_i\le x_i\}\right)=\prod_{i\in J}\mathbb{P}(X_i\le x_i)\quad\text{for all } x_i\in\mathbb{R}, i\in J.$$

Before relating this to Definitions D.35 and D.61, we record a consequence of Theorem D.46.

Lemma E.11. *If E_n, $n\ge1$, is a decreasing sequence of events in the sense that $E_n\supseteq E_{n+1}$ for all $n\ge1$, then*

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}}E_n\right)=\lim_{n\to\infty}\mathbb{P}(E_n).$$

Proof. Let $A_n=E_n^c=\Omega\setminus E_n$, $n\ge1$. Then A_n, $n\ge1$, is an increasing sequence of events and Theorem D.46 yields

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}}E_n\right)=1-\mathbb{P}\left(\bigcup_{n\in\mathbb{N}}A_n\right)=1-\lim_{n\to\infty}\mathbb{P}(A_n)=\lim_{n\to\infty}\mathbb{P}(E_n).\qquad\square$$

Remark E.12. Definition E.10 is compatible with Definition D.61 if the marginal densities are continuous (up to finitely many discontinuities), by integrating the factorising joint densities on the one hand, and on the other hand by differentiating the display here w.r.t. all x_i, $i\in J$.

We can also see that this is compatible with Definition D.35 since we can choose $A_i=(-\infty,x_i]\subseteq\mathbb{R}$ in Definition D.35, and it was shown on problem sheets (for $\#J=2$, which generalises easily) that it suffices to check the condition in Definition D.35 for $A_i = \{b_i\}$ for all $b_i \in \mathbb{R}$, $i \in J$. This check can be done in a few steps, by passing from $A_i = (-\infty, x_i]$ to sets of the form $B_i = (-a_i, b_i]$, replacing the A_i one by one by the B_i taking set differences, then noting that $\bigcap_{n\in\mathbb{N}}\{X_i \in (b_i - 1/n, b_i], i \in J\} = \{X_i = b_i, i \in J\}$, then applying Lemma E.11 ensures that the probability of the infinite intersection is $\lim_{n\to\infty} \mathbb{P}(X_i \in (b_i - 1/n, b_i], i \in J)$. We leave the details to the reader.

Alternative proofs are available after a more systematic development of measure theory.

Theorem E.13. *If X and Y are independent random variables with finite expectation, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, i.e.* $\operatorname{cov}(X, Y) = 0$.

Proof. First note that the independence of X and Y entails the independence of X^+ or X^- on the one hand and of Y^+ or Y^- on the other. By linearity of expectation, it suffices to consider the case of non-negative X and Y. Now proceed as in the proof of Theorem E.8 to write $X_nY_n \le (XY)_{2n} \le (X_n+2^{-n})(Y_n+2^{-n})$ and taking expectations, factorising the outer terms by Theorem D.30 and then letting $n \to \infty$, we find $\mathbb{E}[X]\mathbb{E}[Y] \le \mathbb{E}[XY] \le \mathbb{E}[X]\mathbb{E}[Y]$. □

Corollary E.14. *If $r_i \in \mathbb{N}$, $1 \le i \le k$, and if W_i, $1 \le i \le k$, are independent random variables so that W_i has finite r_ith moment $\mathbb{E}[W_i^{r_i}]$, $1 \le i \le k$, then*

$$\mathbb{E}\left[\prod_{i=1}^k W_i^{r_i}\right] = \prod_{i=1}^k \mathbb{E}\left[W_i^{r_i}\right].$$

Proof. The case $r_1 = \cdots = r_k = 1$ follows by the argument of Theorem E.13. It therefore suffices to show that $W_i^{r_i}$ are independent. If all r_i are odd, this follows from the definition of independence which entails that for all $J \subseteq \{1, \ldots, k\}$,

$$\mathbb{P}\left(\bigcap_{i\in J}\left\{W_i^{r_i} \le z_i\right\}\right) = \mathbb{P}\left(\bigcap_{i\in J}\left\{W_i \le z_i^{1/r_i}\right\}\right) = \prod_{i\in J}\mathbb{P}\big(W_i \le z_i^{1/r_i}\big) = \prod_{i\in J}\mathbb{P}\big(W_i^{r_i} \le z_i\big).$$

If some of the r_i are even, we note that

$$\left\{W_i^{r_i} \le z_i\right\} = \left\{-z_i^{1/r_i} \le W_i \le z_i^{1/r_i}\right\} = \left\{W_i \le z_i^{1/r_i}\right\} \setminus \left\{W_i < z_i^{1/r_i}\right\},$$

so it suffices to note that factorisations as in the display also hold if either one of the inequalities is strict (by Theorem D.46) or if the set is replaced by what is needed here. Inductively, this follows if this replacement is done for more than one i. □

Remark E.15. These results and the methods of proof are not very satisfactory. Methods from Part A Integration will give access to more efficient and more general proofs of the claim that functions of independent random variables are independent and that expectations of independent random variables factorise.

However, at this point, we have reduced the technical gaps to three items. The first is the existence of probability spaces beyond (finitely many) discrete random variables. The second is proving generalisations of formulas for expectations of functions of one or more (jointly) continuous random variables beyond powers and beyond requirements that variables are independent with piecewise continuous densities. The third is lifting the finite variance condition in the proof of the weak law of large numbers. We will deduce the first two from Part A Integration, but leave the third to Part B Probability, Measure and Martingales. See also the discussion of technical gaps in Part A Probability.

F

Part A Analysis: Relevant material

In this appendix we provide a preview of Part A Integration and some relevant material from Part A Metric Spaces and Complex Analysis. Most importantly, we introduce the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}$ and Lebesgue measure with its main properties, and we introduce the Lebesgue integral and main results. The relevance of the former for us is that it provides a probability space for all random variables in Prelims and Part A Probability. The latter provides a theorem analogous to Proposition C.7 that allows interchanging the order of integration. This is the key to rigorously proving the remaining results about continuous random variables stated in Prelims Probability. We record those consequences in Appendix G.

F.1 Lebesgue measure on $\mathbb{R}$

The aim of Lebesgue measure is to extend the notion of the length of an interval to other subsets of $\mathbb{R}$.

Definition F.1. For any interval with endpoints a and b, define the length as $m(I) = b - a$. For $A \subseteq \mathbb{R}$, we define the *outer measure* of A to be

$$m^*(A) = \inf\left\{\sum_{n=1}^{\infty} m(I_n)\colon I_n \text{ intervals}, A \subseteq \bigcup_{n=1}^{\infty} I_n\right\}.$$

Remark F.2. This definition assigns with *every* subset of $\mathbb{R}$ the smallest sum of lengths of intervals that allows to cover A (in the sense of the approximation property of the infimum that any value slightly larger than $m^*(A)$ allows such a cover).

Proposition F.3.

1. *$m^*(\emptyset) = 0$ and $m^*(\{x\}) = 0$;*
2. *m^* assigns the length to intervals: $m^*(I) = b - a$ for intervals I with endpoints a, b;*
3. *m^* is translation invariant: $m^*(A + x) = m^*(A)$*
4. *m^* behaves well under scaling: $m^*(\alpha A) = |\alpha| m^*(A)$*
5. *m^* is increasing: $m^*(A) \leq m^*(B)$ if $A \subseteq B$;*
6. *m^* is* sub*additive: $m^*(A \cup B) \leq m^*(A) + m^*(B)$ and $m^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} m^*(A_n)$.*

Definition F.4. A subset E of $\mathbb{R}$ is said to be *null* if $m^*(E) = 0$.

Corollary F.5. *1. Any subset of a null set is null.*

2. If E_n is a null set for all $n \geq 1$, then $\bigcup_{n=1}^{\infty} E_n$ is null.

3. Any countable subset of $\mathbb{R}$ is null.

Using the Axiom of Choice, it can be shown that the outer measure m^* is not countably additive. Indeed, Lebesgue measure will be defined as the restriction of m^* to a smaller collection of subsets of $\mathbb{R}$ on which it can then be shown that countable additivity holds.

Definition F.6. A subset E of $\mathbb{R}$ is said to be measurable if

$$m^*(A) = m^*(A \cap E) + m^*(A \setminus E)$$

We denote by $\mathcal{M}_{\text{Leb}}(\mathbb{R})$, or just $\mathcal{M}_{\text{Leb}}$ the set of all measurable subsets of $\mathbb{R}$.

Proposition F.7. *1. If E is null, then $E \in \mathcal{M}_{\text{Leb}}$.*

2. If I is any interval, then $I \in \mathcal{M}_{\text{Leb}}$.

3. If $E \in \mathcal{M}_{\text{Leb}}$, then $\mathbb{R} \setminus E \in \mathcal{M}_{\text{Leb}}$.

4. If $E_n \in \mathcal{M}_{\text{Leb}}$ for all $n \geq 1$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{M}_{\text{Leb}}$.

5. If $E_n \in \mathcal{M}_{\text{Leb}}$, $n \geq 1$, and $E_n \cap E_k = \emptyset$ when $n \neq k$, then $m^*(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} m^*(E_n)$.

Remark F.8. In particular, $\mathcal{M}_{\text{Leb}}$ satisfies the σ-algebra axioms of Definition D.1 and the restriction of m^* to $\mathcal{M}_{\text{Leb}}$ satisfies the measure axioms that $m^*(\emptyset) = 0$, that $m^*(E) \geq 0$ for all $E \in \mathcal{M}_{\text{Leb}}$ and that *countable additivity* in the sense of 5. above holds when m^* is restricted to $\mathcal{M}_{\text{Leb}}$.

Definition F.9. We call $m \colon \mathcal{M}_{\text{Leb}} \to [0, \infty]$ given by $m(E) = m^*(E)$, for $E \in \mathcal{M}_{\text{Leb}}$, *Lebesgue measure on* $\mathbb{R}$.

Corollary F.10. *Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{M}_{\text{Leb}}|_{[0,1]} := \{E \in \mathcal{M}_{\text{Leb}} \colon E \subseteq [0, 1]\}$ and let $\mathbb{P}$ be the restriction $m|_{\mathcal{F}}$ of Lebesgue measure m to $\mathcal{F}$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space in the sense of Definition D.1.*

Lemma F.11. *Let $\mathcal{B} \subset \mathcal{P}(\mathbb{R})$. Then there is a unique σ-algebra $\mathcal{F}_{\mathcal{B}}$ on $\mathbb{R}$ which is generated by $\mathcal{B}$ in the following sense:*

(i) $\mathcal{F}_{\mathcal{B}}$ is a σ-algebra and $\mathcal{B} \subseteq \mathcal{F}_{\mathcal{B}}$.

(ii) If $\mathcal{F}$ is a σ-algebra on $\mathbb{R}$ and $\mathcal{B} \subseteq \mathcal{F}$ then $\mathcal{F}_{\mathcal{B}} \subseteq \mathcal{F}$.

Definition F.12. The σ-algebra $\mathcal{M}_{\text{Bor}}(\mathbb{R})$ generated by the intervals is the *Borel σ-algebra.*

Proposition F.13. *1. Let $\mathcal{B}$ be any one of the following classes of subsets of $\mathbb{R}$.*

(i) All intervals of the form (a, ∞).

(ii) All intervals of the form $[a, b]$.

(iii) All open sets.

Then $\mathcal{M}_{\text{Bor}}(\mathbb{R})$ is is generated by $\mathcal{B}$, i.e. is the smallest σ-algebra on $\mathbb{R}$ containing $\mathcal{B}$.

2. $\mathcal{M}_{\text{Bor}}(\mathbb{R})$ is a strict subset of $\mathcal{M}_{\text{Leb}}(\mathbb{R})$.

F.2 Lebesgue integral

Definition F.14. A function $f\colon \mathbb{R} \to \mathbb{R}$ is (Lebesgue) *measurable* if $f^{-1}(I) \in \mathcal{M}_{\text{Leb}}$ for each interval I. It is *Borel measurable* if $f^{-1}(I) \in \mathcal{M}_{\text{Bor}}$ for each interval I.

A function $\phi\colon \mathbb{R} \to \mathbb{R}$ is *simple* if it is measurable and it takes only finitely many real values, i.e. if it can be written as

$$\phi = \sum_{i=1}^{k} c_i \mathbf{1}_{B_i}$$

for some distinct non-zero $c_i \in \mathbb{R}$ and disjoint $B_i \in \mathcal{M}_{\text{Leb}}$, $1 \le i \le k$.

Proposition F.15. *Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be measurable and $h\colon \mathbb{R} \to \mathbb{R}$ continuous (or Borel measurable). Then $f + g$, fg, $\max\{f, g\}$ and $h \circ f$ are measurable.*

Definition F.16 (Lebesgue integral)**.** For a non-negative simple function $\phi = \sum_{i=1}^{k} c_i \mathbf{1}_{B_i}$,

$$\int_{\mathbb{R}} \phi := \int_{-\infty}^{\infty} \phi(x)dx := \sum_{i=1}^{k} c_i m(B_i) \in [0, \infty].$$

For a non-negative measurable function $f\colon \mathbb{R} \to [0, \infty)$,

$$\int_{\mathbb{R}} f := \sup\left\{\int_{\mathbb{R}} \phi\colon \phi \text{ simple}, 0 \le \phi \le f\right\} \in [0, \infty] \quad \text{and} \quad \int_E f := \int_{\mathbb{R}} f\mathbf{1}_E, \ E \in \mathcal{M}_{\text{Leb}}.$$

A measurable function $f\colon \mathbb{R} \to \mathbb{R}$ is called *Lebesgue integrable over E* if $\int_E |f| < \infty$, and then

$$\int_E f := \int_E f^+ - \int_E f^- \quad \text{where } f^+ := \max\{f, 0\} \text{ and } f^- := \max\{-f, 0\}.$$

Theorem F.17 (Monotone Convergence Theorem)**.** *If (f_n) is an increasing sequence of non-negative measurable functions and $f = \lim_{n\to\infty} f_n$, then $\int_{\mathbb{R}} f = \lim_{n\to\infty} \int_{\mathbb{R}} f_n$.*

Corollary F.18. *Let $f\colon [a, b] \to [0, \infty)$ be continuous. Then the Lebesgue integral $\int_{[a,b]} f$ as defined above equals the Riemann integral $\int_a^b f$ as defined in Prelims Analysis.*

Proposition F.19. *The set $\mathcal{L}^1(\mathbb{R})$ of Lebesgue integrable functions on $\mathbb{R}$ forms a vector space and the Lebesgue integral is a linear functional on it: $\int_{\mathbb{R}}(\alpha f + \beta g) = \alpha \int_{\mathbb{R}} f + \beta \int_{\mathbb{R}} g$.*

Proposition F.20. *Riemann integrable functions on $[a, b]$ are Lebesgue integrable over $[a, b]$.*

Theorem F.21 (Dominated Convergence Theorem)**.** *If (f_n) is a sequence of integrable functions, A a null set and f such that $f_n(x) \to f(x)$ for all $x \notin A$, and such that $|f_n| \le g$ for an integrable function g, then f is integrable and $\int f = \lim_{n\to\infty} \int f_n$.*

Theorem F.22 (Substitution)**.** *Let $g\colon I \to \mathbb{R}$ be a monotonic function with a continuous derivative on an interval I, and let J be the interval $g(I)$. A measurable function $f\colon J \to \mathbb{R}$ is Lebesgue integrable over J if and only if $(f \circ g)g'$ is Lebesgue integrable over I. Then*

$$\int_J f(x)dx = \int_I f(g(y))|g'(y)|dy.$$

Theorem F.23 (Differentiability Lemma)**.** *For intervals I, J and $f\colon I \times J \to \mathbb{R}$ such that $x \mapsto f(x, y)$ is integrable over I for all $y \in J$ and $y \mapsto f(x, y)$ is differentiable with $|\frac{\partial f}{\partial y}(x, y)| \le g(x)$ for all $x \in I$ and integrable g, $F(y) = \int_I f(x, y)dx$ is differentiable with $F'(y) = \int_J \frac{\partial f}{\partial y}(x, y)dx$.*

F.3 Double integrals

The class $\mathcal{L}^1(\mathbb{R}^2)$ of Lebesgue integrable functions $f\colon \mathbb{R}^2 \to \mathbb{R}$ is defined in exactly the same way as $\mathcal{L}^1(\mathbb{R})$, except that intervals (a,b), and their lengths $b-a$, are replaced by rectangles $(a,b)\times(c,d)$ and their areas $(b-a)(d-c)$. Then one defines outer measure, measurable sets, measurable functions, simple functions, Lebesgue integrable functions and (double) integrals just as in Sections F.1–F.2. The (double) Lebesgue integral of a Lebesgue integrable function f over $\mathbb{R}^2$ will be denoted by either of the following

$$\int_{\mathbb{R}^2} f, \qquad \int_{\mathbb{R}^2} f(x,y)d(x,y).$$

Theorem F.24 (Fubini's Theorem)**.** *Let $f \in \mathcal{L}^1(\mathbb{R}^2)$. Then, there is a null set A such that for all $y \notin A$, the function $x \mapsto f(x,y)$ is Lebesgue integrable. Moreover, if $F(y)$ is defined by $F(y) = \int_{\mathbb{R}} f(x,y)dx$ for $y \notin A$ and by $F(y) = 0$ for $y \in A$, then F is Lebesgue integrable, and*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dx \right) dy = \int_{\mathbb{R}^2} f(x,y)d(x,y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dy \right) dx,$$

where the last repeated integral also exists in the sense described above.

Theorem F.25 (Tonelli's Theorem)**.** *Let $f\colon \mathbb{R}^2 \to [0,\infty)$ be a measurable function, and suppose that either of the following repeated integrals is finite:*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)|dx \right) dy, \qquad \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)|dy \right) dx,$$

Then f is integrable. Hence, Fubini's Theorem is applicable to both f and $|f|$.

If E is a measurable subset of $\mathbb{R}^2$ and $f\colon E \to \mathbb{R}$ is any function, then F is said to be Lebesgue integrable over E if $\tilde{f}$ is integrable over $\mathbb{R}^2$ where $\tilde{f}(x,y) = f(x,y)$ if $(x,y) \in E$, $\tilde{f}(x,y) = 0$ otherwise. Then $\int_E f$ is defined to be $\int_{\mathbb{R}^2} \tilde{f}$.

For a function $T\colon E' \to \mathbb{R}^2$ defined on an open subset of $\mathbb{R}^2$, let $T(u,v) = (x(u,v), y(u,v))$. We consider *partial derivatives*, as follows. If $u \mapsto x(u,v)$ is differentiable, we write this derivative as $\frac{\partial x}{\partial u}(u,v)$. With analogous notation $\frac{\partial x}{\partial v}$, $\frac{\partial y}{\partial u}$, $\frac{\partial y}{\partial v}$, let J_T be the Jacobian matrix:

$$J_T = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

Theorem F.26. *Let E' be an open subset of $\mathbb{R}^2$, $T\colon E' \to \mathbb{R}^2$ be a one-to-one differentiable function of E' onto a subset E of $\mathbb{R}^2$, and $f\colon E \to \mathbb{R}$ be a function. Then f is Lebesgue integrable over E if and only if $(f \circ T)|\mathrm{det} J_T|$ is Lebesgue integrable over E'. In that case,*

$$\int_E f = \int_{E'} (f \circ T)|\mathrm{det} J_T|.$$

Remark F.27. A 15-page proof is provided for the k-dimensional case assuming that both T and T^{-1} are continuously differentiable. Part A Multidimensional Analysis and Geometry establishes a chain rule, but this does not entail the substitution rule as in one dimension.

Remark F.28. Section F.3 extends from $\mathbb{R}^2$ to $\mathbb{R}^n$. Moreover, for any (σ-finite) measure spaces $(\Omega_1, \mathcal{F}, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$, one can define a product $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$ such that Fubini's and Tonelli's theorems hold. (Here, μ is called a σ-finite measure on $(\Omega, \mathcal{F})$ if there are $A_n \in \mathcal{F}$ with $\bigcup_{n\geq 1} A_n = \Omega$ and $\mu(A_n) < \infty$ for all $n \geq 1$.)

F.4 Metric spaces

This section is only relevant for Sections F.5–F.6, which we only use to prove the uniqueness and convergence theorems for moment generating functions and to compute the characteristic function of the Cauchy distribution, which are stated without proof in Part A Probability.

Definition F.29. Let X be a set. Then a *distance function* on X is a function $d\colon X \times X \to \mathbb{R}$ with the following properties:

(i) (positivity) $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$;

(ii) (symmetry) $d(x, y) = d(y, x)$;

(iii) (triangle inequality) if $x, y, z \in X$ then we have $d(x, z) \leq d(x, y) + d(y, z)$.

The pair (X, d) consisting of a set X together with a distance function d on it is called a *metric space*.

Definition F.30 (Norms)**.** Let V be any vector space (over the reals). A function $\|\cdot\|\colon V \to [0, \infty)$ is called a *norm* if the following are all true:

- $\|x\| = 0$ if and only if $x = 0$;
- $\|\lambda x\| = |\lambda|\, \|x\|$ for all $\lambda \in \mathbb{R}$, $x \in V$;
- $\|x + y\| \leq \|x\| + \|y\|$ whenever $x, y \in V$.

We call a vector space endowed with a norm $\|\cdot\|$ a *normed space*. Whenever we talk about normed spaces it is understood that we are also thinking of them as metric spaces with the metric being defined by $d(v, w) = \|v - w\|$.

Example F.31. Take $X = \mathbb{R}^n$. Then each of the following functions define metrics on X.

$$d_1(v, w) = \sum_{j=1}^{n} |v_j - w_j|, \quad d_2(v, w) = \left(\sum_{j=1}^{n} (v_j - w_j)^2 \right)^{1/2}, \quad d_\infty(v, w) = \max_{j \in \{1,2,\ldots,n\}} |v_j - w_j|.$$

These are called the ℓ^1-, ℓ^2- (or Euclidean) and ℓ^∞-distances, repectively. The *Euclidean norm* $\|v\|_2$ of a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ is

$$\|v\|_2 := \left(\sum_{j=1}^{n} |v_j|^2 \right)^{1/2}.$$

Lemma F.32. *If $x, y \in \mathbb{R}^n$, then $\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$.*

Suppose that (X, d) is a metric space and let Y be a subset of X. Then the restriction of d to $Y \times Y$ gives a metric so that $(Y, d|_{Y \times Y})$ is a metric space. We call Y equipped with this metric a *subspace*.

Definition F.33 (Continuity)**.** Let (X, d_X) and (Y, d_Y) be metric spaces. We say a function $f\colon X \to Y$ is continuous at $a \in X$ if for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any $x \in X$ with $d_X(a, x) < \delta$ we have $d_Y(f(x), f(a)) < \varepsilon$.

Definition F.34 (Balls)**.** Let X be a metric space. If $a \in X$ and $\varepsilon > 0$ then we define the *open ball of radius* ε *to be the set*

$$B(a, \varepsilon) = \{x \in X : d(x, a) < \varepsilon\}.$$

Definition F.35 (Open sets)**.** If X is a metric space then we say a subset $U \subseteq X$ is *open* if for each $y \in U$ there is some $\delta > 0$ such that the open ball $B(y, \delta)$ is contained in U.

Lemma F.36. *Every open ball in a metric space is an open set.*

Definition F.37. We say that a metric space is *disconnected* if we can write it as the disjoint union of two nonempty open sets. We say that a space is *connected* if it is not disconnected.

Theorem F.38. *A subset of* $\mathbb{R}$ *is connected if and only if it is an interval.*

Definition F.39. Let X be a metric space. Then we say that X is path-connected if the following is true: for any $a, b \in X$ there is a continuous map $\gamma\colon [-0, 1] \to X$ with $\gamma(0) = a$ and $\gamma(1) = b$.

Theorem F.40. *A path-connected metric space is connected.*

Theorem F.41. *A connected open subset of a normed space is path-connected.*

F.5 Holomorphic functions on domains

We can identify $\mathbb{C}$ with the plane $\mathbb{R}^2$ by recording real and imaginary parts. Thus we have mutually inverse bijections

$$z \mapsto (\Re(z), \Im(z)) \qquad \text{and} \qquad (x, y) \mapsto x + iy,$$

respectively from $\mathbb{C}$ to $\mathbb{R}^2$ and from $\mathbb{R}^2$ to $\mathbb{C}$. The metric induced by the Euclidean norm on $\mathbb{R}^2$ also gives a metric on $\mathbb{C}$ by the identification of $\mathbb{C}$ and $\mathbb{R}^2$ described by the mutually inverse bijections above.

If $z = \Re(z) + i\Im(z)$ is a complex number, we write $|z|$ (called the *modulus*) for this Euclidean norm, that is,

$$|z| = \sqrt{(\Re(z))^2 + (\Im(z))^2}.$$

The distance between two points $z, w \in \mathbb{C}$ is then $|z - w|$.

Definition F.42. A connected open subset $D \subseteq \mathbb{C}$ of the complex plane is called a *domain.*

Suppose that $a \in \mathbb{C}$, and that U is a *neighbourhood of* a, that is, U contains some ball $B(a, \eta)$, $\eta > 0$, but U itself need not be open. Suppose $F\colon U \setminus \{a\} \to \mathbb{C}$ is a function. Then we say that $\lim_{z \to a} F(z) = L$ if the following is true: for all $\varepsilon > 0$, there is some $\delta > 0$ such that if $0 < |z - a| < \delta$ then $|F(z) - L| < \varepsilon$.

Definition F.43 (Complex differentiability)**.** Let $a \in \mathbb{C}$, and suppose that $f\colon U \to \mathbb{C}$ is a function, where U is a neighbourhood of a. In particular, f is defined on some ball $B(a, \eta)$. Then we say that f is (complex) differentiable at a if

$$\lim_{z \to a} \frac{f(z) - f(a)}{z - a}$$

exists. If the limit exists, we write $f'(a)$ for it and call this the derivative of f at a.

Lemma F.44. *Let $a \in \mathbb{C}$, let U be a neighbourhood of a and let $f, g\colon U \to \mathbb{C}$.*

(1) (Sums, products) If f, g are differentiable at a then $f+g$ and fg are differentiable at a and $(f+g)'(a) = f'(a) + g'(a)$, $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$.

(2) (Quotients) If f, g are differentiable at a and $g(a) \neq 0$ then f/g is differentiable at a and
$$(f/g)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$

(3) (Chain rule) If $V \supset f(U)$ is a neighbourhood of $f(a)$ and $h\colon V \to \mathbb{C}$ is a function that is differentiable at $f(a)$, then $h \circ f$ is differentiable at a with
$$(h \circ f)'(a) = h'(f(a))f'(a).$$

Proposition F.45 (Differentiation of power series)**.** . *Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series, with radius of convergene R. Let $s(z)$ be the function to which this series converges on $B(0, R)$. Then the power series $t(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}$ also has radius of convergence R and on $B(0, R)$ the power series s is complex differentiable with $s'(z) = t(z)$. In particular, a power series is infinitely complex differentiable within its radius of convergence.*

Definition F.46. Let $U \subseteq \mathbb{C}$ be an open set (for example, a domain). Let $f\colon U \to \mathbb{C}$ be a function. If f is complex differentiable at every $a \in U$, we say that f is *holomorphic* on U.

A function $f\colon U \to \mathbb{C}$ is called *holomorphic at $z_0 \in U$* if f is holomorphic on a ball $B(z_0, r) := \{z \in \mathbb{C}\colon |z - z_0| < r\}$ around z_0 of radius $r > 0$.

Note that power series are holomorphic on the open disk inscribed by their radius of convergence.

Theorem F.47 (Identity theorem)**.** *Let U be a domain and suppose that f_1, f_2 are holomorphic functions defined on U. Then if $S = \{z \in U\colon f_1(z) = f_2(z)\}$ has a limit point in U, we must have $S = U$, that is $f_1(z) = f_2(z)$ for all $z \in U$.*

F.6 Contour integration and the Residue theorem

A *path* in the complex plane is a continuous function $\gamma\colon [a, b] \to \mathbb{C}$. A path is said to be *closed* if $\gamma(a) = \gamma(b)$. If γ is a path, we will write γ^* for its impage, that is
$$\gamma^* = \{\gamma(t), t \in [a, b]\}.$$

Definition F.48. We will say that a path $\gamma\colon [a, b] \to \mathbb{C}$ is *differentiable* if its real and imaginary parts are differentiable as real-valued functions. Equivalently, γ is differentiable at $t_0 \in [a, b]$ if
$$\lim_{t \to t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0}$$
exists, and the we denote this limit as $\gamma'(t_0)$. (If $t = a$ or $t = b$, then we interpret the above as a one-sided limit.) We say that a path is C^1 if it is differentiable and its derivative γ' is continuous.

We will say that a path is *piecewise-C^1* if it is continuous on $[a, b]$ and the interval $[a, b]$ can be divided into subintervals on each of which γ is C^1. That is, there is a finite sequence $a = a_0 < a_1 < \cdots < a_m = b$ such that $\gamma|_{[a_{j-1}, a_j]}$ is C^1 for all $1 \leq j \leq m$. Thus in particular, the left-hand and right-hand derivatives of γ at a_j, $1 \leq j \leq m-1$ may not be equal.

If $\gamma_1 \colon [a, b] \to \mathbb{C}$ and $\gamma_2 \colon [c, d] \to \mathbb{C}$ are two paths such that $gamma_1(b) = \gamma_2(c)$ then they can be concatenated to give a path $\gamma_1 \star \gamma_2$ which traverses first γ_1 and then γ_2. Formally, $\gamma_1 \star \gamma_2 \colon [a, b + d - c] \to \mathbb{C}$ where

$$\gamma_1 \star \gamma_2(t) = \begin{cases} \gamma_1(t), & a \leq t \leq b, \\ \gamma_2(b + t - c), & b \leq t \leq b + d - c. \end{cases}$$

So a piecewise C^1 path is precisely a finite concatenation of C^1 paths.

To define the integral of functions $F \colon [a, b] \to \mathbb{C}$, we write $F(t) = G(t) + iH(t)$, where $G, H \colon [a, b] \to \mathbb{R}$. Then we say that F is Riemann integrable if both G and H are, and we define

$$\int_a^b F(t)dt = \int_a^b G(t)dt + i \int_a^b H(t)dt.$$

It is easy to check that the integral is then complex-linear, that is, if $F_1, F_2 \colon [a, b] \to \mathbb{C}$ are both Riemann integrable on $[a, b]$ and $\alpha, \beta \in \mathbb{C}$, then $\alpha F_1 + \beta F_2$ is Riemann integrable and

$$\int_a^b (\alpha F_1(t) + \beta F_2(t))dt = \alpha \int_a^b F_1(t)dt + \beta \int_a^b F_2(t)dt.$$

Lemma F.49. *Let $[a, b]$ be a closed and bounded interval and $S \subset [a, b]$ a finite set. If f is a bounded continuous function (taking real or complex values) on $[a, b] \setminus S$ then it is Riemann integrable on $[a, b]$.*

Lemma F.50. *Suppose that $F \colon [a, b] \to \mathbb{C}$ is a function. Then we have*

$$\left| \int_a^b F(t)dt \right| \leq \int_a^b \left| F(t) \right| dt.$$

Definition F.51. If $\gamma \colon [a, b] \to \mathbb{C}$ is a piecewise-C^1 path and $f \colon \mathbb{C} \to \mathbb{C}$, then we define the integral of f along γ to be

$$\int_\gamma f(z)dz = \int_a^b f(\gamma(t))\gamma'(t)dt.$$

Proposition F.52. *Let $f \colon U \to \mathbb{C}$ be a continuous function on an open subset $U \subseteq \mathbb{C}$ and $\gamma_1 \colon [a, b] \to \mathbb{C}$ and $\gamma_2 \colon [c, d] \to \mathbb{C}$ be piecewise-C^1 paths whose images lie in U. Then we have*

$$\int_{\gamma_1 \star \gamma_2} f(z)dz = \int_{\gamma_1} f(z)dz + \int_{\gamma_2} f(z)dz.$$

Lemma F.53. *Let $\gamma \colon [0, 1] \to \mathbb{C} \setminus \{0\}$ be a path. Then there is a continuous function $a \colon [0, 1] \to \mathbb{R}$ such that*

$$\gamma(t) = |\gamma(t)| e^{2\pi i a(t)}$$

Moreover, if a and b are two such functions, then there exists $n \in \mathbb{Z}$ such that $a(t) = b(t) + n$ for all $t \in [0, 1]$.

Definition F.54. If $\gamma\colon [0,1] \to \mathbb{C} \setminus \{0\}$ is a closed path and $\gamma(t) = |\gamma(t)|e^{2\pi i a(t)}$ as in the previous lemma, then since $\gamma(0) = \gamma(1)$, we must have $a(1) - a(0) \in \mathbb{Z}$. This integer is called the *winding number* $I(\gamma, 0)$ *of* γ *around* 0. *If* γ *is a closed path with* $z_0 \notin \gamma^*$, *then let* $T\colon \mathbb{C} \to \mathbb{C}$ *be given by* $T(z) = z - z_0$ *and define* $I(\gamma, z_0) = I(T \circ \gamma, 0)$.

Theorem F.55 (Riemann's removable singularity theorem)**.** *Suppose that* U *is an open subset of* $\mathbb{C}$ *and* $z_0 \in U$. *If* $f\colon U \setminus \{z_0\} \to \mathbb{C}$ *is holomorphic on* $U \setminus \{z_0\}$ *and bounded in a neighbourhood of* z_0, *then* f *extends to a holomorphic function on all of* U.

Definition F.56. Let $f\colon U \to \mathbb{C}$ be a function, where U is open. We say that $z_0 \in \overline{U}$ is a *regular* point of f if f is holomorphic at z_0. Otherwise we say that z_0 is *singular*.

We say that z_0 is an *isolated singularity of* f if f is holomorphic on $B(z_0, r) \setminus \{z_0\}$ for some $r > 0$.

Suppose that z_0 is an isolated singularity of f. If f is bounded near z_0, we say that f has a *removable singularity* at z_0. If f is not bounded near z_0, but the function $1/f(z)$ has a removable singularity at z_0, then we say that f has a *pole* at z_0. If $1/f(z) = (z - z_0)^m g(z)$ for $z \in B(z_0, r) \setminus \{z_0\}$ for some $r > 0$, some $m \geq 1$ and some holomorphic $g\colon B(z_0, r) \to \mathbb{C}$ with $g(z_0) \neq 0$. If $m = 1$ we say that f has a *simple pole* at z_0.

Lemma F.57. *Let* f *be a holomorphic function with a pole of order* m *at* z_0. *then there is an* $r > 0$ *such that for all* $z \in B(z_0, r) \setminus \{z_0\}$, *we have*

$$f(z) = \sum_{n=-m}^{\infty} c_n (z - z_0)^n.$$

Definition F.58. The series $\sum_{n=-m}^{\infty} c_n (z - z_0)^n$ is called the *Laurent series* for f at z_0. A function on an open set U which has only isolated singularities all of which are poles is called a *meromorphic function* on U.

Definition F.59. Let $z_0 \in \mathbb{C}$, $k \in \mathbb{N}$ and $c_n \in \mathbb{C}$, $n \geq -k$. Suppose that the series

$$f(z) = \sum_{n=-k}^{\infty} c_n (z - z_0)^n$$

converges on $\{z \in \mathbb{C}\colon 0 < |z - z_0| < r\}$ for some $r > 0$. The *residue* of f at z_0 is defined to be the coefficient c_{-1} and denoted $\mathrm{Res}_{z_0}(f)$.

Theorem F.60 (Residue theorem)**.** *Suppose that* U *is an open set in* $\mathbb{C}$ *and* γ *is a path whose inside is contained in* U, *so that for all* $z \notin U$ *we have* $I(\gamma, z) = 0$. *Then if* $S \subset U$ *is a finite set such that* $S \cap \gamma^* = \emptyset$ *and* f *is a holomprphic function on* $U \setminus S$ *we have*

$$\int_\gamma f(z)dz = 2\pi i \sum_{a \in S} I(\gamma, a)\mathrm{Res}_a(f).$$

Lemma F.61 (Jordan's Lemma)**.** *Let* $f\colon \mathbb{H} \to \mathbb{C}_\infty$ *be a meromorphic function on the upper-half plane* $\mathbb{H} = \{z \in \mathbb{C}\colon \Im(z) > 0\}$. *Suppose that* $f(z) \to 0$ *as* $z \to \infty$ *in* $\mathbb{H}$. *Then if* $\gamma_R(s) = Re^{is}$ *for* $s \in [0, \pi]$ *we have*

$$\forall t > 0 \quad \int_{\gamma_R} f(z)e^{i\alpha z}dz \to 0 \quad \text{as } R \to \infty.$$

Example F.62. $\int_{-r}^{r} \frac{\sin(x)}{x} dx \to \pi$ as $r \to \infty$.

G

Part A Analysis: Relevant consequences

This appendix contains further developments that build on Part A Integration and are relevant for rigorously following the general theory developed in Prelims and Part A Probability. This material is non-examinable. Indeed, particularly Section G.2 interacts quite strongly with subtle methods as well as definitions and theorems from Part A Integration.

G.1 Sequences of independent random variables

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of Corollary F.10. Specifically, $\Omega = [0,1]$, the σ-algebra $\mathcal{F} = \mathcal{M}_{\text{Leb}}|_{[0,1]}$ of Lebesgue-measurable sets contains all intervals (and their countable unions etc.), and $\mathbb{P} = m|_{\mathcal{F}}$ is the (unique) probability measure on $\mathcal{F}$ that, when applied specifically to intervals, assigns their length.

Proposition G.1. *The function $U\colon \Omega \to \mathbb{R}$, $U(\omega) = \omega$, is a random variable that is uniformly distributed on* $[0,1]$.

Proof. For all $x \in \mathbb{R}$, we have

$$\{U \leq x\} = \begin{cases} \emptyset & \text{for } x < 0, \\ [0,x] & \text{for } x \in [0,1], \\ [0,1] & \text{for } x \geq 1, \end{cases} \qquad \text{and} \quad \mathbb{P}(U \leq x) = \begin{cases} 0 & \text{for } x < 0, \\ x & \text{for } x \in [0,1], \\ 1 & \text{for } x > 1. \end{cases}$$

We recognise this as the cumulative distribution function of the uniform distribution on $[0,1]$. Hence, X is a random variable that is uniformly distributed on $[0,1]$. □

Corollary G.2. *Consider any right-continuous function $F\colon \mathbb{R} \to \mathbb{R}$ that satisfies conditions 1, 3 and 4 of Theorem D.44. Let $Q(u) = \inf\{x \in \mathbb{R}\colon F(x) > u\}$. In the setting of Proposition G.1, $X := Q(U)$ is a random variable that has c.d.f. F.*

Proof. For $x \in \mathbb{R}$ and $u \in [0,1]$, we have implications $F(x) > u \Rightarrow Q(u) \leq x \Rightarrow F(x) \geq u$, since F is non-decreasing and right-continuous. Hence $\{Q(U) \leq x\} = \{u \in [0,1]\colon Q(u) \leq x\}$ is an interval, and therefore in $\mathcal{F}$. Since, $\mathbb{P}(U = F(x)) = 0$, by Theorem D.49, we have

$$F(x) = \mathbb{P}(U \leq F(x)) = \mathbb{P}(F(x) > U) \leq \mathbb{P}(Q(U) \leq x) \leq \mathbb{P}(F(x) \geq U) = F(x).$$

Hence $Q(U)$ is a random variable and has c.d.f. F, as required. □

This means we have constructed one random variable with any given distribution.

Lemma G.3. *Consider the setting of Proposition G.1. For $n \geq 1$, let $B_n = 1$ if $\lfloor 2^n U \rfloor$ is an odd integer and $B_n = 0$ otherwise. Then B_n, $n \geq 1$, is a sequence of independent Bernoulli random variables with $\mathbb{P}(B_n = 0) = \mathbb{P}(B_n = 1) = 1/2$.*

This is left as an optional exercise in Part A Probability.

Corollary G.4. *In the setting of Lemma G.3, we have $U = \sum_{n\geq 1} B_n 2^{-n}$, on the event $\{U < 1\}$ of probability 1. Now consider the strictly increasing sequence $(p_k)_{k\geq 1}$ of all prime numbers ($p_1 = 2$, $p_2 = 3$, $p_3 = 5$ etc.) and let $U_k = \sum_{n\geq 1} B_{p_k^n} 2^{-n}$ for all $k \geq 1$. Then U_k, $k \geq 1$, is a sequence of independent uniform random variables on $[0, 1]$.*

This is left as an optional exercise in Part A Probability.

Corollary G.5. *Consider any sequence of right-continuous functions $F_k : \mathbb{R} \to \mathbb{R}$, $k \geq 1$, that satisfy conditions 1, 3 and 4 of Theorem D.44. Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which we can define a sequence $(X_k)_{k\geq 1}$ of independent random variables such that X_k has c.d.f. F_k for each $k \geq 1$.*

Proof. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{M}_{\text{Leb}}|_{[0,1]}, m|_{\mathcal{F}})$, and the family $\{U_k, k \geq 1\}$ of independent uniform random variables of the preceding corollary. Let $Q_k(u) = \inf\{x \in \mathbb{R}: F_k(x) > u\}$, $u \in [0, 1)$. Then the argument of the proof of Corollary G.2 shows that $Q_k(U_k)$ has c.d.f. F_k for all $k \geq 1$. Furthermore, for any finite $J \subset \mathbb{N}$ and $x_k \in \mathbb{R}$, $k \in J$,

$$\mathbb{P}(X_k \leq x_k, \forall k \in J) = \mathbb{P}(U_k \leq F_k(x_k), \forall k \in J) = \prod_{k\in J} \mathbb{P}(U_k \leq F_k(x_k)) = \prod_{k\in J} \mathbb{P}(X_k \leq x_k),$$

i.e. independence of U_k, $k \geq 1$, entails the independence of X_k, $k \geq 1$. □

G.2 Random variables with continuous distributions

As we will use the theory of Lebesgue integration to rigorously establish results about expectations of continuous random variables, it is natural to generalise the notion of a continuous random variable to allow Lebesgue integrable (rather than Riemann integrable) densities.

Definition G.6. A *continuous random variable* X *is a random variable whose c.d.f. satisfies*

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{(-\infty,x]} f_X(u)du,$$

where $f_X : \mathbb{R} \to \mathbb{R}$ is a function such that

(a) $f_X(u) \geq 0$ for all $u \in \mathbb{R}$,

(b) the Lebesgue integral $\int_{\mathbb{R}} f_X(u)du$ exists and equals 1.

Then f_X is called probability density function (p.d.f.) *of X or, sometimes, just its* density.

Theorem G.7. *Let X be a continuous random variable with p.d.f. f_X. Then for all Borel sets $A \in \mathcal{M}_{\text{Bor}}(\mathbb{R})$, the set $\{X \in A\} = \{\omega \in \Omega: X(\omega) \in A\}$ is an event (in $\mathcal{F}$) and we have*

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx.$$

Proof. By definition, the claims hold when $A = (-\infty, x]$. This extends to closed intervals as in Theorem D.49. By Proposition F.13, the Borel σ-algebra is generated by the class of closed intervals. To show that our claims holds for all $A \in \mathcal{M}_{\rm Bor}(\mathbb{R})$, we consider the collection $\mathcal{A}$ of sets A for which the claims hold. The proof is complete if we can show that $\mathcal{A}$ is a σ-algebra. First note that $\{X \in \mathbb{R}\} = \Omega \in \mathcal{F}$, $\{X \in A\} \in \mathcal{F} \Rightarrow \{X \in A^c\} = \{X \in A\}^c \in \mathcal{F}$, and $\{X \in A_n\} \in \mathcal{F}$, $n \geq 1$, implies $\{X \in \bigcup_{n=1}^\infty A_n\} = \bigcup_{n=1}^\infty \{X \in A_n\} \in \mathcal{F}$, since $\mathcal{F}$ is a σ-algebra. It remains to study the representations of probabilities as integrals.

- Clearly, $\mathbb{P}(X \in \mathbb{R}) = 1 = \int_\mathbb{R} f_X(x)dx$, hence $\mathbb{R} \in \mathcal{A}$.
- If $A \in \mathcal{A}$, then $\mathbb{P}(X \in A^c) = 1 - \mathbb{P}(X \in A) = \int_\mathbb{R} f_X(x)dx - \int_A f_X(x)dx = \int_{A^c} f_X(x)dx$, by linearity of the Lebesgue integral, hence $A^c \in \mathcal{A}$.
- Before handling countable unions, consider the special case of *disjoint* $B_1, \ldots, B_n \in \mathcal{A}$, then $\bigcup_{k=1}^n B_k \in \mathcal{A}$ by additivity on both sides of the claim. In the general case of $A_n \in \mathcal{A}$, $n \geq 1$, the sets $B_1 := A_1$ and, inductively, $B_{n+1} := A_{n+1} \setminus \bigcup_{k=1}^n B_k$ are in $\mathcal{A}$, by what we have already shown and by the argument of the previous bullet point, for all $n \geq 2$. Furthermore, we have $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$. By countable additivity on the left and the Monotone Convergence Theorem on the right hand side, we conclude that $\mathbb{P}\big(X \in \bigcup_{n=1}^\infty A_n\big) = \sum_{n=1}^\infty \mathbb{P}(X \in B_n) = \sum_{n=1}^\infty \int_{B_n} f_X(x)dx = \int_{\bigcup_{n=1}^\infty A_n} f_X(x)dx$. □

Remark G.8. A quicker proof of Theorem E.1 than the Riemann integration proof we provided was sketched in Prelims Probability, interchanging the order of integrals. Formally, we apply Tonelli's theorem (Theorem F.25) to the non-negative measurable function $(x, y) \mapsto f_X(x)\mathbf{1}_S(x, y)$ where $S = \{(x, y) \in [0, \infty)^2 \colon x \leq y\}$. This argument also allows us to lift the piecewise continuity of f_X. Let X be any non-negative continuous random variable in the sense of Definition G.6. Then

$$\begin{aligned}\mathbb{E}[X] &= \int_{x\in[0,\infty)} x f_X(x)dx = \int_{x\in[0,\infty)} \int_{y\in[0,x]} f_X(x)dydx = \int_{y\in[0,\infty)} \int_{x\in[y,\infty)} f_X(x)dxdy \\ &= \int_0^\infty \mathbb{P}(X > y)dy\end{aligned}$$

where the final integral is still an improper Riemann integral since $y \mapsto \mathbb{P}(X > y)$ is monotonic and hence Riemann integrable on $[0, x]$ for all $x \geq 0$, the limit $x \to \infty$ exists as an increasing limit and coincides with the Lebesgue integral by the Monotone Convergence Theorem.

We can now rephrase and prove Theorem E.3:

Theorem G.9. *Let X be a continuous random variable with p.d.f. f_X, and let $h\colon \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. Then*

$$\mathbb{E}[h(X)] = \int_\mathbb{R} h(x) f_X(x)dx$$

provided that the integral $\int_\mathbb{R} |h(x)| f_X(x)dx$ is finite.

Proof. Again, the proof sketched in Prelims Probability for the case of non-negative h interchanges the order of integrals. Using Theorem E.1 and Theorem G.7, and applying

Tonelli's theorem to the measurable function $(x,y) \mapsto f_X(x)\mathbf{1}_H(x,y)$ with $H = \{(x,y) \in \mathbb{R}\times[0,\infty)\colon h(x) > y\}$, we get

$$\mathbb{E}[h(X)] = \int_0^\infty \mathbb{P}(h(X) > y)dy = \int_0^\infty \mathbb{P}(X \in h^{-1}(y,\infty)) = \int_{y\in[0,\infty)}\int_{x\in h^{-1}(y,\infty)} f_X(x)dxdy$$
$$= \int_{x\in\mathbb{R}}\int_{y\in[0,h(x))} f_X(x)dydx = \int_{x\in\mathbb{R}} h(x)f_X(x)dx.$$

Note that $h(X)$ does not have to be a continuous random variable if we use the first step as the definition of $\mathbb{E}[h(X)]$. For general h, the above applies with h replaced by $h^+ = \max\{h,0\}$ and $h^- = \max\{-h,0\}$. Since $h = h^+ - h^-$, this completes the proof by linearity of expectation and integration. □

Remark G.10. The same proof also establishes Theorem D.62 and its generalisation to Borel measurable $h\colon \mathbb{R}^2 \to \mathbb{R}$, i.e. for functions h such that $h^{-1}(I)$ is in the σ-algebra on $\mathbb{R}^2$ generated by rectangles in $\mathbb{R}^2$. Specifically, we establish as in Theorem G.7 that

$$\mathbb{P}(h(X,Y) > z) = \int_{(x,y)\in h^{-1}(z,\infty)} f_{X,Y}(x,y)d(x,y),$$

and Tonelli's theorem, which generalises to $\mathbb{R}^3$ (or to $\mathbb{R}\times\mathbb{R}^2$), applies to the resulting triple integrals.

G.3 Some integrals

The final remaining detail from Prelims Probability that we postponed is the identification of the normalisation of the standard normal distribution. In Corollary C.10, we showed that

$$\int_{-\infty}^\infty e^{-x^2/2}dx = \sqrt{2}\Gamma\left(\frac{1}{2}\right), \qquad \text{where } \Gamma\left(\frac{1}{2}\right) = \int_0^\infty u^{-1/2}e^{-u}du.$$

In Prelims Probability, the standard expression $\sqrt{2\pi}$ was proved using a change to polar coordinates. Since we have not yet rigorously established the two-dimensional substitution rule, we instead give a proof based on Fubini's theorem.

Theorem G.11. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Proof. By Tonelli's theorem, $f(u,v) = u^{-1/2}e^{-u}v^{-1/2}e^{-v}$ is Lebesgue integrable over $(0,\infty)^2$. Writing this as a repeated integral, we can apply the one-dimensional substitution rule with $g_1\colon (u,\infty) \to (0,\infty)$, $v = g_1(w) = w - u$, $g_1'(w) = 1$. As this preserves integrability, Fubini's theorem then allows us to change the order of integration to obtain

$$\left(\Gamma\left(\frac{1}{2}\right)\right)^2 = \int_{u\in(0,\infty)}\int_{v\in(0,\infty)} f(u,v)dvdu = \int_{u\in(0,\infty)}\int_{w\in(u,\infty)} f(u,w-u)dwdu$$
$$= \int_{w\in(0,\infty)}\int_{u\in(0,w)} u^{-1/2}(w-u)^{-1/2}e^{-w}dwdu.$$

Applying the one-dimensional substitution rule with $g_2\colon (0,1) \to (0,w)$, $u = g_2(s) = sw$, $g_2'(s) = w$, the integrand factorises and we obtain

$$\left(\Gamma\left(\frac{1}{2}\right)\right)^2 = \int_{w\in(0,\infty)}\int_{s\in(0,1)} s^{-1/2}(1-s)^{-1/2}e^{-w}dwds$$

$$= \left(\int_{(0,1)} s^{-1/2}(1-s)^{-1/2} ds \right) \left(\int_{(0,\infty)} e^{-w} dw \right).$$

By the Monotone Convergence Theorem, we can view both integrals as improper Riemann integrals and apply the fundamental theorem of calculus to find that

$$\left(\Gamma\left(\frac{1}{2}\right) \right)^2 = \lim_{\varepsilon \downarrow 0} \lim_{t \uparrow 1} \left(\arcsin \sqrt{t} - \arcsin \sqrt{\varepsilon} \right) \lim_{z \to \infty} (1 - e^{-z}) = \pi.$$

Since $\Gamma(\frac{1}{2}) \geq 0$, this completes the proof. □

Together with Corollary C.10, this immediately yields the desired normalisation constant of the standard normal distribution.

Corollary G.12. *We have* $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$.

G.4 Integration theory for σ-finite measure spaces

Picking up on Remark F.28, we consider a σ-finite measure space $(\Omega, \mathcal{F}, \mu)$, i.e. a set Ω with a σ-algebra $\mathcal{F}$ and a set function $\mu \colon \mathcal{F} \to [0, \infty]$ satisfying

1. $\mu(\emptyset) = 0$.
2. If $E_n \in \mathcal{F}$, $n \geq 1$, and $E_n \cap E_k = \emptyset$ when $n \neq k$, then $\mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$.
3. There are $A_n \in \mathcal{F}$ with $\bigcup_{n=1}^{\infty} A_n = \Omega$ and $\mu(A_n) < \infty$ for all $n \geq 1$.

Example G.13. So far, we have introduced the measure space $(\mathbb{R}, \mathcal{M}_{\text{Leb}}, m)$, which is σ-finite, as we can choose $A_n = [-n, n] \in \mathcal{M}_{\text{Leb}}$, $n \geq 1$. We have also considered general probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$, as well as specific ones to establish the existence of random variables with given distributions. Implicitly, we have also studied the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \#)$ of natural numbers $\mathbb{N}$ equipped with the power set $\mathcal{P}(\mathbb{N})$ and *counting measure* $\#$, for which $\#(A) = \#A$ is the number of elements of $A \subseteq \mathbb{N}$. This measure space is also σ-finite as we can choose $A_n = \{1, \ldots, n\}$, $n \geq 1$.

While we will not require any other σ-finite measure spaces, several product spaces combining these examples will be relevant. The general setup of σ-finite measure spaces is both economical and can be read with particular special cases in mind.

Mimicking Sections F.2 and F.3, we obtain the following theory, which we will specialise to provide justifications to interchange expectation with series and other limits. This is relevant to formalise the (non-examinable) proof of the strong law of large numbers given in Part A Probability, and also to prove the uniqueness and convergence theorems for moment generating functions and characteristic functions.

Definition G.14. A function $f \colon \Omega \to \mathbb{R}$ is *$\mathcal{F}$-measurable* if $f^{-1}(I) \in \mathcal{F}$ for each interval I.

A function $\phi \colon \Omega \to \mathbb{R}$ is *simple* if it is $\mathcal{F}$-measurable and it takes only finitely many real values, i.e. if it can be written as

$$\phi = \sum_{i=1}^{k} c_i \mathbf{1}_{B_i}$$

for some distinct non-zero $c_i \in \mathbb{R}$ and disjoint $B_i \in \mathcal{F}$, $1 \leq i \leq k$.

Proposition G.15. *Let $f, g\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable and $h\colon \mathbb{R} \to \mathbb{R}$ continuous (or Borel measurable). Then $f+g$, fg, $\max\{f,g\}$ and $h \circ f$ are $\mathcal{F}$-measurable.*

Definition G.16 (μ-integral)**.** For a non-negative simple function $\phi = \sum_{i=1}^k c_i \mathbf{1}_{B_i}$,

$$\int_\Omega \phi d\mu := \int_\Omega \phi(\omega)\mu(d\omega) := \sum_{i=1}^k c_i \mu(B_i) \in [0,\infty].$$

For a non-negative $\mathcal{F}$-measurable function $f\colon \Omega \to [0,\infty)$,

$$\int_\Omega f d\mu := \sup\left\{\int_\Omega \phi\colon \phi \text{ simple, } 0 \le \phi \le f\right\} \in [0,\infty] \quad \text{and} \quad \int_E f d\mu := \int_\Omega f \mathbf{1}_E d\mu,\ E \in \mathcal{F}.$$

An $\mathcal{F}$-measurable function $f\colon \Omega \to \mathbb{R}$ is called *μ-integrable over E* if $\int_E |f| d\mu < \infty$, and then

$$\int_E f d\mu := \int_E f^+ d\mu - \int_E f^- d\mu \quad \text{where } f^+ := \max\{f,0\} \text{ and } f^- := \max\{-f,0\}.$$

Theorem G.17 (Monotone Convergence Theorem)**.** *If (f_n) is an increasing sequence of non-negative $\mathcal{F}$-measurable functions and $f = \lim_{n\to\infty} f_n$, then $\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu$.*

Proposition G.18. *The set $\mathcal{L}^1(\mu) := \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ of μ-integrable functions on Ω forms a vector space and the μ-integral is a linear functional on it: $\int_\Omega (\alpha f + \beta g) = \alpha \int_\mathbb{R} f + \beta \int_\mathbb{R} g$.*

Theorem G.19 (Dominated Convergence Theorem)**.** *If (f_n) is a sequence of μ-integrable functions, $A \in \mathcal{F}$ with $\mu(A) = 0$ and f such that $f_n(\omega) \to f(\omega)$ for all $\omega \notin A$, and such that $|f_n| \le g$ for a μ-integrable function g, then f is μ-integrable and $\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu$.*

Theorem G.20 (Differentiability Lemma)**.** *For an interval J, a set $I \in \mathcal{F}$ and $f\colon I \times J \to \mathbb{R}$ such that $\omega \mapsto f(\omega, y)$ is integrable over I for all $y \in J$ and $y \mapsto f(\omega, y)$ is differentiable with $|\frac{\partial f}{\partial y}(\omega, y)| \le g(\omega)$ for all $\omega \in I$ and a μ-integrable function $g\colon \Omega \to [0,\infty)$, the function $F(y) = \int_I f(\omega, y)\mu(d\omega)$ is differentiable on J with $F'(y) = \int_I \frac{\partial f}{\partial y}(\omega, y)\mu(d\omega)$, $y \in J$.*

For σ-finite measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, we define the Cartesian product $\Omega_1 \times \Omega_2$ equipped with the σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ generated by the set $\mathcal{R}$ of all rectangles of the form $A_1 \times A_2$ with $A_i \in \mathcal{F}_i$, $i = 1, 2$. For $A \subseteq \Omega_1 \times \Omega_2$, we define the $(\mu_1 \times \mu_2)$-outer measure of A

$$(\mu_1 \times \mu_2)^*(A) = \inf\left\{\sum_{n=1}^\infty \mu_1(I_1^{(n)})\mu_2(I_2^{(n)})\colon I_i^{(n)} \in \mathcal{F}_i,\ i = 1,2,\ A \subseteq \bigcup_{n=1}^\infty \left(I_1^{(n)} \times I_2^{(n)}\right)\right\}.$$

Then we can show as in the construction of Lebesgue measure from the (Lebesgue) outer measure in Section F.1 that the restriction $\mu_1 \times \mu_2$ of $(\mu_1 \times \mu_2)^*$ to $\mathcal{F}_1 \otimes \mathcal{F}_2$ is a σ-finite measure, which we call the *product measure* of μ_1 and μ_2.

Theorem G.21 (Fubini's Theorem)**.** *Let $f \in \mathcal{L}^1(\mu_1 \times \mu_2)$. Then, there is $A \in \mathcal{F}_2$ with $\mu_2(A) = 0$ such that for all $\omega_2 \notin A$, the function $\omega_1 \mapsto f(\omega_1, \omega_2)$ is μ_1-integrable. Moreover, if $F(\omega_2)$ is defined by $F(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)$ for $\omega_2 \notin A$ and by $F(\omega_2) = 0$ for $\omega_2 \in A$, then F is μ_2-integrable, and*

$$\begin{aligned}\int_{\Omega_2}\left(\int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)\right)\mu_2(d\omega_2) &= \int_{\Omega_1 \times \Omega_2} f d(\mu_1 \times \mu_2) \\ &= \int_{\Omega_1}\left(\int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2)\right)\mu_1(d\omega_1),\end{aligned}$$

where the last repeated integral also exists in the sense described above.

Theorem G.22 (Tonelli's Theorem)**.** *Let $f\colon \Omega_1 \times \Omega_2 \to [0,\infty)$ be a $(\mathcal{F}_1 \otimes \mathcal{F}_2)$-measurable function, and suppose that either of the following repeated integrals is finite:*

$$\int_{\Omega_2} \left(\int_{\Omega_1} |f(\omega_1,\omega_2)| \mu_1(d\omega_1) \right) \mu_2(d\omega_2), \qquad \int_{\Omega_1} \left(\int_{\Omega_2} |f(\omega_1,\omega_2)| \mu_2(d\omega_2) \right) \mu_1(d\omega_1),$$

Then f is $(\mu_1 \times \mu_2)$-integrable. Hence, Fubini's Theorem is applicable to both f and $|f|$.

G.5 Integration theory and expectations

Specifically, we can apply Fubini's and Tonelli's Theorems with $\Omega_1 = \mathbb{N}$, $\mathcal{F}_1$ the power set, μ_1 counting measure $\mu_1(A) = \#A$ and $(\Omega_2, \mathcal{F}_2, \mu_2) = (\Omega, \mathcal{F}, \mathbb{P})$. Integration with respect to counting measure is summation $\int_{\mathbb{N}} f(n)\mu_1(dn) = \sum_{n\in\mathbb{N}} f(n)$. Integration with respect to $\mathbb{P}$ is an alternative approach to expectation setting for $\mathbb{P}$-integrable $X\colon \Omega \to \mathbb{R}$

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \int_\Omega X(\omega)\mathbb{P}(d\omega).$$

In the present notation, the definitions in the preceding section give for simple functions

$$X = \sum_{i=1}^k c_i \mathbf{1}_{B_i}, \qquad B_i \in \mathcal{F}, 1 \le i \le k, \text{ disjoint},$$

the integral $\mathbb{E}[X] := \displaystyle\int_\Omega X(\omega)\mathbb{P}(d\omega) = \sum_{i=1}^k c_i \mathbb{P}(B_i) \in [0,\infty)$, and generalising to non-negative random variables by defining

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(d\omega) := \sup \left\{ \int_\Omega \phi(\omega)\mathbb{P}(d\omega) \colon \phi \text{ simple, } 0 \le \phi \le X \right\} \in [0,\infty].$$

This notion of expectation is clearly compatible with our definition of expectation of discrete random variables for simple functions since $B_i = \{X = c_i\}$, if we choose a representation of X with minimal k, so that c_i, $1 \le i \le k$, are distinct. We claim compatibility for all non-negative random variables. Integration theory provides a Monotone Convergence Theorem, linearity of expectation, order properties and allows to adapt the sandwich arguments of Section E.2 to show our claim for bounded non-negative random variables. Then for all non-negative random variables, $\mathbb{E}[X \wedge m]$ increases to $\mathbb{E}[X]$ in the new sense, by the Monotone Convergence Theorem. In the old sense, order properties yield $\lim_{m\to\infty} \mathbb{E}[X \wedge m] \le \mathbb{E}[X]$. But if this was a strict inequality, there would be a strict inequality for $X_n = 2^{-n}\lfloor 2^n X \rfloor$ instead of X for n sufficiently large. But we can write this limit as a series

$$\lim_{m\to\infty} \mathbb{E}[X_n \wedge m] = \sum_{j=1}^\infty \sum_{k=(j-1)2^n+1}^{j2^n} k2^{-n}\mathbb{P}(X_n = k2^{-n}) = \sum_{k=0}^\infty \mathbb{P}(X_n = k2^{-n}) = \mathbb{E}[X_n]$$

by Corollary C.8, so there cannot be a strict inequality and the general case of our claim follows from the bounded case. We could now restate all propositions and theorems of the preceding section in expectation notation. Here is an example. Others are left to the reader.

Theorem G.23 (Dominated Convergence Theorem)**.** *If $X_n\colon \Omega \to \mathbb{R}$, $n \ge 1$, is a sequence of random variables, $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ and Y such that $X_n(\omega) \to Y(\omega)$ for all $\omega \notin A$, and such that $|X_n| \le Z$ for a random variable Z with finite expectation $\mathbb{E}[Z] < \infty$, then the expectation of Y exists and $\mathbb{E}[X_n] \to \mathbb{E}[Y]$, as $n \to \infty$.*