Extreme Value Distribution Project

Kalle Kytölä

May 31, 2025

This is a blueprint for a project which forms a part of the course *MS-EV0029 Introduction to Formalized Mathematics in Lean* at Aalto University, running from February to June, 2025. The goal is to formalize the Fisher-Tippett-Gnedenko theorem, classifying univariate extreme value distributions. Contributions are welcome by course participants!

WARNING: The blueprint is still work in progress! Contributions by the course participants to improve the blueprint and add details (which are helpful for formalization) are very welcome!

Chapter 1

Cumulative distribution functions

Definition 1.1. A function $F\colon \mathbb{R} \to \mathbb{R}$ is a *cumulative distribution function* (c.d.f.) if

(i) $x \mapsto F(x)$ is increasing;

(ii) $x \mapsto F(x)$ is right-continuous;

(iii) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$.

Lemma 1.2. *If X is a real-valued random variable, then the function $F\colon \mathbb{R} \to \mathbb{R}$ given by $F(x) = \mathsf{P}[X \leq x]$ is a c.d.f.*

Proof. Property (1.) in Definition 1.1 is obvious (by monotonicity of measures) and properties (2.) and (3.) are simple consequences of monotone convergence theorems for probability measures. □

1.1 Degenerate distributions

Definition 1.3. A c.d.f. F is said to be *degenerate* if for every $x \in \mathbb{R}$ we have either $F(x) = 0$ or $F(x) = 1$. Otherwise F is said to be *nondegenerate*.

Lemma 1.4. *F is a degenerate c.d.f. if and only if there exists a $x_0 \in \mathbb{R}$ such that*

$$F(x) = \begin{cases} 0 & \text{for } x < x_0 \\ 1 & \text{for } x \geq x_0. \end{cases}$$

Proof. The "if" direction is clear. To prove the "only if" direction, assume that F is a degenerated c.d.f., and let $x_0 = \inf\{x \in \mathbb{R} \mid F(x) = 1\}$. Then it is straightforward to show by properties of a c.d.f. that F has the asserted form. □

Lemma 1.5. *The c.d.f. of Dirac delta mass δ_{x_0} at $x_0 \in \mathbb{R}$ is degenerate.*

Proof. ... □

Lemma 1.6. *If a c.d.f. F is degenerate, then it is the c.d.f. of a Dirac delta mass δ_{x_0} at some point $x_0 \in \mathbb{R}$.*

Proof. ... □

1.2 Distributions of maxima of independent random variables

Lemma 1.7. *Let X and Y be two independent real-valued random variables with respective cumulative distribution functions F and G, i.e. $F(x) = \mathsf{P}[X \le x]$ and $G(x) = \mathsf{P}[Y \le x]$. Then the c.d.f. of $M = \max(X, Y)$ is $x \mapsto F(x)\, G(x)$.*

Proof. Fix $x \in \mathbb{R}$. Note that $\max(X, Y) \le x$ if and only if both $X \le x$ and $Y \le x$. Calculate, using independence,

$$\mathsf{P}\big[\max(X,Y) \le x\big] \;=\; \mathsf{P}\big[X \le x \;\; Y \le x\big] \;=\; \mathsf{P}\big[X \le x\big]\,\mathsf{P}\big[Y \le x\big] \;=\; F(x)\,G(x).$$

□

Lemma 1.8. *Let $X_0, X_1, \ldots, X_{n-1}$ be independent identically distributed real-valued random variables with cumulative distribution functions F, i.e. $F(x) = \mathsf{P}[X_j \le x]$ for every j. Then the c.d.f. of*

$$M_n = \max_{0 \le j < n} X_j$$

is the function $x \mapsto \big(F(x)\big)^n$.

Proof. Induction on n using 1.7. □

Lemma 1.9. *Let $X_0, X_1, \ldots, X_{n-1}$ be independent identically distributed real-valued random variables with cumulative distribution functions F, i.e. $F(x) = \mathsf{P}[X_j \le x]$ for every j, and let $a > 0$ and $b \in \mathbb{R}$. Then the c.d.f. of*

$$\hat{M}_n = \frac{\max_{0 \le j < n} X_j - b}{a}$$

is the function $x \mapsto \big(F(ax + b)\big)^n$.

Proof. Use 1.8 and do a change of variables. □

1.3 Distributions of minima of independent random variables

1.4 Equivalence classes modulo affine transformations

Definition 1.10. The collection of all transformations $\mathbb{R} \to \mathbb{R}$ of the form $x \mapsto ax + b$, where $a > 0$, $b \in \mathbb{R}$, forms a group. We call this the *orientation preserving affine isomorphism group* and denote it by $\mathrm{Aff}^+_{\mathbb{R}}$.

Definition 1.11. The action of an orientation preserving affine isomorphism $A \in \mathrm{Aff}^+_{\mathbb{R}}$ on a cumulative distribution function F is defined so that $A.F \colon \mathbb{R} \to \mathbb{R}$ is given by $(A.F)(x) = F(A^{-1}(x))$. Then $A.F$ is also a c.d.f.

Lemma 1.12. *The actions of orientation preserving affine isomorphisms on a cumulative distribution functions is a group action, i.e., $1.F = F$ and $(AB).F = A.(B.F)$ for any c.d.f. F and any $A, B \in \mathrm{Aff}^+_{\mathbb{R}}$.*

Proof. Direct calculations. □

Lemma 1.13. *Let F be a cumulative distribution function and $A \in \mathrm{Aff}^+_{\mathbb{R}}$ an orientation preserving affine isomorphism. Then $A.F$ is degenerate if and only if F is degenerate.*

Proof. Straightforward from the definitions. □

Lemma 1.14. *Let F be a cumulative distribution function, and $A \in \mathrm{Aff}^+_{\mathbb{R}}$ an orientation preserving affine isomorphism. If a point $x \in \mathbb{R}$ is a continuity point of F, then the point $A(x) \in \mathbb{R}$ is a continuity point of $A.F$.*

Proof. Straightforward. □

1.5 Miscellaneous results on cumulative distribution functions

Lemma 1.15 (Continuity points of c.d.f.s are those which carry no point mass)**.** *Let F be cumulative distribution function of a probability measure μ on $\mathbb{R}$. A point $x \in \mathbb{R}$ is a continuity point of F if and only if $\mu[\{x\}] = 0$.*

Proof. A c.d.f. is always continuous from the right.

Continuity of F from the left at x means that for any sequence $(x_n)_{n\in\mathbb{N}}$ increasing to x (i.e., $x_n \leq x_{n+1} < x$ for all $n \in \mathbb{N}$)

$$F(x_n) \to F(x),$$

or equivalently in terms of measures

$$\mu[(-\infty, x_n]] \to \mu[(-\infty, x]].$$

But by monotone convergence of measures, we always have

$$\begin{aligned} \mu[(-\infty, x_n]] \to\ & \mu[(-\infty, x)] \\ &= \mu[(-\infty, x]] - \mu[\{x\}]. \end{aligned}$$

A comparison of these conditions shows that F is also continuous from the left at x if and only if $\mu[\{x\}] = 0$. □

Lemma 1.16 (A pair of nontrivial continuity points of nondegenerate c.d.f.)**.** *Let G be a nondegenerate c.d.f. Then there exists continuity points $x_1 < x_2$ of G such that $0 < G(x_1) \leq G(x_2) < 1$.*

Proof. Since G is nondegenerate, there exists some $x_0 \in \mathbb{R}$ such that $0 < G(x_0) < 1$. Since G is continuous from the right, for some small $\delta > 0$ we have that $0 < G(x) < 1$ for all $x \in [x_0, x_0 + \delta)$. Since the continuity points of G are dense, from any nonempty open interval we may pick a continuity point. First pick a continuity point $x_1 \in (x_0, x_0 + \delta)$, and then pick another continuity point $x_2 \in (x_1, x_0 + \delta)$. □

Lemma 1.17 (Equality of c.d.f.s on a dense set suffices)**.** *Suppose that F, G are two c.d.f.s and $S \subseteq \mathbb{R}$ is a dense subset of the real line. If $F(\xi) = G(\xi)$ for all $\xi \in S$, then we have $F = G$.*

Proof. We must prove that for any $x \in \mathbb{R}$ we have $F(x) = G(x)$. But by right-continuity of c.d.f.s, density of S, and coincidence of F and G on S, we have

$$\begin{aligned} F(x) &= \lim_{\xi \to x^+ \text{ along } S} F(\xi) \\ &= \lim_{\xi \to x^+ \text{ along } S} G(\xi) = G(x). \end{aligned}$$

□

1.6 Topology on orientation-preserving affine isomorphisms

Definition 1.18. We equip the space $\mathrm{Aff}^+_{\mathbb{R}}$ of orientation-preserving affine isomorphisms with the topology of pointwise convergence, i.e., with the coarsest topology which makes the evaluations $A \mapsto A(x)$ continuous for all $x \in \mathbb{R}$.

Lemma 1.19 (The coefficients of affine map depend continuously on the map)**.** *The coefficients a and b of an orientation-preserving affine isomorphism $A(x) = ax + b$ depend continuously on A.*

Proof. We may first write $a = A(1) - A(0)$ and $b = A(0)$. These depend continuously on A, since the evaluations $A(1)$ and $A(0)$ do. □

Lemma 1.20 (Metrizability of the topology on oriented affine isomorphisms)**.** *The topology of pointwise convergence makes $\mathrm{Aff}^+_{\mathbb{R}}$ homeomorphic to $\mathbb{R}^2$, and in particular metrizable.*

Proof. The essential claim is that the function

$$\mathrm{cfs}\colon \mathrm{Aff}^+_{\mathbb{R}} \to (0, +\infty) \times \mathbb{R}$$

obtained by mapping $A(x) = ax + b$ to its coefficients (a, b) is a homeomorphism. (The homeomorphism to $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ follows by combining with the homeomorphism $(a, b) \mapsto (\log a, b)$.)

The continuity of cfs follows from Lemma 1.19. For the continuity of the inverse, we must only check that for any $x \in \mathbb{R}$, its composition with the point evaluation at x is continuous. But the composition is $(a, b) \mapsto ax + b$, and the continuity is clear. □

Lemma 1.21 (Inversion of orientation preserving affine isomorphisms is continuous)**.** *The map $A \mapsto A^{-1}$ is continuous on $\mathrm{Aff}^+_{\mathbb{R}}$. In particular, inversion defines a homeomorphism of $\mathrm{Aff}^+_{\mathbb{R}}$ to itself.*

Proof. Calculate and use Lemma 1.19. □

Lemma 1.22 (The action of oriented affine transforms on c.d.f.s is continuous)**.** *The action $A.F$ of $A \in \mathrm{Aff}^+_{\mathbb{R}}$ on a c.d.f F depends jointly continuously on A and F.*

(The topology on c.d.f.s is the topology of convergence in distribution, i.e., convergence at all continuity points of the limit cdf.)

Proof. The spaces are metrizable, so it suffices to check sequential continuity.

Suppose that $A_n \to B$ (oriented affine isomophisms) and $F_n \xrightarrow{\mathrm{d}} G$ (c.d.f.s) as $n \to \infty$. Let $x \in \mathbb{R}$ be a continuity point of $B.G$. Let $\varepsilon > 0$. Note that $B^{-1}(x)$ is a continuity point of G. Then there exists a $\delta > 0$ such that $|G(y) - G\big(B^{-1}(x)\big)| < \frac{\varepsilon}{2}$ when $|y - B^{-1}(x)| < \delta$.

By density of continuity points of G, pick continuity points y_-, y_+ such that

$$B^{-1}(x) - \delta \ < \ y_- \ < \ B^{-1}(x) \ < \ y_+ \ < \ B^{-1}(x) + \delta.$$

Since $F_n \xrightarrow{\mathrm{d}} G$, we have that $F_n(y_\pm) \to G(y_\pm)$ as $n \to \infty$, and in particular there exists some N such that for $n \geq N$ we have $\left|F_n(y_\pm) - G(y_\pm)\right| < \frac{\varepsilon}{4}$ for both $y_\pm$.

By Lemma 1.21 and $A_n \to B$, we get that $A_n^{-1} \to B^{-1}$, and in particular there exists some N' such that for $n \geq N'$ we have

$$y_- < A_n^{-1}(x) < y_+.$$

For $n \geq \max\{N, N'\}$, we then have

$$\begin{aligned}
(A_n.F_n)(x) &= F_n(A_n^{-1}(x)) \\
&\leq F_n(y_+) \\
&< G(y_+) + \frac{\varepsilon}{4} \\
&\leq G(B^{-1}(x)) + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} \\
&= (B.G)(x) + \frac{3\varepsilon}{4} \\
&< (B.G)(x) + \varepsilon
\end{aligned}$$

and similarly

$$\begin{aligned}
(A_n.F_n)(x) &= F_n(A_n^{-1}(x)) \\
&\geq F_n(y_-) \\
&> G(y_-) - \frac{\varepsilon}{4} \\
&\geq G(B^{-1}(x)) - \frac{\varepsilon}{2} - \frac{\varepsilon}{4} \\
&= (B.G)(x) - \frac{3\varepsilon}{4} \\
&> (B.G)(x) + \varepsilon,
\end{aligned}$$

which together yield $\left|(A_n.F_n)(x) - (B.G)(x)\right| < \varepsilon$.

Since x was an arbitrary continuity point of $B.G$ and $\varepsilon > 0$ was arbitrary, this proves that $A_n.F \xrightarrow{\mathrm{d}} B.G$. □

Chapter 2

Extreme value distributions

2.1 Definition of extreme value distributions

Lemma 1.9 motivates the following definition.

Definition 2.1. A c.d.f. G is said to be an *extreme value distribution* if G is nondegenerate and there exists a c.d.f. F and a sequence $(A_n)_{n\in\mathbb{N}}$ of orientation preserving affine isomorphisms $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$, such that for every continuity point $x \in \mathbb{R}$ of G we have

$$\lim_{n\to\infty} \big((A_n.F)(x)\big)^n \;=\; G(x).$$

Lemma 2.2. *Let G be an extreme value distribution and and $A \in \mathrm{Aff}^+_{\mathbb{R}}$ an orientation preserving affine isomorphism. Then also $A.G$ is an extreme value distribution.*

Proof. Straightforward using Lemmas 1.14 and 1.13. □

2.2 Three types of extreme value distributions

Definition 2.3. *The standard Gumbel distribution* is the c.d.f. Λ given by

$$\Lambda(x) = \exp\big(-\exp(-x)\big).$$

(In the parametrization of extreme value distribution types by one index $\gamma \in \mathbb{R}$, this case corresponds to $\gamma = 0$.)

Definition 2.4. *The standard (reverse) Weibull distribution* of parameter $\alpha > 0$ is the c.d.f. Ψ_α given by

$$\Psi_\alpha(x) = \begin{cases} \exp\big(-(-x)^\alpha\big) & \text{for } x < 0 \\ 1 & \text{for } x \geq 0. \end{cases}$$

(In the parametrization of extreme value distribution types by one index $\gamma \in \mathbb{R}$, this case corresponds to $\gamma < 0$ via $\gamma = -1/\alpha$.)

Definition 2.5. *The standard Fréchet distribution* of parameter $\alpha > 0$ is the c.d.f. Φ_α given by

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \exp\big(-x^{-\alpha} \big) & \text{for } x > 0. \end{cases}$$

(In the parametrization of extreme value distribution types by one index $\gamma \in \mathbb{R}$, this case corresponds to $\gamma > 0$ via $\gamma = 1/\alpha$.)

Theorem 2.6. *The standard Gumbel distribution Λ is an extreme value distribution.*

Proof. Set $A_n(x) = x - \log(n)$ for $n \in \mathbb{N}$. Then $A_n^{-1}(x) = x + \log(n)$ and for any $n \geq 1$ and $x \in \mathbb{R}$ we get

$$\begin{aligned} \big((A_n.\Lambda)(x)\big)^n &= \big(\Lambda(x + \log(n))\big)^n \\ &= \Big(\exp\big(-\exp(-(x + \log n))\big)\Big)^n \\ &= \exp\big(-n \exp(-x - \log n)\big) \\ &= \exp\big(-n\, e^{-x}\, e^{-\log n}\big) \\ &= \exp\big(-e^{-x}\big) \\ &= \Lambda(x). \end{aligned}$$

Since the above is true for each n, we in particular have

$$\lim_{n\to\infty} \big((A_n.\Lambda)(x)\big)^n = \Lambda(x)$$

for all $x \in \mathbb{R}$. Since Λ is also nondegenerate, this shows that it is an extreme value distribution. □

Theorem 2.7. *For any $\alpha > 0$, the standard Weibull distribution Ψ_α is an extreme value distribution.*

Proof. ... □

Theorem 2.8. *For any $\alpha > 0$, the standard Fréchet distribution Φ_α is an extreme value distribution.*

Proof. ... □

2.3 Equivalent formulations of the limit relation

Lemma 2.9 (Logarithmic version of the limit relation)**.** *Let F and G be c.d.f.s, and $(A_n)_{n\in\mathbb{N}}$ a sequence of orientation preserving affine isomorphisms $A_n \in \mathrm{Aff}_{\mathbb{R}}^{+}$. Then for any $x \in \mathbb{R}$ such that $0 < G(x) < 1$, the two conditions*

$$\textit{(i)} \qquad \lim_{n\to\infty} \big((A_n.F)(x)\big)^n = G(x)$$

$$\textit{(ii)} \qquad \lim_{n\to\infty} n \log F(A_n^{-1}(x)) = \log G(x)$$

are equivalent.

Proof. Recall that $(A_n.F)(x) = F(A_n^{-1}(x))$. Then just take logarithms (and use continuity) to get from (i) to (ii), and take exponentials (and use continuity) to get from (ii) to (i). $\square$

Lemma 2.10 (Relation implies F tending to one)**.** *Let F and G be c.d.f.s, and $(A_n)_{n\in\mathbb{N}}$ a sequence of orientation preserving affine isomorphisms $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$. Then for any $x \in \mathbb{R}$ such that $0 < G(x) < 1$, if*

$$\text{(i)} \qquad \lim_{n\to\infty} \big((A_n.F)(x)\big)^n = G(x)$$

holds, then necessarily

$$\lim_{n\to\infty} F(A_n^{-1}(x)) \;=\; 1.$$

Proof. Otherwise $\big(F(A_n^{-1}(x))\big)^n$ would have $0 \neq G(x)$ as an accumulation point, contradicting the assumed limit (i).

To wit, if for some $\delta > 0$ we would have $F(A_n^{-1}(x)) \leq 1-\delta$ for infinitely many n, then $0 \leq \big(F(A_n^{-1}(x))\big)^n \leq (1-\delta)^n$ for those n, and since $(1-\delta)^n \to 0$, we would get $\big(F(A_n^{-1}(x))\big)^n \to 0 \neq G(x)$ along the subsequence of those n; a contradiction. $\square$

Lemma 2.11 (Taylor expansion limit modification)**.** *Let $S \subset \mathbb{R}$ be a subset with $0 \in S$, and let $f_1, f_2 \colon S \to \mathbb{R}$ be functions. Let also $(t_n)_{n\in\mathbb{N}}$ be a sequence in S, and let $(m_n)_{n\in\mathbb{R}}$ be a sequence of real numbers tending to infinity, $\lim_{n\to\infty} m_n = +\infty$. Assume that for $j = 1,2$*

- $f_j(0) = 0$*;*
- *the derivative $f_j'(0)$ exists (derivative taken within the set S)*

Assume further about $j = 1$ that

- $f_1'(0) \neq 0$*;*
- *the limit $\lim_{n\to\infty} (m_n\, f_1(t_n))$ exists;*
- *for any $\delta > 0$ there exists an $\varepsilon > 0$ such that if $|f_1(t)| < \varepsilon$ (with $t \in S$) then $|t| < \delta$.*

Denote $c = \frac{1}{f_1'(0)} \lim_{n\to\infty} (m_n\, f_1(t_n))$. Then we have $\lim_{n\to 0} t_n = 0$ and

$$\lim_{n\to\infty} \Big(m_n\, f_2(t_n)\Big) \;= c\, f_2'(0).$$

Proof. This is in principle straightforward: the assumptions are first checked to imply that $\lim_{n\to 0} t_n = 0$, and then one can just consider the first order Taylor expansions of the functions f_j at 0 given by the assumed existence of the derivatives (the key is $t_n = \frac{c}{m_n} + \mathfrak{o}(\frac{1}{m_n})$). $\square$

Lemma 2.12 (Taylored version of the limit relation)**.** *Let F and G be c.d.f.s, and $(A_n)_{n\in\mathbb{N}}$ a sequence of orientation preserving affine isomorphisms $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$. Then for any $x \in \mathbb{R}$ such that $0 < G(x) < 1$, the two conditions*

$$\text{(ii)} \qquad \lim_{n\to\infty} n \,\log F(A_n^{-1}(x)) \;=\; \log G(x)$$

$$\text{(iii)} \qquad \lim_{n\to\infty} \Big(n\,(1 - F(A_n^{-1}(x)))\Big) \;=\; -\log G(x)$$

are equivalent.

Proof. Both implications (ii) $\Rightarrow$ (iii) and (iii) $\Rightarrow$ (ii) are proven similarly using Lemma 2.11.

Assume (ii). Let $f_1(t) = -\log(1-t)$ and $f_2(t) = t$ and $S = [0,1) \subset \mathbb{R}$, and $m_n = n$ and $t_n = 1 - F(A_n^{-1}(x))$. It is straightforward to check the assumptions of Lemma 2.11, with

$$f_1'(0) \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\Big(-\log(1-t)\Big) \;=\; 1$$

and $f_2'(0) = \mathrm{id}'(0) = 1$. The key assumption about the existence of the limit $\lim_{n\to\infty}(m_n\, f_1(t_n))$ is given by (ii), and the conclusion is (iii).

Similarly assuming (iii) we derive (ii) with Lemma 2.11 just interchanging the roles of the two functions, i.e., now using $f_1(t) = t$ and $f_2(t) = -\log(1-t)$ instead. □

Lemma 2.13 (Inverted Taylored version of the limit relation)**.** *Let F and G be c.d.f.s, and $(A_n)_{n\in\mathbb{N}}$ a sequence of orientation preserving affine isomorphisms $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$. Then for any $x \in \mathbb{R}$ such that $0 < G(x) < 1$, the two conditions*

(iii)
$$\lim_{n\to\infty}\Big(n\,(1 - F(A_n^{-1}(x)))\Big) \;=\; -\log G(x)$$

(iv)
$$\lim_{n\to\infty} \frac{1}{n\,(1 - F(A_n^{-1}(x)))} \;=\; \frac{1}{-\log G(x)}$$

are equivalent.

Proof. By assumption $G(x) \in (0,1)$ we have $-\log G(x) > 0$.

The implication (iii) $\Rightarrow$ (iv) can therefore be seen by applying $t \mapsto \frac{1}{t}$ and using its continuity at $t = -\log G(x)$, and the converse implication (iv) $\Rightarrow$ (iii) similarly by using continuity of $t \mapsto \frac{1}{t}$ at $t = \frac{1}{-\log G(x)}$. □

Lemma 2.14 (Transformed version of the limit relation)**.** *Let F and G be c.d.f.s, and $(A_n)_{n\in\mathbb{N}}$ a sequence of orientation preserving affine isomorphisms $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$. Then for any $x \in \mathbb{R}$ such that $0 < G(x) < 1$, the two conditions*

(iv)
$$\lim_{n\to\infty} \frac{1}{n\,(1 - F(A_n^{-1}(x)))} \;=\; \frac{1}{-\log G(x)}$$

(v)
$$\lim_{n\to\infty} \frac{1}{n}\,\frac{\mathbf{1}}{\mathbf{1} - \widetilde{A_n.F}}(x) \;=\; \frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{G})}(x)$$

are equivalent.

(See Definitions 5.3 and 5.5 for the transforms involved in condition (v)).

Proof. This is in principle straightforward, although certain cases need to be checked separately and the continuity of various natural extensions need to addressed. □

Theorem 2.15 (Equivalent versions of the limit relation)**.** *Let F and G be c.d.f.s, and $(A_n)_{n\in\mathbb{N}}$ a sequence of orientation preserving affine isomorphisms $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$. Then for any $x \in \mathbb{R}$ such* *that* $0 < G(x) < 1$, *the conditions*

$$\begin{aligned}
&\textit{(i)} && \lim_{n\to\infty} \left((A_n.F)(x)\right)^n &&= G(x)\\
&\textit{(ii)} && \lim_{n\to\infty} n\,\log F(A_n^{-1}(x)) &&= \log G(x)\\
&\textit{(iii)} && \lim_{n\to\infty} \Big(n\,(1-F(A_n^{-1}(x)))\Big) &&= -\log G(x)\\
&\textit{(iv)} && \lim_{n\to\infty} \frac{1}{n\,(1-F(A_n^{-1}(x)))} &&= \frac{1}{-\log G(x)}\\
&\textit{(v)} && \lim_{n\to\infty} \frac{1}{n}\frac{\mathbf{1}}{\mathbf{1}-\widetilde{A_n.F}}(x) &&= \frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{G})}(x)
\end{aligned}$$

are equivalent.

(See Definitions 5.3 and 5.5 for the transforms involved in condition (v)).

(More equivalent conditions are to be added; this is just a theorem to collect various equivalent phrasings.)

Proof. This is just a combination of earlier results. □

Chapter 3

Classification of extreme value distributions

3.1 Auxiliary classification I

Lemma 3.1 (Second order differential equation for Q)**.** *Suppose that*

$$Q\colon \mathbb{R} \to \mathbb{R}$$

is differentiable and satisfies

$$Q(0) = 0 \qquad \textit{and} \qquad Q'(0) = 1$$

and

$$Q(h + s) = Q(h)\alpha(s) + Q(s)$$

for some $\alpha\colon \mathbb{R} \to \mathbb{R}$ and every $s, h \in \mathbb{R}$. Then Q is twice continuously differentiable and satisfies

$$Q''(s) = Q'(s)\, Q''(0) \qquad \textit{for every } s \in \mathbb{R}. \tag{3.1}$$

Proof. Note first that the equation implies (rearranging and dividing by h), for any s and $h \neq 0$,

$$\frac{Q(h+s) - Q(s)}{h} = \frac{Q(h)}{h}\alpha(s).$$

Taking the limit as $h \to 0$ and using $Q(0) = 0$ yields $Q'(s) = Q'(0)\,\alpha(s) = \alpha(s)$, where we also took into account $Q'(0) = 1$. Therefore necessarily $\alpha = Q'$, and the equation can be rewritten in the form

$$Q(h + s) = Q(h)\, Q'(s) + Q(s) \qquad \text{for } s, h \in \mathbb{R}.$$

Rearranging the equation, we find

$$Q(h + s) - Q(s) = Q(h)\, Q'(s)$$

and interchanging the role of s and h also

$$Q(h + s) - Q(h) = Q(s)\, Q'(h).$$

Subtracting the last two equations yields

$$Q(s) - Q(h) = Q(s)\,Q'(h) - Q(h)\,Q'(s),$$

which by rearranging and dividing by $h \neq 0$ yields

$$\frac{Q(h)}{h}(Q'(s) - 1) = Q(s)\frac{Q'(h) - 1}{h}.$$

Taking the limit $h \to 0$, recalling $Q(0) = 1$ and $Q'(0) = 1$, gives the existence of the second derivative $Q''(0)$ and the equation

$$Q'(s) - 1 = Q(s)Q''(0).$$

Solving $Q'(s) = 1 + Q''(0)\,Q(s)$ and recalling that Q is differentiable shows that Q' is also differentiable, so Q is indeed twice differentiable. Differentiating, we get the asserted equation

$$Q''(s) = Q'(s)\,Q''(0).$$

Since Q' is differentiable and in particular continuous, this also shows that Q'' is continuous, i.e., that Q is twice continuously differentiable. □

Lemma 3.2 (Solution for Q)**.** *Suppose that $Q\colon \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable and Q' is positive and Q and satisfies $Q(0) = 0$, $Q'(0) = 1$, and the equation concluded in Lemma 3.1 with $\gamma = Q''(0)$, i.e.,*

$$Q''(s) = \gamma\,Q'(s) \qquad \text{for every } s \in \mathbb{R}. \tag{3.2}$$

Then Q is given by

$$Q(s) = \begin{cases} \frac{e^{\gamma s}-1}{\gamma} & \text{if } \gamma \neq 0 \\ s & \text{if } \gamma = 0 \end{cases} \qquad \text{for } s \in \mathbb{R}.$$

Proof. Since Q is differentiable and $Q'(s) > 0$ for any $s \in \mathbb{R}$, we can write (3.2) as

$$\frac{\mathrm{d}}{\mathrm{d}s}\log Q'(s) = \frac{Q''(s)}{Q'(s)} = \gamma.$$

Integrating and noting $\log Q'(0) = \log 1 = 0$, this yields $\log Q'(s) = \gamma s$ for all $s \in \mathbb{R}$, or equivalently $Q'(s) = e^{\gamma s}$. Integrating a second time and noting $Q(0) = 0$, this yields the desired formulas in the two cases $\gamma \neq 0$ and $\gamma = 0$, respectively, as follows.

If $\gamma \neq 0$, then we get

$$Q(s) = \int_0^s e^{\gamma u}\,\mathrm{d}u = \frac{e^{\gamma s} - 1}{\gamma}.$$

If $\gamma = 0$, then we get

$$Q(s) = \int_0^s e^0\,\mathrm{d}u = s.$$

□

Theorem 3.3 (Monotone functions are a.e. differentiable)**.** *If $f : \mathbb{R} \to \mathbb{R}$ is nondecreasing, then the derivative $f'(x)$ exists at almost every $x \in \mathbb{R}$. In particular there exists points x where $f'(x)$ exists.*

Proof. (The proof is already in Mathlib.) $\square$

Lemma 3.4 (Solution for E)**.** *Suppose that $E\colon (0, \infty) \to \mathbb{R}$ is nondecreasing and nonconstant function which satisfies $E(1) = 0$ and*

$$E(\lambda\sigma) = E(\lambda)A(\sigma) + E(\sigma)$$

for some $A\colon (0, \infty) \to (0, \infty)$ and all $\lambda, \sigma > 0$. Then, denoting $c = E'(1)$ and $\gamma = \frac{1}{E'(1)} \frac{\mathrm{d}^2}{\mathrm{d}s^2} E(e^s)\big|_{s=0}$, for all $\lambda \in \mathbb{R}$ we have

$$E(\lambda) = \begin{cases} c(\lambda^\gamma - 1) & \text{if } \gamma \neq 0. \\ c\,\log(\lambda) & \text{if } \gamma = 0. \end{cases}$$

Proof. Denote $H(s) = E(e^s)$ for $s \in \mathbb{R}$. Then $H(0) = E(1) = 0$ and H is also nondecreasing and nonconstant. By Theorem 3.3, there exists some $s_0 \in \mathbb{R}$ such that the derivative $H'(s_0)$ exists.

The equation for E yields an equation for H,

$$H(h + s) = H(h)A(e^s) + H(s) \qquad \text{for any } s, h \in \mathbb{R}.$$

We can rearrange this equation and divide by h, and we get for any $s, h \in \mathbb{R}$, $h \neq 0$,

$$\frac{H(h + s) - H(s)}{h} = \frac{H(h)}{h} A(e^s).$$

If $s = s_0$, then the LHS tends to $H'(s_0)$ as $h \to 0$. The RHS must therefore also have a limit as $s \to 0$, and observing that $\frac{H(h)}{h} = \frac{H(h)-H(0)}{h}$, that limit is $H'(0)A(e^{s_0})$, which shows that the derivative $H'(0)$ exists. Then applying the same equation at general $s \in \mathbb{R}$ shows that $H'(s) = H'(0)A(e^s)$ must exist, so $H\colon \mathbb{R} \to \mathbb{R}$ is in fact everywhere differentiable.

Note that we must have $H'(0) > 0$, because if $H'(0) = 0$ then by the above equation $H'(s) = 0$ for every s and then H is a constant function, which is a contradiction. Denote $c = H'(0) = \frac{\mathrm{d}}{\mathrm{d}s} E(e^s)\big|_{s=0} = E'(1)$.

Also the equation and positivity of the function A give $H'(s) > 0$ for all s.

Now consider $Q(s) = \frac{H(s)}{c}$. This H is obviously also differentiable, since H is. The equation for H yields the following equation for Q,

$$Q(h + s) = Q(h)\alpha(s) + Q(s) \qquad \text{for any } s, h \in \mathbb{R},$$

where $\alpha(s) = A(e^s)$. By Lemma 3.1 Q is then twice continuously differentiable and satisfies

$$Q'(s)Q''(0) = Q''(s) \qquad \text{for all } s \in \mathbb{R}. \tag{3.3}$$

By Lemma 3.2 we then get a formula for Q, which involves

$$\gamma = Q''(0) = \frac{H''(0)}{c} = \frac{1}{c} \frac{\mathrm{d}^2}{\mathrm{d}s^2} E(e^s)\big|_{s=0}.$$

If $\gamma \neq 0$ then the formula reads $Q(s) = \frac{e^{\gamma s}-1}{\gamma}$. Now just tracing the definitions, we get the asserted formula

$$E(\lambda) = H(\log\lambda) = c\,Q(\log\lambda) = c\,(e^{\gamma\,\log(\lambda)} - 1) = c\,(\lambda^\gamma - 1).$$

If $\gamma = 0$ then the formula reads $Q(s) = s$. Now just tracing the definitions, we get instead

$$E(\lambda) = H(\log\lambda) = c\,Q(\log\lambda) = c\,\log(\lambda),$$

and the proof is complete. $\square$

3.2 Inverting to recover a cumulative distribution function

Chapter 4

Convergence in distribution with cdfs

This chapter provides a standard characterization of convergence in distribution (weak convergence of probability measures) on the real line in terms of cumulative distribution functions.

4.1 Convergence in distribution

Convergence in distribution for random variables can be defined when the random variables take values in a topological space, and it amounts to the weak convergence of the probability measures that are the laws of those random variables. In the special case of real-valued random variables, or probability measures on the real line, the definition reads:

Definition 4.1 (Weak convergence of probability measures)**.** A sequence $(\mu_n)_{n\in\mathbb{N}}$ of Borel probability measures on $\mathbb{R}$ converges weakly to a Borel probability measure μ on $\mathbb{R}$ if for all bounded continuous functions $f\colon \mathbb{R} \to [0, +\infty)$ we have

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_n(x) = \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu(x).$$

4.2 Auxiliary results

Lemma 4.2 (Monotone real functions have only countably many points of discontinuity)**.** *A monotone function $f\colon \mathbb{R} \to \mathbb{R}$ can have at most countably many points of discontinuity. In particular the set $D \subset \mathbb{R}$ of continuity points of f is dense in $\mathbb{R}$.*

Proof. (The proof should already be in Mathlib.) □

Lemma 4.3 (Tightness of a cumulative distribution function)**.** *Let F be a cumulative distribution function. Then for any $\varepsilon > 0$ there exists points $a, b \in \mathbb{R}$ with $a < b$ such that $F(b)-F(a) > 1-\varepsilon$ and F is continuous at the points a and b.*

Proof. Cumulative distribution functions satisfy $F(x) \downarrow 0$ as $x \downarrow -\infty$ and $F(x) \uparrow 1$ as $x \uparrow +\infty$. The required large difference $F(b)-F(a)$ is obtained by choosing a small enough so that $F(a) < \frac{\varepsilon}{2}$ and b large enough so that $F(b) > 1 - \frac{\varepsilon}{2}$. In order to guarantee that $a < b$ and that a and b are continuity points of F, we recall that continuity points of the monotone function F are dense by Lemma 4.2, so we may decrease a and increase b as appropriate. □

Lemma 4.4 (Subdivision with small mesh and within dense set)**.** *Let $D \subset \mathbb{R}$ be a dense set and $a, b \in D$ with $a < b$. Then for any $\delta > 0$ there exists a $k \in \mathbb{N}$ and $a = c_0, c_1, \dots, c_{k-1}, c_k = b \in D$ such that $|c_j - c_{j-1}| < \delta$ for all $j = 1, \dots, k$.*

Proof. ... □

Lemma 4.5 (Subdivision for continuous function approximation)**.** *Let $D \subset \mathbb{R}$ be a dense set, let $f : \mathbb{R} \to \mathbb{R}$ be continuous, let $a, b \in D$ with $a < b$, and let $\varepsilon > 0$. Then there exists a $k \in \mathbb{N}$ and points $a = c_0 < c_1 < \dots < c_{k-1} < c_k = b$ such that for each $j = 1, \dots, k$ we have $c_j \in D$ and*

$$\left|f(x) - f(c_j)\right| < \varepsilon \qquad \text{for} \qquad x \in [c_{j-1}, c_j].$$

Proof. On the compact interval $[a, b] \subset \mathbb{R}$, the continuous function f is uniformly continuous, so for some $\delta > 0$ we have $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$ and $x, y \in [a, b]$. Now apply Lemma 4.4 to choose k and points $a = c_0 < c_1 < \dots < c_{k-1} < c_k$ such that $c_j - c_{j-1} < \delta$ and $c_j \in D$ for all $j = 1, \dots, k$. Now for any $j = 1, \dots, k$, since for $x \in [c_{j-1}, c_j]$ we have $|x - c_j| < \delta$, we get

$$\left|f(x) - f(c_j)\right| < \varepsilon$$

as desired. □

Lemma 4.6 (Simple function integral as linear combination of cdf differences)**.** *Let $a = c_0 < c_1 < \dots < c_k = b$ and consider the linear combination of indicator functions*

$$h(x) = \sum_{j=1}^{k} \alpha_j \, \mathbb{1}_{(c_{j-1}, c_j]}(x).$$

Then the integral of h with respect to a Borel probability measure μ on $\mathbb{R}$ whose can be written as

$$\int_{\mathbb{R}} h(x) \, \mathrm{d}\mu(x) = \sum_{j=1}^{k} \alpha_j \left(F(c_j) - F(c_{j-1})\right),$$

where F is the c.d.f. of μ.

Proof.

$$\begin{aligned}
\int_{\mathbb{R}} h \, \mathrm{d}\mu &= \int_{\mathbb{R}} \Big(\sum_{j=1}^{k} \alpha_j \, \mathbb{1}_{(c_{j-1}, c_j]}(x) \Big) \, \mathrm{d}\mu(x) \\
&= \sum_{j=1}^{k} \alpha_j \int_{\mathbb{R}} \mathbb{1}_{(c_{j-1}, c_j]}(x) \, \mathrm{d}\mu(x) \\
&= \sum_{j=1}^{k} \alpha_j \, \mu\big[(c_{j-1}, c_j]\big] \\
&= \sum_{j=1}^{k} \alpha_j \left(F_n(c_j) - F_n(c_{j-1})\right)
\end{aligned}$$

□

Lemma 4.7 (One of the portmanteau implications)**.** *Weak convergence of probability measures implies that if the boundary of a Borel set carries no probability mass under the limit measure, then the limit of the measures of the set equals the measure of the set under the limit probability measure.*

In other words, if $\lim_{n\to\infty} \mu_n = \mu$ *in the sense of weak convergence of measures, Definition 4.1, and if* $A \subset \mathbb{R}$ *is a Borel set such that* $\mu[\partial A] = 0$*, then*

$$\lim_{n\to\infty} \mu_n[A] = \mu[A].$$

Proof. (The proof is in Mathlib.) □

4.3 Convergence in distribution from pointwise convergence of cdfs

Theorem 4.8 (Sufficient condition for convergence in distribution with cdfs)**.** *Let* F *and* F_n*,* $n \in \mathbb{N}$*, be cumulative distribution functions of probability measures* μ *and* μ_n*,* $n \in \mathbb{N}$*, respectively, i.e.,*

$$F(x) = \mu[(-\infty, x]] \qquad \text{for } x \in \mathbb{R}$$
$$F_n(x) = \mu_n[(-\infty, x]] \qquad \text{for } x \in \mathbb{R} \text{ and } n \in \mathbb{N}.$$

If $\lim_{n\to\infty} F_n(x) = F(x)$ *for all continuity points* x *of* F*, then* $\lim_{n\to\infty} \mu_n = \mu$ *in the sense of weak convergence of measures, Definition 4.1.*

Proof. Let $D \subset \mathbb{R}$ denote the set of continuity points of F. By Lemma 4.2, D is dense in $\mathbb{R}$. Assume that $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in D$.

Let $\varepsilon > 0$. Choose, by Lemma 4.3, points $a, b \in D$, $a < b$, such that $F(b) - F(a) > 1 - \varepsilon$.

Observe also that since $\lim_{n\to\infty} F_n(a) = F(a)$ and $\lim_{n\to\infty} F_n(b) = F(b)$, there exists some N_1 such that we have

$$F_n(b) - F_n(a) > 1 - 2\varepsilon \qquad \text{for all } n \geq N_1.$$

Let $f \colon \mathbb{R} \to \mathbb{R}$ be bounded and continuous. By Lemma 4.5 we can choose points $a = c_0 < c_1 < \cdots < c_{k-1} < c_k = b$ such that for all $j = 1, \ldots, k$ we have $c_j \in D$ and

$$\left|f(x) - f(c_j)\right| < \varepsilon \qquad \text{for} \qquad x \in [c_{j-1}, c_j].$$

Define the simple function $h \colon \mathbb{R} \to \mathbb{R}$ by

$$h(x) = \sum_{j=1}^{k} f(c_j)\, \mathbb{1}_{(c_{j-1}, c_j]}(x)$$

The above estimate shows that $|f(x) - h(x)| < \varepsilon$ for all $x \in [a, b]$. By boundedness of f, there exists a constant $K > 0$ such that $|f(x)| \leq K$ for all $x \in \mathbb{R}$. Since h vanishes outside $(a, b]$, the triangle inequality for integral with respect to μ_n gives

$$\left| \int_{\mathbb{R}} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}} h \, \mathrm{d}\mu_n \right| \leq \underbrace{\int_{(a,b]} |f - h| \, \mathrm{d}\mu_n}_{\leq \varepsilon} + \underbrace{\int_{\mathbb{R}\setminus(a,b]} |f| \, \mathrm{d}\mu_n}_{\leq K\, \mu_n\left[\mathbb{R}\setminus(a,b]\right]} .$$

When $n \geq N_1$, we have $\mu_n[\mathbb{R} \setminus (a,b]] = 1 - \mu_n[(a,b]] = 1 - (F_n(b) - F_n(a)) < 2\varepsilon$, and thus the triangle inequality implies

$$\Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}} h \, \mathrm{d}\mu_n \Big| \leq \varepsilon + K \, 2\varepsilon = (1 + 2K) \, \varepsilon.$$

Similarly, integrating now with respect to μ instead, one shows that

$$\Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu - \int_{\mathbb{R}} h \, \mathrm{d}\mu \Big| \leq (1 + K) \, \varepsilon.$$

It remains to consider the integrals of the function h with respect to both μ_n and μ. By Lemma 4.6, these integrals are expressible in terms of the cumulative distribution functions,

$$\int_{\mathbb{R}} h \, \mathrm{d}\mu_n = \sum_{j=1}^{k} f(c_j) \left(F_n(c_j) - F_n(c_{j-1}) \right)$$

and

$$\int_{\mathbb{R}} h \, \mathrm{d}\mu = \sum_{j=1}^{k} f(c_j) \left(F(c_j) - F(c_{j-1}) \right).$$

The difference of the integrals of h with respect to these two can therefore be estimated as

$$\begin{aligned}
\Big| \int_{\mathbb{R}} h \, \mathrm{d}\mu - \int_{\mathbb{R}} h \, \mathrm{d}\mu_n \Big| &= \Big| \sum_{j=1}^{k} f(c_j) \left(F(c_j) - F_n(c_j) - F(c_{j-1}) + F_n(c_{j-1}) \right) \Big| \\
&\leq \sum_{j=1}^{k} |f(c_j)| \Big(\big|F(c_j) - F_n(c_j)\big| + \big|F(c_{j-1}) + F_n(c_{j-1})\big| \Big) \\
&\leq 2kK \max_{j=0,\ldots,k} \big|F(c_j) - F_n(c_j)\big|.
\end{aligned}$$

By our assumption (ii), we have $\lim_{n\to\infty} F_n(c_j) = F(c_j)$ for each $j = 1, \ldots, k$, so there exists N_2 such that for $n \geq N_2$ we have $\max_{j=1,\ldots,k} |F(c_j) - F_n(c_j)| < \frac{\varepsilon}{k}$, and thus

$$\Big| \int_{\mathbb{R}} h \, \mathrm{d}\mu - \int_{\mathbb{R}} h \, \mathrm{d}\mu_n \Big| \leq 2K\varepsilon.$$

Combining the estimates we have obtained, for $n \geq \max(N_1, N_2)$, we have

$$\begin{aligned}
& \Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu - \int_{\mathbb{R}} f \, \mathrm{d}\mu_n \Big| \\
\leq \; & \underbrace{\Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu - \int_{\mathbb{R}} h \, \mathrm{d}\mu \Big|}_{\leq (1+K)\varepsilon} + \underbrace{\Big| \int_{\mathbb{R}} h \, \mathrm{d}\mu - \int_{\mathbb{R}} h \, \mathrm{d}\mu_n \Big|}_{\leq 2K\varepsilon} + \underbrace{\Big| \int_{\mathbb{R}} h \, \mathrm{d}\mu_n - \int_{\mathbb{R}} f \, \mathrm{d}\mu_n \Big|}_{\leq (1+2K)\varepsilon} \\
\leq \; & (2 + 5K)\varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this shows that $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$ as $n \to \infty$, so we have established the weak convergence $\mu_n \to \mu$ according to Definition 4.1. $\square$

Lemma 4.9 (Necessary condition for convergence in distribution with cdfs)**.** *Let μ and μ_n, $n \in \mathbb{N}$, be Borel probability measures on $\mathbb{R}$, and let F and F_n, $n \in \mathbb{N}$, be their cumulative distribution functions, respectively, i.e.,*

$$\begin{aligned} F(x) &= \mu[(-\infty, x]] && \text{for } x \in \mathbb{R} \\ F_n(x) &= \mu_n[(-\infty, x]] && \text{for } x \in \mathbb{R} \text{ and } n \in \mathbb{N}. \end{aligned}$$

If $\lim_{n\to\infty} \mu_n = \mu$ in the sense of weak convergence of measures, Definition 4.1, then for all continuity points x of F we have $\lim_{n\to\infty} F_n(x) = F(x)$.

Proof. Let $x \in \mathbb{R}$ be a continuity point of F. Then we have $\mu[\{x\}] = 0$, by Lemma 1.15. Note that the boundary of the Borel set $(-\infty, x] \subset \mathbb{R}$ is the singleton $\partial(-\infty, x] = \{x\}$. Therefore the assumption $\lim_{n\to\infty} \mu_n = \mu$ implies that

$$\mu_n[(-\infty, x]] \to \mu[(-\infty, x]],$$

by a general fact (Lemma 4.7) about weakly converging sequences of measures that a for Borel sets whose boundary carries no mass in the limit measure. In terms of the c.d.f.s, the above reads

$$F_n(x) \to F(x)$$

as asserted. □

Chapter 5

Transforms of cumulative distribution functions

In this part, we introduce certain transforms and extensions of cumulative distribution functions, which are used in the classification calculation of the extreme value distributions.

5.1 Extended cumulative distribution function

Definition 5.1 (Extended cumulative distribution function)**.** The extension $\widetilde{F}$ of a c.d.f. F is the function

$$\widetilde{F} \colon [-\infty, +\infty] \to [0, 1]$$

given by

$$\widetilde{F}(x) = \begin{cases} 0 & \text{if } x = -\infty \\ F(x) & \text{if } -\infty < x < +\infty \\ 1 & \text{if } x = +\infty. \end{cases}$$

Lemma 5.2 (Continuity points of extended c.d.f.)**.** *The extension $\widetilde{F}$ of a c.d.f. F is continuous at $x = -\infty$, $x = +\infty$, and at any $x \in (-\infty, +\infty)$ where F is continuous.*

Proof. Since $\lim_{x \to +\infty} F(x) = 1$ by properties of c.d.f.s and $\widetilde{F}(+\infty) = 1$ by definition of the extension, continuity at $x = +\infty$ follows. Continuity at $x = -\infty$ is similar.

Suppose F is continuous at $x \in \mathbb{R}$. Then since $\widetilde{F}$ coincides with F in a neighborhood of x (indeed on the open set $\mathbb{R} \subsetneq [-\infty, +\infty]$), the continuity of F at $x \in \mathbb{R}$ implies continuity of $\widetilde{F}$ at x. $\square$

5.2 One over one minus cumulative distribution function

Definition 5.3 (One over one minus cumulative distribution function)**.** The transform $\frac{\mathbf{1}}{\mathbf{1} - \widetilde{F}}$ of a c.d.f. F is the function

$$\frac{\mathbf{1}}{\mathbf{1} - \widetilde{F}} \colon [-\infty, +\infty] \to [1, +\infty]$$

given by

$$\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}(x)=\begin{cases}1 & \text{if } \widetilde{F}(x)=0\\ \frac{1}{1-\widetilde{F}(x)} & \text{if } 0<\widetilde{F}(x)<1\\ +\infty & \text{if } \widetilde{F}(x)=1,\end{cases}$$

where $\widetilde{F}\colon [-\infty,+\infty]\to[0,1]$ is the extension of the c.d.f. F.

Lemma 5.4 (Continuity points of one over one minus c.d.f.)**.** *The transform $\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}$ of a c.d.f. F is continuous at $x=-\infty$, $x=+\infty$, and at any $x\in(-\infty,+\infty)$ where F is continuous.*

Proof. Since $\lim_{x\to+\infty}\widetilde{F}(x)=\widetilde{F}(+\infty)=1$ by Lemma 5.2 and the continuous extension of $p\mapsto\frac{1}{1-p}$ to a function $[0,1]\to[0,+\infty]$ tends to $+\infty$ as $p\to1$, we have

$$\lim_{x\to+\infty}\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}(x)\;=\;+\infty\;=\;\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}(+\infty).$$

Therefore $\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}$ is continuous at $+\infty$. Continuity at $-\infty$ similarly follows from $\lim_{x\to-\infty}\widetilde{F}(x)=\widetilde{F}(-\infty)=0$ and $\frac{1}{1-p}$ tending to 1 as $p\to0$, which give

$$\lim_{x\to-\infty}\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}(x)\;=\;1\;=\;\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}(-\infty).$$

Suppose F is continuous at $x\in\mathbb{R}$, and recall from Lemma 5.2 that $\widetilde{F}$ is then also continuous at x. Now $\frac{\mathbf{1}}{\mathbf{1}-\widetilde{F}}$ is a composition of the continuous function $p\mapsto\frac{1}{1-p}\colon[0,1]\to[0,+\infty]$ with $\widetilde{F}$, and as such also becomes continuous at x. □

5.3 One over negative logarithm cumulative distribution function

Definition 5.5 (One over negative logarithm cumulative distribution function)**.** The transform $\frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}$ of a c.d.f. F is the function

$$\frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}\colon[-\infty,+\infty]\to[0,+\infty]$$

given by

$$\frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}(x)=\begin{cases}0 & \text{if } \widetilde{F}(x)=0\\ \frac{1}{\log\left(1/\widetilde{F}(x)\right)} & \text{if } 0<\widetilde{F}(x)<1\\ +\infty & \text{if } \widetilde{F}(x)=1,\end{cases}$$

where $\widetilde{F}\colon[-\infty,+\infty]\to[0,1]$ is the extension of the c.d.f. F.

Lemma 5.6 (Continuity points of one over negative logarithm c.d.f.)**.** *The transform $\frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}$ of a c.d.f. F is continuous at $x=-\infty$, $x=+\infty$, and at any $x\in(-\infty,+\infty)$ where F is continuous.*

Proof. Since $\lim_{x\to+\infty} \widetilde{F}(x) = \widetilde{F}(+\infty) = 1$ by Lemma 5.2 and the continuous extension of $p \mapsto \frac{1}{\log(1/p)}$ to a function $[0,1] \to [0,+\infty]$ tends to $+\infty$ as $p \to 1$, we have

$$\lim_{x\to+\infty} \frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}(x) \;=\; +\infty \;=\; \frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}(+\infty).$$

Therefore $\frac{\mathbf{1}}{\widetilde{\log}\left(1/\widetilde{F}\right)}$ is continuous at $+\infty$. Continuity at $-\infty$ similarly follows from $\lim_{x\to-\infty} \widetilde{F}(x) = \widetilde{F}(-\infty) = 0$ and $\frac{1}{\log(1/p)}$ tending to 0 as $p \to 0$, which give

$$\lim_{x\to-\infty} \frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}(x) \;=\; 0 \;=\; \frac{\mathbf{1}}{\widetilde{\log}(1/\widetilde{F})}(-\infty).$$

Suppose F is continuous at $x \in \mathbb{R}$, and recall from Lemma 5.2 that $\widetilde{F}$ is then also continuous at x. Now $\frac{\mathbf{1}}{\widetilde{\log}\left(1/\widetilde{F}\right)}$ is a composition of the continuous function $p \mapsto \frac{1}{\log(1/p)} : [0,1] \to [0,+\infty]$ with $\widetilde{F}$, and as such also becomes continuous at x. $\square$

Chapter 6

One-parameter subgroups of affine isomorphisms

Lemma 6.1 (Functional equation in one parameter subgroups of affine isomorphisms)**.** *Suppose that $t \mapsto A_t$ is a homomorphism of multiplicative groups $(0,+\infty) \to \mathrm{Aff}^+_{\mathbb{R}}$, i.e., for any $s,t > 0$ we have*

$$A_{st} = A_s \circ A_t.$$

Write $A_t(x) = a_t x + b_t$, with $a_t > 0$ and $b_t \in \mathbb{R}$. Then we have, for any $s,t > 0$,

$$\begin{aligned} a_{ts} &= a_t\, a_s \qquad \text{and} \\ b_{ts} &= a_t\, b_s + b_t. \end{aligned}$$

(Also by symmetry $b_{ts} = a_s\, b_t + b_s$.)

Proof. □

Lemma 6.2 (Functional equation scaling coefficient solution)**.** *Suppose that $a\colon (0,+\infty) \to (0,+\infty)$ is a measurable function satisfying, for any $s,t > 0$,*

$$a(ts) = a(t)\, a(s).$$

Then there exists a $\rho \in \mathbb{R}$ such that for all $t > 0$,

$$a(t) = t^{\rho}.$$

Proof. □

Lemma 6.3 (Functional equation translation coefficient solution with $\rho = 0$)**.** *Suppose that $b\colon (0,+\infty) \to \mathbb{R}$ is a measurable function satisfying, for any $s,t > 0$,*

$$b(ts) = b(s) + b(t).$$

Then there exists a constant c such that for $t \in (0,+\infty)$ we have

$$b(t) = -c\, \log(t).$$

Proof. □

Lemma 6.4 (Functional equation translation coefficient solution with $\rho \neq 0$)**.** *Suppose that* $\rho \in \mathbb{R} \setminus \{0\}$ *and* $b \colon (0, +\infty) \to \mathbb{R}$ *is a measurable function satisfying, for any* $s, t > 0$,

$$b(ts) = t^{\rho}\, b(s) + b(t).$$

Then there exists a constant c *such that for* $t \in (0, +\infty) \setminus \{1\}$ *we have*

$$b(t) = c(1 - t^{-\rho}).$$

Proof. □

The above solutions to functional equations can be used to classify all one-parameter subgroups of the group of oriented affine isomorphisms of $\mathbb{R}$. Such a subgroup can be given in a parametrized form as a group homomorphism $\mathbb{R} \to \mathrm{Aff}^{+}_{\mathbb{R}}$ (from the additive group $\mathbb{R}$) or alternatively as group homomorphisms $(0, +\infty) \to \mathrm{Aff}^{+}_{\mathbb{R}}$ (from the multiplicative group $(0, +\infty)$). The additive and multiplicative versions are related by the change of variable $\mathbb{R} \ni t \leftrightarrow \lambda := e^{t} \in (0, +\infty)$ (conversely, $t = \log(\lambda)$). *(In Lean the type* $\mathbb{R}$ *is more convenient than the type* $(0, +\infty)$*, so in formal statements we prefer the former choice.)*

Theorem 6.5 (One-parameter subgroups of affine isomorphisms of $\mathbb{R}$)**.** [TODO: Switch to additive notation and $\mathbb{R}$ rather than multiplicative notation and $(0, +\infty)$, to match the most convenient formalized statements.]

Suppose that $t \mapsto A_t$ *is a measurable homomorphism of multiplicative groups* $(0, +\infty) \to \mathrm{Aff}^{+}_{\mathbb{R}}$, *i.e., for any* $s, t > 0$ *we have*

$$A_{st} = A_s \circ A_t$$

and $A_t(x) = a_t x + b_t$, *with* $t \mapsto a_t$ *and* $t \mapsto b_t$ *measurable functions. Then either*

(0) *there exists a* $\beta \in \mathbb{R}$ *such that for all* $t > 0$ *and* $x \in \mathbb{R}$

$$A_t(x) = x + \beta \log(t);$$

or

(1) *there exists a* $\rho \neq 0$ *and* $c \in \mathbb{R}$ *such that for all* $t > 0$ *and* $x \in \mathbb{R}$

$$A_t(x) = t^{\rho}(x - c) + c.$$

Proof. □

Chapter 7

Types

Definition 7.1 (Type (of distribution on $\mathbb{R}$))**.** Two c.d.f.s F, G are said to be of the same type, if there exists an order-preserving affine isomorphism $A \in \mathrm{Aff}_{\mathbb{R}}^{+}$ such that $G = A.F$.

Being of the same type is an equivalence relation, and the equivalence classes are called types (of distributions on $\mathbb{R}$).

7.1 Convergence to types

The notion of convergence of cumulative distribution functions considered here is always taken to be pointwise convergence on the set of continuity points of the limit c.d.f. By Theorem 4.8 and Lemma 4.9, this corresponds to convergence in distribution (weak convergence of probability measures). Below, when we write $F_n \xrightarrow{\mathrm{d}} F$ for c.d.f.s F_n, $n \in \mathbb{N}$, and F, this is always what we mean.

For affine maps $A \colon \mathbb{R} \to \mathbb{R}$, we use the topology of pointwise convergence of functions. Equivalently, convergence of affine maps means the convergence of the coefficients $a, b \in \mathbb{R}$ in the expression $x \mapsto ax + b$ of the affine functions (so $A_n \to A$ if and only if the functions are of the form $A_n(x) = a_n x + b_n$ and $A(x) = ax + b$, and $a_n \to a$ and $b_n \to b$).

Lemma 7.2 (Unique affine relation among two nondegenerate c.d.f.s)**.** *Let F, G be two c.d.f.s of the same type, and $A \in \mathrm{Aff}_{\mathbb{R}}^{+}$ an affine isomorphism such that $G = A.F$. If F is nondegenerate, then A is the only element of $\mathrm{Aff}_{\mathbb{R}}^{+}$ for which the relation $G = A.F$ holds.*

Proof. Since F is nondegenerate, we can find two different points $x_1 < x_2$ such that $0 < F(x_1) < F(x_2) \leq 1$. By right continuity of F, we can assume these points to be taken minimal with the given values, i.e., $x_j = \inf\{x \in \mathbb{R} \mid F(x) = F(x_j)\}$ for $j = 1, 2$.

The assumption $G = A.F$ means $G(x) = F\big(A^{-1}(x)\big)$ for all $x \in \mathbb{R}$. Therefore $G\big(A(x_1)\big) = F(x_1) < F(x_2) = G\big(A(x_2)\big)$. We also get $A(x_j) = \inf\{y \in \mathbb{R} \mid G(y) = F(x_j)\}$ for $j = 1, 2$ by strict monotonicity and bijectivity of A (if, for example, there would exist a $y_2' < A(x_2)$ such that $G(y_2') = F(x_2)$, then the point $x_2' = A^{-1}(y_2') < x_2$ would be such that $F\big(A^{-1}(y_2')\big) = G(y_2') = F(x_2)$, contradicting the minimality of x_2).

If $\widetilde{A} \in \mathrm{Aff}_{\mathbb{R}}^{+}$ is also such that $G = \widetilde{A}.F$, then the same holds for it: $\widetilde{A}(x_j) = \inf\{y \in \mathbb{R} \mid G(y) = F(x_j)\}$ for $j = 1, 2$. We conclude that

$$\begin{aligned}
\widetilde{A}(x_1) &= \inf\{y \in \mathbb{R} \mid G(y) = F(x_1)\} = A(x_1) \\
\widetilde{A}(x_2) &= \inf\{y \in \mathbb{R} \mid G(y) = F(x_2)\} = A(x_2).
\end{aligned}$$

But an affine map of $\mathbb{R}$ is determined by its values at two distinct points: from $ax_1 + b = y_1$ and $ax_2 + b = y_2$ with $x_1 \neq x_2$ one can solve a, b. Therefore we must have $\widetilde{A} = A$. $\square$

Lemma 7.3 (Degeneration by shrinking affine transformations)**.** *Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of c.d.f.s which converges to a c.d.f. G, $F_n \xrightarrow{\mathrm{d}} G$. Consider affine transformations of the form $A_n(x) = a_n x + b_n$, with $a_n > 0$ and $b_n \in \mathbb{R}$, such that $a_n \to 0$ and $b_n \to \beta \in \mathbb{R}$ as $n \to \infty$. Then $A_n.F_n \xrightarrow{\mathrm{d}} \widetilde{G}$, where $\widetilde{G}$ is the degenerate c.d.f. of the delta mass at β.*

Proof. It suffices to prove that for any $x < \beta$ we have $\widetilde{G}(x) = 0$ and for any $x > \beta$ we have $\widetilde{G}(x) = 1$.

Let us focus on the latter, so let $x > \beta$. Let also $\varepsilon > 0$; we will prove that $\widetilde{G}(x) > 1 - \varepsilon$, and the claim will follow.

By density of continuity points of $\widetilde{G}$, we can choose x' such that $\beta < x' < x$ and $\widetilde{G}$ is continuous at x'. Then the assumed convergence $A_n.F_n \xrightarrow{\mathrm{d}} \widetilde{G}$ implies that $(A_n.F_n)(x') \to \widetilde{G}(x')$. Since $\widetilde{G}(x) \geq \widetilde{G}(x')$, it suffices to prove that $\widetilde{G}(x') > 1 - \varepsilon$.

Since G is a c.d.f., we can choose a continuity point z of G large enough so that $G(z) > 1 - \varepsilon$. Then by the assumed convergence $F_n \xrightarrow{\mathrm{d}} G$, we have $F_n(z) \to G(z)$. By definition we have

$$(A_n.F_n)(A_n(z)) \;=\; F_n\big(A_n^{-1}(A_n(z))\big) \;=\; F_n(z) \;\longrightarrow\; G(z).$$

Note that $A_n(z) = a_n z + b_n \to \beta$ as $n \to \infty$ by the assumptions $a_n \to 0$ and $b_n \to \beta$. In particular, for n large enough, we have $A_n(z) < x'$. Therefore, for n large enough

$$F_n(z) \;=\; (A_n.F_n)(A_n(z)) \;\leq\; (A_n.F_n)(x').$$

The LHS tends to $G(z)$ and the RHS tends to $\widetilde{G}(x')$, showing

$$1 - \varepsilon \;<\; G(z) \;\leq\; \widetilde{G}(x') \;\leq\; \widetilde{G}(x).$$

This concludes the proof that $\widetilde{G}(x) = 1$ for all $x > \beta$.

The proof that $\widetilde{G}(x) = 0$ for all $x < \beta$ is similar. $\square$

Lemma 7.4 (Impossibility of expanding affine transformations)**.** *Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of c.d.f.s which converges to a nondegenerate c.d.f. G, $F_n \xrightarrow{\mathrm{d}} G$. Consider affine transformations of the form $A_n(x) = a_n x + b_n$, with $a_n > 0$ and $b_n \in \mathbb{R}$, such that $a_n \to +\infty$ as $n \to \infty$. Then $A_n.F_n$ cannot converge to any c.d.f.*

Proof. Since G is assumed nondegenerate, by Lemma 1.16 we can pick two continuity points $x_1 < x_2$ of G such that $0 < G(x_1) \leq G(x_2) < 1$. Then from the assumption $F_n \xrightarrow{\mathrm{d}} G$ we get $F_n(x_1) \to G(x_1)$ and $F_n(x_2) \to G(x_2)$.

Assume, by way of contradiction, that we have convergence $A_n.F \xrightarrow{\mathrm{d}} \widetilde{G}$ to some c.d.f. $\widetilde{G}$.

We claim that then $\big(A_n(x_1)\big)_{n\in\mathbb{N}}$ is bounded from below and $\big(A_n(x_2)\big)_{n\in\mathbb{N}}$ is bounded from above. Since

$$a_n = \frac{A_n(x_2) - A_n(x_1)}{x_2 - x_1},$$

this will show that $(a_n)_{n\in\mathbb{N}}$ is bounded from above, contradicting the assumption $a_n \to +\infty$, and finishing the proof.

To show that $(A_n(x_1))_{n\in\mathbb{N}}$ is bounded from below, choose a continuity point z of $\widetilde{G}$ such that $\widetilde{G}(z) < G(x_1)$. Then the assumed convergence $A_n.F \xrightarrow{\mathrm{d}} \widetilde{G}$ implies $(A_n.F)(z) \to \widetilde{G}(z)$. On the other hand, if $(A_n(x_1))_{n\in\mathbb{N}}$ is not bounded from below, then along some subsequence $(n_k)_{k\in\mathbb{N}}$ of indices we have $A_{n_k}(x_1) < z$, and for those indices we then have

$$F_{n_k}(x_1) \;=\; (A_{n_k}.F_{n_k})(A_{n_k}(x_1)) \;\leq\; (A_{n_k}.F_{n_k})(z).$$

The LHS tends to $G(x_1)$ as $k \to \infty$, whereas the RHS tends to $\widetilde{G}(z)$. We get $G(x_1) \leq \widetilde{G}(z)$, contradicting the choice of z. This shows that $(A_n(x_1))_{n\in\mathbb{N}}$ must in fact be bounded from below.

The proof that $(A_n(x_2))_{n\in\mathbb{N}}$ must be bounded from above is similar. □

Theorem 7.5 (Convergence to types)**.** *Suppose that $(F_n)_{n\in\mathbb{N}}$ is a sequence of c.d.f.s which converges to a nondegenerate c.d.f. G, i.e., $F_n \xrightarrow{\mathrm{d}} G$ as $n \to \infty$. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of oriented affine isomorphisms of $\mathbb{R}$, $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$ such that $A_n.F_n \xrightarrow{\mathrm{d}} \widetilde{G}$ for some c.d.f. $\widetilde{G}$.*

If we write $A_n(x) = a_n x + b_n$, then $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are bounded sequences.

If $\widetilde{G}$ is nondegenerate, then $A_n \to A \in \mathrm{Aff}^+_{\mathbb{R}}$ and $A.G = \widetilde{G}$. In particular G and $\widetilde{G}$ are of the same type. Moreover, A is the unique affine transformation for which the equality $A.G = \widetilde{G}$ holds.

Proof. Let us first argue that $(a_n)_{n\in\mathbb{N}}$ are bounded. If not, then by passing to a subsequence, we have $a_{n_k} \to +\infty$. But since $F_{n_k} \xrightarrow{\mathrm{d}} G$ and G is nondegenerate, it contradicts Lemma 7.4 to have $A_{n_k}.F_{n_k} \xrightarrow{\mathrm{d}} \widetilde{G}$. Therefore $(a_n)_{n\in\mathbb{N}}$ must be bounded: there exists some $M > 0$ such that $a_n \leq M$ for all $n \in \mathbb{N}$.

Let us then argue that $(b_n)_{n\in\mathbb{N}}$ is bounded. If not, then we can extract a subsequence such that either $b_{n_k} \to -\infty$ or $b_{n_k} \to +\infty$. Let us prove the impossibility of the second one, the first is similar. So assume that $b_{n_k} \to +\infty$. Since G is nondegenerate, we may pick a continuity point x_0 of G such that $0 < G(x_0) < 1$. Then we have $F_n(x_0) \to G(x_0)$ by the assumption $F_n \xrightarrow{\mathrm{d}} G$. We may also pick a continuity point z of $\widetilde{G}$ such that $\widetilde{G}(z) > G(x_0)$. Then we have $(A_n.F_n)(z) \to \widetilde{G}(z)$ by the assumption $A_n.F_n \xrightarrow{\mathrm{d}} \widetilde{G}$. But $A_{n_k}(x_0) = a_{n_k}x_0 + b_{n_k} \to +\infty$, since $0 < a_{n_k} \leq M$ and $b_{n_k} \to +\infty$. Therefore we have for all large enough k that $A_{n_k}(x_0) > z$. And then

$$(A_{n_k}.F_{n_k})(z) \;\leq\; (A_{n_k}.F_{n_k})(A_{n_k}(x_0)) \;=\; F_{n_k}(x_0).$$

The LHS tends to $\widetilde{G}(z)$ as $k \to \infty$, and the RHS tends to $G(x_0)$. Therefore we get $\widetilde{G}(z) \leq G(x_0)$, contradicting the choice of z. This shows that we cannot have $b_{n_k} \to +\infty$. Similarly one proves that we cannot have $b_{n_k} \to -\infty$. We conclude that $(b_n)_{n\in\mathbb{N}}$ is indeed bounded.

From now on, suppose furthermore that also $\widetilde{G}$ is nondegenerate. We claim that then $(a_n)_{n\in\mathbb{N}}$ is bounded away from 0: for some $\varepsilon > 0$ we have $a_n \geq \varepsilon$ for all $n \in \mathbb{N}$. If not, then we could extract a subsequence such that $a_{n_k} \to 0$ and also $b_n \to \beta$ (since b_n is bounded). But since $F_{n_k} \xrightarrow{\mathrm{d}} G$ and G is nondegenerate, from Lemma 7.3 we would get $A_{n_k}.F_{n_k} \xrightarrow{\mathrm{d}} \widetilde{G}_0$, where $\widetilde{G}_0$ is degenerate. But by assumption $A_{n_k}.F_{n_k} \xrightarrow{\mathrm{d}} \widetilde{G}$, where $\widetilde{G}$ is nondegenerate; this is impossible by uniqueness of limits for convergence in distribution. Therefore $(a_n)_{n\in\mathbb{N}}$ must indeed be bounded away from 0.

Note that since $(a_n)_{n\in\mathbb{N}}$ is bounded away from 0 and $+\infty$, and $(b_n)_{n\in\mathbb{N}}$ is bounded, we can extract a subsequence such that $a_{n_k} \to \alpha \in (0,+\infty)$ and $b_{n_k} \to \beta \in \mathbb{R}$. By assumption, we have $A_{n_k}.F_{n_k} \xrightarrow{\mathrm{d}} \widetilde{G}$. But since $A_{n_k} \to A$ where $A(x) = \alpha x + \beta$ and we have also assumed $F_{n_k} \xrightarrow{\mathrm{d}} G$, this implies by continuity (Lemma 1.22) that $A_{n_k}.F_{n_k} \xrightarrow{\mathrm{d}} A.G$. By uniqueness of limits, we get $A.G = \widetilde{G}$.

To prove that in fact $A_n \to A$, not just along a subsequence, note the following. From any subsequence A_{n_k}, we can extract a further convergent subsequence of values of $a_{n_{k_\ell}}$ and $b_{n_{k_\ell}}$ values as above, with limits $\alpha' \in (0,+\infty)$ and $\beta' \in \mathbb{R}$. The same argument as above shows that $A'.G = \widetilde{G}$ where $A'(x) = \alpha' x + \beta'$. Lemma 7.2 then says that we must have $A' = A$, i.e., $\alpha' = \alpha$ and $\beta' = \beta$. Since any subsequence has a convergent further subsequence with the same limit, the entire sequence must converge, $A_n \to A$. The proof is complete. □

Theorem 7.6 (Convergence to types again)**.** *Let $(A_n)_{n\in\mathbb{N}}$ and $(\widetilde{A}_n)_{n\in\mathbb{N}}$ be two sequences of oriented affine isomorphisms of $\mathbb{R}$, $A_n, \widetilde{A}_n \in \mathrm{Aff}^+_{\mathbb{R}}$. Write $A_n(x) = a_n x + b_n$ and $\widetilde{A}_n(x) = \tilde{a}_n x + \tilde{b}_n$, and for the inverses $A_n^{-1}(x) = c_n x + d_n$ and $\widetilde{A}_n^{-1}(x) = \tilde{c}_n x + \tilde{d}_n$.*

Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of c.d.f.s such that $A_n.F_n \xrightarrow{\mathrm{d}} G$, with G a nondegenerate c.d.f.

Then the convergence of also $\widetilde{A}_n.F_n \xrightarrow{\mathrm{d}} G$ holds if and only if the coefficients of the affine maps satisfy the relations

$$\frac{\tilde{a}_n}{a_n} \to 1 \quad \text{and} \quad \frac{a_n \tilde{b}_n - \tilde{a}_n b_n}{a_n} \to 0,$$

or equivalently,

$$\frac{\tilde{c}_n}{c_n} \to 1 \quad \text{and} \quad \frac{\tilde{d}_n - d_n}{c_n} \to 0.$$

Proof. (This is actually just a special case of what is stated as Corollary 7.7 below. The better organization is to prove that corollary directly, and obtain this lemma as a special case.)

We will apply the convergence to types with the reference sequence $(A_n.F_n)_{n\in\mathbb{N}}$, which by assumption tends to a nondegenerate G.

To express the other sequence in terms of the reference sequence, we write

$$\widetilde{A}_n.F_n \;=\; (\widetilde{A}_n A_n^{-1}).(A_n.F_n).$$

By assumption this also tends to G.

Theorem 7.5 applies, and guarantees convergence $\widetilde{A}_n A_n^{-1} \to A$ to some $A \in \mathrm{Aff}^+_{\mathbb{R}}$, and it also implies $A.G = G$ (note that the other limit is also G by our assumptions). However, the unique A for which we have $A.G = G$ is $A = \mathrm{id}$. We therefore get $\widetilde{A}_n A_n^{-1} \to \mathrm{id}$. To explicitly see the coefficients, note that

$$A_n^{-1}(x) \;= a_n^{-1}(x - b_n)$$

and

$$\begin{aligned} \widetilde{A}_n(A_n^{-1}(x)) \;&= \tilde{a}_n(a_n^{-1}(x - b_n)) + \tilde{b}_n \\ &= \tilde{a}_n a_n^{-1} x - \tilde{a}_n a_n^{-1} b_n + \tilde{b}_n. \end{aligned}$$

The convergence $\widetilde{A}_n A_n^{-1} \to \mathrm{id}$ is equivalent to the convergence of the coefficients,

$$\begin{aligned} \tilde{a}_n a_n^{-1} &\longrightarrow 1 \\ -\tilde{a}_n a_n^{-1} b_n + \tilde{b}_n &\longrightarrow 0. \end{aligned}$$

The second one can be rewritten as

$$\frac{a_n \tilde{b}_n - \tilde{a}_n b_n}{a_n} \longrightarrow 0.$$

□

Corollary 7.7 (Convergence to types with different limits)**.** *Let $(A_n)_{n\in\mathbb{N}}$ and $(\widetilde{A}_n)_{n\in\mathbb{N}}$ be two sequences of oriented affine isomorphisms of $\mathbb{R}$, $A_n, \widetilde{A}_n \in \mathrm{Aff}_{\mathbb{R}}^{+}$. Write $A_n(x) = a_n x + b_n$ and $\widetilde{A}_n(x) = \tilde{a}_n x + \tilde{b}_n$, and for the inverses $A_n^{-1}(x) = c_n x + d_n$ and $\widetilde{A}_n^{-1}(x) = \tilde{c}_n x + \tilde{d}_n$.*

Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of c.d.f.s such that $A_n.F_n \xrightarrow{\mathrm{d}} G$ and $\widetilde{A}_n.F_n \xrightarrow{\mathrm{d}} \widetilde{G}$, with G and $\widetilde{G}$ nondegenerate c.d.f.s. Then for some $\alpha > 0$ and $\beta \in \mathbb{R}$ we have

$$\frac{\tilde{a}_n}{a_n} \to \alpha \quad \text{ and } \quad \frac{a_n \tilde{b}_n - \tilde{a}_n b_n}{a_n} \to \beta,$$

and we have

$$A.G = \widetilde{G} \qquad \text{where } \; A(x) = \alpha x + \beta.$$

Equivalently, with $\gamma = \alpha^{-1}$ and $\delta = -\alpha^{-1}\beta$ so that $A^{-1}(x) = \gamma x + \delta$, we have

$$\frac{\tilde{c}_n}{c_n} \to \gamma \quad \text{ and } \quad \frac{\tilde{d}_n - d_n}{c_n} \to \delta.$$

In particular, G and $\widetilde{G}$ have the same type.

Proof. We will apply the convergence to types with the reference sequence $(A_n.F_n)_{n\in\mathbb{N}}$, which by assumption tends to a nondegenerate G.

To express the other sequence in terms of the reference sequence, we write

$$\widetilde{A}_n.F_n \;=\; (\widetilde{A}_n A_n^{-1}).(A_n.F_n).$$

By assumption this tends to a nondegenerate $\widetilde{G}$.

Theorem 7.5 applies, and guarantees convergence $\widetilde{A}_n A_n^{-1} \to A$ to some $A \in \mathrm{Aff}_{\mathbb{R}}^{+}$, and it also implies $A.G = \widetilde{G}$.

Write $A(x) = \alpha x + \beta$. To explicitly see the coefficients of $\widetilde{A}_n A_n^{-1}$, note that

$$A_n^{-1}(x) \;= a_n^{-1}(x - b_n)$$

and

$$\begin{aligned} \widetilde{A}_n(A_n^{-1}(x)) \;&= \tilde{a}_n(a_n^{-1}(x - b_n)) + \tilde{b}_n \\ &= \tilde{a}_n a_n^{-1} x - \tilde{a}_n a_n^{-1} b_n + \tilde{b}_n. \end{aligned}$$

The convergence $\widetilde{A}_n A_n^{-1} \to A$ is equivalent to the convergence of the coefficients,

$$\begin{aligned} \tilde{a}_n a_n^{-1} &\longrightarrow \alpha \\ -\tilde{a}_n a_n^{-1} b_n + \tilde{b}_n &\longrightarrow \beta. \end{aligned}$$

The second one can be rewritten as

$$\frac{a_n \tilde{b}_n - \tilde{a}_n b_n}{a_n} \longrightarrow \beta.$$

$\square$

Lemma 7.8 (A choice of normalizing constants for convergence to types)**.** *(It is possible to choose normalization constants for the affine transformations using the left-continuous inverses of the c.d.f.s. TODO: Precise statement.)*

Proof. ... $\square$

7.2 One-parameter subgroups of affine transformations

Definition 7.9 (Subgroup of translations)**.** The mapping $s \mapsto A_s$ with

$$A_s(x) = x + s$$

is a homomorphism $\mathbb{R} \to \mathrm{Aff}^+_{\mathbb{R}}$. The image of this homomorphism is the *subgroup of translations* in $\mathrm{Aff}^+_{\mathbb{R}}$.

Lemma 7.10 (Only translations have no fixed points)**.** *If $A \in \mathrm{Aff}^+_{\mathbb{R}}$ has no fixed points (no $x \in \mathbb{R}$ such that $A(x) = x$) then A belongs to the subgroup of translations, i.e., $A(x) = x + s$ for some $s \in \mathbb{R}$ (in fact $s \neq 0$).*

Proof. Let us prove this by contrapositive: that any element A which is not a translation must have a fixed point. So assume that A is not a translation, i.e., $A(x) = ax + b$ with some $a \neq 1$. Then the fixed point equation $A(x) = x$ reads

$$ax + b = x,$$

and it has a solution $x = \frac{-b}{a-1} \in \mathbb{R}$, which then is a fixed point of A. $\square$

Lemma 7.11 (Conjugate of translation is translation)**.** *Let $A_s^{(\beta)} = x + \beta s$ for $s, \beta \in \mathbb{R}$ as in Definition 7.9. Let also $B \in \mathrm{Aff}^+_{\mathbb{R}}$ be given by $B(x) = ax + b$. Then*

$$B\, A_s^{(\beta)}\, B^{-1} = A_s^{(a\beta)}.$$

Proof. Calculate, for $x \in \mathbb{R}$

$$\begin{aligned} (BA_s^{(\beta)}B^{-1})(x) &= (BA_s^{(\beta)})(\frac{x-b}{a}) \\ &= B(\frac{x-b}{a} + \beta s) \\ &= a(\frac{x-b}{a} + \beta s) + b \\ &= x - b + a\beta s + b \\ &= x + a\beta s \;=\; A_s^{(a\beta)}(x). \end{aligned}$$

$\square$

Definition 7.12 (Subgroup fixing a point)**.** The mapping $s \mapsto A_s$ with

$$A_s(x) = e^s(x - c) + c$$

is a homomorphism $\mathbb{R} \to \mathrm{Aff}^+_{\mathbb{R}}$. The image of this homomorphism is the *subgroup fixing* c in $\mathrm{Aff}^+_{\mathbb{R}}$.

Lemma 7.13 (Characterization of the subgroup fixing a point)**.** *An orientation-preserving affine transformation $A \in \mathrm{Aff}^+_{\mathbb{R}}$ belongs to the subgroup fixing $c \in \mathbb{R}$ if and only if $A(c) = c$.*

(Note that the subgroup is a priori defined as the image of a homomorphism, so the statement indeed requires a proof.)

Proof. Suppose first that A is an element of the said subgroup, i.e., $A(x) = e^s(x - c) + c$ for some $s \in \mathbb{R}$. Then clearly $A(c) = c$.

Suppose then that $A(c) = c$. Write $A(x) = ax + b$ for $a > 0$ and $b \in \mathbb{R}$. Plug in $x = c$ in the assumed fixed point property to obtain

$$ac + b = c.$$

The above can be solved to give $b = (1 - a)c$. Also since $a > 0$, we can write $a = e^s$ with $s \in \mathbb{R}$. With these, the formula for A simplifies to

$$A(x) = e^s x + c(1 - e^s) = e^s(x - c) + c.$$

This shows $A = A_s$ as desired (with A_s as in Definition 7.12). □

Lemma 7.14 (Conjugate of fixing is fixing image)**.** *Let $A_s^{(\alpha;c)} = e^{\alpha s}(x - c) + c$ for $\alpha, c \in \mathbb{R}$ as in Definition 7.12. Let also $B \in \mathrm{Aff}^+_{\mathbb{R}}$ be given by $B(x) = ax + b$. Then*

$$B\, A_s^{(\alpha;c)}\, B^{-1} = A_s^{(\alpha;B(c))}.$$

Proof. Calculate, for $x \in \mathbb{R}$

$$\begin{aligned}
(BA_s^{(\alpha;c)}B^{-1})(x) \;&= (BA_s^{(\alpha;c)})(\frac{x-b}{a}) \\
&= B(e^{\alpha s}(\frac{x-b}{a} - c) + c) \\
&= ae^{\alpha s}(\frac{x-b}{a} - c) + ac + b \\
&= e^{\alpha s}(x - (ac + b)) + (ac + b) \\
&= e^{\alpha s}(x - B(c)) + B(c) \\
&= A_s^{(\alpha;B(c))}(x).
\end{aligned}$$

□

7.3 Self-similarity characterizations of the extreme value distributions

Lemma 7.15 (Continuous parameter extreme value limit relation)**.** *Let F be a c.d.f.*

(Note that below we use the sequence $(F^n)_{n \in \mathbb{N}}$ of nth powers of a fixed c.d.f., not a sequence of arbitrary c.d.f.s. Recall that the nth power F^n is the c.d.f. of the maximum of n independent random variables with the distribution F.)

Suppose that for a sequence $(A_n)_{n\in\mathbb{N}}$ of oriented affine isomorphisms of $\mathbb{R}$*,* $A_n \in \mathrm{Aff}^+_{\mathbb{R}}$*, we have*

$$A_n.F^n \xrightarrow{\mathrm{d}} G,$$

where G is a c.d.f.

Then, for any $t > 0$, denoting by G^t the c.d.f. given by $G^t(x) = \big(G(x)\big)^t$, we have

$$A_n.F^{\lfloor nt \rfloor} \xrightarrow{\mathrm{d}} G^t,$$

where, for $x \in \mathbb{R}$, the floor notation $\lfloor x \rfloor$ stands for the greatest integer $k \in \mathbb{Z}$ such that $k \le x$.

Proof. Let $t > 0$ and let $x \in \mathbb{R}$ be a continuity point of G. For $n \in \mathbb{N}$, calculate

$$\big((A_n.F)(x)\big)^{\lfloor nt \rfloor} = \Big(\big((A_n.F)(x)\big)^n\Big)^{\lfloor nt \rfloor / n}.$$

By assumption, we have $\big((A_n.F)(x)\big)^n \to G(x)$ as $n \to \infty$. Also $\lfloor nt \rfloor / n \to t$ as $n \to \infty$. By (joint) continuity of the power function $(x, y) \mapsto x^y = \exp\big(y \log(x)\big)$, we get that the expression above tends to $G(x)^t$.

Finally noting that the continuity points of G^t are the same as the continuity points of G, the above in fact proves the asserted $A_n.F^{\lfloor nt \rfloor} \xrightarrow{\mathrm{d}} G^t$. □

Lemma 7.16 (Self-similarity of extreme value distributions)**.** *Suppose that G is an extreme-value distribution. Then there exists a family $(A_t)_{t>0}$ of oriented affine isomorphisms of* $\mathbb{R}$*,* $A_t \in \mathrm{Aff}^+_{\mathbb{R}}$*, such that for any $t > 0$*

$$G^t = A_t.G.$$

Moreover, $t \mapsto A_t$ is a measurable homomorphism of multiplicative groups $(0, +\infty) \to \mathrm{Aff}^+_{\mathbb{R}}$.

Proof. ... □

Lemma 7.17 (Self-similar continuous c.d.f. family characterization $\gamma = 0$)**.** *Suppose that G is a nondegenerate c.d.f. such that*

$$G^t = A_t.G \qquad \text{for any } t > 0,$$

where

$$A_t(x) = x + \beta \log t$$

with $\beta > 0$.

Then with $d = \log\big(-\log G(0)\big)$, for all $x \in \mathbb{R}$ we have

$$G(x) = \exp\Big(-\exp\big(-\beta^{-1}x + d\big)\Big).$$

(In particular, G is of Gumbel type: there exists an $A \in \mathrm{Aff}^+_{\mathbb{R}}$ such that $G = A.\Lambda$.)

Proof. Since G is nondegenerate, there exists an $x_0 \in \mathbb{R}$ with $0 < G(x_0) < 1$. Write $q = -\log G(x_0) > 0$, so that $G(x_0) = e^{-q}$.

For $t > 0$, from the equation $G^t = A_t.G$ we get for any $x \in \mathbb{R}$

$$G(x)^t = G(A_t^{-1}(x)) = G(x - \beta \log(t)).$$

In particular, with $x = x_0 + \beta \log(t)$, we get

$$G(x_0 + \beta \log(t))^t = G(x_0) = e^{-q},$$

from which we can solve

$$G(x_0 + \beta \log(t)) = e^{-q/t}.$$

The above holds for any $t > 0$, and any $x \in \mathbb{R}$ can be written as $x = x_0 + \beta \log\big(e^{(x-x_0)/\beta}\big)$. We therefore get, for any $x \in \mathbb{R}$,

$$G(x) = \exp\Big(-q \exp\big(-(x-x_0)/\beta\big)\Big).$$

This is of the desired form, with $d = \frac{x_0}{\beta} + \log(q)$. Plugging in $x = 0$ then shows $d = \log\big(-\log G(0)\big)$. $\square$

Lemma 7.18 (Self-similar continuous c.d.f. family characterization $\gamma > 0$)**.** *Suppose that G is a nondegenerate c.d.f. such that*

$$G^t = A_t.G \qquad \textit{for any } t > 0,$$

where

$$A_t(x) = t^\alpha x + c\,(1 - t^\alpha) = t^\alpha (x - c) + c$$

with $c \in \mathbb{R}$ and $\alpha > 0$.

Then with $\sigma = \big(-\log G(c+1)\big)^\alpha$, for all $x \geq c$ we have

$$G(x) = \exp\Big(-\big(\frac{x-c}{\sigma}\big)^{-1/\alpha}\Big).$$

(It easily follows that G is of Fréchet type: there exists an $A \in \mathrm{Aff}^+_{\mathbb{R}}$ such that $G = A.\Phi_\alpha$.)

Proof. Since G is nondegenerate, there exists an $x_0 \in \mathbb{R}$ with $0 < G(x_0) < 1$. Write $q = -\log G(x_0) > 0$, so that $G(x_0) = e^{-q}$.

For $t > 0$, from the equation $G^t = A_t.G$ we get for any $x \in \mathbb{R}$

$$G(x)^t = G(A_t^{-1}(x)).$$

Note that for $x \leq c$ and $t = 2$ we have $A_2^{-1}(x) = 2^{-\alpha}(x-c) + c \geq x$, so the above implies $G(x)^2 = G(A_2^{-1}(x)) \geq G(x)$. This not possible if $0 < G(x) < 1$, so we must in particular have $x_0 > c$.

With $x = A_t(x_0)$ in the above equation, we get

$$G(A_t(x_0))^t = G(x_0) = e^{-q},$$

from which we can solve

$$G(x) = G(A_t(x_0)) = e^{-q/t}.$$

Any $x \geq c$ can be written as $x = A_t(x_0) = t^{\alpha}(x_0 - c) + c$ with $t = \big(\frac{x-c}{x_0-c}\big)^{1/\alpha}$ (recall that $x_0 - c > 0$). We therefore get, for any $x \geq c$,

$$G(x) = \exp\Big(- q\,\big(\frac{x-c}{x_0-c}\big)^{-1/\alpha}\Big).$$

This is of the form

$$G(x) = \exp\Big(- \big(\frac{x-c}{\sigma}\big)^{-1/\alpha}\Big),$$

and plugging in $x = c+1$ yields $\sigma = \big(- \log G(c+1)\big)^{\alpha}$. ☐

Lemma 7.19 (Self-similar continuous c.d.f. family characterization $\gamma < 0$)**.** *Suppose that G is a nondegenerate c.d.f. such that*

$$G^t = A_t.G \qquad \textit{for any } t > 0,$$

where

$$A_t(x) = t^{-\alpha}x + c\,(1 - t^{-\alpha}) = t^{-\alpha}(x - c) + c$$

with $c \in \mathbb{R}$ and $\alpha > 0$.

Then with $\sigma = \big(- \log G(c-1)\big)^{-\alpha}$, for all $x \leq c$ we have

$$G(x) = \exp\Big(- \big(\frac{c-x}{\sigma}\big)^{1/\alpha}\Big).$$

(It easily follows that G is of Weibull type: there exists an $A \in \mathrm{Aff}^{+}_{\mathbb{R}}$ such that $G = A.\Psi_{\alpha}$.)

Proof. (Note: The Lean formalized statement uses the opposite sign of α: it is assumed that $\alpha < 0$ and $A_t(x) = t^{+\alpha}(x - c) + c$.)

Since G is nondegenerate, there exists an $x_0 \in \mathbb{R}$ with $0 < G(x_0) < 1$. Write $q = -\log G(x_0) > 0$, so that $G(x_0) = e^{-q}$.

For $t > 0$, from the equation $G^t = A_t.G$ we get for any $x \in \mathbb{R}$

$$G(x)^t = G(A_t^{-1}(x)).$$

Note that for $x \geq c$ and $t = 2$ we have $A_2^{-1}(x) = 2^{\alpha}(x - c) + c \geq x$, so the above implies $G(x)^2 = G(A_2^{-1}(x)) \geq G(x)$. This not possible if $0 < G(x) < 1$, so we must in particular have $x_0 < c$.

With $x = A_t(x_0)$ in the above equation, we get

$$G\big(A_t(x_0)\big)^t = G(x_0) = e^{-q},$$

from which we can solve

$$G(x) = G(A_t(x_0)) = e^{-q/t}.$$

Any $x \leq c$ can be written as $x = A_t(x_0) = t^{-\alpha}(x_0 - c) + c$ with $t = \big(\frac{c-x}{c-x_0}\big)^{-1/\alpha}$ (recall that $c - x_0 > 0$). We therefore get, for any $x \leq c$,

$$G(x) = \exp\Big(-q\,\big(\frac{c-x}{c-x_0}\big)^{1/\alpha}\Big).$$

This is of the form

$$G(x) = \exp\Big(-\big(\frac{c-x}{\sigma}\big)^{1/\alpha}\Big),$$

and plugging in $x = c-1$ yields $\sigma = \big(-\log G(c-1)\big)^{-\alpha}$. □

Theorem 7.20 (Three types of extreme value distributions [Fisher-Tippett-Gnedenko)**.** *] For any extreme value distribution G, one of the following holds:*

(Λ) $G = A.\Lambda$ for some $A \in \mathrm{Aff}^+_{\mathbb{R}}$;

(Φ) $G = A.\Phi_\alpha$ for some $A \in \mathrm{Aff}^+_{\mathbb{R}}$ and $\alpha > 0$;

(Ψ) $G = A.\Psi_\alpha$ for some $A \in \mathrm{Aff}^+_{\mathbb{R}}$ and $\alpha > 0$.

In particular, the only three possible types of extreme value distributions are the type of the Gumbel c.d.f., the type of the Fréchet c.d.f. Φ_α for $\alpha > 0$, and the type of the Weibull c.d.f. Ψ_α for $\alpha > 0$.

Proof. ... □

Chapter 8

Left-continuous inverses

Let R and S be complete linear orders, for example $[0,1]$, $[0,+\infty]$, or $[-\infty,+\infty] =: \overline{\mathbb{R}}$.

Definition 8.1. Let $f\colon R \to S$ be a function (usually assumed nondecreasing). The left-continuous inverse of f is the function $f^{\to 1}\colon S \to R$ given by

$$f^{\to 1}(y) := \inf\{x \in R \mid f(x) \geq y\} \qquad \text{for } y \in S.$$

The right-continuous inverse $f^{\leftarrow 1}\colon S \to R$ is analoguously defined by

$$f^{\leftarrow 1}(y) := \sup\{x \in R \mid f(x) \leq y\} \qquad \text{for } y \in S.$$

Chapter 9

Cauchy-Hamel functional equation

9.1 Positive measure additive subgroups of the reals

Lemma 9.1 (Countably many connected components for an open set)**.** *Let X be a locally connected separable space. Then any open subset $U \subseteq X$ has at most countably many connected components.*

Proof. (The proof is already formalized, see: `IsOpen.countable-setOf-connectedComponentIn`.) $\square$

For subsets $A, B \subseteq \mathbb{R}$, we use the following notation for pointwise sum sets and difference sets:

$$A + B = \{a + b \mid a \in A,\, b \in B\},$$
$$A - B = \{a - b \mid a \in A,\, b \in B\}.$$

We denote by Λ the Lebesgue measure on $\mathbb{R}$.

Lemma 9.2 (Finding an interval with high overlap)**.** *Let $A \subset \mathbb{R}$ be a measurable set such that $0 < \Lambda[A] < +\infty$. Then for any $r \in [0, 1)$, there exists a nontrivial interval $J \subset \mathbb{R}$ (a subset of the real line which is connected and has nonempty interior) such that*

$$\Lambda[A \cap J] \ > \ r\,\Lambda[J].$$

Proof. Assume, without loss of generality, $0 < r < 1$. Since the Lebesgue measure is outer regular, we can find an open set $U \subset \mathbb{R}$ such that $A \subseteq U$ and $\Lambda[U] < r^{-1}\,\Lambda[A]$.

The open set U has at most countably many connected components (which are in fact open intervals); denote by $(U_i)_{i \in I}$ the indexed collection of them.

Note that for at least one index $j \in I$ we have $\Lambda[A \cap U_j] > r\Lambda[U_j]$, because otherwise we get

$$\begin{aligned}\Lambda[A] = \Lambda[A \cap U] &= \sum_{i \in I} \Lambda[A \cap U_i] \\ &\leq r \sum_{i \in I} \Lambda[U_i] \\ &= r\Lambda[U],\end{aligned}$$

contradicting the choice of U.

Now $J = U_j$ has the desired properties. $\square$

Lemma 9.3 (Shifts of a smaller interval contained in a larger interval)**.** *Let I, J be nontrivial intervals, whose length satisfy*

$$0 \; < \; \Lambda[J] \; < \; \Lambda[I] \; < +\infty.$$

Then there exists an open interval Δ of length $\Lambda[\Delta] = \Lambda[I] - \Lambda[J] > 0$ such that

$$t + J \; \subset \; I \qquad \text{for any } t \in \Delta.$$

Proof. ... $\square$

Lemma 9.4 (Overlapping union of copies of an interval)**.** *Let J be a nontrivial interval of finite length ($0 \; < \; \Lambda[J] \; < +\infty$). Let $c < 1$ and denote $\delta = c\Lambda[J] \; < \; \Lambda[J]$. Then for any $t \in (-\delta, \delta)$, the set $J' = (t + J) \cup J$ is an interval (connected set with nonempty interior) whose length satisfies the bound $\Lambda[J'] \; < \; (1 + c)\, \Lambda[J]$.*

Proof. Denote $a = \inf J$ and $b = \sup J$. We have $-\infty < a < b < +\infty$ and

$$(a, b) \; \subseteq \; J \; \subseteq \; [a, b]$$

and we have $\Lambda[J] = b - a$.

Let $t \in (+\delta, \delta)$, with $\delta = c\, \Lambda[J] = c(b - a)$. Then we have the containments

$$\big(\min\{a, a + t\}, \max\{b, b + t\}\big) \; \subseteq \; (t + J) \cup J \; \subseteq \; \big[\min\{a, a + t\}, \max\{b, b + t\}\big].$$

It in particular follows that $J' := (t + J) \cup J$ is an interval.

If $t \geq 0$, then J' is contained in $[a, b+t]$ which has length $b+t-a = (b-a)+t < \Lambda[J]+c\, \Lambda[J] = (1 + c)\Lambda[J]$. If $t < 0$, then J' is contained in $[a + t, b]$ and one similarly gets a length bound. $\square$

Lemma 9.5 (Difference set of positive measure set contains an interval)**.** *Let $A \subset \mathbb{R}$ be a measurable set of positive Lebesgue measure. Then there exists a $\delta > 0$ such that*

$$(-\delta, \delta) \; \subseteq \; A - A.$$

Proof. Pick a measurable subset $A_0 \subseteq A$ such that $0 < \Lambda[A_0] < +\infty$.

By Lemma 9.2, we can find a nontrivial interval J such that $\Lambda[A_0 \cap J] > \frac{3}{4}\Lambda[J]$. Let $\delta = \frac{1}{2}\Lambda[J]$. We claim that $(-\delta, \delta) \subseteq A_0 - A_0$ (which then clearly implies the assertion of the lemma). Indeed, suppose $t \in (-\delta, \delta)$. Then by Lemma 9.4, $(t + J) \cup J$ is an interval of length less than $\frac{3}{2}\Lambda[J]$. Moreover, we have $A_0 \cap J \subseteq (t + J) \cup J$ and $t + (A_0 \cap J) \subseteq (t + J) \cup J$. Now note that

$$\Lambda[t + (A_0 \cap J)] = \Lambda[A_0 \cap J] > \frac{3}{4}\Lambda[J].$$

If the sets $t + (A_0 \cap J)$ and $(A_0 \cap J)$ were disjoint, then the measure of $(A_0 \cap J) \cup (t + (A_0 \cap J))$ would thus be greater than $\frac{3}{2}\Lambda[J]$, which is impossible given the length of the interval $(t + J) \cup J$ is less than $\frac{3}{2}\Lambda[J]$. Therefore there exists a point $a \in (t+(A_0 \cap J)) \cap (A_0 \cap J)$. Denoting $a' = a-t$, we have $a, a' \in A_0 \cap J \subseteq A_0$ and $t = a - a' \in A_0 - A_0 \subseteq A - A$. Since $t \in (-\delta, \delta)$ was arbitrary, this proves the assertion. $\square$

Lemma 9.6 (Dividing a high overlap interval)**.** *Let A be a measurable. Suppose that for some $a < b$, the interval $J = (a, b)$ satisfies $\Lambda[A \cap J] > r\,\Lambda[J]$. Let $m \in \mathbb{N}_+$, and consider the subintervals*

$$J_i = \Big(a + \frac{i}{(b-a)m},\ a + \frac{i+1}{(b-a)m}\Big) \qquad \text{for } i = 0, \dots, m-1.$$

Then for some i we have $\Lambda[A \cap J_i] > r\,\Lambda[J_i]$.

Proof. □

Lemma 9.7 (Difference of two positive measure sets contains an interval)**.** *Let $A, B \subset \mathbb{R}$ be two measurable sets of positive Lebesgue measure. Then the difference set $A - B$ contains a nontrivial open interval.*

Proof. Pick measurable subsets $A_0 \subseteq A$ and $B_0 \subseteq B$ such that $0 < \Lambda[A_0] < +\infty$ and $0 < \Lambda[B_0] < +\infty$.

By Lemma 9.2, we can find nontrivial intervals J, I such that $\Lambda[A_0 \cap J] > \frac{3}{4}\Lambda[J]$ and $\Lambda[B_0 \cap I] > \frac{3}{4}\Lambda[I]$.

By Lemma 9.6 for any $n, m \in \mathbb{N}$ we can find intervals $I' \subset I$ and $J' \subset J$ with $\Lambda[I'] = \frac{1}{n}\Lambda[I]$ and $\Lambda[J'] = \frac{1}{m}\Lambda[J]$ and such that $\Lambda[A_0 \cap J'] > \frac{3}{4}\Lambda[J']$ and $\Lambda[B_0 \cap I'] > \frac{3}{4}\Lambda[I']$. By choosing n and m suitably, we can ensure that

$$\frac{1}{2}\Lambda[I'] \;\le\; \Lambda[J'] \;<\; \Lambda[I'].$$

Then by Lemma 9.3 there exists an open interval Δ of length $\Lambda[\Delta] > \Lambda[I'] - \Lambda[J']$ such that for all $t \in \Delta$ we have $t + J' \subset I'$.

Consider a fixed $t \in \Delta$. Observe that $t + (B_0 \cap J) \subset I'$ and

$$\Lambda[t + (B_0 \cap J')] \;=\; \Lambda[B_0 \cap J'] \;>\; \frac{3}{4}\Lambda[J'] \;\ge\; \frac{3}{8}\Lambda[I'].$$

We now claim that $A_0 \cap I'$ and $t + (B_0 \cap J')$ intersect. Their measures add up to at least

$$\begin{aligned}
\Lambda[A_0 \cap I'] + \Lambda[t + (B_0 \cap J')] &= \Lambda[A_0 \cap I'] + \Lambda[B_0 \cap J'] \\
&> \frac{3}{4}\Lambda[I'] + \frac{3}{4}\Lambda[J'] \\
&\ge \frac{3}{4}\Lambda[I'] + \frac{3}{8}\Lambda[I'] \;=\; \frac{9}{8}\Lambda[I'].
\end{aligned}$$

and both have been shown to be subsets of I'; therefore they cannot be disjoint. In particular there is a point $z \in (A_0 \cap I') \cap (t + B_0 \cap J')$, which means that

$$a = z = t + b$$

for some $a \in A_0 \cap I'$ and $b \in B_0 \cap J'$. We solve $t = a - b \in A_0 - B_0 \subseteq A - B$. Since $t \in \Delta$ was arbitrary, we have shown that the nontrivial open interval Δ is contained in $A - B$. □