Records
Charles M. Goldie

TU Dresden, 6–9 October 2008

Contents

1. Record values 1
2. Record times 11
3. Limits 19
4. Extensions 24
References 34
Appendix: solutions to exercises 36

1. Record values

1.1. Basics. This is an expansion of the first half of the survey [Bunge & Goldie, 2001], plus some more recent material. Its view of the subject is almost disjoint from that of other treatments such as [Arnold, Balakrishnan & Nagaraja, 1998] or [Nevzorov, 2001].

General notation & abbreviations.
• \( \mathbb{N} = \{1, 2, \ldots \} \), the set of natural numbers; \( \mathbb{N}_0 := \mathbb{N} \cup \{0\} \).
• \( x > 0 \) is positive; \( x \geq 0 \) is non-negative.
• \((\Omega, \mathcal{A})\) is a measurable space and \( P \) is a probability measure on it, making \((\Omega, \mathcal{A}, P)\) a probability space. The elements of the \( \sigma \)-algebra \( \mathcal{A} \) are the events.
• For an event \( A \) the indicator random variable (r.v.) \( 1_A \) is given by \( 1_A(\omega) := 1 \) for \( \omega \in A \), \( 1_A(\omega) := 0 \) otherwise.
• A filtration in \( \mathcal{A} \) is a non-decreasing sequence \((\mathcal{G}_n)_{n \in \mathbb{N}_0}\) of sub-\( \sigma \)-algebras of \( \mathcal{A} \).
• A stopping time for a filtration \((\mathcal{G}_n)\) is a \( \mathbb{N}_0 \cup \{\infty\}\)-valued r.v. \( L \) with the property that \( \{L = n\} \in \mathcal{G}_n \) for all \( n \in \mathbb{N}_0 \) (where \( \{L = n\} \) is an abbreviated notation for the event \( \{\omega : L(\omega) = n\} \)).
For such a stopping time the *up-to-* $L$ $\sigma$-algebra is

$$G_L := \{ A \in \mathcal{A} : A \cap \{ L = n \} \in \mathcal{G}_n \ \forall n \in \mathbb{N}_0 \}.$$  

It is a $\sigma$-algebra.

**Records: basic setup.**

- $F$ is the distribution function (d.f.) of a r.v. $X$ on $\mathbb{R}$.
- $F(x) := P(X \leq x)$, right-continuous (rt-cts).
- $x_* := \sup \{ x : F(x) < 1 \}$.
- $\bar{F}(x) := 1 - F(x)$.
- The probability space is assumed to support $X, X_1, X_2, \ldots$, which are i.i.d. (independent and identically distributed) $\sim F$.
- For the sequence $(X_n)$ the *natural filtration* is $(\mathcal{A}_n)$ defined by $\mathcal{A}_0 := \{ \emptyset, \Omega \}$ and $\mathcal{A}_n := \sigma(X_1, \ldots, X_n)$ for $n \in \mathbb{N}$.
- $(X_n)$ is *strong Markov*: if $L$ is a finite stopping time (for the natural filtration) then the sequence $(X_{L+n})_{n \in \mathbb{N}}$ is i.i.d. $\sim F$ and is independent of $\mathcal{A}_L$.
- The *order statistics* $X_1 \geq X_2 \geq \cdots \geq X_n$ are $X_1, \ldots, X_n$ in order.

**Exercise 1.1.** Prove that $(X_n)$ is strong Markov\(^1\).

**Definition 1.2.** The *initial rank* of $X_n$ is $\rho_n := \sum_{i=1}^n 1_{X_i \geq X_n}$.

Figure 1.3 shows initial ranks for some integer-valued observations. A sequence $(X_n)$ that starts 1, $-2$, 4, 1, 1, … is shown. The sequence $(\rho_n)$ of initial ranks then begins 1, 2, 1, 3, 4, ….

**Figure 1.3.** Initial ranks when observation values can coincide.

**Definition 1.4.** $X_n$ is a *$k$-record* if $\rho_n = k$.

The *record values* are the 1-records.

- $X_1$ is a record value.
- For $k > 1$ the first $k$-record occurs at or after time $k$.
- These are *upper* records: one can alternatively work with lower records.

\(^1\)Solutions are at the end.
Fig. 1.5 shows 1-, 2-, 3- and 4-records from some actual data, the rainfall amounts for Vancouver in January, in inches for the years 1906–1966 inclusive [Glick, 1978, Data Set 1]. Were we to include data from later years, more values might be added above those shown, but the 1- to 4-records up to level 13.28 (the highest 4-record in the figure) are necessarily complete.

![Figure 1.5. The 1-, 2-, 3- and 4-records for Vancouver rainfall.](image)

**Definition 1.6.** Let $R^k_1 < R^k_2 < \cdots$ denote the successive $k$-records (note the strict inequalities). The whole sequence is denoted $R^k = (R^k_1, R^k_2, \ldots)$.

**1.2. Structure.** We determine the structure of $R^1$. Later this will yield the structure of $R^k$ for every $k$.

If $x_+ < \infty$ and $P(X = x_+) > 0$ there will be a final $k$-record; otherwise not.

Because the (finite or infinite) sequence $R^1$ is strictly increasing we may regard it as a (random) set.

**Lemma 1.7.** Let $E$ be any finite union of (disjoint) intervals $(u, v]$ in $(-\infty, x_+]$. Then

$$P(R^1 \cap E = \emptyset) = e^{-\eta(E)} \quad (1.1)$$

where $\eta$ is the measure on $(-\infty, x_+]$ defined by

$$\eta(-\infty, x] := -\ln F(x).$$
Proof. Suppose (1.1) holds for $E$ a union of $k$ disjoint intervals in $(-\infty, x_+)$ (note that this interval excludes $x_+$). Induct on $k$. Thus pick $v, w$ such that $\sup E \leq v < w < x_+$, and let $A$ denote the event that no records among $X_1, \ldots, X_{n-1}$ are in $E$. Then

$$P(R_1 \cap E = \emptyset, \ R_1 \cap (v, w] = \emptyset)$$

$$= \sum_{n=1}^{\infty} P(R_1 \cap E = \emptyset, \bigvee_{1}^{n-1} X_i \leq v, X_n > w)$$

$$= P(A \cap \bigvee_{1}^{n-1} X_i \leq v) \ P(X_n > w)$$

$$= \frac{F(w)}{F(v)} P \left( A \cap \left\{ \bigvee_{1}^{n-1} X_i \leq v \right\} \right) \ P(X_n > v).$$

Repeat the first four lines of this calculation with $v$ replacing $w$, obtaining

$$P(R_1 \cap E = \emptyset) = \sum_{n=1}^{\infty} P(R_1 \cap E = \emptyset, \bigvee_{1}^{n-1} X_i \leq v, X_n > v)$$

$$= P \left( A \cap \bigvee_{1}^{n-1} X_i \leq v, X_n > v \right)$$

$$= P \left( A \cap \bigvee_{1}^{n-1} X_i \leq v \right) \ P(X_n > v).$$

Comparing these two calculations, you find that

$$P(R_1 \cap E = \emptyset, \ R_1 \cap (v, w] = \emptyset) = \frac{F(w)}{F(v)} P(R_1 \cap E = \emptyset).$$

By the induction hypothesis and the definition of $\eta$ the right-hand side equals $e^{-\eta(v,w)} e^{-\eta(E)}$, and that in turn equals $e^{-\eta(E\cup(v,w))}$. So (1.1) is proved for $E \subseteq (-\infty, x_+)$. An easy calculation allows the exclusion of $x_+$ to be removed. \[\Box\]

If $x_+ < \infty$ and $P(X = x_+) > 0$ then $\eta(-\infty, x_+) < \infty$ and $\eta\{x_+\} = \infty$. Otherwise, and in particular if $x_+ = \infty$, $\eta(-\infty, x_+) = \infty$ and $\eta\{x_+\} = 0$. Always, $\eta(-\infty, x) < \infty$ for all $x < x_+$.

Definition 1.8. $\eta$ is the avoidance measure of $R_1$.

Theorem 1.9. The law of $R_1$ is the unique law of a simple\(^2\) point process such that (1.1) holds for all finite unions $E$ of intervals.

\(^2\)Simple: no multiple points
Let $D$ be the set of points where $F$ is discontinuous. $D$ is also the set of atoms of $\eta$, i.e. the set of points where $\eta(-\infty, x]$ is discontinuous.

Because $\bar{F}(x) = F(x, \infty) = e^{-\eta(-\infty, x]}$,

\[
\therefore \quad F(x, \infty) = e^{-\eta(-\infty, x)};
\]

\[
\therefore \quad \frac{F(x, \infty)}{F(x, \infty)} = e^{-\eta(x)} = P(R^1 \cap \{x\} = \emptyset).
\]

**Definition 1.10.** The discrete part of $\eta$ is the measure $\eta_d(E) := \sum_{x \in D \cap E} \eta\{x\}$.

The continuous part of $\eta$ is the measure $\eta_c := \eta - \eta_d$.

**Theorem 1.11 (Shorrock, 1972, Shorrock, 1974).** $R^1$ is composed of

- a Poisson process $R^1_c$ of characteristic measure $\eta_c$,
- independently of $R^1_c$ and of each other,

- at each $x \in D$, a demon who
  - with probability $1 - e^{-\eta(x)}$ gives $R^1$ a point at $x$,
  - or with probability $e^{-\eta(x)}$ does not.

Also $R^1$ is completely random (= independent increments) and satisfies (1.1) for all Borel sets $E$.

**Note.** A Poisson process as referred to above is more precisely an inhomogeneous Poisson process of continuous characteristic measure $\nu = \eta_c$. This is a simple, completely random point process $N$ with, for any Borel set $B$, $N(B) \sim \text{Pois}(\nu(B))$.

**Proof of Theorem 1.11.** An elaboration of that of Theorem 1.9. From (1.1) the demons have the probabilities stated and are independent of each other and of the process with them removed. The latter is a simple completely random point process with no certain points, so (cf. Daley & Vere-Jones, 2003, Theorem 2.4.V]) is a Poisson process. As (1.1) holds for it, for all $E$, so it does for $R^1$ (re-insert the demons). \hfill \Box

### 1.3. Conditioning.

**Definition 1.12.** Let $Y$ be an integrable r.v. and $\mathcal{F}$ a $\sigma$-algebra of events. Any $\mathcal{F}$-measurable r.v. $Z$ that satisfies

\[
E(1_F Z) = E(1_F Y) \quad \forall F \in \mathcal{F}
\]

is a version of the conditional expectation of $Y$ given $\mathcal{F}$, written $E(Y|\mathcal{F})$. Any two versions are equal a.s. (almost surely).

For an event $A$ the conditional probability of $A$ given $\mathcal{F}$ is $P(A|\mathcal{F}) := E(1_A|\mathcal{F})$. 

---

**Proof.** Daley & Vere-Jones, 2008, Theorem 9.2.XII. \hfill \Box
Write $E(Y|\mathcal{F}, \mathcal{G})$ as shorthand for $E(Y|\sigma(\mathcal{F}, \mathcal{G}))$, and similarly for conditioning on more than two sub-$\sigma$-algebras. Use similar shorthand for conditional probability.

Two standard theorems on integration extend to conditional expectations:

**Theorem 1.13 (Monotone Convergence Theorem for Conditional Expectation).** Let $\mathcal{F}$ be a $\sigma$-algebra of events and $(Y_n)$ a sequence of r.v.s such that $0 \leq Y_n \leq Y_{n+1}$ a.s. Let $Y := \lim_{n \to \infty} Y_n$ a.s. Then $E(Y_n|\mathcal{F}) \to E(Y|\mathcal{F})$ a.s. as $n \to \infty$.

**Theorem 1.14 (Dominated Convergence Theorem for Conditional Expectation).** Let $\mathcal{F}$ be a $\sigma$-algebra of events and $Y, Y_1, Y_2, \ldots$ r.v.s such that $Y_n \to Y$ a.s. as $n \to \infty$. Suppose there exists a r.v. $Z$ such that $|Y_n| \leq Z$ a.s. for all $n$, and $EZ < \infty$. Then $E(Y_n|\mathcal{F}) \to E(Y|\mathcal{F})$ a.s. as $n \to \infty$.

**Proofs.** See e.g. [Williams, 1991].

The latter reference has a handy list of properties of conditional expectation in its inner back cover. We shall need two more of these, both proved in [Williams, 1991]:

**Smoothing:** If $\mathcal{G}$ is a sub-$\sigma$-algebra of $\mathcal{F}$ then
\[
E(E(X|\mathcal{F})|\mathcal{G}) = E(X|\mathcal{G}) \quad \text{a.s.} \quad (1.2)
\]

**Pull-through:** If $Y$ is $\mathcal{F}$-measurable and bounded then
\[
E(YX|\mathcal{F}) = YE(X|\mathcal{F}) \quad \text{a.s.} \quad (1.3)
\]

**Definition 1.15.** Let $\mathcal{G}, \mathcal{H}, \mathcal{K}$ be sub-$\sigma$-algebras of $\mathcal{A}$. $\mathcal{H}$ and $\mathcal{K}$ are conditionally independent given $\mathcal{G}$ if
\[
P(H \cap K|\mathcal{G}) = P(H|\mathcal{G})P(K|\mathcal{G}) \quad \forall \ H \in \mathcal{H}, \ K \in \mathcal{K}. \quad (1.4)
\]

**Exercise 1.16.** Let $X \sim \text{Bernoulli}(1/2)$, i.e. $X := 1$ with probability $1/2$, and otherwise $X := 0$. If $X = 0$ let $Y, Z$ be independent Bernoulli(1/4), while if $X = 1$ let $Y, Z$ be independent Bernoulli(3/4). Let $\mathcal{G} := \sigma(X), \mathcal{H} := \sigma(Y), \mathcal{K} := \sigma(Z)$. Check that $\mathcal{H}, \mathcal{K}$ are not independent, but are conditionally independent given $\mathcal{G}$.

**Proposition 1.17.** $\mathcal{H}$ and $\mathcal{K}$ are conditionally independent given $\mathcal{G}$ iff
\[
P(H|K, \mathcal{G}) = P(H|\mathcal{G}) \quad \text{a.s.} \quad \forall \ H \in \mathcal{H}. \quad (1.5)
\]

(1.5) is called redundant conditioning. Thus Prop. 1.17 says that conditional independence is equivalent to redundant conditioning. The definition (1.4) is symmetric as between $\mathcal{H}$ and $\mathcal{K}$, while (1.5) is not, but obviously there is an equivalent form of (1.5) obtained by interchange of symbols $H, \mathcal{H}$ with $K, \mathcal{K}$.

**Proof.** See, e.g. [Chow & Teicher, 1997, Theorem 7.3.1].
It is easy to extend conditional independence to many σ-algebras, in the obvious way.

The next result needs the following (as does the last result above).

**Lemma 1.18 (Dynkin’s Theorem).** Let $\mathcal{C}$ and $\mathcal{D}$ be classes of subsets of a fixed set $T$, satisfying $\mathcal{C} \subseteq \mathcal{D}$, and suppose $\mathcal{C}$ is closed under pairwise intersection while $\mathcal{D}$ contains $T$ and is closed under proper differences $(A, B \in \mathcal{D}, A \subseteq B$ imply $B \setminus A \in \mathcal{D})$ and non-decreasing limits. Then $\mathcal{D} \supseteq \sigma(\mathcal{C})$.

**Proof.** See e.g. [Williams, 1991].

**Proposition 1.19.** Let $(\mathcal{G}_n)_{n \in \mathbb{N}_0}$ be a filtration in $\mathcal{A}$. Let $\mathcal{H}_1, \mathcal{H}_2, \ldots$ be sub-σ-algebras of $\mathcal{A}$, and set $\mathcal{K}_n := \sigma(\mathcal{H}_1, \ldots, \mathcal{H}_n)$. Suppose that for all $n \in \mathbb{N}$,

(i) $\mathcal{H}_{n+1}$ is conditionally independent of $\mathcal{K}_n$, given $\mathcal{G}_{n+1}$, and

(ii) $\mathcal{G}_{n+1}$ is conditionally independent of $\mathcal{K}_n$, given $\mathcal{G}_n$.

Then $(\mathcal{H}_n)_{n \in \mathbb{N}}$ are conditionally independent, given $\mathcal{G}_\infty := \sigma(\mathcal{G}_1, \mathcal{G}_2, \ldots)$, and

$$P(H_n|\mathcal{G}_\infty) = P(H_n|\mathcal{G}_n) \text{ a.s. } \forall H_n \in \mathcal{H}_n, \ n \in \mathbb{N}. \quad (1.6)$$

**Proof. Step 1: preliminaries.** Understand “a.s.” to apply in what follows, and that $\mathcal{G}_n, \mathcal{H}_{n+1}, \ldots$ represent generic elements of the respective classes $\mathcal{G}_n, \mathcal{H}_{n+1}, \ldots$ etc.

Write the conditions in terms of redundant conditioning:

(i)' $P(H_{n+1}|\mathcal{K}_n, \mathcal{G}_{n+1}) = P(H_{n+1}|\mathcal{G}_{n+1})$

(ii)' $P(G_{n+1}|\mathcal{K}_n, \mathcal{G}_n) = P(G_{n+1}|\mathcal{G}_n)$.

**Step 2: we prove by induction that**

$$P(G_{n+i}|\mathcal{K}_n, \mathcal{G}_n) = P(G_{n+i}|\mathcal{G}_n) \quad (\forall G_{n+i} \in \mathcal{G}_{n+i}) \quad (1.7)$$

holds for all $i \in \mathbb{N}$. For suppose that (1.7) holds for some $i \in \mathbb{N}$. By building up through simple functions and employing the Monotone Convergence Theorem for Conditional Expectations (Theorem 1.13) you can extend this to the statement that

$$E(Y|\mathcal{K}_n, \mathcal{G}_n) = E(Y|\mathcal{G}_n) \ \forall Y \in L^1 \text{ that are } \mathcal{G}_{n+i}-\text{measurable}. \quad (1.8)$$

In (ii)' replace $n$ by $n+i$: $P(G_{n+i+1}|\mathcal{K}_{n+i}, \mathcal{G}_{n+i}) = P(G_{n+i+1}|\mathcal{G}_{n+i})$. Apply $E(\cdot|\mathcal{K}_n, \mathcal{G}_{n+i})$ to both sides and use ‘Smoothing’ (1.2):

$$P(G_{n+i+1}|\mathcal{K}_n, \mathcal{G}_{n+i}) = P(G_{n+i+1}|\mathcal{G}_{n+i}). \quad (1.9)$$

Then

$$P(G_{n+i+1}|\mathcal{K}_n, \mathcal{G}_n) = E(P(G_{n+i+1}|\mathcal{K}_n, \mathcal{G}_{n+i})|\mathcal{K}_n, \mathcal{G}_n) \quad \text{by (1.2)}$$

$$= E(P(G_{n+i+1}|\mathcal{G}_{n+i})|\mathcal{K}_n, \mathcal{G}_n) \quad \text{by (1.9)}$$

$$= E(P(G_{n+i+1}|\mathcal{G}_{n+i})|\mathcal{G}_n) \quad \text{by (1.8)}$$

$$= P(G_{n+i+1}|\mathcal{G}_n) \quad \text{by (1.2)}.$$
Thus (1.7) is proved by induction, since (ii)' is its $i = 1$ case.

Step 3: we deduce

$$P(H_1 \cap \cdots \cap H_n|G_\infty) = P(H_1 \cap \cdots \cap H_n|G_n)$$  \hspace{1cm} (1.10)

and

$$P(H_1 \cap \cdots \cap H_n|G_{n+1}) = P(H_1 \cap \cdots \cap H_n|G_n).$$  \hspace{1cm} (1.11)

First note that the conclusion of the previous Step may be expressed as saying that (1.7) holds on $\bigcup_{i=1}^\infty G_{n+i}$. We may apply Dynkin's Theorem (Lemma 1.18) because $\bigcup G_{n+i}$ contains pairwise unions of its elements, since $G_n$ increases with $n$. Hence

$$P(G_\infty|K_n, G_n) = P(G_\infty|G_n).$$

Thus $G_\infty$ is conditionally independent of $K_n$, given $G_n$; hence by the same property $P(K_n|G_\infty) = P(K_n|G_n)$ (for all $K_n \in K_n$), and in particular we conclude (1.10). On applying $E(\cdot|G_{n+1})$ to both sides and making use of ‘Smoothing’ (1.2) we deduce (1.11).

Step 4: we prove by induction that

$$P(H_1 \cap \cdots \cap H_n|G_n) = \prod_{i=1}^n P(H_i|G_i) \quad \forall H_i \in \mathcal{H}, \; i = 1, \ldots, n$$  \hspace{1cm} (1.12)

holds for all $n \in \mathbb{N}$. For suppose that (1.12) holds for some $n \in \mathbb{N}$. Then

$$P(H_1 \cap \cdots \cap H_{n+1}|G_{n+1}) = E\left(E(1_{H_1} \cdots 1_{H_{n+1}}|K_n, G_{n+1})|G_{n+1}\right)$$

(by Smoothing (1.2))

$$= E\left(1_{H_1} \cdots 1_{H_n} E(1_{H_{n+1}}|K_n, G_{n+1})|G_{n+1}\right)$$

(by Pull-through (1.3))

$$= E\left(1_{H_1} \cdots 1_{H_n} E(1_{H_{n+1}}|G_{n+1})|G_{n+1}\right)$$

(by (i)')

$$= P(H_1 \cap \cdots \cap H_n|G_{n+1}) P(H_{n+1}|G_{n+1})$$

$$= P(H_1 \cap \cdots \cap H_n|G_n) P(H_{n+1}|G_{n+1})$$

(by (1.11))

$$= \prod_{i=1}^{n+1} P(H_i|G_i)$$

(by the induction hypothesis).

So (1.12) is indeed proved by induction.
Step 5: conclusion. With (1.10), (1.12) gives
\[ P(H_1 \cap \cdots \cap H_n | G_\infty) = \prod_{i=1}^{n} P(H_i | G_i). \] (1.13)

From this (1.6) follows as a special case.
Substitute (1.6) into (1.13):
\[ P(H_1 \cap \cdots \cap H_n | G_\infty) = \prod_{i=1}^{n} P(H_i | G_\infty), \]
i.e. conditional independence. \qed

1.4. Ignatov’s Theorem.

THEOREM 1.20 (‘Ignatov’). \( R^1, R^2, \ldots \) are i.i.d.

Proof history.
- [Ignatov, 1976/77], submitted 1978, appeared 1986; continuous case
- [Deheuvels, 1983], continuous case; incomplete
- [Goldie, 1983]
- [Goldie & Rogers, 1984]
- [Stam, 1985], continuous case
- [Engelen, Tommassen & Vervaat, 1988]
- [Samuels, 1992]
- [Yao, 1997]
- [Gnedin, 2008], continuous case

Proof. Assume \( F \) has no atom at \( x_+ \) (for how to remove that assumption, see [Goldie & Rogers, 1984]).

Step 1. Let \( U \) be a finite union of intervals in \((-\infty, x_+]\) and fix \( z \leq x_+ \). For \( a \leq b \leq x_+ \) and for any sequence \((x_n) \subset (-\infty, x_+]\) let
\[ h(a, b, x_1, x_2, \ldots) := \]
\[ 1 \{ \text{no records of the sequence } (x_n) \text{ are in } ((a, b] \cap U) \cup (b, z] \}. \]
The convention that \((b, z] = \emptyset \) if \( b \geq z \) is in use here. So
\[ Eh(a, b, X_1, X_2, \ldots) = P(R^1 \cap ((a, b] \cap U) \cup (b, z]) = \emptyset \]
\[ = \exp \left( -\eta ((a, b] \cap U) \cup (b, z] \right) \]
(by Theorem 1.11)
\[ = e^{-\eta((a,b]\cap U)} e^{-\eta(b,z]} \] (1.14)

Step 2. Fix \( k \in \mathbb{N} \). Define stopping times \( T_1 := k \),
\[ T_{n+1} := \min \{ j : j > T_n, \ X_j > X_{T_n}^k \} = \min \{ j : j > T_n, \ \rho_j \leq k \}. \]
Let 
\[ Y_0 := -\infty, \quad Y_n := X_{T_n}^k \quad \text{for} \quad n = 1, 2, \ldots. \]

Step 3. The points of \( \mathbb{R}^{k+1} \) that fall in \( (Y_{n-1}, Y_n] \) are those of \( X_{T_{n+j}} \) for \( j \in \mathbb{N} \) such that
\[ \bigvee_{i=1}^{j-1} X_{T_{n+i}} < X_{T_{n+j}} \leq Y_n. \]

Thus they are the records of the sequence \( x_1 = X_{T_{n+1}}, x_2 = X_{T_{n+2}}, \ldots \) that fall in \( (Y_{n-1}, Y_n] \). They cease when \( X_{T_{n+1}} \) occurs, and \( X_{T_{n+1}} \) is just the first \( x_j = X_{T_{n+j}} \) that exceeds \( Y_n \). So
\[ 1_{\{ \mathbb{R}^{k+1} \cap (Y_{n-1}, Y_n] \cap U = \emptyset, \ X_{T_{n+1}} > z \}} = h(Y_{n-1}, Y_n, X_{T_{n+1}}, X_{T_{n+2}}, \ldots). \]

Set
\[ C_n := \{ \mathbb{R}^{k+1} \cap (Y_{n-1}, Y_n] \cap U = \emptyset \}. \]

Thus
\[ P(C_n \cap \{ X_{T_{n+1}} > z \} | A_{T_n}) = E(h(Y_{n-1}, Y_n, X_{T_{n+1}}, X_{T_{n+2}}, \ldots ) | A_{T_n}). \]

and since \( Y_{n-1}, Y_n \) are \( A_{T_n} \)-measurable, the Strong Markov Property and (1.14) give that the right-hand side evaluates as
\[ e^{-\eta(Y_{n-1}, Y_n] \cap U)} e^{-\eta(Y_n, z]} \quad \text{a.s.} \]

Thus
\[ P(C_n \cap \{ X_{T_{n+1}} > z \} | A_{T_n}) = e^{-\eta(Y_{n-1}, Y_n] \cap U)} e^{-\eta(Y_n, z]} \quad \text{a.s.} \quad (1.15) \]

Step 4. In Proposition 1.19 we shall take
\[ G_n := \sigma(X_1, \ldots, X_k-1, X_{T_1}, \ldots, X_{T_n}) = \sigma(\{ X_j : j \leq T_n, \ \rho_j \leq k \}), \]
\[ H_n := \{ \emptyset, \Omega, C_n, C_n^C \}, \]

and recall that \( K_n := \sigma(H_1, \ldots, H_n) \). Since \( \sigma(K_n, G_n) \subseteq A_{T_n}, (1.15) \) gives
\[ P(C_n \cap \{ X_{T_{n+1}} > z \} | K_{n-1}, G_n) = e^{-\eta(Y_{n-1}, Y_n] \cap U)} e^{-\eta(Y_n, z]} \]

Put \( z := -\infty: \)
\[ P(C_n | K_{n-1}, G_n) = e^{-\eta(Y_{n-1}, Y_n] \cap U)} . \]

The right-hand side is \( G_n \)-measurable, so
\[ P(C_n | G_n) = e^{-\eta(Y_{n-1}, Y_n] \cap U)} . \quad (1.17) \]

Also the last two left-hand sides are equal, so we gain condition (i) of Prop. 1.19. i.e. the conditional independence of \( H_n \) and \( K_{n-1} \) given \( G_n \).

Step 5. Take \( U := \emptyset \) in (1.16):
\[ P(X_{T_{n+1}} > z | K_{n-1}, G_n) = e^{-\eta(Y_n, z]} . \quad (1.18) \]

So (1.16) says that the event \{ \( X_{T_{n+1}} > z \) \} is conditionally independent of \( C_n \), hence of \( H_n \), given \( \sigma(K_{n-1}, G_n) \). This may be re-expressed as
\[ P(X_{T_{n+1}} > z | H_n, K_{n-1}, G_n) = P(X_{T_{n+1}} > z | K_{n-1}, G_n) , \]
and note that by (1.18) the conditioning on $K_{n-1}$ in the right-hand side is redundant, so
\[ P(X_{T_{n+1}} > z|K_n, G_n) = P(X_{T_{n+1}} > z|G_n). \]
Thus $X_{T_{n+1}}$ is conditionally independent of $K_n$, given $G_n$. Hence obviously $\sigma(X_{T_{n+1}}, G_n)$ is conditionally independent of $K_n$, given $G_n$. That is, we have condition (ii) of Prop. 1.19, that $G_{n+1}$ is conditionally independent of $K_n$, given $G_n$.

**Step 6.** So we have Prop. 1.19, and by it, the $H_n$ are conditionally independent, given $G_\infty$. Therefore
\[
P(R_{k+1} \cap U = \emptyset|G_\infty) = P\left( \bigcup_{n=1}^{\infty} C_n |G_\infty \right)
= \lim_{m \to \infty} P\left( \bigcup_{n=1}^{m} C_n |G_\infty \right) \text{ by Theorem 1.14}
= \lim_{m \to \infty} \prod_{1}^{m} P(C_n |G_n) \text{ by (1.6)}
= \prod_{1}^{\infty} P(C_n |G_n)
= \prod_{1}^{\infty} e^{-\eta(Y_{n-1}, Y_n \cap U)} \text{ by (1.17)}
= e^{-\eta(U)}.
\]

By Theorem 1.9, the conditional law of $R_{k+1}$ given $G_\infty$ is that of a Shorrock process of avoidance measure $\eta$. So $R_{k+1}$ is a $\eta$-Shorrock process, independent of $G_\infty$. But $G_\infty$ is the $\sigma$-algebra generated by all $X_n$ with $\rho_n \leq k$, so $G_\infty = \sigma(R^1, \ldots, R^k)$. The result follows. □

2. Record times

2.1. Sojourns. Assume $P(X = x_+) = 0$, so each $R^k$ is an infinite sequence.

Fix $k \in \mathbb{N}$.

Arrange $Q^k := R^1 \cup \cdots \cup R^k$ in a sequence in increasing order:
\[ Q^k = \{Q^k_1 \leq Q^k_2 \leq \cdots \}. \]
So $Q^k$ is the set of all observations of initial rank up to $k$, arranged in increasing order. If $F$ is not continuous this sequence can contain repeats.

Let $L^k_1 := k$ and
\[ L^k_{j+1} := \min\{n : n > L^k_j, \rho_n \leq k\}. \]
These are the times when the $k^{th}$ order statistic $X^k$ steps to the next point of $Q^k$: 

$$X^k_n = Q^k_j \text{ for all } n \text{ with } L^k_j \leq n < L^k_{j+1}.$$ 

**Definition 2.1.** The *sojourn* of the $k^{th}$ order statistic at $Q^k_j$, the time it spends there, is 

$$\Delta^k_j := L^k_{j+1} - L^k_j \quad (j = 1, 2, \ldots).$$ 

**Figure 2.2.** The 3$^{rd}$ order statistic for Vancouver rainfall.

In Fig. 2.2 we plot the 3$^{rd}$ order statistic ($m = 3$) for the data used for Fig. 1.5. The observations having initial rank at most 3 are plotted as $(n, X_n)$. Each such $X_n$ eventually becomes a value taken by the third order statistic, with a sojourn that is shown as a horizontal line segment at the same height as $X_n$. There are no coincident values here, so to illustrate this effect we use in Fig. 2.3 the precipitation data for Vancouver in July in reverse time order (1966 back to 1906) and again display the 3$^{rd}$ order statistic. The times $n = 1$ through to $n = 61$ correspond to years 1966 down to 1906. One can see that the two values of 3-36, which occurred at times $n = 1$ and $n = 25$, give rise to end-to-end sojourns from $n = 25$ to $n = 35$ and from $n = 35$ to $n = 51$, at the same height.

**Theorem 2.4.** The sojourns $\Delta^k_1, \Delta^k_2, \ldots$ are conditionally independent given $Q^k$, with geometric distributions 

$$P(\Delta^k_j = l|Q^k) = \left(1 - F(Q^k_j)\right)F(Q^k_{j+1}) \quad (l = 1, 2, \ldots).$$
Figure 2.3. Vancouver in July in reverse time: 3rd order statistic.

Proof. Fix \( l \in \mathbb{N}, z \leq x_+ \), and define
\[
f(b, x_1, \ldots, x_l) := \mathbf{1}\left\{ \bigvee_{i=1}^{l-1} x_i \leq b, \ x_l > b \lor z \right\}
\]
for \( b, x_1, \ldots, x_l \leq x_+ \); then
\[
E f(b, X_1, \ldots, X_l) = F(b)^{l-1} \bar{F}(b \lor z).
\]
Now \( L^k_j \) is the \( T_j \) in the Ignatov proof (proof of Theorem 1.20) above, so
\[
\mathbf{1}\{\Delta^k_n = l, X_{T_n+1} > z\} = f(X^k_{T_n}, X_{T_n+1}, \ldots, X_{T_n+l}).
\]
Therefore
\[
P(\Delta^k_n = l, X_{T_n+1} > z | A_{T_n}) = E\left( f(X^k_{T_n}, X_{T_n+1}, \ldots, X_{T_n+l}) | A_{T_n} \right)
\]
by the Strong Markov Property (Exercise 1.1) and (2.1). Now use Proposition 1.19 as in the Ignatov proof, with the same \( G_n \) but now with \( H_n := \sigma(\Delta^k_n) \). The conditions hold, so the Proposition gives conditional independence of \( \Delta^k_1, \Delta^k_2, \ldots \) given \( G_\infty \), with respective laws
\[
P(\Delta^k_j = l | G_\infty) = F(X^k_{T_j})^{l-1} \bar{F}(X^k_{T_j})
\]
As \( \sigma(Q^k) \subseteq \sigma(G_\infty) \) and these right-hand sides are \( \sigma(Q^k) \)-measurable, we may take the conditional expectation given \( Q^k \) of the above, and obtain conditional laws given \( Q^k \) and conditional independence given \( Q^k \).

Definition 2.5. Let \( X^{m-\cdot}(\cdot) \) be the left-continuous inverse of \( X^m \):
\[
X^{m-\cdot}(x) := \inf\{n \geq m : X^m_n \geq x\} \quad (x \leq x_+).
\]
Left-continuity yields the convenient relationship
\[
X^{m-\cdot}(x) \leq n \iff x \leq X^m_n.
\]
Then
\[
X^{m\leftarrow}[x, y] := X^{m\leftarrow}(y) - X^{m\leftarrow}(x)
= \#\{n : X_n^m \in [x, y]\}.
\]

**Theorem 2.6.** Fix \(m \in \mathbb{N}\). The process \(X^{m\leftarrow}\) has independent increments: for any disjoint intervals \(I_1, \ldots, I_k \subset (-\infty, x_+],\)

\[
P(X^{m\leftarrow} I_1 = n_1, \ldots, X^{m\leftarrow} I_k = n_k) = \prod_{l=1}^k P(X^{m\leftarrow} I_l = n_l),
\]

and

\[
P(X^{m\leftarrow}[x, y] = n)
= \begin{cases} 
(F_{y, \infty}/F_{x, \infty})^m & \text{for } n = 0, \\
\left(\sum_{k=1}^{m\wedge n} \binom{m}{k} \binom{n-1}{k-1} F[x, y]^k F(-\infty, y)^{n-k}\right) & \text{for } n = 1, 2, \ldots.
\end{cases}
\tag{2.2}
\]

**Proof.** \(X^{m\leftarrow} = \sum_{r \in Q^m I} \Delta_r^m\), and between different \(I\) the \(Q^m I\) are independent and the \(\Delta_r^m\) conditionally independent given \(Q^m\), so the \(X^{m\leftarrow}\) are independent.

To prove (2.2) for \(n = 0\), note that \(X^{m\leftarrow}[x, y] = 0\) if and only if all of the first \(m\) of the \(X_i\) that are \(\geq x\) are also \(\geq y\). Hence the result.

Fix \(n \geq 1\) and consider the event that \(X^{m\leftarrow}[x, y] = n\). It needs that of the first \(m\) observations \(\geq x\), \(k\) must be \(< y\), where \(1 \leq k \leq m\) (i.e. not \(k = 0\)). That has probability

\[
\left(\binom{m}{k} \frac{F[x, y]}{F[x, \infty]}^k \frac{F[y, \infty]}{F[x, \infty]}^{m-k}\right)
\]

and accounts for 1 time unit spent by \(X^m\) in \([x, y]\).

Then for \(X^m\) to have sojourn time \(n\) in \([x, y]\),

(a) of the next \(n - 1\) observations, exactly \(n - k\) must be \(< y\), and

(b) the next observation after that must be \(\geq y\).

Now (a) has probability \(\binom{n-1}{k-1} F[y, \infty]^{k-1} F(-\infty, y)^{n-k}\) and requires \(k \leq n\), while (b) has probability \(F[y, \infty]\). So

\[
P(X^{m\leftarrow}[x, y] = n) = \sum_{k=1}^{m\wedge n} \left(\binom{m}{k} \frac{F[x, y]}{F[x, \infty]}^k \frac{F[y, \infty]}{F[x, \infty]}^{m-k}. \right)
\cdot \binom{n-1}{k-1} F[y, \infty]^{k-1} F(-\infty, y)^{n-k} F[y, \infty],
\]

hence (2.2). \(\square\)
2.2. Times.

Theorem 2.7. Assume $F$ is continuous. Then $\rho_1, \rho_2, \ldots$ are independent and $\rho_n \sim \text{Unif}\{1, \ldots, n\}$.

The uniform distribution here is discrete uniform. The result is known as the Dwass-Rényi Lemma [Dwass, 1960], [Rényi, 1962].

Proof. As is conventional, let $S_n$ denote the set of permutations on the set $\{1, \ldots, n\}$, i.e. injective maps $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$. Any $n$-tuple $(x_1, \ldots, x_n)$ of unequal values generates both $\pi \in S_n$ via $x_{\pi(1)} < x_{\pi(2)} < \cdots < x_{\pi(n)}$, and an $n$-tuple of initial ranks $r = (r_1, \ldots, r_n) \in \{1\} \times \{1, 2\} \times \cdots \times \{1, \ldots, n\}$, and there is a 1–1 correspondence between the $\pi$ and $r$.

As $F$ is continuous, the random permutation $\Pi$ given by $(X_1, \ldots, X_n)$ is uniform on $S_n$, so $(\rho_1, \ldots, \rho_n)$ is uniform on its space, i.e. $\rho_1, \ldots, \rho_n$ are independent, with $\rho_k \sim \text{Unif}\{1, \ldots, k\}$. □

Aside: number of records. Let $N_n$ be the number of records among $X_1, \ldots, X_n$:

$$N_n := \sum_{k=1}^n I_k \quad \text{where} \quad I_k := 1\{\rho_k = 1\}.$$ 

By the Dwass-Rényi Lemma the $I_k$ are independent with $P(I_k = 1) = \frac{1}{k}$, $P(I_k = 0) = 1 - \frac{1}{k}$.

So, with $\gamma = 0.57721 \cdots$ the Euler-Mascheroni constant,

$$EN_n = \sum_{k=1}^n \frac{1}{k} = \ln n + \gamma + O\left(\frac{1}{n}\right);$$

$$\text{var} \ N_n = \sum_{k=1}^n \left(\frac{1}{k} - \frac{1}{k^2}\right) = \ln n + \gamma - \frac{\pi^2}{6} + O\left(\frac{1}{n}\right).$$

One may prove (see e.g. [Resnick, 1999, §7.4, §9.8])

$$\frac{N_n}{\ln n} \overset{a.s.}{\longrightarrow} 1, \quad \frac{N_n - \ln n}{\sqrt{\ln n}} \overset{d}{\longrightarrow} N(0, 1),$$

etc.

Definition 2.8. The values of $n$ when $\rho_n = 1$ are the record times $1 = L_1 < L_2 < \cdots$.

If $P(X = x_\ast) = 0$ this is an infinite sequence.

We restrict attention to record times but all that follows has versions for $k$-record times, suitably defined.
Theorem 2.9. Assume $F$ continuous. Then $(L_n)_{n \geq 1}$ is a Markov chain with $L_1 = 1$ and stationary transition laws

$$P(L_{n+1} = l | L_n = j) = \frac{j}{(l-1)l} \quad (l = j + 1, j + 2, \ldots)$$

(2.3)

Proof. It suffices to show that for $1 < l_2 < \cdots < l_n$,

$$P(L_2 = l_2, \ldots, L_n = l_n) = \frac{1}{(l_2 - 1) \cdots (l_n - 1) \cdot l_n} \quad \forall n,$$

because then

$$P(L_{n+1} = l_{n+1} | L_2 = l_2, \ldots, L_n = l_n) = \frac{(\frac{1}{(l_2-1) \cdots (l_{n+1}-1) \cdot 1_{n+1}})}{(\frac{1}{(l_2-1) \cdots (l_n-1) \cdot 1_n})} \frac{l_n}{(l_{n+1}-1)l_{n+1}},$$

whence the Markov property and (2.3).

By the Dwass-Rényi Lemma 2.7,

$$P(L_2 = l_2, \ldots, L_n = l_n) = P(\rho_k = 1 \text{ when } k \in \{l_2, \ldots, l_n\}, \rho_k \neq 1 \text{ when } k \in \{2, \ldots, l_n\} \setminus \{l_2, \ldots, l_n\})$$

$$= \left(1 - \frac{1}{2}\right) \left(1 - \frac{1}{3}\right) \cdots \left(1 - \frac{1}{l_2 - 1}\right) \frac{1}{l_2} \left(1 - \frac{1}{l_2 + 1}\right) \cdots \left(1 - \frac{1}{l_n - 1}\right) \frac{1}{l_n}$$

$$= \frac{1}{2} \frac{2}{3} \frac{l_2 - 2}{l_2 - 1} \frac{1}{l_2} \frac{l_3 - 2}{l_3 - 1} \frac{1}{l_3} \frac{l_4 - 2}{l_4 - 1} \cdots \frac{l_n - 2}{l_n - 1} \frac{1}{l_n}$$

$$= \frac{1}{l_2 - 1} \cdots \frac{1}{l_n - 1} \frac{1}{l_n}$$

as claimed. \(\square\)

Lemma 2.10. Assume $F$ continuous. Let $W_1, W_2, \ldots$ be i.i.d. Unif$(0,1)$, independent of $(L_j)_{j \geq 1}$. Define

$$Y_n := -\ln \left((1 - W_n) \frac{L_n}{L_{n+1}} + W_n \frac{L_n}{L_{n+1} - 1}\right), \quad (n = 1, 2, \ldots).$$

Then $Y_1, Y_2, \ldots$ are i.i.d. Expon$(1)$ r.v.s.

Proof. We prove, equivalently, that the random variables $V_n := e^{-Y_n}$ are independent Unif$(0,1)$.

Given $(L_n)_{n \in \mathbb{N}} = (l_n)_{n \in \mathbb{N}}$ the $V_n$ are independent with conditional laws

$$\text{Unif} \left(\frac{l_n}{l_{n+1}}, \frac{l_n}{l_{n+1} - 1}\right)$$
respectively. So for \( v_1, \ldots, v_n \) irrational in \((0,1)\), and \(1 = l_1 < l_2 < \cdots < l_{n+1}\),

\[
P(V_1 \in dv_1, \ldots, V_n \in dv_n | L_1 = l_1, \ldots, L_{n+1} = l_{n+1}) = \prod_{j=1}^{n} \frac{dv_j}{l_j l_{j+1} - l_j l_{j+1} - 1} 1 \left\{ \frac{l_j}{l_{j+1} - 1} < v < \frac{l_j}{l_{j+1} + 1} \right\}
\]

Sum out \( L_{n+1} \) by taking \( E(\cdot | L_1 = l_1, \ldots, L_n = l_n) \). The left-hand side becomes \( P(V_1 \in dv_1, \ldots, V_n \in dv_n | L_1 = l_1, \ldots, L_n = l_n) \), while on the right we multiply by \( P(L_{n+1} = l_{n+1} | L_1 = l_1, \ldots, L_n = l_n) \), and add over all values of \( l_{n+1} \), so the last term in the product becomes just \( dv_n \).

Then take \( E(\cdot | L_1 = l_1, \ldots, L_{n-1} = l_{n-1}) \), and so on, hence

\[
P(V_1 \in dv_1, \ldots, V_n \in dv_n) = dv_1 \cdots dv_n,
\]

which shows that the \( V_n \) are indeed independent Unif\((0,1)\).

**Theorem 2.11** ([Williams-Pfeifer Strong Approximation for Record Times, [Williams, 1973], [Pfeifer, 1987]). Assume \( F \) continuous. Use the probability space extended by the \( W_n \) as above. Then

\[
L_{n+1} = \lceil L_n e^{Y_n} \rceil \quad \text{for } n = 1, 2, \ldots,
\]

where

\[
[x] := \min\{n \text{ integer}, n \geq x\}.
\]

**Proof.** From the previous proof,

\[
\frac{L_n}{L_{n+1}} < V_n < \frac{L_n}{L_{n+1} - 1} \quad \text{a.s.}
\]

As \( 1/V_n = e^{Y_n} \),

\[
L_{n+1} - 1 < L_n e^{Y_n}, \quad L_{n+1} > L_n e^{Y_n} \quad \text{a.s.,}
\]

hence the result.

**Theorem 2.12** ([Pfeifer, 1987]). Assume \( F \) continuous. Use the probability space extended by the \( W_n \) as in Lemma 2.10. Set \( S_n := \sum_1^n Y_j \). Then there exists \( Z > 0 \) with \( E(Z^k) < \infty \) for all \( k \), such that \( Z \) and \((S_n - n)/\sqrt{n}\) are asymptotically independent, and

\[
\ln L_n = Z + S_{n-1} + o(e^{-n/2}) \quad \text{a.s.} \quad (n \to \infty).
\]

**Proof.** Step 1: existence of \( Z \). Rearrange the definition of \( Y_n \) to give

\[
\frac{L_{n+1}}{L_n} = e^{Y_n} \left( 1 + \frac{W_n}{L_{n+1} - 1} \right),
\]
whence

\[ \ln L_{n+1} - \ln L_n = Y_n + \ln \left(1 + \frac{W_n}{L_{n+1} - 1}\right). \]

Add; hence \( \ln L_n = S_{n-1} + Z_n \) where

\[ Z_n := \sum_{j=1}^{n-1} \ln \left(1 + \frac{W_j}{L_{j+1} - 1}\right) =: \sum_{j=1}^{n-1} C_j, \]

say. Then
\[
0 < Z_n \uparrow Z := \sum_{j=1}^{\infty} C_j < \infty \sum_{j=1}^{\infty} \frac{1}{L_{j+1} - 1}.
\]

Now \( S_n/n \overset{a.s.}{\to} 1 \), so w.p. 1 (with probability 1) \( S_{n-1} > \frac{3}{4} n \) for all large \( n \), and \( L_n > e^{S_{n-1}} > e^{3n/4} \) so \( Z < \infty \) a.s.

**Step 2: the final claim.** From the above, \( \ln L_n = S_{n-1} + Z - R_n \) where \( R_n := \sum_{j=n}^{\infty} C_j > 0 \). Now w.p. 1, for all large \( n \),
\[
R_n < \sum_{j=n}^{\infty} \frac{1}{e^{3j/4} - 1} < 2 \sum_{j=n}^{\infty} e^{-3j/4} = \frac{2e^{-3n/4}}{1 - e^{-3/4}} = o(e^{-n/2}),
\]
proving the final claim.

**Step 3: \( E(Z^p) < \infty \).** Now
\[
0 < C_j < \frac{1}{L_{j+1} - 1} \leq 2e^{-S_j},
\]
so
\[
E|C_j|^p \leq 2^p e^{-pS_j} = \frac{2^p}{(1 + p)^j}
\]
since \( S_j = \sum Y_i \) and \( Y_i \sim \text{Expon}(1) \). Thus \( \|C_j\|_p \leq 2/(1 + p)^j/p \), and as this is summable in \( j \) it follows that \( \|Z\|_p < \infty \).

**Step 4: asymptotic independence.** Asymptotic independence of \( Z \) and \( (S_n - n)/\sqrt{n} \) is the claim that
\[
P\left( Z \leq z, \frac{S_n - n}{\sqrt{n}} \leq s \right) \to P(Z \leq z)\Phi(s) \quad (n \to \infty). \quad (2.4)
\]

By the Williams-Pfeifer Strong Approximation, i.e. \( L_{n+1} = \lceil L_ne^{Y_n} \rceil \), we may regard \( L_n \) as \( \sigma(Y_1, \ldots, Y_{n-1}) \)-measurable. Then by the defining formula for \( Y_n \) we may regard \( W_n \) as \( \sigma(Y_1, \ldots, Y_n) \)-measurable. Thus \( Z_m \) is \( \sigma(Y_1, \ldots, Y_m) \)-measurable, so is independent of \( (Y_j)_{j>n} \). So for fixed \( m \),
\[
P\left( Z_m \leq z, \frac{S_n - S_m - (n-m)}{\sqrt{n-m}} \leq s \right) \to P(Z_m \leq z)\Phi(s) \quad (n \to \infty).
\]

By Dini’s Theorem this holds uniformly in \( s \). Since
\[
\frac{S_n - n}{\sqrt{n}} = \frac{S_n - S_m - (n-m)}{\sqrt{n-m}}(1 + o(1)) + o(1) \quad \text{a.s.,}
\]
it follows that
\[ P\left(Z_m \leq z, \frac{S_n - n}{\sqrt{n}} \leq s\right) \to P(Z_m \leq z)\Phi(s). \]

We may make \(m\) so large that \(P(|Z_m - Z| > \varepsilon) < \varepsilon\). Then
\[ P\left(Z \leq z, \frac{S_n - n}{\sqrt{n}} \leq s\right) \leq P\left(Z_m \leq z + \varepsilon, \frac{S_n - n}{\sqrt{n}} \leq s\right) + \varepsilon \]
\[ \to P(Z_m \leq z + \varepsilon)\Phi(s) + \varepsilon \]
\[ \leq P(Z \leq z + 2\varepsilon)\Phi(s) + 2\varepsilon. \]

Hence
\[ \limsup_{n \to \infty} P\left(Z \leq z, \frac{S_n - n}{\sqrt{n}} \leq s\right) \leq P(Z \leq z)\Phi(s). \]

Similarly for the lower bound. So (2.4) holds at all \(z\) that are not atoms of \(Z\). This gives the result, i.e. that \((Z, (S_n - n)/\sqrt{n}) \Rightarrow (Z, \Phi)\). \(\square\)

3. Limits

3.1. Limit laws for record values.

**Definition 3.1.** For laws \(F_n, F\) on \(\mathbb{R}\), say \(F_n \Rightarrow F\) (‘weak’ convergence) if
\[ \int f \, dF_n \to \int f \, dF \quad \forall f \in C_b^+(\mathbb{R}), \]
(where \(C_b^+(\mathbb{R})\) is the set of all continuous bounded non-negative functions on \(\mathbb{R}\)).

**Definition 3.2.** R.v.s \(X, Y\), or equivalently their laws \(F, G\), are of the same type if there exist \(a \in (0, \infty)\) and \(b \in \mathbb{R}\) so that
\[ Y \overset{L}{=} aX + b, \text{ equivalently } G(y) = F\left(\frac{y - b}{a}\right) \forall y. \]

**Exercise 3.3.** Show this is an equivalence relation on laws on \(\mathbb{R}\).

**Definition 3.4.** The equivalence classes are the types.

**Exercise 3.5.** Name or describe the types containing the laws (a) \(N(0, 1)\), (b) \(\text{Unif}(0, 1)\), (c) \(\text{Expon}(1)\), (d) \(\text{Unif}\{1, \ldots, n\}\), (e) degenerate at 0.

**Theorem 3.6 (Convergence of Types).** Let \(X, X_n\) be r.v.s, \(a_n > 0, b_n \in \mathbb{R}\) (norming and centring constants, or scale and location constants), such that
\[ \frac{X_n - b_n}{a_n} \Rightarrow X \quad (n \to \infty), \]
with $X$ non-degenerate. Let $Y$ be a r.v., $\alpha_n > 0$, $\beta_n \in \mathbb{R}$ constants. Then

$$(i) \quad \frac{X_n - \beta_n}{\alpha_n} \Rightarrow Y$$

iff

$$(ii) \quad a_n \rightarrow \alpha \in [0, \infty), \quad \frac{b_n - \beta_n}{\alpha_n} \rightarrow \beta \in \mathbb{R} \quad (n \rightarrow \infty).$$

In that case $Y \overset{d}{=} \alpha X + \beta$, and $\alpha \geq 0$, $\beta$ are the unique constants for which this holds.

When (i) or (ii) holds, $Y$ is non-degenerate iff $\alpha > 0$, and $X$ and $Y$ are then of the same type.

**Proof.** [Gnedenko & Kolmogorov, 1954]. Later books, such as [Billingsley, 1995], give a slightly less general version of the Theorem in which $Y$ is also assumed non-degenerate, though it is not hard to extend the proof to the full version. □

**Theorem 3.7 ([Resnick, 1973]).** Assume $F$ continuous. Then the possible limit laws for $(R_n - b_n)/a_n$ are those in the type of one of

$$(i) \quad \tilde{\Phi}_\alpha(x) := \begin{cases} 0 & \text{if } x \leq 0, \\ \Phi(\alpha \ln x) & \text{if } x > 0; \end{cases}$$

$$(ii) \quad \tilde{\Psi}_\alpha(x) := \begin{cases} \Phi(-\alpha \ln(-x)) & \text{if } x < 0, \\ 1 & \text{if } x \geq 0; \end{cases}$$

$$(iii) \quad \Phi,$$

where $\Phi$ is the $N(0, 1)$ d.f. and $\alpha > 0$ is constant.

The proof needs some more constructions and two lemmas.

**Definition 3.8.** $H(x) := H(-\infty, x]$ where $H$ is the hazard measure corresponding to $F$, defined by

$$H(A) := \int_A \frac{dF(x)}{F(x, \infty)}$$

for Borel sets $A$ in $\mathbb{R}$.

**Proposition 3.9.** $H$ is the intensity measure of the point process $\mathbb{R}^1$: $H(A) = E\#(\mathbb{R}^1 \cap A)$. 


PROOF. Note that we do not assume $F$ continuous. We have

$$
E\#(\mathbf{R}^1 \cap (-\infty, u]) = E \sum_{n=1}^{\infty} 1\{X_n \in \mathbf{R}^1, X_n \leq u\}
= \sum_{n=1}^{\infty} P\left( \bigvee_{i=1}^{n-1} X_i < X_n \leq u \right)
= \sum_{n=1}^{\infty} \int_{(-\infty,u]} F^{n-1}(-\infty, x) \, dF(x)
= \int_{(-\infty,u]} dF(x) \left( 1 - F(-\infty, x) \right) = \int_{(-\infty,u]} \frac{dF(x)}{F[x, \infty]},
$$

as desired. \hfill \Box

**Definition 3.10.** The associated law for $F$ is $A(x) := 1 - e^{-\sqrt{H(x)}}$.

It is a d.f.

**Definition 3.11.** For $\tilde{G}$ one of $\tilde{\Phi}_\alpha$, $\tilde{\Psi}_\alpha$, $\Phi$ we say $F$ is in the domain of attraction for record values of $\tilde{G}$, notation $F \in DR(\tilde{G})$, if there exist $a_n > 0$, $b_n$ so that $(R_n^1 - b_n)/a_n \Rightarrow \tilde{G}$.

**Exercise 3.12.** Find the densities of $\tilde{\Phi}_\alpha$ and $\tilde{\Psi}_\alpha$. Note that $\tilde{\Phi}_1$ is the log-normal law.

**Lemma 3.13.** For any $r$ and $n$, $P(R_n \leq r) = P(H(R_n) \leq H(r))$. (This does not need $F$ to be continuous.)

**Proof.** If $R_n \leq r$ then $H(R_n) \leq H(r)$ as $H$ is non-decreasing. So

$$\{H(R_n) \leq H(r)\} = \{R_n \leq r\} \cup \{R_n > r, H(R_n) \leq H(r)\}.$$

The latter event comes down to $R_n$ lying in an open interval $I$ to the right of $r$, this interval being one where $H$ is flat. As $H$ is right-continuous, $H(I) = 0$, so $F(I) = 0$. So also $P(R_n \in I) = 0$, hence the result. \hfill \Box

**Lemma 3.14.** Assume $F$ continuous. Then

$$
(H(R_n))_{n \geq 1} \xrightarrow{L} \left( \sum_{i=1}^{n} E_i \right)_{n \geq 1},
$$

where $E_1, E_2, \ldots$ are independent Expon(1) r.v.s.

**Proof.** Write $R_n := R_n^1$. Note $R_1 = X_1 \sim F$, and

$$R_{n+1}|(R_1, \ldots, R_n) \sim X|X > R_n. \tag{3.1}$$

As $F$ is continuous, $(F(X_n))_{n \in \mathbb{N}}$ is a sequence of independent Unif(0, 1) r.v.s, and

$$H(x) = \int_{-\infty}^{x} \frac{dF(u)}{1 - F(u)} = -\ln(1 - F(x)),$$

RECORDS 21
so \((H(X_n))_{n \in \mathbb{N}}\) is a sequence of independent \text{Expon}(1) r.v.s. Further, 
\((H(R_n))_{n \in \mathbb{N}}\) is the record-value sequence of \((H(X_n))_{n \in \mathbb{N}}\) almost surely. From (3.1), this record-value sequence has the claimed representation. □

**Proof of Theorem 3.7.** Suppose \((R_n - b_n)/a_n \Rightarrow \tilde{G}\) where \(\tilde{G}\) is non-degenerate. Choose \(x\) where \(\tilde{G}\) is continuous. Then
\[
\tilde{G}(x) \leftarrow P(R_n \leq a_n x + b_n) = P(H(R_n) \leq H(a_n x + b_n))
\]
by Lemma 3.13
\[
= P\left(\frac{\sum_{i=1}^{n} E_i - n}{\sqrt{n}} \leq \frac{H(a_n x + b_n) - n}{\sqrt{n}}\right)
\]
by Lemma 3.14. But
\[
P\left(\frac{\sum_{i=1}^{n} E_i - n}{\sqrt{n}} \leq y\right) \to \Phi(y) \quad \forall \ y \in [-\infty, \infty].
\]
Because \(\Phi\) is strictly increasing, there must exist \(g(x) \in \mathbb{R}\) such that
\[
\frac{H(a_n x + b_n) - n}{\sqrt{n}} \to g(x) \quad (n \to \infty)
\]
and
\[
\Phi(g(x)) = \tilde{G}(x).
\]
(3.3)
Now \(g\) inherits monotonicity from \(H\). Thus (3.2–3.3) hold for all continuity points of \(\tilde{G}\), where \(g\) is a non-decreasing \([-\infty, \infty]\)-valued function.

Since \(\Phi\) is strictly increasing and continuous, \(g\) and \(\tilde{G}\) have the same continuity points. Make \(g\) right-continuous, then (3.3) holds for all \(x\). Since \(\tilde{G}\) is non-degenerate, there exist continuity points where \(g\) is finite.

We use the notation \([\cdot]\) for integer part. Now (3.2) is equivalent to
\[
\lim_{t \to \infty} \frac{H(a_{[t]} x + b_{[t]}) - t}{\sqrt{t}} \to g(x) \quad \forall \ x \text{ where } g \text{ cts},
\]
(3.4)
since \(t/\lfloor t \rfloor \to 1\) and \((t - \lfloor t \rfloor)/\sqrt{t} \to 0\). If also \(g(x)\) is finite, this implies
\[
H(a_{[t]} x + b_{[t]}) \sim t \quad (t \to \infty),
\]
(3.5)
hence
\[
\sqrt{H(a_{[t]} x + b_{[t]})} - \sqrt{t} = \frac{H(a_{[t]} x + b_{[t]}) - t}{\sqrt{t}} \cdot \frac{\sqrt{t}}{\sqrt{H(a_{[t]} x + b_{[t]})} + \sqrt{t}} \to g(t) \cdot \frac{1}{2}
\]
(3.6)
Conversely, (3.6) \(\Rightarrow\) (3.5) \(\Rightarrow\) (3.4).
Set $u := e^{\sqrt{t}}$, $a_{[t]} = \alpha(u)$, $b_{[t]} = \beta(u)$. Recall $A(x) = 1 - e^{-\sqrt{H(x)}}$, so (3.6) becomes

$$u(1 - A(\alpha(u)x + \beta(u))) \rightarrow e^{-\frac{1}{2}g(x)} \quad (u \rightarrow \infty)$$

for all $x$ where the right-hand side is continuous and in $(0, \infty)$.

Since $1 - A(\cdots) \rightarrow 0$,

$$1 - A(\cdots) \sim -\ln(1 - (1 - A(\cdots))) = -\ln A(\cdots).$$

So finally (3.2) is equivalent to

$$A^u(\alpha(u)x + \beta(u)) \rightarrow e^{-e^{-\frac{1}{2}g(x)}} \quad (u \rightarrow \infty)$$

(3.7)

for all $x$ where the right-hand side is continuous and in $(0, 1)$.

The left-hand side of (3.7) is the d.f. of

$$\max(Y_1, \ldots, Y_u) - \beta(u)$$

where $Y_1, Y_2, \ldots$ are i.i.d. $\sim A$. So $\exp(-e^{-\frac{1}{2}g(x)})$ must be an extreme-value limit law. The result follows by extreme-value theory (see §3.2).

**3.2. Extreme-value theory.** Let $Y_1, Y_2, \ldots$ be i.i.d. $\sim A$.

Set $M_n := \max\{Y_1, \ldots, Y_n\}$.

¿When does $(M_n - \beta_n)/\alpha_n \Rightarrow G$ non-degenerate?

**Theorem 3.15 (B. V. Gnedenko).** The possible $G$ are those in the types of

(i) $\Phi_\alpha(x) := \begin{cases} 0 & (x \leq 0) \\ e^{-x^{-\alpha}} & (x > 0) \end{cases}$ the Fisher-Tippett type,

(ii) $\Psi_\alpha(x) := \begin{cases} e^{-(x)\alpha} & (x < 0) \\ 1 & (x \geq 0) \end{cases}$ the Fréchet type,

(iii) $\Lambda(x) := e^{-e^{-x}} \quad (x \in \mathbb{R})$, the Gumbel type,

where $\alpha > 0$.

**Proof.** see, e.g. [Bingham, Goldie & Teugels, 1989, Theorem 8.13.1].

**Definition 3.16.** When $(M_n - \beta_n)/\alpha_n \Rightarrow G$, the law $A$ of $Y_1, Y_2, \ldots$ is in the domain of attraction for maxima of $G$, notation $A \in DM(G)$.

**Corollary 3.17.** Let $F$ be continuous and $A$ its associated law. Then

(i) $F \in DR(\Phi_\alpha)$ iff $A \in DM(\Phi_{\alpha/2})$;
(ii) $F \in DR(\Psi_\alpha)$ iff $A \in DM(\Psi_{\alpha/2})$;
(iii) $F \in DR(\Phi)$ iff $A \in DM(\Lambda)$. 

The constants in
\[ \frac{R_n - b_n}{a_n} \Rightarrow \tilde{G}, \quad \frac{M_n - \beta_n}{\alpha_n} \Rightarrow G \]
can be taken to be related by
\[ a_n = \alpha(e\sqrt{n}), \quad b_n = \beta(e\sqrt{n}). \]

**Proof.** By (3.2) and (3.7),
\[ \frac{R_n - b_n}{a_n} \Rightarrow \tilde{G} \iff \frac{M_n - \beta(n)}{\alpha(n)} \Rightarrow e^{-e^{-\frac{1}{2}g(x)}} \]
where \( \Phi(g(x)) = \tilde{G}(x) \) and the \( a_n, \alpha_n, b_n, \beta_n \) are related as claimed.

When \( \tilde{G}(x) = \Phi(\ln(x)) \) we have \( g(x) = \ln(x) \), so \( G(x) = \exp(-e^{-x/2}) = \Phi_{\alpha/2}(x) \). Similarly for \( \tilde{\Psi}_\alpha(x) \).

When \( \tilde{G}(x) = \Phi(\ln(-\ln G(x))) \) when \( \tilde{G} = \tilde{\Phi}_\alpha \) and \( G = \Phi_\alpha \), and similarly when \( \tilde{G} = \tilde{\Psi}_\alpha \) and \( G = \Psi_\alpha \), and when \( \tilde{G} = \Phi \) and \( G = \Lambda \).

**Exercise 3.18.** Check that the laws \( \Phi_\alpha, \Psi_\alpha, \Lambda \) are max-stable: if \( Y_1, Y_2, \ldots \) are i.i.d. \( \sim A \), where \( A \) is in one of these types, then each \( M_n \) has the same type as \( A \).

**Exercise 3.19.** Check that \( \tilde{G}(x) = \Phi(\ln(-\ln G(x))) \) when \( \tilde{G} = \tilde{\Phi}_\alpha \) and \( G = \Phi_\alpha \), and similarly when \( \tilde{G} = \tilde{\Psi}_\alpha \) and \( G = \Psi_\alpha \), and when \( \tilde{G} = \Phi \) and \( G = \Lambda \).

### 4. Extensions

#### 4.1. Records in a partially ordered set.

Let \( S \) be a set with a partial order \( < \).

That is, the graph \( G_< := \{(x, y) \in S \times S : x < y \} \) has properties
\begin{enumerate}
  \item[(i)] antisymmetry: \( x < x \) for no \( x \in S \),
  \item[(ii)] transitivity: \( x < y, y < z \Rightarrow x < z \).
\end{enumerate}

Let \( S \) be a \( \sigma \)-algebra of subsets of \( S \). Assume
\[ G_< \in S \times S \quad \text{(the graph is product-measurable)}. \]

Let \( \mu \) be a probability law on \((S, S)\).

Let \( X_1, X_2, \ldots \) be i.i.d. \( \sim \mu \).

Adjoin extra points \( -\infty, \infty \) with the properties
\[ -\infty < x < \infty \quad \forall x \in S. \]

Let \( S^* := S \cup \{-\infty, \infty\} \).

Define *intervals* \((x, y) := \{z \in S : x < z < y\} \) for \( x, y \in S^* \).

Let
\[ S_\mu := \{y \in S : \mu(\{-\infty, y\}) < 1\}. \]

Note first that Fubini’s Theorem gives that
\[ \int_S 1_{G_<}(x, y) d\mu(x) = \mu(\{-\infty, y\}) \]
is measurable as a function of \( y \). So \( S_\mu \in S \).
Proposition 4.1. \( \mu(S_\mu) = 1 \).

Proof. Set \( C := S_\mu^c \). For \( x \in C \) the set \((-\infty, x)\) is \( \mu \)-full so \( \mu(C \cap (-\infty, x)) = \mu(C) \). Integrating this gives \( \int_C \mu(C \cap (-\infty, x)) \, d\mu(x) = \mu^2(C) \), i.e.

\[
\int_C \int_C 1\{y < x\} \, d\mu(y) \, d\mu(x) = \mu^2(C).
\] (4.1)

Fubini’s Theorem allows interchange of order of integration:

\[
\int_C \int_C 1\{y < x\} \, d\mu(x) \, d\mu(y) = \mu^2(C),
\]
and on interchanging the dummy variables \( x \) and \( y \) we conclude

\[
\int_C \int_C 1\{x < y\} \, d\mu(y) \, d\mu(x) = \mu^2(C).
\] (4.2)

The sum of the left-hand sides of (4.1) and (4.2) is at most \( \mu^2(C) \). So \( 2\mu^2(C) \leq \mu^2(C) \). Therefore \( \mu^2(C) = 0 \), whence the result. \( \square \)

So \( S_\mu \) functions as the ‘support’ of \( \mu \).

Definition 4.2. Hazard measure \( H \) on \((S, \mathcal{S})\):

\[
H(A) := \int_{A \cap S_\mu} \frac{1}{\mu((-\infty, x)^c)} \, d\mu(x).
\]

Definition 4.3. \( X_n \) is a record if \( X_k < X_n \) for \( k = 1, \ldots, n - 1 \).

Let \( R \) denote the set of records.

Theorem 4.4 ([Goldie & Resnick, 1989]). Let \( A \in \mathcal{S} \) and define events \( A_n := \{X_n \in A \cap R\} \). Then \( \sum_{n=1}^\infty P(A_n) = H(A) \). Further, \( P(\#(R \cap A) = \infty) = 1 \) or 0 according as \( \sum_{n=1}^\infty P(A_n) = H(A) = \infty \) or \( < \infty \).

The proof will use the two standard results following. For proofs, see e.g. [Chow & Teicher, 1997], Exercise 4.2.16 and Corollary 7.3.8 respectively.

Lemma 4.5 (Kochen-Stone extension of (Borel-)Cantelli). If for some \( c > 0 \),

\[
P(A_i \cap A_j) \leq cP(A_i)(P(A_{j-i}) + P(A_j))
\]
and \( \sum P(A_n) = \infty \), then \( P(\sum 1_{A_n} = \infty) > 0 \).

A map \( \pi := (\pi_1, \pi_2, \ldots) \) from \( \mathbb{N} \) onto itself is called a finite permutation if \( \pi \) is one-to-one and \( \pi_n = n \) for all but a finite number of integers. Considering \( X := (X_1, X_2, \ldots) \) as a random element of \( \mathbb{R}^\infty := \mathbb{R} \times \mathbb{R} \times \cdots \) with its class \( \mathcal{B}^\infty \) of Borel sets, define \( \pi X := (X_{\pi_1}, X_{\pi_2}, \ldots) \). An event \( \{X \in B\} \), where \( B \in \mathcal{B}^\infty \), is called a permutable event for \( X \) if \( P(\{X \in B\} \triangle \{\pi X \in B\}) = 0 \) for all finite permutations \( \pi \).
Theorem 4.6 (Hewitt-Savage 0–1 Law). The σ-algebra of permutable events for the i.i.d. sequence \( X \) is degenerate.

Proof of Theorem 4.4. Note that \( A_n = \{ X_n > X_k \text{ for } k = 1, \ldots, n-1, X_n \in A \} \). For \( m > n \),

\[
A_m \subset \{ X_m > X_j \text{ for } j = n + 1, \ldots, m - 1, X_m \in A \},
\]

and the right-hand side has the same probability as \( A_{m-n} \) and is independent of \( A_n \). So

\[
P(A_n \cap A_m) \leq P(A_n)P(A_{m-n}).
\]

By Borel’s Lemma (i.e. the easy half of the Borel-Cantelli Lemma) and the Kochen-Stone Lemma 4.5,

\[
P(\sum 1_{A_n} = \infty) = 0 \text{ iff } \sum P(A_n) < \infty.
\]

Now

\[
\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \int_{A} \mu^{n-1}(-\infty, x) \, d\mu(x)
\]

\[
= \sum_{n=1}^{\infty} \int_{A \cap S_{\mu}} \mu^{n-1}(-\infty, x) \, d\mu(x) \text{ by Prop. 4.1}
\]

\[
= \int_{A \cap S_{\mu}} \frac{d\mu(x)}{\mu((-\infty, x)^c)} = H(A).
\]

Thus \( P(\#(R \cap A) = \infty) = 0 \) if and only if \( H(A) < \infty \).

Finally the Hewitt-Savage Law (Theorem 4.6) gives \( P(\#(R \cap A) = \infty) = 0 \) or 1.

\( \square \)

4.2. Strict multivariate records. Continue with the setup and notation of the last section, but now specialise to \( \mathbb{R}^d \) with \( d > 1 \), define \( x < y \) component-wise, and take \( A := \mathbb{R}^d \).

Theorem 4.7 ([Gnedin, 1998]). If \( F \) is a non-singular Gaussian law on \( \mathbb{R}^d \), with correlation matrix \( \Lambda \), then there exist \( \alpha > 1 \) and \( \beta \in \{2, \ldots, d\} \), both depending on \( \Lambda \), so that

\[
P(A_n) \propto n^{-\alpha}(\ln n)^{(\alpha-\beta)/2}.
\]

Consequently \( P(\#R < \infty) = 1 \) for all non-singular Gaussian laws. The same holds for singular Gaussian laws unless all correlation coefficients are +1.

For \( d = 2 \) and correlation coefficient \( \rho \in (-1, 1) \), more precisely,

\[
P(A_n) \propto n^{-2/(1+\rho)}(\ln n)^{-\rho/(1+\rho)}.
\]

Proof. Omitted.

\( \square \)
4.3. Chain records. Let $X, X_1, X_2, \ldots$ be i.i.d. in $\mathbb{R}^d$. Define $x < y$ component-wise. Define a form of lower record as follows:

- $X_n$ is a chain record if it is below the previous chain record.

**Definition 4.8.** Set $T_1 := 1$, and for $k = 2, 3, \ldots$,

$$T_k := \min\{n > T_{k-1} : X_n < X_{T_{k-1}}\}.$$ 

The chain records are $R_k := X_{T_k}$.

Let $N_n := \sum_{j=1}^n 1\{X_j \text{ is a chain record}\}$.

**Theorem 4.9** ([Gnedin, 2007]). Suppose $X$ has a continuous product distribution. Let $W$ be the product of $d$ independent $\text{Unif}(0, 1)$ r.v.s, so that $m := E(-\ln W) = d$ and $\sigma^2 := \text{var}(-\ln W) = d$. Then $N_n \sim m^{-1} \ln n$ a.s. and

$$\frac{N_n - m^{-1} \ln n}{\sigma^2 m^{-3} \ln n} \Rightarrow N(0, 1).$$

**Note 1.** To calculate the mean and variance use the fact that $W$ has density $(-\ln w)^{d-1}/(d-1)!$ on $(0, 1)$.

**Note 2.** The $d = 1$ case is included! For $d = 1$, $W \sim \text{Unif}(0, 1)$ so $-\ln W \sim \text{Expon}(1)$, so $m = 1 = \sigma^2$. As in §2.2,

$$N_n \sim \ln n \text{ a.s., } \frac{N_n - \ln n}{\sqrt{\ln n}} \Rightarrow N(0, 1).$$

**Sketch proof of Theorem 4.9.** $X \sim F^{(d*)}$ where $F$ is a continuous d.f. on $\mathbb{R}$. As

$$X = (X^{(1)}, \ldots, X^{(d)}) < (Y^{(1)}, \ldots, Y^{(d)}) = Y$$

if and only if

$$(F(X^{(1)}), \ldots, F(X^{(d)})) < (F(Y^{(1)}), \ldots, F(Y^{(d)})) \text{ a.s.,}$$

without loss of generality we may assume $X \sim \text{Unif}(0, 1)^d$. We make that assumption from now on.

Define the lower section of $x = (x^{(1)}, \ldots, x^{(d)}) \in (0, 1)^d$ to be \{y \in (0, 1)^d : y < x\}, and define the height $h(x)$ of $x$ to be the area of the lower section, i.e. $h(x) := x^{(1)} \ldots x^{(d)}$.

Let $W := h(X_1) = h(R_1)$. Given $R_1, R_2$ has the law of $X$ conditional on it being in the lower section of $R_1$. So the conditional law of $h(R_2)$ given $R_1 = r$ is the law of $Wh(r)$. Thus let $W_1, W_2, \ldots$ be i.i.d., then we can consider $H(R_2)$ as being $W_1 W_2$.

Similarly for $h(R_3)$, and so on. Thus for general $n$, $H_n := h(R_n) = W_1 \cdots W_n$ where $W_1, W_2, \ldots$ are i.i.d.

Given $R_k, R_{k+1}$ is the next $X$ to fall in the lower section of $R_k$, so

$$P(T_{k+1} - T_k = n) = (1 - H_k)^n H_k \text{ for } n = 1, 2, \ldots.$$ 

So the sojourns $T_{k+1} - T_k$ are conditionally independent given the random sequence $(H_n)$, with the conditional laws just mentioned.
Set \( N_n := \max\{ k : T_k \leq n \} \). A.V. Gnedin shows that \( N_n \) is approximately \( K_n := \max\{ k : H_k > 1/n \} \). As \(-\ln H_n = \sum_{i=1}^{n} -\ln W_i\), \( K_n \) is the renewal function at \( \ln n \) of this, hence the results by renewal theory.

### 4.4. Longest increasing sequences.

In \( \mathbb{R}^2 \) use the notation \( x = (x_1, x_2) \), define \( x < y \) component-wise, take \( X_1, X_2, \ldots \) i.i.d. \( \sim F \), and as above define hazard measure \( H \) by

\[
H(dx) := \frac{F(dx)}{1 - F(x_\bot)} = \frac{P(X_1 \in dx)}{P(X_1 < x)^c}.
\]

Let \( A \) be an interval \([a, b]\) in \( \mathbb{R}^2 \). Considering strict records (in both coordinates simultaneously) we know from Theorem 4.4 that the number \( N_A \) of records falling in \( A \) is finite a.s. if and only if \( H(A) < \infty \). In this section we will find out about the r.v. \( N_A \), when it is finite.

Given points \( x_1 < \cdots < x_n \) in \( A \), join \( a < x_1 < \cdots < x_n < b \) by straight lines to form a path.

**Theorem 4.10** ([Goldie & Resnick, 1995]). Assume \( H(A) < \infty \), that \( H \) has a bounded density on \( A \) and that the distribution \( G \) on \( A \) given by

\[
G(x) := \frac{H(x)}{H(A)} \quad (x \in A)
\]

satisfies the conditions of either Theorem 4.12 or Theorem 4.20 below. Then, given \( N_A = n \), as \( n \to \infty \) the path joining the records converges in probability to a non-random limit curve which maximises the Deuschel-Zeitouni functional \( J(\phi) \) or the Goldie-Resnick functional \( T(f) \) (see below for both of these) respectively.

**Sketch proof.** The avoidance probability of a Borel set \( B \) in \( \mathbb{R}^2 \) is \( Q(B) := P(N_B = 0) \), and when \( B \) is an open box \((x, y)\) we write \( Q(x, y) := Q((x, y)) \) for short.

The records in \( A = [a, b] \) are \( R_1, \ldots, R_{N_A} \). As

\[
dH(x) = P(X \in du \mid X \not< x),
\]

we have

\[
P(N_A = n, R_1 \in dx_1, \ldots, R_n \in dx_n) = Q(a, x_1)dH(x_1)Q(x_1, x_2)dH(x_2)\cdots dH(x_n)Q(x_n, b). \quad (4.3)
\]

As \( F \) is continuous, \( Q(B) \sim 1 \) for small \( B \), so can we neglect \( Q \) in the above? Yes, because

\[
H(B) = EN_B = P(N_B \geq 1) = 1 - Q(B), \quad (4.4)
\]
so
\[
1 \geq \prod_{0}^{n} Q(x_i, x_{i+1}) \geq \prod_{0}^{n} (1 - H(x_i, x_{i+1})) \\
\geq 1 - \sum_{0}^{n} H(x_i, x_{i+1}) \\
= 1 - H \left( \bigcup_{0}^{n} (x_i, x_{i+1}) \right)
\]
and the latter converges to 1 if the area of \( \bigcup_{0}^{n} (x_i, x_{i+1}) \) tends to 0. Thus the left-hand side of (4.3) is approximated by \( dH(x_1) \cdots dH(x_n) \), and the result follows. \( \square \)

**Definition 4.11.** Let \( B^\uparrow \) be the set of non-decreasing right-continuous functions \( \phi : [a_1, a_2] \to [b_1, b_2] \). For \( \phi \in B^\uparrow \), \( \phi(x) = \int_{a}^{x} \dot{\phi}(t) \, dt + \phi_s(x) \) where \( \phi_s \) is singular. Assuming \( G \) has a density \( g \), define \( J : B^\uparrow \to \mathbb{R} \) by
\[
J(\phi) := \int_{a_1}^{a_2} \sqrt{\dot{\phi}(x) g(x, \phi(x))} \, dx.
\]

**Theorem 4.12 ([Deuschel & Zeitouni, 1995]).** Let \( Z_1, \ldots, Z_n \) be i.i.d. \( \sim G \) on an interval \( A \) in \( \mathbb{R}^2 \). Assume that \( G \) has a density \( g \) that is \( C^1_b \) and such that \( \ln g \) is bounded. Assume also that \( J(\phi) \) is maximised on a finite set \( \{\bar{\phi}_1, \ldots, \bar{\phi}_k\} \). (4.5)

On the event \( Z_1 < \cdots < Z_n \) let \( \phi_n \) denote the element of \( B^\uparrow \) formed by joining \( a < Z_1 < \cdots < Z_n < b \) by straight-line segments. Then for each \( \varepsilon > 0 \),
\[
P(\min\{\|\phi_n - \bar{\phi}_1\|_\infty, \ldots, \|\phi_n - \bar{\phi}_k\|_\infty\} > \varepsilon | Z_1 < \cdots < Z_n) \\
\to 0 \quad (n \to \infty).
\]

**Proof.** Omitted. \( \square \)

Set \( \bar{J} := \sup_{\phi \in B^\uparrow} J(\phi) \). When \( G \) is a product distribution the diagonal is the unique maximising curve \( \bar{\phi} \), and then obviously \( J = J(\bar{\phi}) = 1 \).

**Definition 4.13.** For \( y_1, \ldots, y_n \in \mathbb{R} \) an increasing subsequence is \( y_{i_1} < y_{i_2} < \cdots < y_{i_k} \) where \( i_1 < i_2 < \cdots < i_k \).

(That is, in selecting the \( y \) you can miss indices out: the \( y \) selected don’t have to be a run.)

**Theorem 4.14 ([Deuschel & Zeitouni, 1995]).** Let \( Z_1, \ldots, Z_n \) be i.i.d. \( \sim G \) in \([0, 1]^2\). Order them by their \( x \) components and let \( L_n \) be the length of the longest increasing subsequence (of the \( y \) components).
Assume that $G$ has a density $g$ that is $C^1_b$ and such that $\ln g$ is bounded. Then $L_n/\sqrt{n} \xrightarrow{P} 2J$.

**Proof.** Omitted. □

This builds on, and extends, the celebrated solution to Ulam’s problem:

**Theorem 4.15.** Let $l_n$ be the length of the longest increasing subsequence in a random permutation of order $n$. Then $l_n/\sqrt{n} \xrightarrow{P} 2$.

**Proof.** The original proofs are in [Vershik & Kerov, 1977] and [Logan & Shepp, 1977]; later ones are in [Seppäläinen, 1996] and [Aldous & Diaconis, 1995]. See [Aldous & Diaconis, 1999] for a survey. □

To avoid the rather unsatisfactory condition (4.5), different assumptions seem to be needed, and will lead to further conclusions. First, for probability measures $P$, $Q$ on $(\Omega, A)$ recall that $P$ is said to be absolutely continuous with respect to $Q$, notation $P \ll Q$, if for each event $A$, $Q(A) = 0$ implies $P(A) = 0$. The Radon-Nikodým Theorem (see, e.g., [Billingsley, 1995, Theorem 32.2]) gives, when $P \ll Q$, the existence of a non-negative r.v. $dP/dQ$ such that

$$P(A) = \int_A \frac{dP}{dQ} dQ \quad \forall A \in A.$$ 

Now for an important concept from information theory.

**Definition 4.16.** For probability measures $P$, $Q$ on $(\Omega, A)$ the I-divergence (Kullback-Leibler information number, relative entropy) is

$$D(P\|Q) := \begin{cases} \int_{\Omega} (\ln \frac{dP}{dQ}) dP & \text{if } P \ll Q, \\ \infty & \text{if not.} \end{cases}$$

For probability densities $p$, $q$ on $\mathbb{R}$ this reduces to

$$D(p\|q) := \begin{cases} \int_{-\infty}^{\infty} (\ln \frac{p(x)}{q(x)}) p(x) dx & \text{if } p(x) = 0 \text{ whenever } q(x) = 0, \\ \infty & \text{if not.} \end{cases}$$

When $P$ or $Q$ is the law of a random variable we will allow ourselves to use the r.v. in place of its law in the notation $D(\cdot\|\cdot)$.

**Proposition 4.17.** $D(P\|Q)$ is well-defined, has alternative formula

$$D(P\|Q) = \begin{cases} \int_{\Omega} \frac{dP}{dQ} (\ln \frac{dP}{dQ}) dQ & \text{if } P \ll Q, \\ \infty & \text{if not.} \end{cases}$$

and satisfies $0 \leq D(P\|Q) \leq \infty$.

**Proof.** For a careful proof see [Dupuis & Ellis, 1997, §1.4]. There are less careful—incomplete—proofs elsewhere. □
Aside: Statistics.

**Theorem 4.18** (Stein’s Lemma or the ‘Chernoff-Stein Lemma’). For testing

\[ H_0 : \text{the density is } p, \]

against

\[ H_1 : \text{the density is } q, \]

the most powerful level-\(\alpha\) test, based on a random sample of size \(n\), has Type II error probability

\[ \beta_n(\alpha) = e^{-D(p\|q)n(1+o(1))} \text{ as } n \to \infty. \]

**Proof.** See e.g. [Cover & Thomas, 2006]. \(\Box\)

Now back to the main thread (end of the Aside):

**Definition 4.19.** A parametrised curve \( f = (f_1, f_2) \) on \( A = [a, b] \subset \mathbb{R}^2 \) is an element of \( DL \), the space of non-decreasing functions \( f : [0, 1] \to \mathbb{R}^2 \) that are left-continuous on \((0, 1]\) and have \( f(0) = a \) and \( f(1) \leq b \). Assuming \( G \) has a density \( g \), define \( T : DL \to \mathbb{R} \) by

\[ T(f) := \int_0^1 \ln g(f(p)) \, dp - D(f_1(U)\|U) - D(f_2(V)\|V) \quad (4.6) \]

where \( U \sim \text{Unif}(a_1, a_2) \) and \( V \sim \text{Unif}(b_1, b_2) \).

**Theorem 4.20** ([Goldie & Resnick, 1995]). Assume that \( G \) has a density \( g \) that is continuous and has \( \ln g \) bounded on \( A \), and further is such that \( \ln g \) is L-superadditive on \( A^o \):

\[ \frac{\partial^2 \ln g(x, y)}{\partial x \partial y} \geq 0 \quad ((x, y) \in A^o) \quad (4.7) \]

\((g \text{ is ‘humped’}). \) Then there is a unique \( \bar{f} \in DL \) that maximises \( T(f) \).

Let \( Z_1, \ldots, Z_n \) be i.i.d. \( \sim G \) on \( A \). On the event \( Z_1 < \cdots < Z_n \) let \( f_n \) denote the element of \( DL \) formed by joining \( a < Z_1 < \cdots < Z_n < b \) by straight-line segments. Then

\[ \|f_n - \bar{f}\|_\infty \xrightarrow{P} 0 \quad (n \to \infty) \]

(\( \|f\|_\infty := \sup_{p \in [0, 1]} \|f(p)\| \) for \( f : [0, 1] \to \mathbb{R}^2 \), and \( \|\cdot\| \) is any norm on \( \mathbb{R}^2 \)).

The proof of this will be subsumed under that of Theorem 4.22 below.

**Fact 4.21.** \( T(\bar{f}) = 2 \ln \bar{J} \).
Theorem 4.22 ([Goldie & Resnick, 1995]). Under the conditions of Theorem 4.20,

\[ P(Z_1 < \cdots < Z_n) = e^{-n(2\ln n - 2 - \ln |A| - T(\bar{f}) + o(1))} \quad (n \to \infty) \]  

(4.8)

and

\[ P(Z_1, \ldots, Z_n \text{ can be ordered}) = e^{-n(\ln n - 1 - \ln |A| - T(\bar{f}) + o(1))} \quad (n \to \infty), \]

(4.9)

where \(|A| = \text{Leb}(A) = (b_1 - a_1)(b_2 - a_2)|.

Sketch proof of Theorems 4.20 and 4.22. Without loss of generality we may assume \(H(A) = 1\), which makes \(G\) and \(H\) identical. We cut the sides of \(A\) each into \(k \geq 2\) equal-sized pieces. A possible path \(z_1 < \cdots < z_n\) gives a count \(np_{ij}\) of observations in the box

\[ \left( a_1 + \frac{i - 1}{k} (b_1 - a_1), a_1 + \frac{i}{k} (b_1 - a_1) \right) \times \left( a_2 + \frac{j - 1}{k} (b_2 - a_2), a_2 + \frac{j}{k} (b_2 - a_2) \right), \]

and hence marginal frequencies \(np_{j}^{(1)}\) and \(np_{j}^{(2)}\) where

\[ p_{j}^{(1)} := \sum_{k=1}^{n} p_{jk}, \quad p_{j}^{(2)} := \sum_{k=1}^{n} p_{kj}. \]

Let \(\bar{h}_{ij}\) be the supremum of \(h\) over the above box. Then

\[ P(Z_1 < \cdots < Z_n \text{ give marginal frequencies } np_{j}^{(i)}) \]

\[ = \int_{\{z_1 < \cdots < z_n \text{ give marginal frequencies } np_{j}^{(i)}\}} h(z_1) \cdots h(z_n) \, dz_1 \cdots dz_n \]

\[ \leq \left( \prod_{i,j} (\bar{h}_{ij})^{np_{ij}} \right) \int_{\{\cdots\}} \, dz_1 \cdots dz_n \]

\[ = \exp\left( n \int_{0}^{1} \ln \bar{h}_k(e_k(p)) \, dp \right) \int_{\{\cdots\}} \, dz_1 \cdots dz_n, \]

where \(\bar{h}_k\) is \(h\) replaced by its upper bounds at mesh \(1/k\), and \(e_k\) is a piecewise-linear element of \(D_L\) that has the correct \(p_{j}^{(i)}\). The above then equals

\[ \exp\left( n \int_{0}^{1} \ln \bar{h}_k(e_k(p)) \, dp \right) \prod_{i=1}^{k} \prod_{j=1}^{k} \frac{(b_i - a_i)^{np_{j}^{(i)}}}{(np_{j}^{(i)})!} \]

\[ = \exp\left( n \int_{0}^{1} \ln \bar{h}_k(e_k(p)) \, dp + n \ln \frac{|A|}{k^2} - \sum_{i=1}^{k} \sum_{j=1}^{k} \ln(np_{j}^{(i)})! \right). \]
By Stirling’s formula, \( \ln(n p_j^{(i)}!) \approx n p_j^{(i)} \ln(n p_j^{(i)}) - n p_j^{(i)} \). Thus the above is at most
\[
\exp\left(n \int_0^1 \ln \tilde{h}_k(e_k(p)) \, dp \right)
+ n \ln|A| - 2n \ln n - n \sum_{i=1}^{2k} \sum_{j=1}^{k} p_j^{(i)} \ln p_j^{(i)} + 2n
\]
\[
= \exp(n \bar{T}_k(f) + n \ln|A| - 2n \ln n + 2n)
\]
where
\[
\bar{T}_k(f) := \int_0^1 \ln \tilde{h}_k(e_k(p)) \, dp - \sum_{i=1}^{2k} \sum_{j=1}^{k} p_j^{(i)} \ln(p_j^{(i)}(b_i - a_i)).
\]
A similar lower bound is obtained in the same way.

Then prove the lower bound for \( \tilde{f} \) exceeds (actually by orders of magnitude) the upper bound for other \( f \), hence we have concentration at \( \tilde{f} \) and obtain (4.8). The L-superadditivity property (4.7) is used in convexity arguments to prove uniqueness of the limit curve \( \tilde{f} \).

Finally, the left-hand side of (4.9) is \( n! \) times that of (4.8), so (4.9) follows after a further use of Stirling’s formula.

**Theorem 4.23 ([Goldie & Resnick, 1995]).** Let \( A = [a, b] \) be an interval in \( \mathbb{R}^2 \). Let \( X_1, X_2, \ldots \) be i.i.d. \( \sim F \) where \( F \) has continuous density \( f \). Set \( h(x) := f(x)/(1 - F(x)) \), \( H(A) := \int_A h(x) \, dx \), \( g(x) := h(x)/H(A) \) for \( x \in A \), and hence define \( T(\cdot) \) by (4.6). Suppose further that \( F(b) < 1 \), \( \ln f \) is bounded in \( A \), and that \( \ln h \) is L-superadditive (see (4.7)) on \( A^\circ \). Then
\[
P(N_A = n) = \left( \frac{|A| e^{T(\tilde{f})} + o(1)}{(n!)^2} \right)^n (n \to \infty) \quad (4.10)
\]
and
\[
P(N_A \geq n) = \left( \frac{|A| e^{T(\tilde{f})} + o(1)}{(n!)^2} \right)^n (n \to \infty). \quad (4.11)
\]

The proof needs one more lemma.

**Lemma 4.24.** Under the conditions of Theorem 4.23 there exists a constant \( C \) such that \( 0 \leq \ln Q(B) \geq -CH(B) \) for all Borel subsets \( B \) of \( A \).

**Proof.** By its definition (at the start of the proof of Theorem 4.10), \( Q \) is non-increasing: \( Q(B) \geq Q(C) \) if \( B \subseteq C \). So for any Borel set \( B \subseteq A \),
\[
Q(B) \geq Q(\{ x : x \leq b \}) = 1 - F(b) > 0,
\]
the equality being by [Goldie & Resnick, 1989, Corollary 2.5]. Put 
\[ C := - (F(b))^{-1} \ln(1 - F(b)) \geq 0, \]
then \[ \ln x \geq -C(1 - x) \] for \( 1 - F(b) \leq x \leq 1 \), so that from the above, \( 0 \geq \ln(Q(b)) \geq -C(1 - Q(b)) \), and the result follows via (4.4).

**Proof of Theorem 4.23.** Let \( g_n \) be the element of \( D_L \) formed by joining \( a < R_1 < \cdots < R_{N_A} < b \) by straight-line segments. For \( \Gamma \) a measurable subset of \( D_L \) we have
\[
P(g_n \in \Gamma, \ N_A = n) = \int \left( \prod_{j=0}^{n} Q(z_j, z_{j+1}) \right) dH(z_1) \cdots dH(z_n)
\]
where the integral is over the set of \( z_1 < \cdots < z_n \) such that the curve formed by joining \( a < z_1 < \cdots < z_n < b \) by straight-line segments lies in \( \Gamma \). If the integrand is replaced by 1 the right-hand side becomes \( P(f_n \in \Gamma, \ Z_1 < \cdots < Z_n) \). Now by Lemma 4.24,
\[
1 \geq \prod_{j=0}^{n} Q(z_j, z_{j+1}) \geq \exp \left( -C \sum_{j=0}^{n} H((z_j, z_{j+1})) \right)
= \exp \left( -CH \left( \bigcup_{j=0}^{n} (z_j, z_{j+1}) \right) \right)
> e^{-CH(A)},
\]
and so (4.10) follows from (4.8) (and again Stirling’s formula). Decay of the right-hand side of (4.10) is so rapid that the terms, \( t_n \) say, are asymptotic to their tail-sums \( \sum_{j=n}^{\infty} t_j \). Therefore (4.11) follows from (4.10).

**References**


Appendix: solutions to exercises

Solution to Exercise 1.1. Choose an event \( A \in A_L \) and real numbers \( x_1, \ldots, x_k \), then

\[
P(A \cap \{X_{L+1} \leq x_1, \ldots, X_{L+k} \leq x_k\})
\]

\[
= \sum_{l=1}^{\infty} P(\{L = l\} \cap A \cap \{X_{L+1} \leq x_1, \ldots, X_{L+k} \leq x_k\})
\]

(as \( L \) is assumed finite)

\[
= \sum_{l=1}^{\infty} P(\{L = l\} \cap A) F(x_1) \cdots F(x_k)
\]
(as \( \{L = l\} \cap A \in \sigma(X_1, \ldots, X_l) \), of which \( X_{l+1}, \ldots, X_{l+k} \) are independent, and i.i.d. \( \sim F \))
\[
= P(A)F(x_1) \cdots F(x_k),
\]
which gives that \( (X_{L+1}, X_{L+2}, \ldots) \) is i.i.d. \( \sim F \) and is independent of \( A_L \). \( \square \)

**Solution to Exercise 1.16.**
\[
P(Y = Z = 1) = P(Y = Z = 1|X = 0)P(X = 0)
+ P(Y = Z = 1|X = 1)P(X = 1)
= \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{2} = \frac{5}{16},
\]
while
\[
P(Y = 1) = P(Z = 1)
= P(Z = 1|X = 0)P(X = 0) + P(Z = 1|X = 1)P(X = 1)
= \frac{1}{4} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{2} = \frac{1}{2},
\]
so \( P(Y = Z = 1) \neq P(Y = 1)P(Z = 1) \) and thus \( Y \) and \( Z \) are dependent, i.e. \( \mathcal{H} \) and \( \mathcal{K} \) are dependent. However,
\[
P(Y = Z = 1|X) = \frac{1}{4} \cdot \frac{1}{2} \mathbf{1}_{X=0} + \frac{3}{4} \cdot \frac{3}{4} \mathbf{1}_{X=1}
= (\frac{1}{4} \mathbf{1}_{X=0} + \frac{3}{4} \mathbf{1}_{X=1})^2
= P(Y = 1|X)P(Z = 1|X),
\]
and similarly for \( P(Y = Z = 0|X) \), \( P(Y = 0, Z = 1|X) \) and \( P(Y = 1, Z = 0|X) \). Thus \( Y \) and \( Z \) are conditionally independent given \( X \), which is to say \( \mathcal{H} \) and \( \mathcal{K} \) are conditionally independent given \( \mathcal{G} \). \( \square \)

**Solution to Exercise 3.3.** Clearly \( X \) is of the same type as itself (take \( a := 1, b := 0 \)). If \( Y \overset{L}{=} aX + b \) with \( a > 0 \) then \( X \overset{L}{=} a^{-1}Y - ba^{-1} \) with \( a^{-1} > 0 \), so \( Y \) and \( X \) are of the same type if \( X \) and \( Y \) are. If \( Y \overset{L}{=} aX + b \) with \( a > 0 \) and \( Z \overset{L}{=} cY + d \) with \( c > 0 \) then \( Z \overset{L}{=} caX + cb + d \) with \( ca > 0 \), so \( X \) and \( Z \) are of the same type if \( X \) and \( Y \), and \( Y \) and \( Z \), are. \( \square \)

**Solution to Exercise 3.5.** (a) All laws \( N(\mu, \sigma^2) \) with \( \sigma > 0 \), i.e. the non-degenerate Gaussian laws on the line. (b) All non-degenerate uniform laws on the line. (c) Translated exponential laws, i.e. those with density \( f(y) := \lambda e^{-\lambda(y-c)} \mathbf{1}_{y > c} \) for some \( \lambda > 0 \) and real \( c \). (d) All discrete uniform laws on \( n \) equally spaced points. (e) All degenerate laws on the line. \( \square \)

**Solution to Exercise 3.12.** The derivative of \( \Phi \) is the density \( \phi(z) = (2\pi)^{-1/2}e^{-z^2/2} \) of the standard Gaussian law on \( \mathbb{R} \). The density of \( \Phi_\alpha \) is then
\[
\phi_\alpha(x) = \frac{\alpha}{x} \phi(\ln x^\alpha) \mathbf{1}_{x > 0} = \frac{\alpha}{x\sqrt{2\pi}} \exp \left( -\frac{(\ln x^\alpha)^2}{2} \right) \mathbf{1}_{x > 0}.
\]
In particular, \( \phi_1(x) = (x\sqrt{2\pi})^{-1}e^{-(\ln x)^2/2} \mathbf{1}_{x > 0} \).
The density of \( \bar{\psi}_\alpha \) is
\[
\bar{\psi}_\alpha(x) = \frac{\alpha}{-x} \phi(-x^\alpha) 1_{x<0} = \frac{\alpha}{-x\sqrt{2\pi}} \exp\left( -\frac{(\ln(-x^\alpha))^2}{2} \right) 1_{x<0}.
\]

\[\square\]

Solution to Exercise 3.18. Suppose first that \( A \) has Fisher-Tippett type, so for some \( a > 0 \) and \( b \), \( A(y) \equiv \Phi_a((y - b)/a) \). Then
\[
P(M_n \leq y) = \Phi^n_a\left(\frac{y - b}{a}\right) = \exp\left(-n\left(\frac{y - b}{a}\right)^{-\alpha}\right) 1_{y>b}
= \exp\left(-\left(\frac{y - b}{n^{1/\alpha}a}\right)^{-\alpha}\right) 1_{y>b}
= \Phi_a\left(\frac{y - b}{n^{1/\alpha}a}\right),
\]
which is again of Fisher-Tippett type. Similarly, if \( A \) has Fréchet type: \( A(y) \equiv \Psi((y - b)/a) \), then
\[
P(M_n \leq y) = \Psi^n_a\left(\frac{y - b}{a}\right) = \exp\left(-n\left(\frac{y - b}{a}\right)^{-\alpha}\right) 1_{y<b}
= \exp\left(-\left(\frac{y - b}{n^{-1/\alpha}a}\right)^{-\alpha}\right) 1_{y<b}
= \Psi_a\left(\frac{y - b}{n^{-1/\alpha}a}\right),
\]
which is again of Fréchet type. Lastly, if \( A \) has Gumbel type: \( A(y) \equiv \Lambda((y - b)/a) \), then
\[
P(M_n \leq y) = \Lambda^n_a((y - b)/a) = \exp\left(-ne^{-(y-b)/a}\right)
= \exp\left(-e^{-(y-b-a \ln n)/a}\right)
= \Lambda((y - b - a \ln n)/a),
\]
again of Gumbel type.

\[\square\]

Solution to Exercise 3.19. When \( G = \Phi_a \) you have for \( x > 0 \) that
\[-\ln G(x) = x^{-\alpha}, \quad \text{so } -\ln(-\ln G(x)) = -\ln x^{-\alpha} = \alpha \ln x, \quad \text{and} \]
\[
\Phi(-\ln(-\ln G(x))) = \Phi(\alpha \ln x) = \Phi_{\alpha}(x).
\]
Similarly when \( G = \Psi_a \) you have for \( x < 0 \) that \( -\ln G(x) = (-x)^\alpha \), so
\[-\ln(-\ln G(x)) = -\ln((-x)^\alpha) = -\alpha \ln(-x), \quad \text{and} \]
\[
\Phi(-\ln(-\ln G(x))) = \Phi(\alpha \ln(-x)) = \Phi_{\alpha}(x).
\]
Finally when \( G = \Lambda \) you have for all \( x \) that \( -\ln(-\ln G(x)) = x \), so
\[
\Phi(-\ln(-\ln G(x))) = \Phi(x).
\]

\[\square\]