Independent Study 2015/2016

Study Notes June 23, 2015 Kenneth Chu

1 Cumulative distribution functions

Definition 1.1 *Let $X : (\,\Omega, \mathcal{A}, \mu\,) \longrightarrow \mathbb{R}$ be a $\mathbb{R}$-valued random variable. The* ***cumulative distribution function*** *of X is, by definition, the function $F_X : \mathbb{R} \longrightarrow [0,1]$ defined as follows:*

$$F_X(x) \;:=\; P(\,X \leq x\,) \;=\; \mu(\{\,\omega \in \Omega \mid X(\omega) \leq x\,\})\,, \quad \text{for each } x \in \mathbb{R}.$$

Definition 1.2 *A function $f : D \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ is said to be*

- ***non-decreasing*** *if $f(x) \leq f(y)$, for any $x, y \in D$ with $x \leq y$.*
- ***non-increasing*** *if $f(x) \geq f(y)$, for any $x, y \in D$ with $x \leq y$.*
- ***monotone*** *if f is either non-decreasing or non-increasing.*

Theorem 1.3 (Theorem 4.29, [1])
Let $f : (a,b) \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ be a non-decreasing function. Then,

$$f(x-) := \lim_{t \to x^-} f(t) \quad \text{and} \quad f(x+) := \lim_{t \to x^+} f(t)$$

exist for every $x \in (a,b)$. More precisely,

$$f(x-) \;=\; \sup_{a<t<x} f(t) \;\leq\; f(x) \;\leq\; \inf_{x<t<b} f(t) \;=\; f(x+).$$

Furthermore, if $a < x < y < b$, then

$$f(x+) \quad \leq \quad f(y-).$$

Proof First note that, since f is non-decreasing, it immediately follows that:

$$\sup_{a<t<x} f(t) \;\leq\; f(x) \;\leq\; \inf_{x<t<b} f(t).$$

Next, we show that $f(x-) := \lim_{t\to x^-} f(t)$ exists and equals $A := \sup_{a<t<x} f(t)$. Let $\varepsilon > 0$ be given. We need to find $\delta > 0$ such that

$$|\,f(t) - A\,| < \varepsilon, \quad \text{for every } t \in (x-\delta, x).$$

By definition of the supremum, there exists $\delta > 0$ with $x - \delta \in (a,x)$ such that

$$A - \varepsilon \;<\; f(x-\delta) \;\leq\; A.$$

Since f is non-decreasing, we have

$$f(x-\delta) \;\leq\; f(t) \;\leq\; A \;:=\; \sup_{\xi \in (x-\delta, x)} f(\xi), \quad \text{for every } t \in (x-\delta, x).$$

We therefore see that

$$|\,f(t) - A\,| \;<\; \varepsilon,$$

as desired. This proves that $f(x-) := \lim_{t\to x^-} f(t)$ indeed exists and equals $A := \sup_{t\in(a,x)} f(t)$. The proof that $f(x+)$ exists and equals $\inf_{t\in(x,b)} f(t)$ is analogous. Lastly, let $a < x < y < b$. Then, choose some $z \in (x,y) = (x,b) \bigcap (a,y)$. Hence, we have

$$f(x+) = \inf_{t\in(x,b)} f(t) \;\leq\; f(z) \;\leq\; \sup_{t\in(a,y)} f(t) \;=\; f(y-).$$

This proof the Theorem is complete. $\square$

Remark 1.4 *The analogous results of the preceding Theorem for non-increasing functions hold, obviously.*

Definition 1.5 *Let $f : D \subseteq \mathbb{R} \longrightarrow \mathbb{R}$. A point $a \in \mathrm{interior}(D)$ is a* ***jump discontinuity*** *of f if both*

$$\lim_{x \to a^-} f(x) \quad \text{and} \quad \lim_{x \to a^+} f(x)$$

exist but they are unequal.

Corollary 1.6 *A monotone $\mathbb{R}$-valued function defined on an interval of $\mathbb{R}$ can have only jump discontinuities.*

Theorem 1.7 *A function $F : \mathbb{R} \longrightarrow [0,1]$ is a cumulative distribution function of some $\mathbb{R}$-valued random variable if and only if each of following four conditions holds:*

- *F is non-decreasing.*
- *F is right-continuous.*
- $\lim_{x \to -\infty} F(x) = 0$.
- $\lim_{x \to +\infty} F(x) = 1$.

Proof If $F : \mathbb{R} \longrightarrow [0,1]$ is a cumulative distribution function of some $\mathbb{R}$-valued random variable $X : (\,\Omega, \mathcal{A}, \mu\,) \longrightarrow \mathbb{R}$, then the four conditions follow immediately from the property of the probability measure μ. Conversely, suppose the four conditions hold. Let $\Omega := [0,1]$ and $\mathcal{B}(\Omega)$ the Borel subsets of Ω. Let μ be the Lebesgue measure on $(\Omega, \mathcal{B}(\Omega))$, i.e. μ is determined by:

$$\mu(\,[0,\omega]\,) \;:=\; \omega, \quad \text{for each } \omega \in \Omega = [0,1].$$

Define the random variable $X : (\Omega, \mathcal{B}(\Omega), \mu) \longrightarrow \mathbb{R}$ by:

$$X(\omega) \;:=\; \inf\{\, x \in \mathbb{R} \mid \omega \leq F(x) \,\}, \quad \text{for each } \omega \in \Omega = [0,1].$$

Note that X is simply the quantile function of F.

> **Claim:** Suppose $G : \mathbb{R} \longrightarrow [0,1]$ is non-decreasing and right-continuous. Then, for any $\omega \in [0,1]$ and $x \in \mathbb{R}$,
>
> $$\inf\{\, \xi \in \mathbb{R} \mid \omega \leq G(\xi) \,\} \leq x \iff \omega \leq G(x).$$
>
> Proof of Claim: Suppose $\omega \leq G(x)$. Then, $x \in \{\, \xi \in \mathbb{R} \mid \omega \leq G(\xi) \,\}$. Hence, $\inf\{\, \xi \in \mathbb{R} \mid \omega \leq G(\xi) \,\} \leq x$. Conversely, suppose $\inf\{\, \xi \in \mathbb{R} \mid \omega \leq G(\xi) \,\} \leq x$. Since G is non-decreasing and right-continuous, we have:
>
> $$\begin{aligned} \inf\{\, \xi \in \mathbb{R} \mid \omega \leq G(\xi) \,\} \leq x &\implies \text{for any } \varepsilon > 0,\ \exists\ \xi \in \mathbb{R}, \text{ satisfying } \omega \leq G(\xi), \text{ such that } \xi \leq x + \varepsilon \\ &\implies \text{for any } \varepsilon > 0,\ \exists\ \xi \in \mathbb{R}, \text{ satisfying } \omega \leq G(\xi) \text{ and } G(\xi) \leq G(x+\varepsilon) \\ &\implies \text{for any } \varepsilon > 0,\ \omega \leq G(x+\varepsilon) \\ &\implies \omega \;\leq\; \lim_{\varepsilon \to 0^+} G(x+\varepsilon) \;=\; G(x). \end{aligned}$$
>
> This completes the proof of the Claim.

Noting that, by hypothesis, F is non-decreasing right-continuous, and invoking the Claim above, we see that

$$\begin{aligned} P(\,X \leq x\,) &= P(\{\,\omega \in \Omega \mid X(\omega) \leq x\,\}) \;=\; P(\{\,\omega \in \Omega \mid \inf\{\,\xi \in \mathbb{R} \mid \omega \leq F(\xi)\,\} \leq x\,\}) \\ &= P(\{\,\omega \in \Omega \mid \omega \leq F(x)\,\}) \;=\; \mu(\{\,\omega \in [0,1] \mid \omega \leq F(x)\,\}) \;=\; \mu(\,[0, F(x)]\,) \\ &= F(x) \end{aligned}$$

This shows that if F satisfies the four conditions, then F is the cumulative distribution function of the random variable X constructed above. The proof of the Theorem is now complete. □

Theorem 1.8 (Darboux-Froda, Theorem 4.30, [1])
The set of discontinuities of a monotone $\mathbb{R}$-valued function defined on an interval of $\mathbb{R}$ is at most countable.

Proof We give the proof for non-decreasing functions; the proof for non-increasing functions is analogous. Let $f : (a,b) \longrightarrow \mathbb{R}$ be non-decreasing, and let $\mathcal{D}(f) \subset (a,b)$ be the set of discontinuities of f. By Corollary 1.6, each $x \in \mathcal{D}(f)$ is a jump discontinuity of f, i.e. both one-sided limits $\lim_{t\to x^-} f(t)$ and $\lim_{t\to x^+} f(t)$ exist, and

$$\lim_{t\to x^-} f(t) \quad < \quad \lim_{t\to x^+} f(t)$$

Thus, for each $x \in \mathcal{D}(f)$, we may choose a rational number $r(x) \in \mathbb{Q}$ such that

$$\lim_{t\to x^-} f(t) \quad < \quad r(x) \quad < \quad \lim_{t\to x^+} f(t).$$

This defines a function $r : \mathcal{D}(f) \longrightarrow \mathbb{Q}$. Note that this function is injective. Indeed, let $x, y \in \mathcal{D}(f)$ with $x < y$. Then, by Theorem 1.3,

$$r(x) \quad < \quad \lim_{t\to x^+} f(t) \quad = \quad f(x+) \quad \leq \quad f(y-) \quad = \quad \lim_{t\to y^-} f(t) \quad < \quad r(y)$$

This shows $r : \mathcal{D}(f) \longrightarrow \mathbb{Q}$ is indeed injective. Since $\mathbb{Q}$ is countable, we may now conclude that $\mathcal{D}(f)$ is at most countable. $\square$

Corollary 1.9 *The cumulative distribution function of an $\mathbb{R}$-valued random variable can have only jump discontinuities, and its set of (jump) discontinuities is at most countable.*

2 The O_P and o_P notations; convergence in distribution implies boundedness in probability

Definition 2.1 (The Big-O_P notation)
Let $\{ X_n : (\Omega_n, \mathcal{A}_n, \mu_n) \longrightarrow \mathbb{R}^k \}_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^k$-valued random variables. Let $\{ a_n \}_{n\in\mathbb{N}}$ be sequence of positive numbers. The notation $X_n = O_p(a_n)$ means:

> *For every $\varepsilon > 0$, there exist $C_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that $P(\,|X_n| \leq C_\varepsilon \cdot a_n\,) > 1 - \varepsilon$, for every $n \geq n_\varepsilon$.*

Proposition 2.2 *The following are equivalent:*

(a) $X_n = O_P(a_n)$.

(b) *For every $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that $P(\,|X_n| \leq C_\varepsilon \cdot a_n\,) > 1 - \varepsilon$, for each $n \in \mathbb{N}$.*

(c) *For every $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that $\limsup_{n\to\infty} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon$.*

(d) *For every $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that $\sup_{n\in\mathbb{N}} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon$.*

(e) $\lim_{C\to\infty} \limsup_{n\to\infty} P(\,|X_n| > C \cdot a_n\,) = 0$.

(f) $\lim_{C\to\infty} \sup_{n\in\mathbb{N}} P(\,|X_n| > C \cdot a_n\,) = 0$.

Proof

<u>(a) $\Longrightarrow$ (b)</u>

Let $\varepsilon > 0$ be given. By (a), there exist $B_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that $P(\,|X_n| \leq B_\varepsilon \cdot a_n\,) > 1 - \varepsilon$, for each $n \geq n_\varepsilon$.

Claim: Let Y be an $\mathbb{R}^k$-valued random variable. Then, for each $\varepsilon > 0$, there exists $A_\varepsilon > 0$ such that $P(\,|Y| \leq A_\varepsilon\,) > 1 - \varepsilon$.

Proof of Claim: Suppose the Claim were false. Then, there exists some $\varepsilon > 0$ such that $P(\,|Y| \leq A\,) \leq 1 - \varepsilon$, for every $A > 0$; equivalently, $P(\,|Y| > A\,) > \varepsilon$, for every $A > 0$. This implies $\lim_{A\to\infty} P(\,|Y| > A\,) = \limsup_{A\to\infty} P(\,|Y| > A\,) \geq \varepsilon > 0$. But this is a contradiction since $\lim_{A\to\infty} P(\,|Y| > A\,) = 0$, for every $\mathbb{R}^k$-valued random variable Y. This proves the Claim.

By the Claim, for each $i = 1, 2, \ldots, n_\varepsilon - 1$, there exists $B_\varepsilon^{(i)} > 0$ such that $P\Big(|X_i| \leq B_\varepsilon^{(i)} \cdot a_i\Big) > 1 - \varepsilon$. Now, let $C_\varepsilon := \max\left\{\, B_\varepsilon^{(1)}, B_\varepsilon^{(1)}, \ldots, B_\varepsilon^{(n_\varepsilon - 1)}, B_\varepsilon \,\right\}$. Then, $P(\,|X_n| \leq C_\varepsilon \cdot a_n\,) > 1 - \varepsilon$, for every $n \in \mathbb{N}$. This proves the implication (a) $\Longrightarrow$ (b).

(b) $\Longrightarrow$ (a) Trivial: Suppose (b) holds. Then (a) immediately follows with $n_\varepsilon = 1$.

(a) $\Longleftrightarrow$ (c) Let $\varepsilon > 0$ be given.

$$
\begin{aligned}
\text{(a)} \quad &\Longleftrightarrow \quad \text{There exist } C_\varepsilon > 0 \text{ and } n_\varepsilon \in \mathbb{N} \text{ such that } P(\,|X_n| \leq C_\varepsilon \cdot a_n\,) > 1 - \varepsilon, \text{ for every } n \geq n_\varepsilon. \\
&\Longleftrightarrow \quad \text{There exist } C_\varepsilon > 0 \text{ and } n_\varepsilon \in \mathbb{N} \text{ such that } P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon, \text{ for every } n \geq n_\varepsilon. \\
&\Longleftrightarrow \quad \text{There exist } C_\varepsilon > 0 \text{ such that } \limsup_{n\to\infty} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon \quad \Longleftrightarrow \quad \text{(c)}
\end{aligned}
$$

(b) $\Longleftrightarrow$ (d) Let $\varepsilon > 0$ be given.

$$
\begin{aligned}
\text{(b)} \quad &\Longleftrightarrow \quad \text{There exists } C_\varepsilon > 0 \text{ such that } P(\,|X_n| \leq C_\varepsilon \cdot a_n\,) > 1 - \varepsilon, \text{ for every } n \in \mathbb{N}. \\
&\Longleftrightarrow \quad \text{There exists } C_\varepsilon > 0 \text{ such that } P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon, \text{ for every } n \in \mathbb{N}. \\
&\Longleftrightarrow \quad \text{There exist } C_\varepsilon > 0 \text{ such that } \sup_{n\in\mathbb{N}} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon \quad \Longleftrightarrow \quad \text{(d)}
\end{aligned}
$$

(d) $\Longleftrightarrow$ (f) Let $\varepsilon > 0$ be given. We first establish that (f) $\Longrightarrow$ (d).

$$
\begin{aligned}
\text{(f)} \quad &\Longleftrightarrow \quad \text{There exists } C_\varepsilon > 0 \text{ such that } \sup_{n\in\mathbb{N}} P(\,|X_n| > C \cdot a_n\,) \leq \varepsilon, \text{ for each } C \geq C_\varepsilon. \\
&\Longrightarrow \quad \text{There exists } C_\varepsilon > 0 \text{ such that } \sup_{n\in\mathbb{N}} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon \quad \Longleftrightarrow \quad \text{(d)}
\end{aligned}
$$

Conversely, suppose (d) holds and $C \geq C_\varepsilon$. Then, $|X_n| > C \cdot a_n \Longrightarrow |X_n| > C_\varepsilon \cdot a_n$. Hence, $\{\,|X_n| > C \cdot a_n\,\} \subseteq \{\,|X_n| > C_\varepsilon \cdot a_n\,\}$, which in turn implies $P(\,|X_n| > C \cdot a_n\,) \leq P(\,|X_n| > C_\varepsilon \cdot a_n\,)$. Thus, we have

$$
\sup_{n\in\mathbb{N}} P(\,|X_n| > C \cdot a_n\,) \;\leq\; \sup_{n\in\mathbb{N}} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \;\leq\; \varepsilon,
$$

i.e. (f) holds.

(c) $\Longleftrightarrow$ (e) Let $\varepsilon > 0$ be given. We first establish that (e) $\Longrightarrow$ (c).

$$
\begin{aligned}
\text{(e)} \quad &\Longleftrightarrow \quad \text{There exists } C_\varepsilon > 0 \text{ such that } \limsup_{n\to\infty} P(\,|X_n| > C \cdot a_n\,) \leq \varepsilon, \text{ for each } C \geq C_\varepsilon. \\
&\Longrightarrow \quad \text{There exists } C_\varepsilon > 0 \text{ such that } \limsup_{n\to\infty} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \leq \varepsilon \quad \Longleftrightarrow \quad \text{(c)}
\end{aligned}
$$

Conversely, suppose (c) holds and $C \geq C_\varepsilon$. Then, $|X_n| > C \cdot a_n \Longrightarrow |X_n| > C_\varepsilon \cdot a_n$. Hence, $\{\,|X_n| > C \cdot a_n\,\} \subseteq \{\,|X_n| > C_\varepsilon \cdot a_n\,\}$, which in turn implies $P(\,|X_n| > C \cdot a_n\,) \leq P(\,|X_n| > C_\varepsilon \cdot a_n\,)$. Thus, we have

$$
\limsup_{n\to\infty} P(\,|X_n| > C \cdot a_n\,) \;\leq\; \limsup_{n\to\infty} P(\,|X_n| > C_\varepsilon \cdot a_n\,) \;\leq\; \varepsilon,
$$

i.e. (e) holds.

This completes the proof of the Proposition. $\square$

Definition 2.3 (Bounded in probability)

A sequence $\{ X_n : (\Omega_n, \mathcal{A}_n, \mu_n) \longrightarrow \mathbb{R}^k \}_{n\in\mathbb{N}}$ *of* $\mathbb{R}^k$*-valued random variables is said to be* ***bounded in probability*** *if* $X_n = O_P(1)$.

Theorem 2.4

If a sequence $\{ X_n : (\Omega_n, \mathcal{A}_n, \mu_n) \longrightarrow \mathbb{R} \}_{n\in\mathbb{N}}$ *of* $\mathbb{R}$*-valued random variables converges in distribution to some random variable* $X : (\Omega, \mathcal{A}, \mu) \longrightarrow \mathbb{R}$*, then the sequence* $\{ X_n \}$ *is bounded in probability.*

PROOF Let $\varepsilon > 0$ be given. We need to show that there exist $C_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that

$$P(\,|X_n| > C_\varepsilon\,) \;\leq\; \varepsilon, \quad \text{for each } n \geq n_\varepsilon.$$

Denote by $F, F_n : \mathbb{R} \longrightarrow [0,1]$ the cumulative distribution functions of X and X_n, respectively. By Theorem 1.7 and the Darboux-Froda Theorem (Theorem 1.8), the cumulative distribution function F satisfies: $\lim_{x\to+\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$, and that F can have at most countably many (jump) discontinuities. Thus for the given $\varepsilon > 0$, we may choose $C_\varepsilon > 0$ sufficiently large such that

$$0 \;\leq\; F(-C_\varepsilon) \;<\; \frac{\varepsilon}{4}, \qquad |\,1 - F(C_\varepsilon)\,| \;<\; \frac{\varepsilon}{4}, \qquad \text{and} \qquad \{\, \pm C_\varepsilon \,\} \;\subset\; \mathcal{C}(F)$$

where $\mathcal{C}(F)$ denotes the continuity set of F. Now, since $\pm C_\varepsilon \in \mathcal{C}(F)$, the convergence in distribution $X_n \xrightarrow{\mathcal{L}} X$ implies that the convergences $F_n(-C_\varepsilon) \longrightarrow F(-C_\varepsilon)$ and $F_n(C_\varepsilon) \longrightarrow F(C_\varepsilon)$ (of sequences of real numbers). Thus, we may choose $n_\varepsilon \in \mathbb{N}$ sufficiently large such that

$$|\,F_n(-C_\varepsilon) - F(-C_\varepsilon)\,| < \frac{\varepsilon}{4}, \quad \text{and} \quad |\,F_n(C_\varepsilon) - F(C_\varepsilon)\,| < \frac{\varepsilon}{4}, \quad \text{for every } n \geq n_\varepsilon.$$

Therefore, for each $n \geq n_\varepsilon$, we have:

$$\begin{aligned}
P(\,|X_n| > C_\varepsilon\,) &= P(\,X_n < -C_\varepsilon\,) + P(\,X_n > C_\varepsilon\,) \;=\; P(\,X_n < -C_\varepsilon\,) + 1 - P(\,X_n \leq C_\varepsilon\,) \\
&\leq P(\,X_n \leq -C_\varepsilon\,) + 1 - P(\,X_n \leq C_\varepsilon\,) \;=\; F_n(\,-C_\varepsilon\,) + 1 - F_n(\,C_\varepsilon\,) \\
&\leq |\,F_n(\,-C_\varepsilon\,) - F(\,-C_\varepsilon\,)\,| + |\,F(\,-C_\varepsilon\,)\,| + |\,1 - F(\,C_\varepsilon\,)\,| + |\,F(\,C_\varepsilon\,) - F_n(\,C_\varepsilon\,)\,| \\
&< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} \;=\; \varepsilon
\end{aligned}$$

This completes the proof that a sequence $\{\, X_n \,\}_{n\in\mathbb{N}}$ of $\mathbb{R}$-valued random variables is bounded in probability whenever it converges in distribution. $\square$

References

[1] RUDIN, W. *Principles of Mathematical Analysis*, third ed. McGraw-Hill, 1976.