Stats 310A Session 1

October 4, 2017

1 Constructing the uniform measure on $[0, 1)$

In this section, we briefly describe how to construct the uniform measure (also the Lebesgue measure) on $[0, 1)$. We will mostly follow Dembo's notes and Sections 1, 2 in Billingsley [1].

Consider sets that are finite disjoint unions of intervals in $[0, 1)$. Let $\mathcal{B}_0$ denote this family of sets:

$$\mathcal{B}_0 = \left\{ A = \bigcup_{k=1}^{n} [a_k, b_k) : 0 \le a_1 < b_1 < \cdots < a_n < b_n \le 1, n \in \mathbb{N} \right\}.$$

It is easy to verify that $\mathcal{B}_0$ is an *algebra*: it is closed under complement and union, and $\emptyset \in \mathcal{B}_0$.

Now, define set function $\lambda : \mathcal{B}_0 \to [0, 1]$ as

$$\lambda(A) = \sum_{k=1}^{n} (b_k - a_k) \quad \text{for} \quad A = \bigcup_{k=1}^{n} [a_k, b_k).$$

We claim that λ is a probability measure on $\mathcal{B}_0$. Clearly, $\lambda(A) \in [0, 1]$ for all $A \in \mathcal{B}_0$, $\lambda(\emptyset) = 0$, and $\lambda([0, 1)) = 1$. It remains to show that λ is *countably additive*, that is,

$$A = \bigcup_{k=1}^{\infty} A_k, \;\; A, A_k \in \mathcal{B}_0, \;\; A_k \text{ disjoint} \;\; \text{implies} \;\; \lambda(A) = \sum_{i=1}^{\infty} \lambda(A_i).$$

To achieve this, we need the following result on the length of intervals. For a (finite) interval $I = [a, b)$, let $|I| = b - a$ denote its length.

Lemma 1.1 (Theorem 1.3, [1])**.** *Let I and $\{I_k\}_{k=0}^{\infty}$ be intervals.*

(i) If $\bigcup_k I_k \subset I$ and the I_k are disjoint, then $\sum_k |I_k| \le |I|$.

(ii) If $I \subset \bigcup_k I_k$, then $|I| \le \sum_k |I_k|$.

(iii) If $I = \bigcup_k I_k$ and the I_k are disjoint, then $|I| = \sum_k |I_k|$.

In particular, (iii) ensures that the length of an interval is not only finitely but also countably additive, which we will now use to show that λ is also countably additive. Let $A = \bigcup_{i=1}^{n} I_k$ and $A_k = \bigcup_{j=1}^{m_k} J_{kj}$ be the disjoint interval representations. Then for all i, we have

$$I_i = I_i \cap A = I_i \bigcap \left(\bigcup_{k=1}^{\infty} \bigcup_{j=1}^{m_k} J_{kj} \right) = \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{m_k} I_i \cap J_{kj}.$$

$I_i \cap J_{kj}$ are disjoint intervals, so we can apply Lemma 1.1(iii) twice to get

$$\lambda(A) = \sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n}\sum_{k=1}^{\infty}\sum_{j=1}^{m_k} |I_i \cap J_{kj}| = \sum_{k=1}^{\infty}\sum_{j=1}^{m_k} |J_{kj}| = \sum_{k=1}^{\infty} \lambda(A_k).$$

This completes the proof.

As λ is a probability measure on the algebra $\mathcal{B}_0$, the Caratheodory extension theorem states that λ has an unique extension onto $\mathcal{B} = \sigma(\mathcal{B}_0)$, giving the Lebesgue measure on Borel sets. Note that λ can be extended onto $\mathcal{G}$, the family of measurable sets, which is strictly larger than $\mathcal{B}$.

Proof of Lemma 1.1 Let $I = [a, b)$ and $I_k = [a_k, b_k)$.

(i) *Finite case.* Suppose there are n intervals. We perform induction on n. The result is obvious when $n = 1$. Assume the result is true for $n - 1$, and let I_k be sorted in the increasing order, then we have $b_k \le a_n < b_n \le b$ for all $k \le n - 1$. Now, the smaller interval $[a, a_n)$ contains $\bigcup_{k=0}^{n-1} I_k$, so by the inductive assumption we have $\sum_{k=1}^{n-1} |I_k| \le a_n - a$. This gives

$$\sum_{k=1}^{n} |I_k| = \sum_{k=1}^{n-1} |I_k| + (b_n - a_n) \le (a_n - a) + (b_n - a_n) = b_n - a \le b - a = |I|,$$

verifying the result for n.

Infinite case. For all n we have $\sum_{k=1}^{n} |I_k| \le |I|$ by the finite case. Letting $n \to \infty$ gives the result.

(ii) *Finite case.* Induction. Assume the result is true for $n-1$. Then, there is at least one interval, WLOG $[a_n, b_n)$, such that $a_n < b \le b_n$. This interval covers the $[a_n, b)$ portion of I, so the rest must cover $[a, a_n)$. By the inductive assumption, $a_n - a \le \sum_{k=1}^{n-1} |I_k|$. This gives

$$|I| = b - a = (a_n - a) + (b - a_n) \le \sum_{k=1}^{n-1} |I_k| + (b_n - a_n) = \sum_{k=1}^{n} |I_k|.$$

Infinite case. By the assumption

$$[a, b) \subset \bigcup_{k=1}^{\infty} [a_k, b_k),$$

we have

$$[a, b - \varepsilon] \subset \bigcup_{k=1}^{\infty} \left(a_k - \frac{\varepsilon}{2^k}, b_k\right) \quad \text{for all } 0 < \varepsilon < b - a,$$

as the LHS is a smaller set and the RHS is a larger set. However, as the interval $[a, b - \varepsilon]$ is compact and the RHS is an open cover, it must have a finite subcover, WLOG $k \in \{1, \dots, n\}$, giving that

$$\bigcup_{k=1}^{n} \left(a_k - \frac{\varepsilon}{2^k}, b_k\right) \supset [a, b - \varepsilon] \supset [a, b - \varepsilon).$$

Applying the finite case, we get

$$b - a \le \varepsilon + \sum_{k=1}^{n} \left(b_k - a_k + \frac{\varepsilon}{2^k}\right) \le \sum_{k=1}^{n} (b_k - a_k) + 2\varepsilon \le \sum_{k=1}^{\infty} (b_k - a_k) + 2\varepsilon.$$

Taking $\varepsilon \to 0$ gives the desired result.

(iii) Follows from (i) and (ii).

□

2 Proof of existence in Caratheodory's extension theorem

In this section, we prove the existence part in Caratheodory's extension theorem: a probability measure P on a field $\mathcal{F}_0$ has an extension to $\sigma(\mathcal{F}_0)$. We will follow Section 3 in Billingsley [1].

For any set $A \subset \Omega$, define its *outer measure* by

$$P^*(A) = \inf_{\{A_n\} \text{ covers } A} \sum_{n=1}^{\infty} P(A_n).$$

$P^*(A)$ measures the size of a set $\mathcal{A}$ by its smallest countable $\mathcal{F}_0$-cover. One can check that it satisfies the following properties:

(i) $P^*(\emptyset) = 0$.

(ii) Nonnegativity: $P^*(A) \geq 0$ for all $A \subset \Omega$.

(iii) Monotonicity: $A \subset B$ implies $P^*(A) \leq P^*(B)$.

(iv) Countable subadditivity: if $A \subset \bigcup_n A_n$, then $P^*(A) \leq \sum_n P^*(A_n)$.

Properties (i) - (iii) are relatively easy to verify; (iv) can be verified by constructing covers of A_n wiithin $\varepsilon/2^n$ of the outer measure. We note that (iv) also implies finite subadditivity, in particular, $P^*(A \cup B) \leq P^*(A) + P^*(B)$.

Now, we define a class of sets

$$\mathcal{G} := \{A \subset \Omega : \mathbb{P}^*(E \cap A) + P^*(E \cap A^c) = P^*(E) \text{ for all } E \subset \Omega\}.$$

Our goal is to show that P^* restricted on $\mathcal{G}$ is the extension of P. The class $\mathcal{G}$ contains $\sigma(\mathcal{F}_0)$ and is what we will later call *measurebale sets*. Also, as a consequence of finite subadditivity, the "$\geq$" direction in the defining equality always holds, so we only need to check the "$\leq$" direction in order to show that a set is in $\mathcal{G}$.

Lemma 2.1. *The class $\mathcal{G}$ is an algebra.*

Proof Clearly $\emptyset \in \mathcal{G}$, and $A \in \mathcal{G}$ implies $\mathcal{A}^c \in \mathcal{G}$ by symmetry of the definition, so it remains to show that $\mathcal{G}$ is closed under intersection. For any $A, B \in \mathcal{G}$ and $E \subset \Omega$, we have

$$\begin{aligned} & P^*(E) = P^*(E \cap A) + P^*(E \cap A^c) \\ &= P^*(E \cap A \cap B) + P^*(E \cap A \cap B^c) + P^*(E \cap A^c \cap B) + P^*(E \cap A^c \cap B^c) \\ &\geq P^*(E \cap A \cap B) + P^*\left(E \cap A \cap B^c \bigcup E \cap A^c \cap B \bigcup E \cap A^c \cap B^c\right) \\ &= P^*(E \cap (A \cap B)) + P^*\left(E \cap (A \cap B)^c\right), \end{aligned}$$

which shows $A \cap B \in \mathcal{G}$. □

Lemma 2.2. *If $A_1, A_2, \ldots$ are disjoint $\mathcal{G}$-sets, then for all $E \subset \Omega$,*

$$P^*\left(E \bigcap \left(\bigcup_n A_n\right)\right) = \sum_n P^*(E \cap A_n).$$

Proof *Finite case.* Suppose there are n sets $A_1, \ldots, A_n$. When $n = 1$ this is obvious. Assume the result holds with $n-1$, then letting $B_k = \bigcup_{i=1}^k A_i$ for all k, we have

$$P^*(E \cap B_n) = P^*(E \cap B_n \cap B_{n-1}) + P^*(E \cap B_n \cap B_{n-1}^c)$$
$$= P^*(E \cap B_{n-1}) + P^*(E \cap A_n) = \sum_{i=1}^{n-1} P^*(E \cap A_i) + P^*(E \cap A_n) = \sum_{i=1}^{n} P^*(E \cap A_i).$$

By induction, the result is true for all finite collections.

Infinite case. As P^* is countably subadditive, we need only show the "$\geq$" direction. By monotonicity and the finite case,

$$P^*\left(E \bigcap \left(\bigcup_i A_i\right)\right) \geq P^*\left(E \bigcap \left(\bigcup_{i=1}^n A_i\right)\right) = \sum_{i=1}^n P^*(E \cap A_n)$$

for all n. Letting $n \to \infty$, we get

$$P^*\left(E \bigcap \left(\bigcup_i A_i\right)\right) \geq \sum_{i=1}^\infty P^*(E \cap A_n).$$

□

Lemma 2.3. *The class $\mathcal{G}$ is a σ-algebra, and P^* restricted on $\mathcal{G}$ is countably additive.*

Proof Suppose $A_1, A_2, \ldots$ are disjoint $\mathcal{G}$-sets. Let $B = \bigcup A_i$ and $B_n = \bigcup_{i=1}^n A_i$. For any $E \subset \Omega$, we have

$$P^*(E) = P^*(E \cap B_n) + P^*(E \cap B_n^c) = \sum_{i=1}^n P^*(E \cap A_i) + P^*(E \cap B_n^c) \geq \sum_{i=1}^n P^*(E \cap A_i) + P^*(E \cap B^c).$$

Letting $n \to \infty$, we obtain

$$P^*(E) \geq \sum_{i=1}^\infty P^*(E \cap A_i) + P^*(E \cap B^c) = P^*(E \cap B) + P^*(E \cap B^c),$$

where we applied Lemma 2.2 to get the last equality. This shows that $B \in \mathcal{G}$, and so $\mathcal{G}$ is a σ-algebra. Taking $E = B$ in the above inequality gives countable additivity. □

Lemma 2.4. *We have $\mathcal{F}_0 \in \mathcal{G}$.*

Proof Let $A \in \mathcal{F}_0$ and take any $E \subset \Omega$. For any $\varepsilon > 0$, there exists a cover $\{A_n\} \subset \mathcal{F}_0$ of E such that $\sum_n P(A_n) \leq P^*(E) + \varepsilon$. Let $B_n = A_n \cap A$ and $C_n = A_n \cap A^c$, these sets are all in $\mathcal{F}_0$ and covers $E \cap A$ and $E \cap A^c$, respectively. So we have

$$P^*(E \cap A) + P^*(E \cap A^c) \leq \sum_n P(A_n \cap A) + \sum_n P(A_n \cap A^c) = \sum_n P(A_n) \leq P^*(E) + \varepsilon.$$

Letting $\varepsilon \to 0$, we get $A \in \mathcal{G}$. □

Lemma 2.5. *P^* restricted on $\mathcal{F}_0$ is equal to P, i.e.*

$$P^*(A) = P(A), \text{ for all } A \in \mathcal{F}_0.$$

Proof Let $A \in \mathcal{F}_0$. Clearly A itself covers A, so $P^*(A) \leq P(A)$. Conversely, if $\{A_n\}$ is a $\mathcal{F}_0$-cover of A, then by the countable subadditivity and monotonicity of P on $\mathcal{F}_0$, we have

$$P(A) \leq \sum_n P(A \cap A_n) \leq \sum_n P(A_n).$$

Taking inf over all covers gives that $P(A) \leq P^*(A)$. $\square$

Proof of existence of extension By Lemmas 2.3, 2.4, and 2.5, the outer measure P^* extends P onto $\mathcal{G}$, which is a σ-algebra that contains $\mathcal{F}_0$. Thus, $\mathcal{G} \supset \sigma(\mathcal{F}_0)$. As P^* is a probability measure on $\mathcal{G}$, it is also a probability measure when restricted to $\sigma(\mathcal{F}_0)$. $\square$

References

[1] P. Billingsley. *Probability and measure.* John Wiley & Sons, 2008.

Stats 310A Session 2

October 6, 2017

1 Characterization of distribution functions

In this section, we give necessary and sufficient conditions for a function $F : \mathbb{R} \to [0, 1]$ to be a distribution function.

Theorem 1 (Thm 1.2.36, Dembo's Notes)**.** *A function $F : \mathbb{R} \to [0, 1]$ is a distribution function of some R.V. if and only if*

(a) F is non-decreasing;

(b) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$;

(c) F is right-continuous, i.e. $\lim_{y\downarrow x} F(y) = F(x)$.

Proof "$\Rightarrow$". Let F be the distribution of some random variable X on a probability space $(\Omega, \mathcal{F}, P)$. Let $x \le y$, then $\{\omega : X(\omega) \le x\} \subseteq \{\omega : X(\omega) \le y\}$, hence $F(x) = \mathbb{P}(X \le x) \le \mathbb{P}(X \le y) = F(y)$. By continuity of P, we have

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} P(\{\omega : X(\omega) \le x\}) = P(\lim_{x\to\infty} \{\omega : X(\omega) \le x\}) = P(\Omega) = 1$$

and similarly $\lim_{x\to 0} F(x) = P(\emptyset) = 0$. Take $x \in \mathbb{R}$, we have

$$\lim_{y\downarrow x} \{\omega : X(\omega) \le y\} = \{\omega : X(\omega) \le x\}.$$

Hence by continuity of P,

$$\lim_{y\downarrow x} F(y) = \lim_{y\downarrow x} P(\{\omega : X(\omega) \le y\}) = P(\{\omega : X(\omega) \le x\}) = F(x).$$

"$\Leftarrow$". We define $X^-(\omega) = \sup\{y : F(y) < \omega\}$ on the probability space $((0, 1], \mathcal{B}_{(0,1]}, U)$, i.e. $(0, 1]$ with the uniform distribution. Note that for all $\omega \in (0, 1)$, as F is non-decreasing and its range contains $(0, 1)$, the set $\{y : F(y) \le \omega\}$ is non-empty and has a finite upper bound. Hence $X^- : (0, 1) \to \mathbb{R}$ is well-defined.

We are going to show that the distribution function of X^- equals to F. We claim that for all $x \in \mathbb{R}$,

$$\big\{\omega : X^-(\omega) \le x\big\} = \{\omega : \omega \le F(x)\}. \tag{1}$$

This implies that the LHS is in $\mathcal{B}_{(0,1]}$ and that

$$U(\big\{\omega : X^-(\omega) \le x\big\}) = U(\{\omega : \omega \le F(x)\}) = U((0, F(x)]) = F(x),$$

so the distribution function of $X^-(\omega)$ is F.

It remains to show (1). Suppose $F(x) \geq \omega$, then by monotonicity $x \geq y$ for all y such that $F(y) < \omega$, giving that $X^-(\omega) = \sup\{y : F(y) < \omega\} \leq x$. Conversely, suppose $X^-(\omega) \leq x$, we claim that $F(x) \geq \omega$ has to be true. If not, then $F(x) < \omega$. By the right continuity of F, there exists some $\varepsilon > 0$ such that $F(x + \varepsilon) < \omega$, giving that

$$X^-(\omega) = \sup\{y : F(y) < \omega\} \geq x + \varepsilon > x,$$

a contradiction. Hence, we must have $F(x) \geq \omega$. □

2 Completion of measure spaces

A nice property about the Lebesgue measure is the following: any subset of a measure-zero set is measurable. To see this, for example on $\mathbb{R}$, let A have measure zero and $B \subset A$. For any $E \subset \mathbb{R}$, we have

$$\begin{aligned} &P^*(E \cap B) + P^*(E \cap B^c) \\ &\leq P^*(E \cap A) + P^*(E \cap B^c \cap A^c) + P^*(E \cap B^c \cap A) \\ &= P^*(E \cap A) + P^*(E \cap A^c) + P^*(E \cap (A \setminus B)) = P^*(E \cap A) + P^*(E \cap A^c) = P^*(E). \end{aligned} \tag{2}$$

The last equality follows as $E \cap (A \setminus B)$ is a subset of A, so $P^*(E \cap (A \setminus B)) \leq P^*(A) = 0$. Hence, B is measurable, and $P(B) \leq P(A) = 0$.

However, such a property might not be present in a general measure space. We are going present a result saying that one can always slightly enlarge the σ-algebra and extend the measure to get this property.

Definition 1 (Def 1.1.34, Dembo's Notes). *We say that a measure space $(\Omega, \mathcal{F}, \mu)$ is* complete *if any subset N of any $B \in \mathcal{F}$ with $\mu(B) = 0$ is also in $\mathcal{F}$.*

Theorem 2 (Thm 1.1.35, Dembo's Notes). *Given a measure space $(\Omega, \mathcal{F}, \mu)$, let*

$$\mathcal{N} = \{N : N \subseteq A \textit{ for some } A \in \mathcal{F} \textit{ with } \mu(A) = 0\}$$

denote the collection of μ-null sets. Then, there exists a complete measure space $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$, called the completion *of the measure space $(\Omega, \mathcal{F}, \mu)$, such that $\overline{\mathcal{F}} = \{F \cup N : F \in \mathcal{F}, N \in \mathcal{N}\}$ and $\overline{\mu} = \mu$ on $\mathcal{F}$.*

Intuitively the result is quite expected: we can add all the μ-null sets into $\mathcal{F}$ and let them have measure zero.

Proof We divide the proof into the following steps.

(1) $\overline{\mathcal{F}}$ is a σ-algebra.

Clearly $\emptyset \in \overline{\mathcal{F}}$. Take any $B \in \overline{\mathcal{F}}$, then $B = F \cup N$ with $F \in \mathcal{F}$ and $N \in \mathcal{N}$. In particular, there exists $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $N \subseteq A$. Thus

$$B^c = F^c \cap N^c = ((F^c \cap A) \cap N^c) \cup ((F^c \cap A^c) \cap N^c) = (F^c \cap A^c) \cup (F^c \cap A \cap N^c).$$

As $F^c \cap A^c \in \mathcal{F}$ and $F^c \cap A \cap N^c \subseteq A$, we have $B^c \in \overline{\mathcal{F}}$. For any $\{B_n\} \in \overline{\mathcal{F}}$, let $B_n = F_n \cup N_n$ and $N_n \subseteq A_n$ be their decompositions, then

$$\bigcup_n B_n = \left(\bigcup_n F_n\right) \cup \left(\bigcup_n N_n\right).$$

As $\bigcup_n N_n \subseteq \bigcup_n A_n$ and $\bigcup_n A_n \in \mathcal{F}$ with $\mu(\bigcup_n A_n) = \sum_n \mu(A_n) = 0$, we have $\bigcup_n N_n \in \mathcal{N}$ and thus $\bigcup_n B_n \in \overline{\mathcal{F}}$.

(2) Define $\overline{\mu}(B) = \mu(F)$ for $B = F \cup N$, $F \in \mathcal{F}$, $N \in \mathcal{N}$. $\overline{\mu}$ is well defined.

We need to verify that if B have two decompositions $B = F_1 \cup N_1$ and $B = F_2 \cup N_2$, then $\mu(F_1) = \mu(F_2)$. Indeed, we have

$$F_1 \subseteq F_1 \cup N_1 = B = F_2 \cup N_2 \subseteq F_2 \cup A_2,$$

where $\mu(A_2) = 0$. Hence $\mu(F_1) \leq \mu(F_2 \cup A_2) \leq \mu(F_2) + \mu(A_2) = \mu(F_2)$. That $\mu(F_2) \leq \mu(F_1)$ follows by exchanging the roles of F_1 and F_2.

(3) $\overline{\mu}$ is a measure on $\overline{\mathcal{F}}$ and agrees with μ on $\mathcal{F}$.

Clearly $\overline{\mu}(\emptyset) = 0$. Let $\{B_n\}$ be a sequence of disjoint sets in $\overline{\mathcal{F}}$ with decompositions $B_n = F_n \cup N_n$. As F_n and N_n are all disjoint, we have

$$\overline{\mu}\left(\bigcup_n B_n\right) = \overline{\mu}\left(\bigcup_n F_n \cup \bigcup_n N_n\right) = \mu\left(\bigcup_n F_n\right) = \sum_n \mu(F_n) = \sum_n \overline{\mu}(B_n).$$

Hence $\overline{\mu}$ is countably additive. For any $F \in \mathcal{F}$, $F = F \cup \emptyset$, so $\overline{\mu}(F) = \mu(F)$.

(4) $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$ is complete.

Take any $B \in \overline{\mathcal{F}}$ with $\overline{\mu}(B) = 0$. Then $B = F \cup N$ for some $F \in \mathcal{F}$ and $N \in \mathcal{N}$. We have $\mu(F) = \overline{\mu}(F) \leq \overline{\mu}(B) = 0$, so F itself is a measure-zero set in $\mathcal{F}$, hence $B \in \mathcal{N}$. ($A \cup F$ contains B for A containing N.) So if $C \subseteq B$, then $C \in \mathcal{N}$, and thus $C \in \overline{\mathcal{F}}$.

□

Remark Another way of constructing the completion is to look at the outer measure μ^* on μ^*-measurable sets $\mathcal{G}$. One can show that this construction coincides with our construction, in particular, $\mathcal{G} = \overline{\mathcal{F}}$. (See Exercise 3.10(c) in Billingsley [1].)

References

[1] P. Billingsley. *Probability and measure.* John Wiley & Sons, 2008.

Stats 310A Session 3

October 12, 2017

1 Comparison of Riemann and Lebesgue integral

It frequently happens that we are required to compute an integral $\int_{(a,b]} f(x)dx$ where f is Lebesgue measurable on $((a,b], \mathcal{G}_{(a,b]}, \lambda)$. How do we compute this integral? Well, we could follow the definition (approximate by simple functions) or use change of variables formula, both still being quite complicated tasks. In practice, however, we often simply compute the Riemannian integral (e.g. by finding the primitive F and computing $F(b) - F(a)$).

We show that any non-negative Riemann integrable function on $(a,b]$ will also be Lebesgue measurable (hence integrable) with coinciding integral values, justifying their relation.

Definition 1. *A function $f : (a,b] \to [0,\infty]$ is* Riemann integrable *with integral $R(f) < \infty$ if for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $|\sum_l f(x_l)\lambda(J_l) - R(f)| \le \varepsilon$ for any $x_l \in J_l$ and $\{J_l\}$ a finite partition of $(a,b]$ into disjoint intervals whose length $\lambda(J_l) \le \delta$.*

Proposition 1 (Proposition 1.3.64, Dembo's Notes)**.** *If $f(x)$ is a non-negative Riemann integrable function on an interval $(a,b]$, then it is also Lebesgue measurable on $(a,b]$ and $\lambda(f) = R(f)$.*

Proof For any $\varepsilon > 0$, there exists some $\delta > 0$, such that for all partition $\{J_l\}$ of size $\le \delta$ and any $x_l \in J_l$

$$R(f) - \varepsilon \le \sum_l f(x_l)\lambda(J_l) \le R(f) + \varepsilon.$$

Define $f_*(J) = \inf\{f(x) : x \in J\}$ and $f^*(J) = \sup\{f(x) : x \in J\}$. Varying x_l in the above bound, we see that

$$R(f) - \varepsilon \le \sum_l f_*(J_l)\lambda(J_l) \le \sum_l f^*(J_l)\lambda(J_l) \le R(f) + \varepsilon.$$

Written differently, if we define for any partition Π

$$\ell(\Pi)(x) = \sum_l f_*(J_l)\mathbf{1}\{x \in J_l\}, \quad u(\Pi)(x) = \sum_l f^*(J_l)\mathbf{1}\{x \in J_l\}.$$

then $\ell(\Pi)$ and $u(\Pi)$ are non-negative simple functions with Lebesgue integrals $\sum_l f_*(J_l)\lambda(J_l)$ and $\sum_l f^*(J_l)\lambda(J_l)$. Consequently, as long as Π has size $\le \delta$, we have $R(f) - \varepsilon \le \lambda(\ell(\Pi)) \le \lambda(u(\Pi)) \le R(f) + \varepsilon$.

Let Π_n be the dyadic partition of $(a,b]$ to 2^n intervals of equal length $(b-a)2^{-n}$. For sufficiently large n, $R(f) - \varepsilon \le \lambda(\ell(\Pi_n)) \le \lambda(u(\Pi_n)) \le R(f) + \varepsilon$. Noting that $u(\Pi_n) \ge u(\Pi_{n+1})$ and so they have a pointwise limit $u(\Pi_n) \downarrow u_\infty$ and similarly $\ell(\Pi_n) \uparrow \ell_\infty$, where u_∞, ℓ_∞ are Lebesgue measurable. By the motonicity of Lebesgue's integral,

$$R(f) - \varepsilon \le \liminf_{n\to\infty} \lambda(\ell(\Pi_n)) \le \lambda(\ell_\infty) \le \lambda(u_\infty) \le \limsup_{n\to\infty} \lambda(u(\Pi_n)) \le R(f) + \varepsilon.$$

Letting $\varepsilon \to 0$ gives $\lambda(\ell_\infty) = \lambda(u_\infty) = R(f)$.

Finally, observe that $\ell_\infty(x) \le f(x) \le u_\infty(x)$, and that

$$\{x : f(x) \ne \ell_\infty(x)\} \subseteq \{x : u_\infty(x) > \ell_\infty(x)\},$$

with the latter a measure-zero set (as $u_\infty \ge \ell_\infty$ and $\int (u_\infty - \ell_\infty)dx = 0$). Hence, by the completeness of the Lebesgue measure, $\{x : f(x) \ne \ell_\infty(x)\}$ is also Lebesgue measurable with measure zero, which implies that f is measurable and $\lambda(f) = \lambda(\ell_\infty)$. □

2 Miscellanous Examples

2.1 Set operations

We have seen some set operations in the the last HW (Exercise 1.2.30, Dembo's Notes). Here we make formal some set operations that will be useful later in the class.

As a motivation, let us think of how we could define the limit of sets (assuming a common superset Ω). Recall that we define the limit of an increasing sequence of sets as $\lim_n A_n = \bigcup_n A_n$ and for a decreasing sequence as $\lim_n A_n = \bigcap_n A_n$. Then, for a general non-monotone sequence, we are going to define the upper and lower limits for the sequence via constructing related monotone sequences. For a sequence of sets $\{A_n\}$, we define

$$\liminf_n A_n = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n = \{\omega : A_n(\omega) \text{ happens for all large } n\},$$
$$\limsup_n A_n = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n = \{\omega : A_n(\omega) \text{ happens infinitely often}\}.$$

It is easy to verify that $\liminf_n A_n \subseteq \limsup_n A_n$, and we say that the limit of A_n exists if $\liminf_n A_n = \limsup_n A_n$.

Let us practice set operations on an example: let $X_1, X_2, \ldots$ be a sequence of R.V.s and X_∞ be an R.V., all defined on some measure space $(\Omega, \mathcal{F})$. Then,

$$\{\omega : X_n(\omega) \to X_\infty(\omega) \text{ as } n \to \infty\} = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{\omega : |X_n(\omega) - X_\infty(\omega)| \le \frac{1}{k}\right\}.$$

As we will see later, this relation is useful for characterizing almost sure convergence, in particular, for showing that $X_n \overset{a.s.}{\to} X_\infty$ implies $X_n \overset{p}{\to} X_\infty$.

To show this, note that $X_n(\omega) \to X_\infty(\omega)$ happens iff for all $k \in \mathbb{N}$, $|X_n(\omega) - X_\infty(\omega)| \le k^{-1}$ for all large n. Hence,

$$\{\omega : X_n(\omega) \to X_\infty(\omega)\} = \bigcap_{k=1}^{\infty} \left\{\omega : |X_n(\omega) - X_\infty(\omega)| \le \frac{1}{k} \text{ for all large } n\right\},$$

which implies the result.

2.2 Generated σ-algebra

The following gives an example in which for an increasing sequence of σ-algebras $\mathcal{F}_n$, $\bigcup_n \mathcal{F}_n$ is a σ-algebra, thereby showing that $\sigma(\bigcup_n \mathcal{F}_n) \supsetneq \bigcup_n \mathcal{F}_n$ in general.

On $\mathbb{R}$, define

$$\mathcal{F}_n = \sigma(\{[a, b) : 2^n a, 2^n b \in \mathbb{Z}\}),$$

i.e. $\mathcal{F}_n$ is generated by intervals whose endpoints are dyadic numbers with no more than n decimal digits. Clearly $\mathcal{F}_n$ is an increasing sequence. We claim that $[0, \frac{1}{3}) \in \sigma(\bigcup_n \mathcal{F}_n) \setminus \bigcup_n \mathcal{F}_n$. Indeed, take x_n to be largest n-digit dyadic number below $\frac{1}{3}$. As dyadic numbers are dense, $x_n \to \frac{1}{3}$. Now, $[0, x_n) \in \mathcal{F}_n$, hence

$$[0, 1/3) = \cup_n [0, x_n) \in \sigma(\cup_n \mathcal{F}_n).$$

However, $[0, \frac{1}{3})$ does not belong to $\mathcal{F}_n$ for all n, as $\frac{1}{3}$ is not dyadic.

References

Stats 310A Session 4

October 18, 2017

1 Uniform integrability

The dominated convergence theorem states that $X_n \overset{a.s.}{\to} X_\infty$ and $|X_n| \leq Y$ for some integrable Y implies that $X_n \overset{L_1}{\to} X_\infty$ and $\mathbb{E}[X_n] \to \mathbb{E}[X_\infty]$. In this section, we explore *uniform integrability*, a useful concept that allows us to relax both conditions assumed in the dominated convergence theorem and still get L_1 convergence. We will follow Section 1.3.4 in Dembo's Notes.

Definition 1 (Uniform integrability)**.** *A collection of R.V.-s $\{X_\alpha, \alpha \in \mathcal{I}\}$ is called* uniformly integrable *(U.I.) if*

$$\lim_{n\to\infty} \sup_{\alpha\in\mathcal{I}} \mathbb{E}[|X_\alpha| \mathbf{1}\{|X_\alpha| > M\}] = 0.$$

Let us show that U.I. is indeed a relaxation of that $|X_\alpha| \leq Y$ for some integrable Y.

Lemma 1.1. *Let Y be integrable and suppose that $|X_\alpha| \leq Y$ for all α, then $\{X_\alpha\}$ is U.I.. In particular, any finite collection of integrable R.V.-s is U.I.. Further, if X_α is U.I. then $\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty$.*

Proof That $\{X_\alpha\}$ is U.I. follows from that

$$\sup_\alpha \mathbb{E}[|X_\alpha| \mathbf{1}\{|X_\alpha| \geq M\}] \leq \mathbb{E}[|Y| \mathbf{1}\{|Y| \geq M\}] \to 0 \text{ as } M \to \infty.$$

If we have a finite collection $\{X_k\}_{k=1}^n$ that are integrable, then $Y = \sum_{k=1}^n |X_k|$ is integrable and dominates X_k. Suppose $\{X_\alpha\}$ is U.I., then for all M we have

$$\sup_\alpha \mathbb{E}[|X_\alpha|] \leq M + \sup_\alpha \mathbb{E}[|X_\alpha| \mathbf{1}\{|X_\alpha| > M\}].$$

The second term goes to zero as $M \to \infty$, so has to be finite for some M. This M yields a finite value on the RHS, so gives a finite upper bound on $\sup_\alpha \mathbb{E}[|X_\alpha|]$. $\square$

To further understand U.I., consider the following example, which shows $\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty$ does not necessarily give U.I..

Example 1: Let X_n be binary R.V.-s with $\mathbb{P}(X_n = 0) = 1 - 1/n$ and $\mathbb{P}(X_n = n) = 1/n$. Then $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = 1$ for all n but X_n is not U.I.: for any M, take $n = \lceil M \rceil$, we have $\mathbb{E}[|X_n| \mathbf{1}\{|X_n| \geq M\}] = \mathbb{E}[|X_n|] = 1$. $\diamond$

We are now ready to state the main convergence theorem.

Theorem 1 (Vitali's convergence theorem)**.** *Suppose $X_n \overset{p}{\to} X_\infty$, then the following are equivalent:*

(a) $\{X_n\}$ is U.I..

(b) $X_n \xrightarrow{L_1} X_\infty$.

(c) X_n *is integrable for all* $n \le \infty$ *and* $\mathbb{E}[|X_n|] \to \mathbb{E}[|X_\infty|]$.

Proof "$(a) \implies (b)$". We first deal with the case that $|X_n| \le M$ for some finite M. For all n and $\varepsilon > 0$, define

$$B_{n,\varepsilon} = \{\omega : |X_n(\omega) - X_\infty(\omega)| > \varepsilon\}.$$

As $X_n \xrightarrow{p} X_\infty$, we have $\mathbb{P}(B_{n,\varepsilon}) \to 0$ as $n \to \infty$ for all ε. In particular, we have $\mathbb{P}(|X_\infty| \ge M+\varepsilon) \le \mathbb{P}(B_{n,\varepsilon})$. Letting $n \to \infty$ gives $|X_\infty| \le M + \varepsilon$ almost surely, which after taking $\varepsilon \to 0$ gives that $|X_\infty| \le M$ almost surely. Hence, $|X_n - X_\infty| \le 2M$, which allows us to bound

$$\mathbb{E}[|X_n - X_\infty|] = \mathbb{E}[|X_n - X_\infty|\mathbf{1}\{B_{n,\varepsilon}\}] + \mathbb{E}[|X_n - X_\infty|\mathbf{1}\{B_{n,\varepsilon}^c\}] \le 2M\mathbb{P}(B_{n,\varepsilon}) + \varepsilon.$$

Taking $n \to \infty$, we conclude that $\limsup_{n\to\infty} \mathbb{E}[|X_n - X_\infty|] \le \varepsilon$, so as ε is arbitrary we get $\mathbb{E}[|X_n - X_\infty|] \to 0$.

We now show the general case where $\{X_n\}$ is U.I. by applying a truncation argument. Define the truncation function

$$\varphi_M(x) = x\mathbf{1}\{|x| \le M\}.$$

As $X_n \xrightarrow{p} X_\infty$ and φ_M is continuous, we have $\varphi_M(X_n) \xrightarrow{p} \varphi_M(X_\infty)$. The R.V.-s $\varphi_M(X_n)$ are bounded in $[-M, M]$, so by the bounded case, we get $\mathbb{E}[|\varphi_M(X_n) - \varphi_M(X_\infty)|] \to 0$.

As $\{X_n\}$ is U.I., we have $\sup_n \mathbb{E}[|X_n|] = c < \infty$ by Lemma 1.1, which gives that

$$c = \sup_n \mathbb{E}[|X_n|] \ge \sup_n \mathbb{E}[|\varphi_M(X_n)|] \ge \lim_{n\to\infty} \mathbb{E}[|\varphi_M(X_n)|] = \mathbb{E}[\varphi_M(X_\infty)].$$

The R.V.-s $|\varphi_M(X_\infty)|$ is an increasing sequence as $M \uparrow \infty$ and converges to $|X_\infty|$. By the monotone convergence theorem, we have

$$\mathbb{E}[|X_\infty|] = \lim_{M\to\infty} \mathbb{E}[|\varphi_M(X_\infty)|] \le c,$$

so X_∞ is integrable.

As $\{X_n\}$ is U.I. and X_∞ is integrable, for any $\varepsilon > 0$, there exists some M such that $\sup_n \mathbb{E}[|X_n|\mathbf{1}\{|X_n| \ge M\}] \le \varepsilon$ and $\mathbb{E}[|X_\infty|\mathbf{1}\{|X_\infty| \ge M\}] \le \varepsilon$. By the triangle inequality, we have

$$\begin{aligned}\mathbb{E}[|X_n - X_\infty|] &\le \mathbb{E}[|X_n - \varphi_M(X_n)|] + \mathbb{E}[|\varphi_M(X_n) - \varphi_m(X_\infty)|] + \mathbb{E}[|\varphi_M(X_\infty) - X_\infty|] \\ &\le 2\varepsilon + \mathbb{E}[|\varphi_M(X_n) - \varphi_M(X_\infty)|].\end{aligned}$$

Letting $n \to \infty$, we get $\limsup_{n\to\infty} \mathbb{E}[|X_n - X_\infty|] \le 2\varepsilon$. Taking $\varepsilon \downarrow 0$ gives that $\mathbb{E}[|X_n - X_\infty|] \to 0$, the desired result.

That $(b) \implies (c)$ is immediate, and we will skip the proof of $(c) \implies (a)$. □

2 Constructing measures from densities

We have seen in the last HW that measures can be defined via densities – indeed, this is one standard way of defining continuous random variables in elementary probability. This section gives a measure-theoretic treatment of such construction and in particular show that it satisfies the composition rule.

Proposition 1. *Fix a measure space* $(\mathbb{S}, \mathcal{F}, \mu)$*. Every* $f \in m\mathcal{F}_+$ *induces a measure* $f\mu$ *on* $(\mathbb{S}, \mathcal{F})$ *via* $f\mu(A) = \mu(f\mathbf{1}_A)$ *for all* $A \in \mathcal{F}$*. These measures satisfy the composition relation* $h(f\mu) = (hf)\mu$ *for all* $f, h \in m\mathcal{F}_+$*. Further,* $h \in L^1(\mathbb{S}, \mathcal{F}, f\mu)$ *if and only if* $fh \in L^1(\mathbb{S}, \mathcal{F}, \mu)$ *and then* $(f\mu)h = \mu(fh)$*.*

Proof We give a proof sketch for the composition rule $h(f\mu) = (hf)\mu$. By definition of the measure operation, this requires showing the identity

$$(f\mu)(h\mathbf{1}_A) = \mu(fh\mathbf{1}_A) \quad \text{for all } A \in \mathcal{F}.$$

We will use the standard machine to show this.

(1) $h = \mathbf{1}_B$ is an indicator ($B \in \mathcal{F}$), then $h\mathbf{1}_A = \mathbf{1}_{A\cap B}$, giving that

$$(f\mu)(h\mathbf{1}_A) = (f\mu)(A \cap B) = \mu(f\mathbf{1}_{A\cap B}) = \mu(fh\mathbf{1}_A).$$

(2) The above identity is linear in the component $\mathbf{1}_B$, so extends to h being simple functions.

(3) For general $h \in m\mathcal{F}_+$, there exists simple functions $h_n \uparrow h$. Part (2) gives that $(f\mu)(h_n\mathbf{1}_A) = \mu(fh_n\mathbf{1}_A)$. Further, $h_n\mathbf{1}_A \uparrow h\mathbf{1}_A$ and $fh_n\mathbf{1}_A \uparrow fh\mathbf{1}_A$ for all $A \in \mathcal{F}$ and $f \in m\mathcal{F}_+$, so by monotone convergence we have that

$$(f\mu)(h\mathbf{1}_A) = \lim_{n\to\infty} (f\mu)(h_n\mathbf{1}_A) = \lim_{n\to\infty} \mu(fh_n\mathbf{1}_A) = \mu(fh\mathbf{1}_A).$$

$\square$

Remark The measure μ is called the *base measure*, the function f the *density*, and $f\mu$ the *induced measure*. Elementary probability theory often uses the Lebesgue measure on $\mathbb{R}$ (or $\mathbb{R}^n$) as the base measure. Given two measures f, g, there exists μ such that $g = f\mu$ iff g is *absolutely continuous* w.r.t. f, in which case μ is called their *Radon-Nikodym derivative*.

Stats 310A Session 5

November 2, 2017

1 Kolmogorov's extension theorem

We state and prove the Kolmogorov's extension theorem when the index set is $T = \{1, 2, 3, \ldots\} = \mathbb{N}$.

Theorem 1 (Theorem 1.4.22, Dembo's Notes). *Suppose we are give probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ that are consistent, that is,*

$$\mu_{n+1}(B_1 \times \cdots \times B_n \times \mathbb{R}) = \mu_n(B_1 \times \cdots \times B_n) \quad \forall B_i \in \mathcal{B},\ i = 1, \ldots, n < \infty. \tag{1}$$

Then, there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_c)$ such that

$$\mathbb{P}(\{\omega : \omega_i \in B_i, i = 1, \ldots, n\}) = \mu_n(B_1 \times \cdots \times B_n) \quad \forall B_i \in \mathcal{B},\ i = 1, \ldots, n < \infty.$$

Remark Kolmogorov's extension theorem builds the foundation on which stochastic processes are defined: namely, for any index set T, to define the distribution of a stochastic process X_T, it suffices to give a consistent collection of joint distributions of $(X_{t_1}, \ldots, X_{t_n})$ on finitely many coordinates. The measure of X_T on $(\mathbb{R}^T, \mathcal{B}_c)$, then, by the extension theorem, is guaranteed to exist and is unique.

The theorem is trivial when $T = \{1, \ldots, n\}$ is finite: just take $\mathbb{P} = \mu_n$. $T = \mathbb{N}$ is the first non-trivial case of the theorem. This case can give us, for example, the probability measure of countably many i.i.d. R.V.-s $(X_1, X_2, \ldots)$.

Proof of Theorem 1 The proof mainly follows that of [1, Chapter 36]. Let $\mathbb{R}_0^{\mathbb{N}}$ be the collection of cylindral sets of the form

$$A = \Big\{x \in \mathbb{R}^{\mathbb{N}} : (x_1, \ldots, x_n) \in H\Big\}, \tag{2}$$

where $n \in \mathbb{N}$ and $H \in \mathcal{B}_{\mathbb{R}^n}$. That is, we consider sets that require the first n coordinates lie in some Borel set $H \subset \mathbb{R}^n$. By definition of the cylindral σ-algebra, we have $\mathcal{B}_c = \sigma(\mathbb{R}_0^{\mathbb{N}})$. On this collection, define the set function

$$\mathbb{P}(A) = \mu_n(H).$$

We are going to use Caratheodory's extension theorem to extend $\mathbb{P}$ to $\mathcal{B}_c$, which we divide into the following steps.

$\mathbb{P}$ is well-defined To show this, we need to verify that if a cylindral set A has two representations of the form (2) then they give coinciding values of $\mathbb{P}(A)$. Consider

$$A = \{x : (x_1, \ldots, x_{n_1}) \in H_1\} = \{x : (x_1, \ldots, x_{n_2}) \in H_2\}$$

for some $n_1 \geq n_2$, then it is easy to see that $H_1 = H_2 \times \mathbb{R}^{n_1-n_2}$. (Check this!) It remains to show that

$$\mu_{n_1}(H_1) = \mu_{n_1}(H_2 \times \mathbb{R}^{n_1-n_2}) = \mu_{n_2}(H_2). \tag{3}$$

Repeating the consistency condition (1) gives that $\mu_{n_1}(B_1 \times \cdots \times B_{n_2} \times \mathbb{R}^{n_1-n_2}) = \mu_{n_2}(B_1 \times \cdots \times B_{n_2})$, and a standard extension argument shows that $\mu_{n_1}(\cdot \times \mathbb{R}^{n_1-n_2}) = \mu_{n_2}(\cdot)$, verifying (3).

$\mathbb{R}_0^{\mathbb{N}}$ is an algebra; $\mathbb{P}$ finitely additive on $\mathbb{R}_0^{\mathbb{N}}$ Clearly $\emptyset \in \mathbb{R}_0^{\mathbb{N}}$. For any cylindral set A, we have $A^c = \{x \in \mathbb{R}^{\mathbb{N}} : (x_1, \ldots, x_n) \in H^c\}$, so $A^c \in \mathbb{R}_0^{\mathbb{N}}$. Let A, B be two cylindral sets:

$$A = \{x : (x_1, \ldots, x_{n_1}) \in H_1\}, \quad B = \{x : (x_1, \ldots, x_{n_2}) \in H_2\}.$$

Without loss of generality, let $n_1 \geq n_2$. We then have

$$A \cup B = \{x : (x_1, \ldots, x_{n_1}) \in H_1 \cup (H_2 \times \mathbb{R}^{n_1-n_2})\} \in \mathbb{R}_0^{\mathbb{N}}. \tag{4}$$

This shows that $\mathbb{R}_0^{\mathbb{N}}$ is an algebra. If A and B are disjoint, then $H_2 \times \mathbb{R}^{n_1-n_2} \cap H_1 = \emptyset$, giving that

$$\mathbb{P}(A \cup B) = \mu_{n_1}(H_1 \cup (H_2 \times \mathbb{R}^{n_1-n_2})) = \mu_{n_1}(H_1) + \mu_1(H_2 \times \mathbb{R}^{n_1-n_2}) = \mathbb{P}(A) + \mathbb{P}(B),$$

so $\mathbb{P}$ is finitely additive.

$\mathbb{P}$ is a probability measure on $\mathbb{R}_0^{\mathbb{N}}$ Clearly $\mathbb{P} \geq 0$ and $\mathbb{P}(\emptyset) = 0$. Let A be a cylindral set, then

$$\mathbb{P}(A^c) = \mu_n(H^c) = 1 - \mu_n(H) = 1 - \mathbb{P}(A).$$

It remains to show countable additivity. As it is finitely additive, it suffices to show that $A_k \in \mathbb{R}_0^{\mathbb{N}}$ with $A_k \downarrow \emptyset$ implies $\mathbb{P}(A_k) \to 0$. (See the Remark in Dembo notes, page 14). As we can always make the defining index non-decreasing, we can let

$$A_k = \{x : (x_1, \ldots, x_{n_k}) \in H_k\}$$

where $n_k \in \mathbb{N}$ is increasing and $H_k \subset \mathbb{R}^{n_k}$.

Suppose $\mathbb{P}(A_k) \not\to 0$, then $\mathbb{P}(A_k) \geq \varepsilon$ holds for all k, for some $\varepsilon > 0$. This means $\mu_{n_k}(H_k) \geq \varepsilon$. Applying [1, Theorem 12.3], there exists compact sets $K_k \subseteq H_k$ such that $\mu_{n_k}(H_k \setminus K_k) \leq \varepsilon/2^{k+1}$. Define

$$B_k = \{x : (x_1, \ldots, x_{n_k}) \in K_k\},$$

then $\mathbb{P}(A_k \setminus B_k) \leq \varepsilon/2^{k+1}$. Define $C_k = \bigcap_{j=1}^k B_j$, then we have $C_k \subset B_k \subset A_k$ and $\mathbb{P}(A_k \setminus C_k) \leq \varepsilon/2$, so $\mathbb{P}(C_k) \geq \varepsilon/2$, and thus C_k is non-empty.

Now, for all k, choose a point $x^{(k)} \in C_k$. As C_k is the intersection of $\{B_j\}_{j \leq k}$, we have $(x_1^{(k)}, \ldots, x_{n_j}^{(k)}) \in K_j$ for all $j \leq k$. In other words, the first n_j indices of $\{x^{(k)}\}_{k \geq j}$ lie in the compact set K_j. Hence, there exists a subsequence k_i such that $(x_1^{(k_i)}, \ldots, x_{n_j}^{(k_i)})$ converges. By the diagonal method, we can find a subsequence k_i such that $(x_1^{(k_i)}, \ldots, x_{n_j}^{(k_i)})$ converges for all j. Let x be the point in $\mathbb{R}^{\mathbb{N}}$ such that $(x_1, \ldots, x_{n_j})$ is the limit of the above sequence (as the limits are consistent, x exists). The closedness of K_j implies that $(x_1, \ldots, x_{n_j}) \in K_j$, so $x \in A_j$. Thus we have found a point $x \in \bigcap_{j=1}^{\infty} A_j$, contradictory to that $A_j \emptyset$. Hence our assumption is wrong so we must have $\mathbb{P}(A_j) \to 0$. $\square$

References

[1] P. Billingsley. *Probability and measure.* John Wiley & Sons, 2008.

Stats 310A Session 6

November 10, 2017

1 The Law of iterated logarithm

In this note we prove the law of iterated logarithm, mainly following [1, Chapter 9]. Let X_i be independent R.V.-s with mean 0 and variance 1. The central limit theorem characterizes the behavior of $S_n = X_1 + \cdots + X_n$ and states that $S_n = O_p(\sqrt{n})$. The law of iterated algorithm refines this result dramatically, precisely characterizing the scalings of the extrema of S_n.

Theorem 1 (Law of iterated logarithm)**.** *We have*

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right) = 1.$$

Equivalently, the theorem states the following: for all $\varepsilon > 0$,

$$\mathbb{P}\left(S_n \geq (1+\varepsilon)\sqrt{2n\log\log n} \text{ i.o.}\right) = 0, \tag{1}$$
$$\mathbb{P}\left(S_n \geq (1-\varepsilon)\sqrt{2n\log\log n} \text{ i.o.}\right) = 1. \tag{2}$$

Hence, showing LIL requires estimating the probability $\mathbb{P}(S_n/\sqrt{n} \geq t)$ very accurately, for t on the order of $\sqrt{\log\log n}$. The following lemma presents such a result.

Lemma 1.1. *Let $a_n \to \infty$ and $a_n/\sqrt{n} \to 0$, then*

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \geq a_n\right) = \exp\left(-\frac{1}{2}a_n^2(1+\xi_n)\right),$$

where $\xi_n \to 0$.

We will also need a variant of Kolmogorov's maximal inequality. Let $M_n = \max_{1\leq k\leq n} S_k$ be the maximum process.

Lemma 1.2. *For $\alpha \geq \sqrt{2}$, we have*

$$\mathbb{P}\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \leq 2\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right).$$

Proof of Theorem 1 We prove the result by looking at a subsequence S_{n_k} where $n_k = \theta^k$ for some carefully chosen $\theta > 1$. We bound the deviation probability carefully and use Borel-Cantelli to show that S_{n_k} exceeds the desired threshold infinitely often with probability zero or one. We then show that S_n has the same behavior as the subsequence.

Proof of (1) Fixing $\varepsilon > 0$, choose θ such that $1 < \theta^2 < 1 + \varepsilon$. Define

$$n_k = \left\lfloor \theta^k \right\rfloor, \quad x_k = \theta\sqrt{2\log\log n_k}.$$

Note that $x_k = (1 + o(1))\theta\sqrt{2\log k}$. Applying Lemmas 1.2, 1.1, we obtain

$$\begin{aligned}
\mathbb{P}\left(\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right) &\leq 2\mathbb{P}\left(\frac{S_{n_k}}{\sqrt{n_k}} \geq x_k - \sqrt{2}\right) \\
&= 2\exp\left(-\frac{1}{2}(x_k - \sqrt{2})^2(1 + \xi_k)\right) \\
&= 2\exp\left(-\frac{1}{2} \cdot 2\theta^2 \log k(1 + o(1))\right) \\
&\leq \frac{2}{k^{\theta^2}},
\end{aligned}$$

the last bound holding for all large k. As $\theta^2 > 1$, the RHS is summable, so by Borel-Cantelli I we have

$$\mathbb{P}\left(\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k \text{ i.o.}\right) = 0.$$

We now argue that $S_n \geq (1 + \varepsilon)\sqrt{2n\log\log n}$ infinitely often will happen with probability zero. Suppose it happens infinitely often, let n be an index where it happens. Let k be such that $n_{k-1} < n \leq n_k$. We then have

$$\begin{aligned}
\frac{M_{n_k}}{x_k\sqrt{n_k}} = \frac{M_{n_k}}{\theta\sqrt{2n_k\log\log n_k}} &\geq \frac{S_n}{\theta\sqrt{2n\log\log n}} \cdot \sqrt{\frac{2n_{k-1}\log\log n_{k-1}}{2n_k\log\log n_k}} \\
&\geq \frac{1+\varepsilon}{\theta} \cdot \sqrt{\frac{2\theta^{k-1} \cdot \log(k-1)}{2\theta^k \log k}}(1 + o(1)) \\
&\geq \frac{1+\varepsilon}{\theta^{3/2}}(1 + o(1)).
\end{aligned}$$

As $1 + \varepsilon > \theta^2 > \theta^{3/2}$, for sufficiently large k, the above quantity will be greater than one. Hence, $M_{n_k}/\sqrt{n_k} \geq x_k$ will happen infinitely often. As this has probability zero, we must have $\mathbb{P}(S_n \geq (1 + \varepsilon)\sqrt{2n\log\log n} \text{ i.o.}) = 0$, thereby showing (1).

Proof of (2) Let θ be an integer such that $3/\sqrt{\theta} < \varepsilon$ and $n_k = \theta^k$. Define

$$a_k = x_k/\sqrt{n_k - n_{k-1}} \quad \text{with } x_k = (1 - \theta^{-1})\sqrt{2n_k\log\log n_k}.$$

As S_n are sums of independent R.V.-s, we can apply Lemma 1.1 to $S_{n_k} - S_{n_{k-1}}$ and get

$$\begin{aligned}
\mathbb{P}\left(S_{n_k} - S_{n_{k-1}} \geq x_k\right) &= \exp\left(-\frac{x_k^2}{2(n_k - n_{k-1})}(1 + \xi_k)\right) \\
&= \exp\left(-\frac{(1-\theta^{-1})^2 2\theta^k \log k}{2(1-\theta^{-1})\theta^k}(1 + o(1))\right) \\
&= \exp\left(-(1 - \theta^{-1})\log k(1 + o(1))\right) \\
&\leq \frac{2}{k^{1-\theta^{-1}}},
\end{aligned}$$

the last bound holding for all large k. As the RHS sums up to infinity and the events are independent, by Borel-Cantelli II we get that

$$\mathbb{P}\left(S_{n_k} - S_{n_{k-1}} \geq x_k \text{ i.o.}\right) = 1.$$

We now argue that the above implies $S_{n_k} > (1-\varepsilon)\sqrt{2n_k \log\log n_k}$ happens infinitely often with probability one, thereby showing the result. Indeed, applying the established result (1) to $-S_{n_k}$ with $\varepsilon = 1$, we get $-S_{n_{k-1}} \leq 2\sqrt{2n_{k-1}\log\log n_{k-1}}$ for all large k. Combined with the above result, we get that with probability one,

$$S_{n_k} \geq x_k - 2\sqrt{2n_{k-1}\log\log n_{k-1}} \geq x_k - \frac{2}{\sqrt{\theta}}\sqrt{2n_k\log\log n_k} = \left(1 - \frac{1}{\theta} - \frac{2}{\sqrt{\theta}}\right)\sqrt{2n_k\log\log n_k}$$
$$\geq \left(1 - \frac{3}{\sqrt{\theta}}\right)\sqrt{2n_k\log\log n_k} \geq (1-\varepsilon)\sqrt{2n_k\log\log n_k}.$$

□

For completeness, we also provide the proof of Lemma 1.2.

Proof of Lemma 1.2 Suppose $M_n/\sqrt{n} \geq \alpha$, then either $S_n/\sqrt{n} \geq \alpha - \sqrt{2}$, or $S_n/\sqrt{n} < \alpha - \sqrt{2}$ and one of the following happens: $M_{j-1} < \alpha\sqrt{n}$ but $M_j \geq \alpha\sqrt{n}$. Defining $A_j = \{M_{j-1} < \alpha\sqrt{n} \leq M_j\}$, then

$$\mathbb{P}\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \leq \mathbb{P}\left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right) + \sum_{j=1}^{n-1}\mathbb{P}\left(A_j \cap \left\{\frac{S_n}{\sqrt{n}} \leq \alpha - \sqrt{2}\right\}\right).$$

On each of the event $A_j \cap \{\cdots\}$, we have $S_j \geq \alpha\sqrt{n}$ and $S_n \leq (\alpha - \sqrt{2})\sqrt{n}$, which implies $(S_n - S_j)/\sqrt{n} \leq -\sqrt{2}$. This event is independent of A_j, and $S_n - S_j$ has variance $n - j$, so we get

$$\mathbb{P}\left(A_j \cap \left\{\frac{S_n}{\sqrt{n}} \leq \alpha - \sqrt{2}\right\}\right) \leq \mathbb{P}\left(A_j \cap \left\{\frac{S_n - S_j}{\sqrt{n}} \leq -\sqrt{2}\right\}\right)$$
$$= \mathbb{P}(A_j)\mathbb{P}\left(\frac{S_n - S_j}{\sqrt{n}} \leq -\sqrt{2}\right) \leq \frac{n-j}{2n}\mathbb{P}(A_j).$$

Plugging into the preceding bound gives

$$\mathbb{P}\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \leq \mathbb{P}\left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right) + \sum_{j=1}^{n-1}\frac{n-j}{2n}\mathbb{P}(A_j) \leq \mathbb{P}\left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right) + \frac{1}{2}\sum_{j=1}^{n-1}\mathbb{P}(A_j).$$

As A_j are disjoint and $\bigcup A_j$ implies $\{M_n/\sqrt{n} \geq \alpha\}$, we get

$$\mathbb{P}\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \leq \mathbb{P}\left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right) + \frac{1}{2}\mathbb{P}\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right),$$

from which the result follows. □

References

[1] P. Billingsley. *Probability and measure.* John Wiley & Sons, 2008.

Stats 310A Session 7

November 16, 2017

1 Characteristic functions

The characteristic function of a real-valued R.V. X is defined as

$$\Phi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} dP_X(x) = \mathbb{E}[e^{i\theta X}] = \mathbb{E}[\cos(\theta X)] + i\mathbb{E}[\sin(\theta X)].$$

As sines and cosines are bounded, the above expectations exist and are finite, so $\Phi_X(\theta)$ is well defined for all $\theta \in \mathbb{R}$. The following result summarizes some basic properties of the characteristic function.

Proposition 1. *We have*

(a) $\Phi_X(0) = 1$.

(b) $\Phi_X(-\theta) = \overline{\Phi_X(\theta)}$.

(c) $|\Phi_X(\theta)| \leq 1$.

(d) $\theta \mapsto \Phi_X(\theta)$ *is a uniformly continuous function on* $\mathbb{R}$.

(e) $\Phi_{aX+b}(\theta) = e^{ib\theta}\Phi_X(a\theta)$.

As characteristic functions offer a way to represent a distribution on $\mathbb{R}$, one naturally wonders if such a representation is one-to-one, i.e. does $\Phi_X(\cdot)$ uniquely determine. the law of X? The answer is yes, which is stated in the following result.

Theorem 1 (Levy's inversion formula, Thm 3.3.12 in Dembo's Notes)**.** *Let X have distribution function F_X and characteristic function Φ_X. For any real numbers $a < b$ and θ, let*

$$\psi_{a,b}(\theta) = \frac{1}{2\pi}\int_a^b e^{-i\theta u} du = \frac{e^{-i\theta a} - e^{-i\theta b}}{i2\pi\theta}.$$

Then,

$$\lim_{T\uparrow\infty} \int_{-T}^{T} \psi_{a,b}(\theta)\Phi_X(\theta)d\theta = \frac{1}{2}[F_X(b) + F_X(b-)] - \frac{1}{2}[F_X(a) + F_X(a-)]. \tag{1}$$

Furthermore, if $\int_R |\Phi_X(\theta)|d\theta < \infty$, then X has the bounded continuous probability density function

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\theta x}\Phi_X(\theta)d\theta. \tag{2}$$

x **Proof**

Proof of (1) The proof follows by carefully computing and swapping the integrals in the inversion formula. Let $J_T(a,b) = \int_{-T}^{T} \psi_{a,b}(\theta)\Phi_X(\theta)d\theta$, we aim to compute the limit of $J_T(a,b)$ as $T \to \infty$. For this end, we define

$$h_{a,b}(x,\theta) = \psi_{a,b}(\theta)e^{i\theta x} = \frac{e^{i\theta(x-a)-e^{i\theta(x-b)}}}{i2\pi\theta}.$$

We have $|h_{a,b}(x,\theta)| = |\psi_{a,b}(\theta)| \le \frac{b-a}{2\pi}$. So on the space $\mathbb{R} \times [-T,T]$ with the product measure of P_X and the Lebsgue measure on $[-T,T]$, the bounded function $h_{a,b}(x,\theta)$ is integrable. So we can apply Fubini theorem to get

$$J_T(a,b) = \int_{-T}^{T} \psi_{a,b}(\theta)\Phi_X(\theta)d\theta = \int_{-T}^{T} \psi_{a,b}(\theta)\left[\int_R e^{i\theta x}dP_X(x)\right]d\theta = \int_{\mathbb{R}\times[-T,T]} h_{a,b}(x,\theta)dP_X(x)d\theta$$
$$= \int_{\mathbb{R}}\left[\int_{-T}^{T} h_{a,b}(x,\theta)d\theta\right]dP_X(x) = \int_{\mathbb{R}}\left[R(x-a,T) - R(x-b,T)\right]dP_X(x),$$

where

$$R(u,T) = \int_{-T}^{T}\frac{e^{i\theta u}}{i2\pi\theta}d\theta = \int_{-T}^{T}\frac{\cos(\theta u) + i\sin(\theta u)}{i2\pi\theta} = \int_0^T \frac{\sin(\theta u)}{\pi\theta}d\theta = \frac{\operatorname{sign}(u)}{\pi}\int_0^{|u|T}\frac{\sin\theta}{\theta}d\theta = \frac{\operatorname{sign}(u)}{\pi}S(|u|T),$$

and $S(t) = \int_0^t \frac{\sin\theta}{\theta}d\theta$. Applying the fact that $\lim_{t\to\infty} S(t) = \pi/2$, we can deduce that

$$\lim_{T\to\infty} R(x-a,T) - R(x-b,T) = g_{a,b}(x) := \begin{cases} 0, & x < a \text{ or } x > b \\ \frac{1}{2}, & x = a \text{ or } x = b \\ 1, & a < x < b. \end{cases}$$

Further, the quantities $S(t)$ are uniformly bounded: $\sup_{t\in\mathbb{R}}|S(t)| \le C < \infty$. By bounded convergence, we get that

$$\lim_{T\to\infty} J_T(a,b) = \lim_{T\to\infty}\int_{\mathbb{R}}[R(x-a,T) - R(x-b,T)]dP_X(x) = \int_{\mathbb{R}}\lim_{T\to\infty}[R(x-a,T) - R(x-b,T)]dP_X(x)$$
$$= \int_{\mathbb{R}} g_{a,b}(x)dP_X(x) = \frac{1}{2}P_X(\{a\}) + P_X((a,b)) + \frac{1}{2}P_X(\{b\})$$
$$= \frac{1}{2}(F_X(a) - F_X(a-) + F_X(b) - F_X(b-)) + F_X(b-) - F_X(a)$$
$$= \frac{1}{2}(F_X(b) + F_X(b-)) - \frac{1}{2}(F_X(a) + F_X(a-)).$$

Proof of the density formula As $\int_{\mathbb{R}}|\Phi_X(\theta)|d\theta < \infty$, the integrand $e^{-i\theta x}\Phi_X(\theta)$ is upper bounded by $|\Phi_X(\theta)$ and so also integrable, therefore $f_X(x)$ is well-defined and finite valued. Taking any x and a point $x+h$ close to x, we have

$$\limsup_{h\to 0}|f_X(x+h) - f_X(x)| \le \limsup_{h\to 0}\frac{1}{2\pi}\int_{\mathbb{R}}|e^{-i\theta h} - 1||\Phi_X(\theta)|d\theta = 0,$$

equality following from the dominated convergence theorem. This shows that $f_X(x)$ is continuous in x. Now, as $\psi_{a,b}(\theta)$ is bounded and $\Phi_X(\theta)$ is integrable, applying dominated convergence to

$\psi_{a,b}(\theta)\Phi_X(\theta)\mathbf{1}\{|\theta| \le T\}$ gives that

$$\lim_{T\to\infty} J_T(a,b) = J_\infty(a,b) = \int_{\mathbb{R}} \psi_{a,b}(\theta)\Phi_X(\theta)d\theta = \frac{1}{2\pi}\int_{\mathbb{R}} \left[\int_a^b e^{-i\theta u}du\right]\Phi_X(\theta)d\theta$$
$$= \int_a^b \int_{\mathbb{R}} \left[\frac{1}{2\pi}e^{-i\theta u}\Phi_X(\theta)d\theta\right]du$$
$$= \int_a^b f_X(u)du.$$

In particular, this shows that $J_\infty(a,b)$ is continuous in a,b. On the other hand, the result (1) gives

$$J_\infty(a,b) = \frac{1}{2}(F_X(b) + F_X(b-)) - \frac{1}{2}(F_X(a) + F_X(a-)),$$

therefore the RHS has to be continuous in a,b. This implies F_X is itself continuous (check this!), and thus $F_X(a) = F_X(a-)$,$F_X(b) = F_X(b-)$ and so

$$\int_a^b f_X(u)du = J_\infty(a,b) = F_X(b) - F_X(a).$$

Hence f_X is the density of X. $\square$

References

Stats 310A Session 8

December 5, 2017

In this session, we will go through some practice problems. These problems fall in the scope of Stats 310A and involves a lot of what we learned comprehensively.

Problem 1 Suppose that you have n standard dice, and roll them all. Remove all those that turn up sixes. Roll the remaning pile again, and repeat the process. Let M_n be the number of rolls needed to remove all the dice.

(a) Produce a sequence of constants a_n such that $M_n/a_n \to 1$ in probability as $n \to \infty$.

Hint: Try to represent M_n in terms of some simpler random variables.

(b) Show that there does not exist sequences b_n and c_n such that $(M_n - b_n)/c_n$ converges in distribution to a non-degenerate limit.

Solution

(a) Let X_i be the number of rolls that dice i required to hit six. Then X_i are i.i.d. Geometrically distributed with parameter $1/6$, and we have $M_n = \max\{X_1, \ldots, X_n\}$. For any integer t, we have

$$\mathbb{P}(M_n \le t) = \mathbb{P}(X_1 \le t)^n = \left(1 - (5/6)^t\right)^n.$$

Intuitively, the problem of choosing a_n is to find a right scaling $t = a_n$ such that the above probability converges to a constant in $(0,1)$. We now show that the sequence $a_n = \log n / \log(6/5)$ satisfies $M_n/a_n \xrightarrow{p} 1$. To this end, it suffices to show that $\mathbb{P}(M_n \le (1+\varepsilon)a_n) \to 1$ and $\mathbb{P}(M_n \le (1-\varepsilon)a_n) \to 0$. For the former, we have

$$\begin{aligned}\mathbb{P}(M_n \le (1+\varepsilon)a_n) &= \mathbb{P}(M_n \le [(1+\varepsilon)\log n/\log(6/5)]) \\ &= \left(1 - (5/6)^{[(1+\varepsilon)\log n/\log(6/5)]}\right)^n \asymp \left(1 - n^{-(1+\varepsilon)}\right)^n \to 1.\end{aligned}$$

Similarly, $\mathbb{P}(M_n \le (1-\varepsilon)a_n) \to 0$.

(b) Suppose there exists b_n, c_n such that $(M_n - b_n)/c_n$ converges to some limiting distribution with c.d.f. F. As F is non-dgenerate, we can choose a continuity point $t \in \mathbb{R}$ such that $F(t) \in (0,1)$. So we have

$$\mathbb{P}\left(\frac{M_n - b_n}{c_n} \le t\right) = \mathbb{P}\left(M_n \le b_n + c_n t\right) = \left(1 - (5/6)^{[b_n + c_n t]}\right)^n \to F(t).$$

In particular, this implies that $[b_n + c_n t] \to \infty$, so we further have (by taking log and using the property of e)

$$\log n - [b_n + c_n t]\log(6/5) \to -\log F(t) \in (0, \infty).$$

Dividing by $\log(6/5)$, we see the sequence $\log n/\log(6/5) - [b_n + c_n t]$ is converging to some limit in $(0, \infty)$. In particular, its fractional part has also converge to some limit in $[0, 1)$ (regarding $1 = 0$ as the fractional part). As $[b_n + c_n t]$ is an integer, the fractional part of $\log n/\log(6/5)$ converges to the same limit, which is impossible as it is actually looping in the interval in $(0, 1)$, as $n \to \infty$.

$\square$

Problem 2 Let $\mathcal{X}$ be a set, $\mathcal{B}$ be a countably generated σ-algebra of subsets of $\mathcal{X}$. Let $\mathcal{P}(\mathcal{X}, \mathcal{B})$ be the set of all probability measures on $(\mathcal{X}, \mathcal{B})$. Make $\mathcal{P}(\mathcal{X}, \mathcal{B})$ into a measurable space by declaring that the map $P \mapsto P(A)$ is Borel measurable for each $A \in \mathcal{B}$. Call the associated σ-algebra $\mathcal{B}^*$.

(a) Show that $\mathcal{B}^*$ is countably generated.

(b) For $\mu \in \mathcal{P}(\mathcal{X}, \mathcal{B})$, show that $\{\mu\} \in \mathcal{B}^*$.

(c) For $\mu, \nu \in \mathcal{P}(\mathcal{X}, \mathcal{B})$, let

$$\|\mu - \nu\| = \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|.$$

Show that the map $(\mu, \nu) \mapsto \|\mu - \nu\|$ is $\mathcal{B}^* \times \mathcal{B}^*$ measurable.

Solution

(a) We have by the definition of $\mathcal{B}^*$ that

$$\mathcal{B}^* = \sigma\left(\{\{P \in \mathcal{P}(\mathcal{X}, \mathcal{B}) : P(A) \leq p\} : A \in \mathcal{B}, p \in [0, 1]\}\right).$$

As $\mathcal{B}$ is countably generated, there exists some countable $\mathcal{B}_0$ such that $\mathcal{B} = \sigma(\mathcal{B}_0)$. Without loss of generality, we can let $\mathcal{B}_0$ be an algebra (if not, consider the smallest algebra containing $\mathcal{B}_0$: this also generates $\sigma(\mathcal{B}_0)$ and is a countable set, see Exercise 1.1.29(b) in Dembo's Notes).

We now claim that

$$\mathcal{B}^* = \sigma\left(\{\{P \in \mathcal{P}(\mathcal{X}, \mathcal{B}) : P(A) \leq p\} : A \in \mathcal{B}_0, p \in [0, 1]\}\right),$$

i.e. the smallest σ-algebra that makes $P \mapsto P(A)$ measurable for all $A \in \mathcal{B}_0$ will make $P \mapsto P(A)$ measurable for al $lA \in \mathcal{B}$. Indeed, by the uniqueness of the Caratheodory extension, we have that $P(A) = P^*(A)$ for all $A \in \mathcal{B}$, therefore

$$P(A) = P^*(A) = \inf_{\{B_j\}_1^k \subset \mathcal{B}_0 \text{ covers } A} \sum_{j=1}^{k} P(B_j).$$

Fixing any $A \in \mathcal{B}$, the collection of all finite $\mathcal{B}_0$- covers of $\mathcal{A}$ is a fixed countable set. Further, each $P \mapsto P(B_j)$ is a measuable function of P, so $P(A)$ being the infimum of countably many measurable function is also measurable.

Finally, we can further reduce the generating set by considering only rational values of p: for any real p, $\{P : P(A) < p\}$ can be written as $\bigcup_{q \leq p} \{P : P(A) \leq q\}$. Hence, we have

$$\mathcal{B}^* = \sigma\left(\{\{P \in \mathcal{P}(\mathcal{X}, \mathcal{B}) : P(A) \leq p\} : A \in \mathcal{B}_0, p \in [0, 1] \cap \mathbb{Q}\}\right),$$

showing that $\mathcal{B}^*$ is countably generated.

(b) Given any $\mu \in \mathcal{P}(\mathcal{X}, \mathcal{B})$, we clearly have

$$\{\mu\} \subseteq \{P : P(A) = \mu(A), \text{ for all } A \in \mathcal{B}_0\} = \bigcap_{A \in \mathcal{B}_0} \{P : P(A) = \mu(A)\}.$$

Our goal is to show the converse direction, thereby showing that $\{\mu\}$ is the intersection of countably many generating sets and thus $\{\mu\} \in \mathcal{B}^*$. This is to say that any two measures that coincide on the generating set $\mathcal{B}_0$ has to coincide on $\mathcal{B}$, which is guaranteed by the uniqueness of the Caratheodory extension.

(c) It suffices to show that for any $t \in \mathbb{R}$,

$$\{(\mu, \nu) : \|\mu - \nu\| \leq t\} = \bigcap_{A \in \mathcal{B}} \{(\mu, \nu) : |\mu(A) - \nu(A)| \leq t\}$$

is a measurable subset of $\mathcal{B}^* \times \mathcal{B}^*$. Note that each set on the RHS is $\mathcal{B}^* \times \mathcal{B}^*$-measuable as the function $(\mu, \nu) \to |\mu(A) - \nu(A)|$ is measurable for all $\mathcal{A}$. Again, the problem is with the uncountable collection, but we claim that it is equal to only the intersection over $A \in \mathcal{B}_0$. In other words, if two measures satisfy $|\mu(A) - \nu(A)| \leq t$ for all $A \in \mathcal{B}_0$, then they satisfy this for all $A \in \mathcal{B}$.

Suppose not, then there is some $A \in \mathcal{B}$ such that $|\mu(A) - \nu(A)| \geq t + \varepsilon$ for some $\varepsilon > 0$. Now, looking at the outer measure, we can find two finite $\mathcal{B}_0$-covers of A such that their summed probabilities are bounded by $\mu(A) + \varepsilon/2$ and $\nu(A) + \varepsilon/2$, respectively. This further gives us (by taking the union) two sets $B, C \in \mathcal{B}_0$ covering A such that $\mu(A) \leq \mu(B) \leq \mu(A) + \varepsilon/2$ and $\nu(A) \leq \nu(C) \leq \nu(A) + \varepsilon/2$. Taking $D = B \cap C$, we have $\mu(A) \leq \mu(D) \leq \mu(A) + \varepsilon/2$ and $\nu(A) \leq \nu(D) \leq \nu(A) + \varepsilon/2$. This means that

$$\mu(D) - \nu(D) \geq \mu(A) - \nu(A) - \varepsilon/2, \quad \nu(D) - \mu(D) \geq \nu(A) - \mu(A) - \varepsilon/2.$$

So we have $|\mu(D) - \nu(D)| \geq |\mu(A) - \nu(A)| - \varepsilon/2 \geq t + \varepsilon/2$, contradicting the fact that $|\mu(A) - \nu(A)| \leq t$ for all $A \in \mathcal{B}_0$. This shows that the intersection can indeed be reduced to a countable intersection, which makes the set $\{(\mu, \nu) : \|\mu - \nu\| \leq t\}$ measurable.

□

Problem 3 Let $X_1, \ldots, X_n$ be independent real valued random variables with a symmetric density $f(x)$ (i.e. $f(-x) = f(x)$). Suppose that there are some $\varepsilon, \delta > 0$ such that $f(x) > \varepsilon$ when $|x| < \delta$. Define the harmonic mean

$$H_n = \frac{n}{\frac{1}{X_1} + \cdots + \frac{1}{X_n}}.$$

Prove that H_n converges in distribution to a Cauchy random variable as $n \to \infty$.
Remark We prove a version assuming the following additional conditions: the density $f(x)$ is bounded ($\sup_{x \in \mathbb{R}} f(x) = B < \infty$) and continuous at 0.

Solution As the Cauchy distribution is invariant to reciprocals (i.e. X is Cauchy then X^{-1} is also Cauchy), it suffices to show that $H_n^{-1} = \sum_{i=1}^n X_i^{-1}/n$ converges to Cauchy. For this end, we compute its characteristic function

$$\mathbb{E}\left[e^{\frac{X_1^{-1}+\cdots+X_n^{-1}}{n}}\right] = \left(\mathbb{E}[e^{itX_1^{-1}/n}]\right)^n = \left(\int_{-\infty}^{\infty} e^{\frac{it}{nx}} f(x)dx\right)^n = \left(1 + \int_{-\infty}^{\infty} (e^{\frac{it}{nx}} - 1) f(x)dx\right)^n$$

$$= \left(1 + 2\int_0^\infty \left(\cos\frac{t}{nx} - 1\right) f(x)dx\right)^n$$

$$= \left(1 + 2\int_0^\infty \left(\cos\frac{|t|}{nx} - 1\right) f(x)dx\right)^n$$

$$= \left(1 + \frac{2|t|}{n}\int_0^\infty \left(\cos\frac{1}{y} - 1\right) f\left(\frac{|t|y}{n}\right) dy\right)^n.$$

We now show that

$$\int_0^\infty \left(\cos\frac{1}{y} - 1\right) f\left(\frac{|t|y}{n}\right) dy \to \int_0^\infty \left(\cos\frac{1}{y} - 1\right) f(0)dy = -c,$$

where $c > 0$, thereby establishing that the characteristic function converges to $e^{-c|t|}$. We have for any $M > 0$ that

$$\left|\int_0^\infty \left(\cos\frac{1}{y} - 1\right) f\left(\frac{|t|y}{n}\right) dy - \int_0^\infty \left(\cos\frac{1}{y} - 1\right) f(0)dy\right|$$

$$\leq \left|\int_0^M \left(\cos\frac{1}{y} - 1\right)\left(f\left(\frac{|t|y}{n}\right) - f(0)\right) dy\right| + \left|\int_M^\infty \left(\cos\frac{1}{y} - 1\right) f\left(\frac{|t|y}{n}\right) dy\right| + \left|\int_M^\infty \left(\cos\frac{1}{y} - 1\right) f(0)dy\right|.$$

The first term is the integral of a function bounded by $4B$ on a finite interval, therefore by the dominated convergence theorem, as $n \to \infty$, $f(|t|y/n) \to f(0)$, so the integral also converges to 0. For the second and third term, note that $|\cos\alpha - 1| \leq \alpha^2$ for small enough α, therefore we have

$$\left|\int_M^\infty \left(\cos\frac{1}{y} - 1\right) f\left(\frac{|t|y}{n}\right) dy\right| \leq \int_M^\infty \frac{1}{y^2} B dy = \frac{B}{M},$$

and the same bound for the third term. Putting together and taking lim sup, we get

$$\limsup_{n\to\infty}\left|\int_0^\infty \left(\cos\frac{1}{y} - 1\right) f\left(\frac{|t|y}{n}\right) dy - \int_0^\infty \left(\cos\frac{1}{y} - 1\right) f(0)dy\right| \leq \frac{2B}{M}.$$

Taking $M \to \infty$, we get that the LHS equals zero, thereby showing the desired convergence.

Finally, as $f(0) > \varepsilon$ and $\cos(1/y) - 1 \leq 0$ and is not always equal to zero, we have $f(0)\int_0^\infty(\cos(1/y) - 1) < 0$, showing our claim. □