1 Preliminaries

1.1 Uniform Convergence

We will develop the idea of something called “continuous convergence” which will be useful to us later on.

Definition 1. Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces and suppose we have a sequence of functions $f_n : \mathcal{X} \to \mathcal{Y}$. We say $f_n$ converges continuously to $f$ if whenever $x_n \to x_0$ in $\mathcal{X}$ we have $f_n(x_n) \to f_0(x_0)$ in $\mathcal{Y}$.

Lemma 2. Suppose $\mathcal{X}$ is compact, and $f_0$ is continuous. Then $f_n \to f_0$ continuously if and only if $f_n \to f_0$ uniformly on $\mathcal{X}$.

Proof. $(\Rightarrow)$ Suppose $f_n \to f_0$ uniformly on $\mathcal{X}$. Take any sequence $x_n \to x_0$ in $\mathcal{X}$. Use $d$ to denote the metric on $\mathcal{Y}$. Have:

$$d(f_n(x_n), f_0(x_0)) \leq d(f_n(x_n), f_0(x_n)) + d(f_0(x_n), f_0(x_0)) \leq \sup_{z \in \mathcal{X}} d(f_n(z), f_0(z)) + d(f_0(x_n), f_0(x_0))$$

The first term goes to 0 since $f_n \to f_0$ uniformly, while the second term goes to zero since $f_0$ is continuous and $x_n \to x_0$.

$(\Rightarrow)$ Suppose by contradiction that $f_n$ does not converge to $f_0$ uniformly. Then there is a subsequence $n_k$ and $\epsilon > 0$ so that:

$$\sup_{z \in \mathcal{X}} d(f_{n_k}(z), f_0(z)) > \epsilon$$

Hence, by definition of sup, for each $n_k$ we can find a $x_k \in \mathcal{X}$ so that $d(f_{n_k}(x_k), f_0(x_k)) > \epsilon/2$. Since the space $\mathcal{X}$ is compact now, there is a limit point $x_0$ for the sequence $x_k$ and a subsequence so that $x_k \to x_0$. But this is a contradiction in the continuous convergence hypothesis, as $\epsilon/2 < d(f_{n_k}(x_k), f_0(x_k)) \leq d(f_{n_k}(x_k), f_0(x_0)) + d(f_0(x_k), f_0(x_0))$ so the latter is bounded away from zero.

Corollary 3. Suppose $U_n : \mathbb{R} \to \mathbb{R}$ are non-decreasing real valued functions on $\mathbb{R}$. Suppose also that $U_0$ is continuous. If $U_n \to U_0$ pointwise as $n \to \infty$ then $U_n \to U_0$ locally uniformly, i.e. for any $a < b$ we have:

$$\sup_{x \in [a,b]} |U_n(x) - U_0(x)| \to 0$$

Proof. We will check that $U_n \to U_0$ continuously on any $[a, b]$. (Rmk: the reason we need to restrict our attention to $[a, b]$ is that the lemma only works on compact spaces $\mathcal{X}$.) Suppose that $x_n \to x_0$. We want to show that $U_n(x_n) \to U_0(x_0)$. It suffices to consider two cases: a) $x_n > x_0$ and b) $x_n < x_0$ (If necessary we can always partition $x_n$ into two subsequences that have this property. Our argument will hold for both subsequences). We will look at the case a), and the case b) is analogous. Given any $\epsilon > 0$, use the continuity of $U_0$ to find $\eta > 0$ such that:

$$|U_0(x_0 + \eta) - U_0(x_0)| < \epsilon$$

Take $n_0$ so large now so that for each $n \geq n_0$ we have $|x_n - x_0| < \eta$ (ok since $x_n \to x$) and $|U_n(x_0 + \eta) - U_0(x_0 + \eta)| < \epsilon$ and $|U_n(x_0) - U_0(x_0)| < \epsilon$. (Ok since $U_n \to U_0$ pointwise) Then we have on one hand that:

$$U_n(x_n) \leq U_n(x_0 + \eta) \text{ since } U \text{ is monotone and } x_n < x_0 + \eta \leq U_0(x_0 + \eta) + \epsilon \leq U_0(x_0) + \epsilon + \epsilon$$

and on the other hand:

$$U_n(x_n) \geq U_n(x_0) \text{ since } x_n > x_0 \text{ and } U \text{ is monotone} \geq U_0(x_0) - \epsilon$$
**Remark 4.** Say $X_n$ are random variables converging $X_n \Rightarrow X_0$ in the weak sense, and suppose $X_0$ has no atoms. Then we know that the distribution functions $F_n$ converge pointwise to $F_0$ pointwise and that $F_0$ is continuous. The corrolary tells us that $F_n \to F_0$ uniformly on compact subsets. We can actually strengthen this a bit to uniform convergence *everywhere* because of the additional fact that $F_n \to 0/1$ as $x \to -\infty/\infty$.

**Corollary 5.** Suppose $X_n$ are random variables and $X_n \Rightarrow X_0$. Suppose further that $X_0$ has no atoms. Then the distribution functions converge uniformly on all of $\mathbb{R}$:

$$
\sup_{x \in \mathbb{R}} |P(X_n \leq x) - P(X_0 \leq x)| \to 0
$$

**Proof.** Fix any $\epsilon > 0$. Choose $[a, b]$ so large now so that $P \left( X_0 \in [a, b] \right) > 1 - \epsilon$. By the corollary/remark we know that $F_n \to F$ uniformly in $[a, b]$ so we know that for $n$ large enough $\sup_{x \in [a, b]} |F_n(x) - F_0(x)| < \epsilon$. Outside of $[a, b]$, consider as follows. By pointwise convergence at $b$, we know that for $n$ large enough $|F_n(b) - F_0(b)| < \epsilon$. Have then for any $x > \epsilon$ and such $n$ sufficiently large:

$$
|F_n(x) - F_n(b)| \leq 1 - F_n(b) \\
\leq 1 - F_0(b) + |F_0(b) - F_n(x)| \\
\leq 2\epsilon
$$

So then:

$$
\sup_{x > b} |F_n(x) - F_n(b)| \leq \sup_{x > b} \{|F_n(x) - F_n(b)| + |F_n(b) - F_0(b)| + |F_0(b) - F_0(x)|\}
\leq 2\epsilon + \epsilon + \epsilon
$$

The same idea works for $x < a$. Combining the convergence in the 3 pieces of $\mathbb{R} = (-\infty, a) \cap [a, b] \cap (b, \infty)$ we get $\sup_{x \in \mathbb{R}} |F_n(x) - F_n(b)| < 5\epsilon$. Since $\epsilon$ arbitrary, we have uniform convergence. 

\[\square\]

### 1.2 Inverses of Monotone Functions

**Definition 6.** If $H$ is a non-decreasing function on $\mathbb{R}$, we define the left-continuos-inverse of $H$ as:

$$
H^{-}(y) = \inf \{ s : H(s) \geq y \}
$$

We take the convention that the infimum over an empty set is $+\infty$.

**Proposition 7.** i) $H^{-}$ is non-decreasing

ii) $H^{-}(y) < z \implies H(z) \geq y$

iii) $H^{-}$ is left continuous at points $y \in \mathbb{R}$

**Proof.** i) For $y_1 < y_2$ we have $\{ s : H(s) \geq y_2 \} \subset \{ s : H(s) \geq y_1 \}$ since $H$ is non-decreasing. Taking inf’s gives:

$$
H^{-}(y_1) = \inf \{ s : H(s) \geq y_1 \} \leq \inf \{ s : H(s) \geq y_2 \} = H^{-}(y_2)
$$

ii) $H^{-}(y) \leq z \implies \inf \{ s : H(s) \geq y \} < z \implies H(z) \geq y$ since $H$ is non-decreasing

ii) Suppose by contradiction that $x_n \uparrow x$ but $H^{-}(x_n) \uparrow H^{-}(x) < H^{-}(x)$. Then there exists $\delta > 0$ so that:

$$
H^{-}(x_n) < y < H^{-}(x) - \delta
$$

By the definition of $H^{-}$, the left inequality tells us $H(y) \geq x_n$ for every $n$. Taking $n \to \infty$ then gives $H(y) \geq x$. By definition of $H^{-}$ again now, $H(y) \geq x$ gives $H^{-}(x) \leq y$. But this is impossible since $H^{-}(x) - \delta > y$! \[\square\]

**Proposition 8.** If $H$ is right continuous (i.e. $x_n \downarrow x_0 \implies H(x_n) \downarrow H(x_0)$) then:

$$
A(y) := \{ s : H(s) \geq y \} \text{ is closed}
$$

$$
H^{-}(y) \leq t \iff y \leq H(t)
$$

$$
t < H^{-}(y) \iff y > H(t)
$$
Proof. i) Since $H$ is non-decreasing, the set $A(y)$ is unbounded on the right (i.e. $s \in A(y) \implies [s, \infty) \subseteq A(y)$) so we only have to check that the left endpoint is a closed endpoint. Indeed, if $s_n \in A(y)$ and $s_n \downarrow s$ then $y \leq H(s_n)$ for each $s_n$ by def’n of $A(y)$. Since $H$ is right continuous, we have $H(s_n) \downarrow H(s)$ and so we conclude $y \leq H(s)$, which shows $s \in A(y)$

ii) Since $A(y)$ is closed, $H^-(y) = \inf A(y)$ is achieved at some point $s^* \in A(y)$. Hence $H(H^-(y)) = H(s^*) \geq y$ since $s^* \in A(y)$.

iii) As in ii), let $s^* = H^+(y) = \inf A(y)$ be the points in the closed set $A(y)$ that achieves the minimum. We use the fact that the entire interval $[s^*, \infty) \subseteq A(y)$ because $H$ is non-decreasing. Consider now:

$$H^-(y) \leq t \iff \inf A(y) \leq t$$
$$\iff t \in A(y) \text{ since } [s^*, \infty) \subseteq A(y)$$
$$\iff H(t) \geq y$$

The strict side is:

$$H^-(y) > t \iff \inf A(y) > t$$
$$\iff t \notin A(y) \text{ since } [s^*, \infty) \subseteq A(y)$$
$$\iff H(t) < y$$

$\square$

Proposition 9. Let $(\Omega, F, P) = ([0, 1], B([0, 1], m)$ be the Lebesgue measure on $[0, 1]$ with the usual Borel sigma algebra. Let $U$ be the identity function on $\Omega$, i.e. $U \sim U\text{ifd}([0, 1])$ is a uniformly distributed random variable. If $F$ is any distribution function on $\mathbb{R}$ with law $df$, then $F^+(-U)$ is a random variable on $[0, 1]$ with law $df$.

Proof. Such $F$ are always right continuous, so by the last proposition, we have:

$$m[F^+(U) \leq u] = m[U \leq F(t)]$$
$$= F(t)$$

$\square$

Definition 10. For any function $H$ denote $\mathcal{C}(H) = \{x \in \mathbb{R} : H \text{ is finite and continuous at } x\}$ (the set of $H$-continuity points).

We define weak convergence of non-decreasing functions by $H_n(x) \to H_0(x)$ for all $x \in \mathcal{C}(H)$.

Proposition 11. If $H_n$ are non-decreasing functions and $H_n \to H_0$ then $H_n^+ \to H_0^+$.

Proof. These proof is along the same lines of the last few...just play with the definition of $^+$ until you get it...I’m going to skip it.

$\square$

Theorem 12. (Skorohod’s Representation Theorem) Suppose $X_n$ on $(\Omega_n, B_n, P_n)$ is a sequence of random variables on a sequence of probability spaces so that $X_n \Rightarrow X_0$. Then there exists random variables $\tilde{X}_n$ on the Lebesgue probability space $(\Omega, B, P)$ (the sequence) so that:

$$\tilde{X}_n \overset{d}{=} X_n \text{ for each } n \in \mathbb{N}$$
$$\tilde{X}_n \Rightarrow \tilde{X}_0 \text{ almost surely with respect to } m$$

Remark 13. The $\tilde{X}_n$’s dont respect possibly dependencies between the $X_n$’s. For example $E(X_1X_2)$ is not the same as $E(\tilde{X}_1\tilde{X}_2)$; the joint distributions are not respected.

Proof. Let $U$ be the identity function on $[0, 1]$ so that $U$ is uniformly distributed. Suppose that the distribution function of $X_n$ is $F_n$. Define:

$$\tilde{X}_n = F_n^+(U)$$

As we have already seen, $\tilde{X}_n \overset{d}{=} X_n$ with this definition. To see the convergence is almost sure now, consider as follows. Firstly, $X_n \Rightarrow X_0$ means that $F_n \to F_0$ (in the sense of weak convergence). By the last proposition, we know that $F_n^+ \to F_0^+$ too then! Therefore:

$$1 \geq m \{0 \leq u \leq 1 : \tilde{X}_n(u) \to \tilde{X}_0(u)\}$$
$$= m \{u : F_n^+(u) \to F_0^+(u)\}$$
$$\geq m(\mathcal{C}(F_0^+))$$
$$= 1$$

The last equality follows by the famous result that a non-decreasing function can only have countably many discontinuities (so in particular the set of discontinuous points, $\mathcal{C}(F_0^+)$, is a null set) $\square$
1.3 Convergence to Types Theorem and Limit Distribution of Maxima

Definition 14. We say two distribution functions $U(x)$ and $V(x)$ are of the same **type** if for some $A > 0$, $B \in \mathbb{R}$ we have:

$$V(x) = U(Ax + B)$$

Example 15. The family of Gaussian distributions scales like this, so they are the same “type”.

Proposition 16. Suppose $U(x)$ and $V(x)$ are two distributions, neither of which concentrates at a point. (i.e. they are not the heaviside step function)

a) Suppose $F_n$ is a sequence of distribution functions, $a_n \geq 0$ and $b_n \in \mathbb{R}$, $\alpha_n > 0$ and $\beta_n \in \mathbb{R}$ are sequences so that:

$$F_n(a_n x + b_n) \to U(x) \text{ and } F_n(\alpha_n x + \beta_n) \to V(x)$$

(where convergence is the weak convergence of distribution functions). Then:

$$\frac{\alpha_n}{a_n} \to A > 0 \text{ and } \frac{\beta_n - b_n}{\alpha_n} \to B \in \mathbb{R}$$

and moreover:

$$V(x) = U(Ax + B)$$

An equivalent way to state this in terms of random variables is as follows:

a') Let $X_n$ be a sequence of random variables and suppose $U, V$ are random variables (neither of which is almost surely constant). If:

$$(X_n - b_n)/a_n \Rightarrow U \text{ and } (X_n - \beta_n)/\alpha_n \Rightarrow V$$

Then:

$$\frac{\alpha_n}{a_n} \to A > 0 \text{ and } \frac{\beta_n - b_n}{\alpha_n} \to B \in \mathbb{R}$$

and:

$$V \overset{d}{=} (U - B)/A$$

Remark 17. This proposition is saying that if you start with some random variables $X_n$, and then you center and scale somehow to get convergence to SOME random variable, say $(X_n - b_n)/a_n \Rightarrow U$, then the ONLY random variable you could hope to converge to are things of the same TYPE as $U$. Indeed, once you have one convergence $(X_n - b_n)/a_n \Rightarrow U$ it’s obvious how changing $a_n$ by a constant factor (or an $o(1)$ factor) and changing $b_n$ by adding a constant (or an $o(1)$ addition) will shift the distribution of $U$ around. The theorem says this is the ONLY thing you can do. Its really a nice uniqueness result.

If you knew before hand that $U$ and $V$ were of the same type, you could play around with the sequences to extract the constant $A$ and $B$. This is how the proof goes. We play around with the convergence until we find $A$ and $B$. Once this is done, we use Skohord to see the convergence in distribution as almost sure convergence in a certain light, and finally conclude that $V \overset{d}{=} (U - B)/A$.

Proof. By using the fact that $H_n \to H_0 \implies H_n^- \to H_0^-$, and other properties we have seen of $\to$, we can take as our hypothesis that:

$$(F_n^{-}(y) - b_n)/a_n \to U^{-}(y) \text{ and } (F_n^{+}(y) - \beta_n)/\alpha_n \to V^{-}(y)$$

weakly. Since neither $U(x)$ nor $V(x)$ are point masses at a single point, we can find $0 < y_1 < y_2 < 1$ so that $y_1, y_2$ are continuity points for both $U^{-}$ and $V^{-}$ so that:

$$U^{-}(y_1) < U^{-}(y_2) \text{ and } V^{-}(y_1) < V^{-}(y_2)$$

(Roughly speaking: this just says that the values $y_1$ and $y_2$ are not both in the same atom of $U$ or $V$). Pluggin in $y = y_1$ and $y = y_2$ into our hypothesis, we have (since these are continuity points) four statements of convergence, one to each of $U^{-}(y_1), U^{-}(y_2), V^{-}(y_1), V^{-}(y_2)$. By subtracting each pair, we get:

$$(F_n^{+}(y_2) - F_n^{+}(y_1))/a_n \to U^{+}(y_2) - U^{+}(y_1) > 0$$

$$(F_n^{+}(y_2) - F_n^{+}(y_1))/\alpha_n \to V^{+}(y_2) - V^{+}(y_1) > 0$$

Dividing the first relation by the second now, we get:

$$\frac{\alpha_n}{a_n} \to \frac{U^{+}(y_2) - U^{+}(y_1)}{V^{+}(y_2) - V^{+}(y_1)} =: A > 0$$

Now, replacing $\alpha_n = Aa_n + o(1)$ we compare our convergences at $y = y_1$:

$$(F_n^{-}(y_1) - b_n)/a_n \to U^{-}(y_1) \text{ and } (F_n^{-}(y_1) - \beta_n)/\alpha_n \to \frac{V^{-}(y_1)}{A}$$

So subtracting these two limits gives:

$$\frac{\beta_n - b_n}{\alpha_n} \to \frac{U^{-}(y_1) - V^{-}(y_1)}{A} =: B$$

So we indeed have desired connection between the $a’s$ and $b’s$ and the $\alpha’s$ and $\beta’s$. The fact that $V(x) = U(Ax + B)$ follows by Skohord’s theorem now. We put this in the next proposition. 

\qed
Proposition 18. If \( \alpha_n/a_n \to A > 0 \) and \( (\beta_n - b_n)/a_n \to B \in \mathbb{R} \), then either of the two conditions: \( (X_n - b_n)/a_n \Rightarrow U \) or 
\( (X_n - \beta_n)/\alpha_n \Rightarrow V \) will imply the other, and the conclusion of the last proposition will hold.

Proof. The main idea is to use Skorohod’s representation theorem to replace the convergence with almost sure convergence, at which point the statement is obvious.

Let \( Y_n = (X_n - b_n)/a_n \) and suppose \( Y_n \Rightarrow U \). By Skorohod’s Representation Theorem we can find \( \tilde{Y}_n, \tilde{U} \) on the probability space \( ([0,1],\mathcal{B}[0,1],m) \) such that \( \tilde{Y}_n \overset{d}{=} Y_n \) and \( \tilde{Y}_n \overset{a.s.}{\to} \tilde{U} \) where the convergence is now almost sure. Define \( \tilde{X}_n := a_n \tilde{Y}_n + b_n \) so that \( \tilde{X}_n \overset{d}{=} X_n \). Then:

\[
\frac{(X_n - \beta_n)}{\alpha_n} \overset{d}{=} \left( \frac{\alpha_n}{\alpha_n} \right) \frac{\tilde{X}_n - \beta_n}{\alpha_n} \overset{a.s.}{\to} \frac{1}{A} \tilde{U} - \frac{1}{A} B \overset{d}{=} \frac{(U - B)}{A}
\]

So we see that \( \frac{(X_n - \beta_n)}{\alpha_n} \Rightarrow \frac{(U - B)}{A} \).

\[\square\]

Remark 19. A nice byproduct of this proof is that if we are given the distribution functions \( F_n^{\pm} \) then the natural constants to try are:

\[
a_n = F_n^{+}(y_2) - F_n^{+}(y_1)
\]
\[
b_n = F_n^{-}(y_1)
\]

(Of course there are many different values of \( y_1 \) and \( y_2 \) one could choose...so the choice of constants is still a bit free. The proposition shows that actually all the choices lead to the same type so all choices are equally valid in some sense)

We will now focus our attention on the theory of the maximum of n i.i.d. random variables.

Proposition 20. Let \( X_1, \ldots \) be iid random variables. Let \( M_n := \max_{1 \leq i \leq n} X_i \). Notice \( M_n \) is non-decreasing as a function of \( n \).

a) The distribution function for \( M_n \) is \( F_n^{\alpha}(x) \) where \( F(x) \) is the distribution function for \( X_i \).

b) Let \( x_0 = \sup \{ x : F(x) < 1 \} \leq \infty \). Then:

\[
\lim_{n \to \infty} M_n = x_0 \ a.s.
\]

Proof. a) This is the standard calculation:

\[
P(M_n \leq x) = P \left( \bigcap_{i=1}^{n} \{ X_i \leq x \} \right) = \prod_{i=1}^{n} P(X_i \leq x) = F^n(x)
\]

b) For any \( x < x_0 \) we have \( F(x) < 1 \). Hence \( \lim_{n \to \infty} F_n^{\alpha}(x) = 0 \). Hence, since the \( M_n \) are non-decreasing, we have:

\[
P \left( \lim_{n \to \infty} M_n < x \right) = P \left( \bigcap_{n=1}^{\infty} \{ M_n < x \} \right) = \lim_{n \to \infty} P(M_n < x) = 0
\]

Hence for all \( x < x_0 \) we know that a.s \( M_n > x \) eventually. On the other hand, by the definition of \( x_0 \) we know that \( M_n \leq x_0 \) always holds. Hence we have that \( M_n \to x_0 \) in probability. Finally, since \( M_n \) are non-decreasing convergence in probability implies convergence almost surely (since the “bad” sets: \( \{ M_n - x_0 \geq \epsilon \} \) \( \{ M_n < x_0 - \epsilon \} \) and a decreasing family of sets as \( \epsilon \to 0 \) because \( M_n \) is non-increasing. Convergence in probability is saying the measure of the sets goes to zero as \( n \to \infty \) while almost sure convergence is saying that their lim-sup (i.e. their intersection) is a null set. Since these sets are decreasing, this is the same thing! \[\square\]

Theorem 21. (The THREE TYPES THEOREM). Let \( X_1, \ldots \) be iid random variables and \( M_n := \max_{1 \leq i \leq n} X_i \). Suppose there exist sequences \( a_n > 0 , b_n \in \mathbb{R} \) so that:

\[
P \left( (M_n - b_n)/a_n \leq x \right) = F^n(a_n x + b_n) \to G(x)
\]

Where the convergence is weakly. Assume that \( G \) is non-degenerate. Then \( G \) is of the type of one of the following three classes:

i) \( \Phi_\alpha(x) = \begin{cases} 0 & x < 0 \\ \exp(-x^\alpha) & x \geq 0 \end{cases} \) for some \( \alpha > 0 \)

ii) \( \Psi_\alpha(x) = \begin{cases} \exp(-|x|^\alpha) & x < 0 \\ 1 & x \geq 0 \end{cases} \) for some \( \alpha > 0 \)

iii) \( \Lambda_\alpha(x) = \exp(-\exp(-x)) \) for \( x \in \mathbb{R} \)
Proof. For \( t \in \mathbb{R} \) define the floor:

\[
|t| = \text{greatest integer less than or equal to } t
\]

For fixed \( t \), the sequence \( |nt| \) is a subsequence of the naturals going to infinity. Hence by the convergence in the hypothesis we have for any \( t \) that:

\[
F^{\lfloor nt \rfloor} (a_{\lfloor nt \rfloor} x + b_{\lfloor nt \rfloor}) \to G(x)
\]

On the other hand, since \( |nt| / n \to t \), we also have:

\[
F^{\lfloor nt \rfloor} (a_n x + b_n) = (F^n (a_n x + b_n))^{\lfloor nt \rfloor} \to G^t(x)
\]

If we think of \( \alpha_n = a_{\lfloor nt \rfloor} \) and \( \beta_n = b_{\lfloor nt \rfloor} \) we see that we have two different limits for the centered/scaled distribution function \( F^{\lfloor nt \rfloor} \), namely: \( F^{\lfloor nt \rfloor} (\alpha_n x + \beta_n) \to G(x) \) and \( F^{\lfloor nt \rfloor} (a_n x + b_n) \to G^t(x) \). By the convergence of types proposition, we know that it must be the case that \( G(x) \) and \( G^t(x) \) are the same type and that there are constants \( A, B \) that relate the \( \alpha \)'s and \( \beta \)'s. We can apply this argument for any \( t \), so we will actually get functions \( A(t) \) and \( B(t) \) so that: (I'm going to actually call them \( \alpha \) and \( \beta \) to match the notation in the book...don't get mixed up with the \( \alpha_n \) and \( \beta_n \) we just had...we are done with those now!)

\[
\lim_{n \to \infty} \frac{a_n}{a_{\lfloor nt \rfloor}} = \alpha(t)
\]

\[
\lim_{n \to \infty} \frac{b_n - b_{\lfloor nt \rfloor}}{a_{\lfloor nt \rfloor}} = \beta(t)
\]

\[
G^t(x) = G(\alpha(t) x + \beta(t))
\]

The functions \( \alpha, \beta \) are measurable with respect to \( t \). This follows because the function \( t \to \frac{a_n}{a_{\lfloor nt \rfloor}} \) is measurable for every \( n \) (its piecewise constant) and limits of measurable functions are still measurable! (Same idea for \( \beta \)). We can now apply a sneaky trick: for any \( t > 0 \) and \( s > 0 \) we will evaluate \( G^{st}(x) \) for \( s, t \in \mathbb{R}^+ \) and we will write it in two ways, namely as \( st \) once and \( t \cdot s \) once. Have:

\[
G^{st}(x) = G(\alpha(ts)x + \beta(ts))
\]

\[
G^{st}(x) = (G^s(x))^t
\]

\[
= (G(\alpha(s)x + \beta(s))^t
\]

\[
= G(\alpha(t)[\alpha(s)x + \beta(s)] + \beta(t))
\]

\[
= G(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t))
\]

Since \( G \) is non-degenerate, by a technical lemma which follows, the arguments must be equal and it must be that \( \alpha(ts)x + \beta(ts) = \alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t) \) for every \( x \). Hence:

\[
\alpha(ts) = \alpha(t)\alpha(s)
\]

\[
\beta(ts) = \alpha(t)\beta(s) + \beta(t)
\]

\[
= \alpha(s)\beta(t) + \beta(s)
\]

(The last line follows by the symmetry \( s \leftrightarrow t \)). The equation \( \alpha(ts) = \alpha(t)\alpha(s) \) has a one parameter family of solutions:

\[
\alpha(t) = t^{-\theta}
\]

for some constant \( \theta \in \mathbb{R} \). (After doing the change of variable \( \varphi(x) = \log(\alpha(e^x)) \), this equation says that \( \varphi(x+y) = \varphi(x) + \varphi(y) \), which is the famous Hamel equation. The pathological cases are excluded because \( \alpha \) is measurable.) There are now three cases that give rise to the three types, either \( \theta < 0 \) or \( \theta = 0 \) or \( \theta > 0 \). Each case is handled individually:

Case 1 (\( \theta = 0 \); this gives rise to the type \( \Lambda(x) = \exp(-e^{-x}) \)) In this case \( \alpha(t) \equiv 1 \) is constant, so the equation for \( \beta \) is:

\[
\beta(ts) = \beta(t) + \beta(s)
\]

After changing variables again, this is again the Hamel equation, so we have:

\[
\beta(t) = -c \log(t) \text{ for some } c \in \mathbb{R}
\]

This yields the following equatation for \( G^t \):

\[
G^t(x) = G(x - c \log t)
\]

We will now argue \( c > 0 \). If \( c = 0 \), we have \( G^t(x) = G(x) \) for every \( t \), which only happens if \( G \) is a degenerate distribution (a point mass at some point...so it takes only values in \( \{0,1\} \)). \( c \) cannot be negative, as \( G^t(x) \) must be a non-increasing function of \( t \) (as \( G(x) \leq 1 \) always), and \( G(x - c \log t) \) is a decreasing function when \( c < 0 \). So it must be the case that \( c > 0 \).
We will now argue that \( G(x) < 1 \) for every \( x \in \mathbb{R} \). Otherwise, if \( G(x_0) = 1 \) for some \( x_0 \) plugging in this value would yield \( 1 = G(x_0 - c \log t) \) for every \( t \). But then \( G(x) = 1 \) for every \( x \) since the range of \( x_0 - c \log t \) is all of \( \mathbb{R} \). Contradiction! So it must be that \( G(x) < 1 \) for all \( x \in \mathbb{R} \). A similar argument shows that \( G(x) > 0 \) for all \( x \).

Define \( p := -\log(-\log(G(0))) \) so that \( G(0) = \exp(-e^p) \). Let \( u = -c \log(t) \) (the range of \( u \) is \((-\infty, \infty)\) as \( t \) ranges in \((0, \infty)\)). Changing variables in the expression \( G'(0) = G(-c \log t) \) gives:

\[
G(u) = \exp(-e^p t) = \exp(-e^{-(c-u)+p}) = \Lambda(c^{-1}u + p)
\]

So indeed, \( G \) is in the same type class as \( \Lambda \) in this case.

**Case 2** \( \theta > 0 \); gives rise to the the type \( \Phi_\alpha = \exp(-x^{-\alpha}) \) supported on \( x > 0 \) Rearranging the equation for \( \beta \) gives:

\[
\frac{\beta(s)}{1 - \alpha(s)} = \frac{\beta(t)}{1 - \alpha(t)}
\]

Hence this value is constant for every value of \( t \). Writing \( "c" \) for this constant, we have then:

\[
\beta(t) = \frac{\beta(t)}{1 - \alpha(t)} (1 - \alpha(t)) = c(1 - \alpha(t)) = c(1 - t^{-\theta})
\]

So the equation for \( G \) is then:

\[
G'(x) = G(t^{-\theta}x + c(1 - t^{-\theta})) = G(t^{-\theta}(x - c) + c)
\]

Let \( y = x - c \) now to get:

\[
G'(y + c) = G(t^{-\theta}y + c)
\]

If we let \( H(x) = G(x + c), \) this is in the same type-class as \( G \), so it suffices to work with \( H \) which satisfies \( H'(y) = H(t^{-\theta}y) \). We claim now that \( H(0) = 0 \). Setting \( y = 0 \) in the above gives \( H'(0) = H(0) \) so it is clear that \( H(0) \) can only be \( 1 \) or \( 0 \). However, \( H(0) = 1 \) is impossible however by the following argument by contradiction. Suppose that \( H(0) = 1 \), then since \( H \) is non-degenerated, then we have an \( x_0 < 0 \) with \( 0 < H(x_0) < 1 \) and so \( H'(x_0) \) is a decreasing function of \( t \). On the other hand, \( H(t^{-\theta}x_0) \) is an increasing function of \( t \) (since \( t^{-\theta}x_0 \) is increasing as \( x_0 < 0 \)), which is a contradiction since these are equal. Hence \( H(x) = 0 \) for \( x < 0 \) since \( H \) is a distribution function.

Now, plugging in \( y = 1 \) we have \( H'(1) = H(t^{-\theta}) \). It must be that \( 0 < H(1) < 1 \) or else \( H \) is identically \( 0 \) or \( 1 \) everywhere. Let \( \alpha := \theta^{-1} \), and define \( p \) so that \( H(1) = \exp(-p^{-\alpha}) \). Under the change of variable \( u = t^{-\theta} \) so that \( t = u^{-\theta^{-1}} = u^{-\alpha} \). Notice that \( u \) ranges in \((0, \infty)\) as \( t > 0 \). We have:

\[
H(u) = H'(1) = \exp(-tp^{-\alpha}) = \exp(-(pu)^{-\alpha}) = \Phi_\alpha(pu)
\]

So indeed \( H \) and therefore \( G \) are in the typeclass of \( \Phi \).

**Case 3** \( \theta < 0 \); gives rise to the the type \( \Psi_\alpha = \exp(-|x|^\alpha) \) supported on \( x < 0 \) This case is very similar to the previous case. As before, we get to (to clarify things I write \( -\theta \) as \(|\theta|\) to remind myself \( \theta < 0 \) here):

\[
\beta(t) = c(1 - t^{\theta})
\]

\[
G'(y + c) = G(t^{\theta}y + c)
\]

\[
H'(y) = H(t^{\theta}y)
\]

We claim now that \( H(0) = 1 \). (This is analogous to the argument that showed \( H(0) = 0 \) in Case 2) Setting \( y = 0 \) in the above gives \( H'(0) = H(0) \) so it is clear that \( H(0) \) can only be \( 1 \) or \( 0 \). However, \( H(0) = 0 \) is impossible however by the following argument by contradiction. Suppose \( H(0) = 0 \), then since \( H \) is non-degenerated, then we have an \( x_0 > 0 \) with \( 0 < H(x_0) < 1 \) and so \( H'(x_0) \) is a decreasing function of \( t \). On the other hand, \( H(t^{\theta}x_0) \) is an increasing function of \( t \) (since \( t^{\theta}x_0 \) is increasing), which is a contradiction since these are equal. Hence \( H(x) = 1 \) for \( x > 0 \) since \( H \) is a distribution function.
Now, plugging in $y = -1$ we have $H^t(-1) = H(-t^{(\theta)})$. It must be that $0 < H(1) < 1$ or else $H$ is identically $0$ or $1$ everywhere. Let $\alpha := |\theta|^{-1}$, and define $p > 0$ so that $H(-1) = \exp(-p^\alpha)$. Under the change of variable $u = -t^{(\theta)}$ so that $t = (-u)^{\theta^{-1}} = |u|^\alpha$. Notice that $u$ ranges in $(-\infty, 0)$ as $t > 0$. We have:

$$
\begin{align*}
    H(u) &= H^t(-1) \\
    &= \exp(-tp^\alpha) \\
    &= \exp(-(p|u|)^\alpha) \\
    &= \Psi_\alpha(pu)
\end{align*}
$$

So indeed $H$ and therefore $G$ are in the typeclass of $\Psi_\alpha$.

\[\square\]

**Lemma 22.** (This was an exercise in Resnick) If $F$ is a non-degenerate distribution and $a > 0, c > 0, b \in \mathbb{R}$ and $d \in \mathbb{R}$ and if:

$$F(ax + b) = F(cx + d)$$

Then $a = c$ and $b = d$.

**Proof.** Under the change of variables $y = cx + d$, $x = c^{-1}(y - d)$ we are left with the hypothesis $F(ac^{-1}y + b - d) = F(y)$. Relabeling these as $A$ and $B$ we have $F(Ay + B) = F(y)$ and we desire to prove that $A = 1$ and $B = 0$.

Firstly notice that if we let $T(x) = Ax + B$ then we can iterate.

We first claim that $B = 0$. Otherwise, plugging in $y = 0$ gives $F(B) = F(0)$ which shows $F$ is constant on the interval $[0, B]$ (or if $B$ is negative, on $[-B, 0]$). From now on we will only consider the case $B > 0$, the other case is very similar). Then for any $-\frac{B}{A} < y < 0$ we have that $0 < Ay + B < B$ so $Ay + B \in [0, B]$ and hence $F(Ay + B) = F(0) = F(B)$ is the same constant. On the other hand, $F(Ay + B) = F(y)$ so $F$ must achieve the same constant value in the range $[-\frac{B}{A}, 0] \cup [0, B]$. The same argument can be repeated to further extend the interval to the left a bit further. For any $-B \left(\frac{1}{A} + \frac{1}{A^2}\right)$.