

# Probability Theory and Example Solution Manual

May 25, 2020

# Contents

<b>1</b>	<b>Measure Theory</b>	<b>1</b>
1.1	Probability Spaces . . . . .	1
1.2	Distributions . . . . .	2
1.3	Random Variables . . . . .	3
1.4	Integration . . . . .	5
1.5	Properties of the integral . . . . .	6
1.6	Expected Value . . . . .	9
1.7	Product Measures, Fubini's Theorem . . . . .	13
<b>2</b>	<b>Law of Large Numbers</b>	<b>16</b>
2.1	Independence . . . . .	16
2.2	Weak Laws of Large Numbers . . . . .	20
2.3	Borel-Cantelli Lemmas . . . . .	23
2.4	Strong Law of Large Numbers . . . . .	26
2.5	Convergence of Random Series* . . . . .	27

# Chapter 1

## Measure Theory

### 1.1 Probability Spaces

**1.1.1.** Let  $\Omega = \mathbb{R}$ ,  $\mathcal{F} =$  all subsets so that  $A$  or  $A^c$  is countable,  $P(A) = 0$  in the first case and  $= 1$  in the second. Show that  $(\Omega, \mathcal{F}, P)$  is a probability space.

*sol.* i)  $\mathcal{F}$  is a  $\sigma$ -algebra on  $\mathbb{R}$ .

$\emptyset \in \mathcal{F}$  since  $\emptyset$  is countable.

By definition,  $\mathcal{F}$  is closed under complementations.

Countable union of countable sets is countable. Union of countable sets and uncountable sets is uncountable. Thus,  $\mathcal{F}$  is closed under countable union.

ii)  $P$  is a probability measure.

$P(\emptyset) = 0$  since  $\emptyset$  is countable. By definition, for any set  $A$ ,  $P(A) \geq 0$ .

Countable union of countable sets is countable. Union of countable sets and uncountable sets is uncountable. Thus,  $P$  has the countable additivity property.

If  $A$  is countable, then  $A^c$  is uncountable since  $\Omega$  is uncountable. Thus,

$$P(\Omega) = P(A) + P(A^c) = 1$$

By (i) and (ii),  $(\Omega, \mathcal{F}, P)$  is a probability space. □

**1.1.2.** Recall the definition of  $\mathcal{S}_d$  from Example 1.1.5. Show that  $\sigma(\mathcal{S}_d) = \mathcal{R}^d$ , the Borel subsets of  $\mathbb{R}^d$

*sol.* i)  $d = 1$ . In  $\mathbb{R}$ , any open set can be represented by countable union of open intervals. Thus, we need to show that any open interval can be represented by elements of  $\mathcal{S}$ . Let  $-\infty < a < b < \infty$ .

$$(a, b) = (-\infty, a]^c \cup (\cup_i (-\infty, b - 1/n])$$

If  $a = -\infty$ , then  $(a, b)$  can be represented by the second term. If  $b = \infty$ , then  $(a, b)$  can be represented by the first term. Thus,  $\sigma(\mathcal{S}) = \sigma(\mathcal{R}) = \mathcal{R}$ .

ii)  $d \geq 2$ .  $\mathcal{S}_d$  is a finite cartesian product of  $\mathcal{S}$ . Similarly,  $\sigma(\mathcal{S}_d) = \sigma(\mathcal{R}_d) = \mathcal{R}_d$ . □

**1.1.3.** A  $\sigma$ -field  $F$  is said to be countably generated if there is a countable collection  $C \subset F$  so that  $\sigma(C) = F$ . Show that  $\mathcal{R}^d$  is countably generated.

*sol.* Let  $C$  be the collection that contains all sets of the form

$$[q_1, \infty) \times \cdots \times [q_d, \infty), (q_1, \dots, q_d) \in \mathbb{Q}^d$$

Then,  $C$  is countable, since it is finite union of countable sets. And  $\sigma(C) = \mathcal{R}^d$  as presented by previous exercise. □

**1.1.4.** (i) Show that if  $F_1 \subset F_2 \subset \dots$  are  $\sigma$ -algebras, then  $\cup_i F_i$  is an algebra. (ii) Give an example to show that  $\cup_i F_i$  need not be a  $\sigma$ -algebra.

*sol.*

(i) By definition,  $\cup_i F_i$  is not empty. Choose  $x \in \cup_i F_i$ . Then, there exists  $F_x$  such that  $x \in F_x$ . Therefore,  $x^c \in F_x \subset \cup_i F_i$ . Thus,  $\cup_i F_i$  is closed under complementations.

Choose,  $x \in F_i, y \in F_j$  and suppose  $i \leq j$ . Then  $x \in F_j$ . Thus,  $x \cup y \in F_j \Rightarrow x \cup y \in \cup_i F_i$ . Thus,  $\cup_i F_i$  is closed under union.

(ii) Let  $F_i = \sigma(\{\{1\}, \{2\}, \dots, \{n\}\})$ . Let  $A = \{\{n\} : n = 3k | k = 1, 2, 3, \dots\}$ . Then for all  $i$ ,  $A \notin F_i$ . Thus,  $A \notin \cup_i F_i$ . However,  $A$  can be represented by countable union. Therefore,  $\cup_i F_i$  is not a  $\sigma$ -algebra.

□

**1.1.5.** A set  $A \subset \{1, 2, \dots\}$  is said to have asymptotic density  $\theta$  if

$$\lim_{n \rightarrow \infty} |A \cap \{1, 2, \dots, n\}|/n = \theta$$

Let  $\mathcal{A}$  be the collection of sets for which the asymptotic density exists. Is  $\mathcal{A}$  a  $\sigma$ -algebra? an algebra?

*sol.* Let  $A$  be the set of even numbers. Next, we construct a set  $B$  in the following way: we begin with  $\{2, 3\}$  and starting with  $k = 2$ , take all even numbers  $2^k < n \leq (3/2) \times 2^k$ , and all odd numbers  $(3/2) \times 2^k < n \leq 2^{k+1}$ . Then, the asymptotic density of  $B$  is 0.5. However, the asymptotic density  $A \cap B$  does not exist.

When  $n = (3/2) \times 2^k$ , then the density is  $1/3$ . When  $n = 2^{k+1}$ , then the density is  $1/4$ .

Thus,  $\mathcal{A}$  is not closed under intersection.  $\mathcal{A}$  is neither  $\sigma$ -algebra nor algebra

□

## 1.2 Distributions

**1.2.1.** Suppose  $X$  and  $Y$  are random variables on  $(\Omega, \mathcal{F}, P)$  and let  $A \in \mathcal{F}$ . Show that if we let  $Z(\omega) = X(\omega)$  for  $\omega \in A$  and  $Z(\omega) = Y(\omega)$  for  $\omega \in A^c$ , then  $Z$  is a random variable.

*sol.* Let the Borel set  $B$  which satisfies that

$$\text{if } \omega \in A, \text{ then } Z(\omega) \in B$$

For arbitrary Borel set  $S$ ,  $S = (S \cap B) \cup (S \cap B^c)$ . Since  $S$  is Borel set,  $\{\omega : X(\omega) \in S\}, \{\omega : Y(\omega) \in S\} \in \mathcal{F}$ .

$$\begin{aligned} \{\omega : Z(\omega) \in S\} &= \{\omega : Z(\omega) \in (S \cap B)\} \cup \{\omega : Z(\omega) \in (S \cap B^c)\} \\ &= \{\omega : X(\omega) \in (S \cap B)\} \cup \{\omega : Y(\omega) \in (S \cap B^c)\} \end{aligned}$$

Since  $\mathcal{F}$  is closed on set operation,  $\{\omega : Z(\omega) \in S\} \in \mathcal{F}$ .

□

**1.2.2.** Let  $\chi$  have the standard normal distribution. Use Theorem 1.2.6 to get upper and lower bounds on  $P(\chi \geq 4)$ .

*sol.*

$$\begin{aligned} P(\chi \geq 4) &= (2\pi)^{-1} \int_4^\infty \exp(-y^2/2) dy \leq (8\pi)^{-1} \exp(-8) \\ &= (2\pi)^{-1} \int_4^\infty \exp(-y^2/2) dy \geq (15/128\pi) \exp(-8) \end{aligned}$$

□

**1.2.3.** Show that a distribution function has at most countably many discontinuities.

*sol.* Let  $D$  be the set of discontinuity points. Choose  $x, y \in D$ . Then we can choose rational number  $q_x \in (F(x-), F(x+))$ . Since  $F$  is increasing, if  $x \neq y$ , then  $q_x \neq q_y$ . Thus  $x \rightarrow q_x$  is one-to-one function. Since  $\mathbb{Q}$  is countable,  $D$  is at most countable.

□

**1.2.4.** Show that if  $F(x) = P(X \leq x)$  is continuous then  $Y = F(X)$  has a uniform distribution on  $(0, 1)$ , that is, if  $y \in [0, 1]$ ,  $P(Y \leq y) = y$ .

sol.

$$\begin{aligned}
\{\omega | Y(\omega) \leq y\} &= \{\omega | F(X(\omega)) \leq y\} \\
&= \{\omega | X(\omega) \leq k\} \quad k = \inf\{x | F(x) \geq y\} \\
P(\{\omega | Y(\omega) \leq y\}) &= P(\{\omega | X(\omega) \leq k\}) \\
&= P(X \leq k) = y
\end{aligned}$$

□

**1.2.5.** Suppose  $X$  has continuous density  $f$ ,  $P(\alpha \leq X \leq \beta) = 1$  and  $g$  is a function that is strictly increasing and differentiable on  $(\alpha, \beta)$ . Then  $g(X)$  has density  $f(g^{-1}(y))/g'(g^{-1}(y))$  for  $y \in (g(\alpha), g(\beta))$  and 0 otherwise. When  $g(x) = ax + b$  with  $a > 0$ ,  $g^{-1}(y) = (y - b)/a$  so the answer is  $(1/a)f((y - b)/a)$ .

sol. Since  $g$  is strictly increasing  $g^{-1}$  exists.

$$\begin{aligned}
P(g(X) \leq y) &= P(X \leq g^{-1}(y)) = \int_{\alpha}^{g^{-1}(y)} f(x) dx \\
\frac{d}{dy} P(g(X) \leq y) &= f(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) = f(g^{-1}(y))/g'(g^{-1}(y))
\end{aligned}$$

□

**1.2.6.** Suppose  $X$  has a normal distribution. Use the previous exercise to compute the density of  $\exp(X)$ .

sol. Let  $g(x) = \exp(x)$ . Then  $g^{-1}(x) = \log(x)$

$$\begin{aligned}
f(g^{-1}(x))/g'(g^{-1}(x)) &= f(\log(x))/\exp(\log(x)) \\
&= f(\log(x))/x = \frac{1}{x\sqrt{2\pi}} \exp(-\log(x)^2/2), \quad x > 0
\end{aligned}$$

□

## 1.3 Random Variables

**1.3.1.** Show that if  $\mathcal{A}$  generates  $\mathcal{S}$ , then  $X^{-1}(\mathcal{A}) \equiv \{\{X \in A\} : A \in \mathcal{A}\}$  generates  $\sigma(X) = \{\{X \in B\} : B \in \mathcal{S}\}$ .

sol. By assumption,  $\mathcal{A} \subset \mathcal{S}$ . Thus,  $X^{-1}(\mathcal{A}) \subset X^{-1}(\mathcal{S})$ . Therefore,  $\sigma(X^{-1}(\mathcal{A})) \subset \sigma(X^{-1}(\mathcal{S})) = \sigma(X)$ .

Choose a set  $B \in \mathcal{S}$ . Since  $\mathcal{A}$  generates  $\mathcal{S}$ ,  $B$  can be represented by sets of  $\mathcal{A}$ . Thus,  $\{X \in B\} \in X^{-1}(\mathcal{A})$ . Therefore,  $\sigma(X) \subset X^{-1}(\mathcal{A})$ . □

**1.3.2.** Prove Theorem 1.3.6 when  $n = 2$  by checking  $\{X_1 + X_2 < x\} \in \mathcal{F}$ .

sol.

$$\{X_1 + X_2 < x\} = \bigcup_{q \in \mathbb{Q}} \{X_1 < q\} \times \{X_2 < x - q\}$$

Both  $\{X_1 < q\}$  and  $\{X_2 < x - q\}$  are open sets. Thus,  $\{X_1 + X_2 < x\} \in \mathcal{R}^2$  since it is represented by countable union. □

**1.3.3.** Show that if  $f$  is continuous and  $X_n \rightarrow X$  almost surely then  $f(X_n) \rightarrow f(X)$  almost surely.

sol. Let  $\Omega_0 = \{\omega : \lim X_n(\omega) = X(\omega)\}$  and  $\Omega_f = \{\omega : \lim f(X_n(\omega)) = f(X(\omega))\}$ .

$$\begin{aligned}
\omega \in \Omega_0 &\Rightarrow \lim X_n(\omega) = X(\omega) \\
&\Rightarrow \lim f(X_n(\omega)) = f(X(\omega)) \quad \because f \text{ is continuous} \\
&\Rightarrow \omega \in \Omega_f \\
&\Rightarrow 1 = P(\Omega_0) \leq P(\Omega_f) \leq 1
\end{aligned}$$

□

**1.3.4.** (i) Show that a continuous function from  $\mathcal{R}^d \rightarrow \mathcal{R}$  is a measurable map from  $(\mathbb{R}^d, \mathcal{R}^d)$  to  $(\mathbb{R}, \mathcal{R})$ .  
(ii) Show that  $\mathcal{R}^d$  is the smallest  $\sigma$ -field that makes all the continuous functions measurable.

*sol.*

- (i)  $\mathcal{R}$  is  $\sigma$ -field which is generated by all open sets. If  $f$  is continuous mapping and image is open, then inverse-image is also open. Thus,  $f$  is a measurable map.
- (ii) Let  $\mathcal{S}$  be the smallest  $\sigma$ -field that makes all the continuous functions measurable. Since  $\mathcal{R}$  is the smallest  $\sigma$ -field which is generated by all open sets,  $\mathcal{S}$  is generated by all open sets in  $\mathbb{R}^d$ . □

**1.3.5.** A function  $f$  is said to be lower semicontinuous or l.s.c. if

$$\liminf_{y \rightarrow x} f(y) \geq f(x)$$

and upper semicontinuous (u.s.c.) if  $-f$  is l.s.c.. Show that  $f$  is l.s.c. if and only if  $\{x : f(x) \leq a\}$  is closed for each  $a \in \mathbb{R}$  and conclude that semicontinuous functions are measurable.

*sol.* Define  $E_a := \{x : f(x) \leq a\}$

if Choose  $x$ . Let  $a = \liminf_{y \rightarrow x} f(y)$ . Then, by definition,  $a \leq f(x)$ . There is the sequence  $\{x_n\}$  such that  $x_n \rightarrow x, f(x_n) \rightarrow a$  as  $n \rightarrow \infty$ . For any  $b > a$ , there exists  $N_b$  such that for all  $n > N_b$ ,  $x_n \in E_b$ . Since  $E_b$  is closed, the limit point  $x \in E_b$  for all  $b > a$ . Thus,  $x \in E_a$ . Since  $f(x) \geq a$  and  $f(x) \leq a$ ,  $f(x) = a = \liminf_{y \rightarrow x} f(y)$ .

only if Construct a sequence  $\{x_n\}$  in  $E_a$  which converges to  $x$ . Since  $f$  is l.s.c

$$f(x) \leq \liminf_{y \rightarrow x} f(y) \leq a$$

Thus,  $x \in E_a$ . Since an arbitrary limit point is in  $E_a$ ,  $E_a$  is closed. □

**1.3.6.** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be an arbitrary function and let  $f^\delta(x) = \sup\{f(y) : |y - x| < \delta\}$  and  $f_\delta(x) = \inf\{f(y) : |y - x| < \delta\}$  where  $|z| = (z_1^2 + \dots + z_d^2)^{1/2}$ . Show that  $f^\delta$  is l.s.c. and  $f_\delta$  is u.s.c. Let  $f^0 = \lim_{\delta \downarrow 0} f^\delta$ ,  $f_0 = \lim_{\delta \downarrow 0} f_\delta$  and conclude that the set of points at which  $f$  is discontinuous  $= \{f^0 \neq f_0\}$  is measurable.

*sol.* By using result of previous exercise, we need to show that  $\{x : f^\delta(x) \leq a\}$  and  $\{x : f_\delta(x) \geq a\}$  are closed for each  $a \in \mathbb{R}$ .

We can construct a sequence  $\{x_n\} \subset \{t : f^\delta(t) \leq (a)\}$  which converges to  $x$ . Choose  $y \in \{t : |t - x| < \delta\}$ . Then, we can find some  $N$  which satisfies for all  $n \geq N$ ,  $|x_n - x| < \delta - |y - x|$ . Thus,  $|x_n - y| \leq |x_n - x| + |y - x| < \delta \Rightarrow f(y) \leq a$ . Since  $y$  is arbitrary chosen,  $\sup\{f(y) : |y - x| < \delta\} \leq a$ . The limit of convergent sequence is in  $\{t : f^\delta(t) \leq (a)\}$ . Thus,  $\{t : f^\delta(t) \leq (a)\}$  is closed. We can prove that  $\{x : f_\delta(x) \geq a\}$  is closed similarly.

Since  $\{f^0 \neq f_0\} = \{x : f^0(x) = f_0(x)\}^c$ , we need to show that  $\{x : f^0(x) = f_0(x)\}$  is measurable. We showed that  $\mathcal{S}$  can be generated by half closed intervals(ex 1.1.2). Thus,  $f^0$  and  $f_0$  are measurable functions. □

**1.3.7.** A function  $\varphi : \Omega \rightarrow \mathbb{R}$  is said to be simple if

$$\varphi(\omega) = \sum_{m=1}^n c_m 1_{A_m}(\omega)$$

where the  $c_m$  are real numbers and  $A_m \in \mathcal{F}$ . Show that the class of  $\mathcal{F}$  measurable functions is the smallest class containing the simple functions and closed under pointwise limits.

*sol.* By definition, simple functions are  $\mathcal{F}$  measurable. □

**1.3.8.** Use the previous exercise to conclude that  $Y$  is measurable with respect to  $\sigma(X)$  if and only if  $Y = f(X)$  where  $f : \mathbb{R} \rightarrow \mathbb{R}$  is measurable

**1.3.9.** To get a constructive proof of the last result, note that  $\{\omega : m2^{-n} \leq Y < (m+1)2^{-n}\} = \{X \in B_{m,n}\}$  for some  $B_{m,n} \in \mathcal{R}$  and set  $f_n(x) = m2^{-n}$  for  $x \in B_{m,n}$  and show that as  $n \rightarrow \infty$ ,  $f_n(x) \rightarrow f(x)$  and  $Y = f(X)$ .

## 1.4 Integration

**1.4.1.** Show that if  $f \geq 0$  and  $\int f d\mu = 0$  then  $f = 0$  a.e.

*sol.* Let  $N$  be the set that satisfies  $x \in N \Rightarrow f(x) = 0$ .

$$\begin{aligned}\int f d\mu &= \int_N f d\mu + \int_{N^c} f d\mu \\ &= \int_{N^c} f d\mu = 0 \\ &\Rightarrow \mu(N^c) = 0\end{aligned}$$

□

**1.4.2.** Let  $f \geq 0$  and  $E_{n,m} = \{x : m/2^n \leq f(x) < (m+1)/2^n\}$ . As  $n \uparrow \infty$ ,

$$\sum_{m=1}^{\infty} \frac{m}{2^n} \mu(E_{n,m}) \uparrow \int f d\mu$$

*sol.* Let  $f_n = \sum_{m=1}^{\infty} 1_{E_{n,m}}$ . Then, for any  $n$ ,  $f_n < f$ . Thus,  $\int f d\mu$  is an upper bound of  $\int f_n d\mu$ . Since  $\int f_n d\mu$  is increasing and bounded, it converges. By definition,

$$\int f - f_n d\mu \leq \int 1/2^n d\mu \rightarrow 0 \text{ as } n \rightarrow \infty$$

Thus,  $\int f_n d\mu$  converges to  $\int f d\mu$

□

**1.4.3.** Let  $g$  be an integrable function on  $\mathbb{R}$  and  $\epsilon > 0$ . (i) Use the definition of the integral to conclude there is a simple function  $\varphi = \sum_k b_k 1_{A_k}$  with  $\int |g - \varphi| dx < \epsilon$ . (ii) Use Exercise A.2.1 to approximate the  $A_k$  by finite unions of intervals to get a step function

$$q = \sum_{j=1}^k c_j 1_{(a_{j-1}, a_j]}$$

with  $a_0 < a_1 < \dots < a_k$ , so that  $\int |\varphi - q| < \epsilon$ . (iii) Round the corners of  $q$  to get a continuous function  $r$  so that  $\int |q - r| dx < \epsilon$ . (iv) To make a continuous function replace each  $c_j 1_{(a_{j-1}, a_j]}$  by a function that is 0  $(a_{j-1}, a_j)^c$ ,  $c_j$  on  $[a_{j-1} + \delta_j, a_j - \delta_j]$ , and linear otherwise. If the  $\delta_j$  are small enough and we let  $r(x) = \sum_{j=1}^k r_j(x)$  then

$$\int |q(x) - r(x)| d\mu = \sum_{j=1}^k \delta_j c_j < \epsilon$$

*sol.*

(i) For partition  $P$ , we set  $f_i(x) = \inf\{f(x) : x \in [x_i, x_{i+1}]\}$ . Then, there exists a partition for every  $\epsilon$  that makes  $\int |\varphi - g| dx < \epsilon$ . ( $\varphi = \sum f_i$ )

□

**1.4.4.** Prove the Riemann-Lebesgue lemma. If  $g$  is integrable then

$$\lim_{n \rightarrow \infty} \int g(x) \cos nx dx = 0$$

*sol.* Let  $\epsilon$  be an arbitrary positive number. We can find a step function  $\varphi$  such that  $\int |g - \varphi| dx < \epsilon/2$ .

$$\begin{aligned}\int \varphi(x) \cos(nx) dx &= \sum_{i=1}^m c_i \int_{a_i}^{b_i} \cos(nx) dx \\ &= \sum_{i=1}^m c_i/n \int_{na_i}^{nb_i} \cos(y) dy \\ &= \sum_{i=1}^m c_i/n (\sin(nb_i) - \sin(na_i)) \rightarrow 0\end{aligned}$$

Thus, we can choose  $N$  which satisfies that for all  $m \geq N$ ,  $\int |\varphi(x) \cos(nx)| dx < \epsilon/2$ . Then,

$$\begin{aligned} \left| \int g(x) \cos(nx) dx \right| &\leq \int |g(x) \cos(nx)| dx \\ &\leq \int |g(x) - \varphi(x)| |\cos(nx)| dx + \int |\varphi(x) \cos(nx)| dx \\ &\leq \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

Since  $\epsilon$  is arbitrary, it converges to 0. □

## 1.5 Properties of the integral

**1.5.1.** Let  $\|f\|_\infty = \inf\{M : \mu(\{x : |f(x)| > M\}) = 0\}$ . Prove that

$$\int |fg| d\mu \leq \|f\|_1 \|g\|_\infty$$

*sol.* By definition,  $g \leq \|g\|_\infty$  a.e.. Thus,

$$\int |fg| d\mu \leq \int |f| \|g\|_\infty d\mu = \|g\|_\infty \int |f| d\mu = \|f\|_1 \|g\|_\infty$$

□

**1.5.2.** Show that if  $\mu$  is a probability measure then

$$\|f\|_\infty = \lim_{p \rightarrow \infty} \|f\|_p$$

*sol.* Since  $\mu$  is a probability measure,  $\mu(\Omega) = 1$ . Suppose two positive numbers,  $n < m$ . Let  $\varphi(x) := x^{m/n}$ . By Jensen's inequality,

$$\|f\|_n^m = \varphi\left(\int |f|^n d\mu\right) \leq \int \varphi(|f|^n) d\mu = \|f\|_m^m$$

Thus,  $\{\|f\|_n\}$  is an increasing sequence.

$$\|f\|_n^n = \int |f|^n d\mu \leq \int (\|f\|_\infty^n) d\mu = \|f\|_\infty^n$$

Thus,  $\|f\|_\infty$  is an upper bound. Therefore, the sequence converges and  $\lim_{p \rightarrow \infty} \|f\|_p \leq \|f\|_\infty$ .

Let  $\epsilon$  be the arbitrary number between 0 and  $\|f\|_\infty$ . Then, there exist a set  $M$  such that  $|f| > \|f\|_\infty - \epsilon$  for all  $x \in M$ .

$$\begin{aligned} \int |f|^p d\mu &\geq \int_M |f|^p d\mu \geq (\|f\|_\infty - \epsilon)^p \mu(M) \\ &\Rightarrow \|f\|_p^p \geq (\|f\|_\infty - \epsilon)^p \mu(M)^{1/p} \\ &\Rightarrow \lim_{p \rightarrow \infty} \|f\|_p \geq \|f\|_\infty - \epsilon \\ &\Rightarrow \lim_{p \rightarrow \infty} \|f\|_p \geq \|f\|_\infty \end{aligned}$$

Thus, the sequences converges to  $\|f\|_\infty$ . □

**1.5.3** (Minkowski's inequality). (i) Suppose  $p \in (1, \infty)$ . The inequality  $|f + g|^p \leq 2^p(|f|^p + |g|^p)$  shows that if  $\|f\|_p$  and  $\|g\|_p$  are  $< \infty$  then  $\|f + g\|_p < \infty$ . Apply Holder's inequality to  $|f||f + g|^{p-1}$  and  $|g||f + g|^{p-1}$  to show  $\|f + g\|_p \leq \|f\|_p + \|g\|_p$  (ii) Show that the last result remains true when  $p = 1$  or  $p = \infty$

*sol.*



(i) By Holder's inequality,

$$\begin{aligned}\int |f||f+g|^{p-1}d\mu &\leq \|f\|_p \|(f+g)^{p-1}\|_{p/(p-1)} \\ &= \left(\int |f|^p d\mu\right)^{1/p} \left(\int |f+g|^p d\mu\right)^{(p-1)/p}\end{aligned}$$

$$\begin{aligned}\text{Similarly, } \int |g||f+g|^{p-1}d\mu &\leq \left(\int |g|^p d\mu\right)^{1/p} \left(\int |f+g|^p d\mu\right)^{(p-1)/p} \\ \int |f+g|^p d\mu &= \int (|f|+|g|)|f+g|^{p-1}d\mu \leq (\|f\|_p + \|g\|_p) \|f+g\|_p^{p-1} \\ &\Rightarrow \|f+g\|_p \leq (\|f\|_p + \|g\|_p)\end{aligned}$$

(ii) When  $p = 1$ , it is trivial.

Let the set  $E = \{M : \mu(\{x : |f(x) + g(x)| > M\}) = 0\}$ . By triangle inequality,

$$|f+g| \leq |f| + |g| \leq \|f\|_\infty + \|g\|_\infty \text{ a.e.}$$

Thus,

$$\mu(\{x : |f(x) + g(x)| > \|f\|_\infty + \|g\|_\infty\}) = 0$$

Since  $\|f\|_\infty + \|g\|_\infty \in E$ ,  $\|f\|_\infty + \|g\|_\infty \geq \inf E = \|f+g\|_\infty$

□

**1.5.4.** If  $f$  is integrable and  $E_m$  are disjoint sets with union  $E$  then

$$\sum_{m=0}^{\infty} \int_{E_m} f d\mu = \int_E f d\mu$$

So if  $f \geq 0$ , then  $\nu(E) = \int_E f d\mu$  defines a measure.

*sol.* If  $E = \emptyset$ ,  $\mu(E) = 0$ . Since  $f \geq 0$ , for any set  $E$ ,  $\nu(E) \geq \nu(\emptyset) = 0$ .

If  $E_i$  are disjoint sets, then

$$\mu(\cup_i E_i) = \int_{\cup_i E_i} f d\mu = \sum_i \int_{E_i} f d\mu = \sum_i \nu(E_i)$$

Thus,  $\nu$  is a measure.

□

**1.5.5.** If  $g_n \uparrow g$  and  $\int g_1^- d\mu < \infty$  then  $\int g_n d\mu \uparrow \int g d\mu$ .

*sol.* Since  $g_n = g_n^+ - g_n^-$  and  $g_n \uparrow g$ ,  $g_n^+ \uparrow g^+$  and  $g_n^- \downarrow g^-$ . By monotone convergence theorem,  $\int g_n^+ d\mu \uparrow \int g^+ d\mu$ . Since  $g_1^-$  is integrable, by dominated convergence theorem,  $\int g_n^- d\mu \rightarrow \int g^- d\mu$ . Thus,  $-\int g_n^- d\mu \uparrow -\int g^- d\mu$ .

$$\therefore \int g_n d\mu = \int g_n^+ - g_n^- d\mu \uparrow \int g d\mu$$

□

**1.5.6.** If  $g_m \geq 0$  then  $\int \sum_{m=0}^{\infty} g_m d\mu = \sum_{m=0}^{\infty} \int g_m d\mu$ .

*sol.* Define  $f_n := \sum_{m=0}^n g_m$  and  $f := \sum_{m=0}^{\infty} g_m$ . Then  $f_n \uparrow f$ . By result of using previous example,  $\int f_n d\mu \uparrow \int f d\mu$ . Thus,

$$\begin{aligned}\int \sum_{m=0}^{\infty} g_m d\mu &= \int \lim_{n \rightarrow \infty} f_n d\mu \\ &= \int f d\mu \\ &= \int \sum_{m=0}^{\infty} g_m d\mu\end{aligned}$$

□

---

**1.5.7.** Let  $f \geq 0$ . (i) Show that  $\int f \wedge n \, d\mu \uparrow \int f \, d\mu$  as  $n \rightarrow \infty$ . (ii) Use (i) to conclude that if  $g$  is integrable and  $\epsilon > 0$  then we can pick  $\delta > 0$  so that  $\mu(A) < \delta$  implies  $\int_A |g| \, d\mu < \epsilon$ .

*sol.*

(i) Define  $E_n := \{x : f(x) > n\}$ . Then, for  $n \leq m$

$$f \wedge m - f \wedge n = (f - n)1_{E_n \cap E_m^c} + (m - n)1_{E_m} \geq 0$$

$$\lim_{n \rightarrow \infty} f \wedge n = f$$

By definition  $f_n \geq 0$ . Thus, by monotone convergence theorem,  $\int f \wedge n \, d\mu \uparrow \int f \, d\mu$ .

(ii) By results of (i), we can find the positive  $N$  for any positive  $\epsilon/2$ ,  $\mu(\{x : |g(x)| > N\}) < \epsilon$ .

$$\begin{aligned} \int_A |g| \, d\mu &\leq \int_{\{x: |g(x)| > N\}} |g| \, d\mu + \int_{\{x \in A: |g(x)| \leq N\}} |g| \, d\mu \\ &\leq \epsilon/2 + N\mu(A) \end{aligned}$$

Thus, if we set  $\delta := \epsilon/(2N)$ , then  $\mu(A) < \delta$  implies  $\int_A |g| \, d\mu < \epsilon$ .

□

**1.5.8.** Show that if  $f$  is integrable on  $[a, b]$ ,  $g(x) = \int_{[a, x]} f(y) \, dy$  is continuous on  $(a, b)$ .

*sol.* Since  $f$  is integrable, there exists a positive number  $M$  such that  $|f| < M$ . For  $s, t \in (a, b)$ ,  $s \leq t$

$$\begin{aligned} -M(t - s) &< g(t) - g(s) = \int_s^t f(y) \, dy < M(t - s) \\ \Rightarrow |g(t) - g(s)| &< M(t - s) \end{aligned}$$

Thus,  $g$  is Lipschitz continuous. Thus,  $g$  is continuous.

□

**1.5.9.** Show that if  $f$  has  $\|f\|_p = (\int |f|^p \, d\mu)^{1/p} < \infty$ , then there are simple functions  $\varphi_n$  so that  $\|\varphi_n - f\|_p \rightarrow 0$ .

*sol.* For  $f$ , we can construct sequence of step functions  $\{\varphi_n\}$  such that  $\varphi_n \uparrow |f|$ . By triangle inequality,  $|f - \varphi_n| \leq |f| + |\varphi_n|$ . Thus, for all  $n$

$$|f - \varphi_n|^p \leq (|f| + |\varphi_n|)^p \leq (2|f|)^p$$

By dominated convergence theorem,

$$\begin{aligned} \lim_{n \rightarrow \infty} \int |f - \varphi_n|^p &= \int \lim_{n \rightarrow \infty} |f - \varphi_n|^p = 0 \\ \Rightarrow \|f - \varphi_n\|_p &= 0 \end{aligned}$$

□

**1.5.10.** Show that if  $\sum_n \int |f_n| \, d\mu < \infty$ , then  $\sum_n \int f_n \, d\mu = \int \sum_n f_n \, d\mu$ .

*sol.*

$$\begin{aligned} \sum_n \int |f_n| \, d\mu &= \sum_n \int f_n^+ \, d\mu + \sum_n \int f_n^- \, d\mu \\ &\Rightarrow \sum_n \int f_n^+ \, d\mu < \infty, \sum_n \int f_n^- \, d\mu < \infty \end{aligned}$$

By the result of (1.5.6),

$$\begin{aligned}
\sum_n \int f_n^+ d\mu &= \int \sum_n f_n^+ d\mu \\
\sum_n \int f_n^- d\mu &= \int \sum_n f_n^- d\mu \\
\sum_n \int f_n d\mu &= \sum_n \int f_n^+ - f_n^- d\mu \\
&= \int \sum_n f_n^+ d\mu - \int \sum_n f_n^- d\mu \\
&= \int \sum_n f_n d\mu
\end{aligned}$$

□

## 1.6 Expected Value

**1.6.1.** Suppose  $\varphi$  is strictly convex, i.e.,  $\varphi(\lambda x + (1-\lambda)y) < \lambda\varphi(x) + (1-\lambda)\varphi(y)$  holds for  $\lambda \in (0,1)$ . Show that, under the assumptions of Theorem 1.6.2,  $\varphi(EX) = E(\varphi X)$  implies  $X = EX$  a.s.

*sol.* Let  $t = EX$ . Since  $\varphi$  is strictly convex, we can find  $a$  which satisfies, for all  $x$ ,  $\varphi(x) \geq \varphi(t) + a(x-t)$ . Thus  $E\varphi(X) \geq \varphi(t) + E[a(X-t)] = \varphi(t)$ . If we add assumption  $\varphi(EX) = E(\varphi X)$ , then we can conclude  $a(X-t) = 0$  a.e. Thus,  $X = t = EX$  a.e. □

**1.6.2.** Suppose  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex. Imitate the proof of Theorem 1.5.1 to show

$$E\varphi(X_1, \dots, X_n) \geq \varphi(EX_1, \dots, EX_n)$$

provided  $E|\varphi(X_1, \dots, X_n)| < \infty$  and  $E|X_i| < \infty$  for all  $i$ .

*sol.* Let  $c = (EX_1, \dots, EX_n)$  and define  $l(x) = a'(x-c) + \varphi(c)$ ,  $a, x \in \mathbb{R}^n$ . Then we can find  $a$  which satisfies that for all  $x$ ,  $\varphi(x) \geq l(x)$  since  $\varphi$  is convex. Thus,

$$E\varphi(X_1, \dots, X_n) \geq aE(X-c) + E\varphi(c) = \varphi(c) = \varphi(EX_1, \dots, EX_n)$$

□

**1.6.3** (Chebyshev's inequality is and is not sharp). (i) Show that Theorem 1.6.4 is sharp by showing that if  $0 < b \leq a$  are fixed there is an  $X$  with  $EX^2 = b^2$  for which  $P(|X| \geq a) = b^2/a^2$ . (ii) Show that Theorem 1.6.4 is not sharp by showing that if  $X$  has  $0 < EX^2 < \infty$  then

$$\lim_{a \rightarrow \infty} a^2 P(|X| \geq a) / EX^2 = 0$$

*sol.*

(i)

$$P(|X| \geq a) = P(X^2 \geq a^2) \leq EX^2/a^2 = b^2/a^2$$

Since  $a, b$  is not arbitrary, we can make the equality hold.

(ii)

$$a^2 P(|X| \geq a) = a^2 P(X^2 \geq a^2) = \int_{a^2}^{\infty} a^2 d\mu \leq \int_{a^2}^{\infty} X^2 d\mu$$

Since  $EX^2$  is finite,

$$\begin{aligned} EX^2 &= \int X^2 d\mu \\ &= \int_0^{a^2} X^2 d\mu + \int_{a^2}^{\infty} a^2 d\mu \\ &\Rightarrow \lim_{a^2 \rightarrow \infty} \int_{a^2}^{\infty} a^2 d\mu = 0 \end{aligned}$$

Thus,  $\lim_{a \rightarrow \infty} a^2 P(|X| \geq a) = 0$ .

□

**1.6.4** (One-sided Chebyshev bound). (i) Let  $a > b > 0$ ,  $0 < p < 1$ , and let  $X$  have  $P(X = a) = p$  and  $P(X = -b) = 1 - p$ . Apply Theorem 1.6.4 to  $\varphi(x) = (x + b)^2$  and conclude that if  $Y$  is any random variable with  $EY = EX$  and  $\text{Var}(Y) = \text{Var}(X)$ , then  $P(Y \geq a) \leq p$  and equality holds when  $Y = X$ . (ii) Suppose  $EY = 0$ ,  $\text{Var}(Y) = \sigma^2$ , and  $a > 0$ . Show that  $P(Y \geq a) \leq \sigma^2/(a^2 + \sigma^2)$ , and there is a  $Y$  for which equality holds.

**1.6.5** (Two nonexistent lower bounds). Show that : (i) if  $\epsilon > 0$ ,  $\inf\{P(|X| > \epsilon) : EX = 0, \text{Var}(X) = 1\} = 0$ . (ii) if  $y \geq 1$ ,  $\sigma^2 \in (0, \infty)$ ,  $\inf\{P(|X| > y) : EX = 1, \text{Var}(X) = \sigma^2\} = 0$

sol. (i)

$$P(|X| > \epsilon) \leq E(X^2)/\epsilon^2 = V(X)/\epsilon^2$$

(ii)

□

**1.6.6** (A useful lower bound). Let  $Y \geq 0$  with  $EY^2 < \infty$ . Apply the Cauchy-Schwarz inequality to  $Y1_{Y>0}$  and conclude

$$P(Y > 0) \geq (EY)^2/EY^2$$

sol.

$$\begin{aligned} E[Y1_{Y>0}] &\leq \sqrt{EY^2 E[1_{Y>0}]} \\ EY - 0 \cdot P(Y = 0) &\leq \sqrt{EY^2 P(Y > 0)} \\ (EY)^2 &\leq EY^2 P(Y > 0) \\ P(Y > 0) &\geq (EY)^2/EY^2 \end{aligned}$$

□

**1.6.7.** Let  $\Omega = (0, 1)$  equipped with the Borel sets and Lebesgue measure. Let  $\alpha \in (1, 2)$  and  $X_n = n^\alpha 1_{(1/(n+1), 1/n)} \rightarrow 0$  a.s. Show that Theorem 1.6.8 can be applied with  $h(x) = x$  and  $g(x) = |x|^{2/\alpha}$ , but the  $X_n$  are not dominated by an integrable function.

sol.

1.  $g \geq 0$ ,  $\because 2/\alpha > 1$ ,  $g(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$
2.  $|h(x)|/g(x) = |x|^{(\alpha-2)/2}$ ,  $\because (\alpha-2) < 1$ ,  $|h(x)|/g(x) \rightarrow 0$  as  $|x| \rightarrow \infty$ .
3.  $Eg_n(X) = n^2 \int 1_{(1/(n+1), 1/n)} d\mu = n/(n+1) < 1$ .

Thus,  $Eh(X_n) \rightarrow Eh(X)$ . However, there is no integrable function which dominates  $X_n$  because there is no upper bound for  $\sup_n |X_n|$ . □

**1.6.8.** Suppose that the probability measure  $\mu$  has  $\mu(A) = \int_A f(x)dx$  for all  $A \in \mathcal{R}$ . Use the proof technique of Theorem 1.6.9 to show that for any  $g$  with  $g \geq 0$  or  $\int |g(x)|\mu(dx) < \infty$  we have

$$\int g(x)\mu(dx) = \int g(x)f(x)dx$$

*sol.*

1. Indicator function

Let  $g = c1_A$ . Then,

$$\int g(x)\mu(dx) = \int_A c\mu(dx) = c\mu(A) = \int_A cf(x)dx = \int g(x)f(x)dx$$

2. Simple function

Let  $g = \sum_{i=1}^n c_i 1_{A_i}$ . Then, by case 1,

$$\int g(x)\mu(dx) = \sum_{i=1}^n c_i \mu(A_i) = \int g(x)f(x)dx$$

3. Nonnegative function

There is a sequence of simple functions  $\{g_n\}$  which satisfies that  $g_n \uparrow g$ . By monotone convergence theorem,

$$\int g(x)\mu(dx) = \lim_{n \rightarrow \infty} \int g_n(x)\mu(dx) = \lim_{n \rightarrow \infty} \int g_n(x)f(x)dx = \int g(x)f(x)dx$$

The proof ends if  $g \geq 0$ .

4. Integrable function( $g$  is integrable)

Let  $g = g^+ - g^-$ . Then

$$\begin{aligned} \int g^+(x)\mu(dx) &= \int g^+(x)f(x)dx, \int g^-(x)\mu(dx) = \int g^-(x)f(x)dx \\ \Rightarrow \int g(x)\mu(dx) &= \int g^+(x)\mu(dx) - \int g^-(x)\mu(dx) \\ &= \int g^+(x)f(x)dx - \int g^-(x)f(x)dx = \int g(x)f(x)dx \end{aligned}$$

□

**1.6.9** (Inclusion-exclusion formula). *Let  $A_1, A_2, \dots, A_n$  be events and  $A = \cup_{i=1}^n A_i$ . Prove that  $1_A = 1 - \prod_{i=1}^n (1 - 1_{A_i})$ . Expand out the right hand side, then take expected value to conclude*

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) \\ &\quad + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(\cap_{i=1}^n A_i) \end{aligned}$$

*sol.*

$$\begin{aligned} \omega \in A &\Leftrightarrow \omega \in \cup_{i=1}^n A_i \Leftrightarrow \omega \notin \cap_{i=1}^n A_i^C \\ \Rightarrow 1_A &= 1 - \prod_{i=1}^n (1 - 1_{A_i}) \\ &= \sum 1_{A_i} - \sum 1_{A_i \cap A_j} + \dots + (-1)^{n-1} \times 1_{\cap_{i=1}^n A_i} \end{aligned}$$

If we take expectation, we obtain the equation.

□

---

**1.6.10** (Bonferroni inequalities). Let  $A_1, A_2, \dots, A_n$  be events and  $A = \cup_{i=1}^n A_i$ . Show that  $1_A \leq \sum_{i=1}^n 1_{A_i}$ , etc. and then take expected values to conclude

$$\begin{aligned} P(\cup_{i=1}^n A_i) &\leq \sum_{i=1}^n P(A_i) \\ P(\cup_{i=1}^n A_i) &\geq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) \\ P(\cup_{i=1}^n A_i) &\leq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) \end{aligned}$$

In general, if we stop the inclusion exclusion formula after an even (odd) number of sums, we get an lower (upper) bound.

*sol.* If  $\omega \notin A$ , then  $\omega \notin A_i$  for all  $i$ . If  $\omega \in A$ , there is at least 1  $A_i$  s.t.  $\omega \in A_i$ . Thus,  $1_A \leq \sum_{i=1}^n 1_{A_i}$ .

If we stop the formula after an even number of sums, □

**1.6.11.** If  $E|X|^k < \infty$  then for  $0 < j < k$ ,  $E|X|^j < \infty$ , and furthermore

$$E|X|^j \leq (E|X|^k)^{j/k}$$

*sol.* Define  $\varphi(x) = x^{k/j}$ . Then  $\varphi$  is a convex function. By Jensen's inequality

$$\begin{aligned} (E|X|^j)^{k/j} &= \varphi(E|X|^j) \leq E\varphi(|X|^j) = E|X|^k \\ \Rightarrow E|X|^j &\leq (E|X|^k)^{j/k} \end{aligned}$$

□

**1.6.12.** Apply Jensen's inequality with  $\varphi(x) = e^x$  and  $P(X = \log y_m) = p(m)$  to conclude that if  $\sum_{m=1}^n p(m) = 1$  and  $p(m), y_m > 0$  then

$$\sum_{m=1}^n p(m)y_m \geq \prod_{m=1}^n y_m^{p(m)}$$

When  $p(m) = 1/n$ , this says the arithmetic mean exceeds the geometric mean.

*sol.*

$$\begin{aligned} E\varphi(X) &= \sum_{n=1}^m p(m) \exp(\log y_m) = \sum_{n=1}^m p(m)y_m \\ \varphi(EX) &= \exp(\sum_{n=1}^m p(m) \log y_m) = \prod_{n=1}^m y_m^{p(m)} \end{aligned}$$

By Jensen's inequality,

$$\sum_{m=1}^n p(m)y_m = E\varphi(X) \geq \varphi(EX) = \prod_{m=1}^n y_m^{p(m)}$$

If  $p(m) = 1/n$ , the left hand becomes arithmetic mean and the right hand becomes geometric mean. □

**1.6.13.** If  $EX_1^- < \infty$  and  $X_n \uparrow X$  then  $EX_n \uparrow EX$ .

*sol.* Same with Exercise 1.5.5. We use the probability measure in this exercise. □

**1.6.14.** Let  $X \geq 0$  but do NOT assume  $E(1/X) < \infty$ . Show

$$\lim_{y \rightarrow \infty} yE(1/X; X > y) = 0, \quad \lim_{y \downarrow 0} yE(1/X; X > y) = 0$$

*sol.*

$$\begin{aligned}
0 \leq yE(1/X; X > y) &= \int_{\{\omega: X(\omega) > y\}} y/X d\mu \\
&\leq \int_{\{\omega: X(\omega) > y\}} 1 d\mu \\
&= \mu(\{\omega : X(\omega) > y\}) \\
\Rightarrow \lim_{y \rightarrow \infty} yE(1/X; X > y) &\leq \lim_{y \rightarrow \infty} \mu(\{\omega : X(\omega) > y\}) = 0
\end{aligned}$$

□

**1.6.15.** If  $X_n \geq 0$  then  $E(\sum_{n=0}^{\infty} X_n) = \sum_{n=0}^{\infty} EX_n$ .

*sol.* Same with Exercise 1.5.10. We use the probability measure in this exercise.

□

**1.6.16.** If  $X$  is integrable and  $A_n$  are disjoint sets with union  $A$  then

$$\sum_{n=0}^{\infty} E(X; A_n) = E(X; A)$$

*i.e., the sum converges absolutely and has the value on the right.*

*sol.* Define  $Y_n := \sum_{i=0}^n X 1_{A_i}$  and  $Y := X 1_A$ . Then,  $Y_n \rightarrow Y$  and  $|Y| \leq |X|$ . Since  $X$  is integrable,  $Y$  is also integrable. For any  $\omega \in \cup_{i=0}^{\infty} A_i$ ,  $Y_n(\omega) = Y(\omega)$ . Thus,  $|Y_n| \leq |Y|$ . By dominated convergence theorem,

$$\lim_{n \rightarrow \infty} EY_n = EY$$

$$\begin{aligned}
\lim_{n \rightarrow \infty} EY_n &= \lim_{n \rightarrow \infty} E(X; A_n) \\
EY &= E(X; A)
\end{aligned}$$

□

## 1.7 Product Measures, Fubini's Theorem

**1.7.1.** If  $\int_X \int_Y |f(x, y)| \mu_2(dy) \mu_1(dx) < \infty$  then

$$\int_X \int_Y |f(x, y)| \mu_2(dy) \mu_1(dx) = \int_{X \times Y} f d(\mu_1 \times \mu_2) = \int_Y \int_X |f(x, y)| \mu_1(dx) \mu_2(dy)$$

**Corollary.** Let  $X = \{1, 2, \dots\}$ ,  $\mathcal{A}$  = all subsets of  $X$ , and  $\mu_1$  = counting measure. If  $\sum_n \int |f_n| d\mu < \infty$  then  $\sum_n \int f_n d\mu = \int \sum_n f_n d\mu$ .

**1.7.2.** Let  $g \geq 0$  be a measurable function on  $(X, \mathcal{A}, \mu)$ . Use Theorem 1.7.2 to conclude that

$$\int_X g d\mu = (\mu \times \lambda)(\{(x, y) : 0 \leq y < g(x)\}) = \int_0^{\infty} \mu(\{x : g(x) > y\}) dy$$

$$\begin{aligned}
\int_X g d\mu &= \int_X \int_{\mathbb{R}} 1_{0 \leq y < g(x)} dy d\mu \\
&= (\mu \times \lambda)(\{(x, y) : 0 \leq y < g(x)\}) \\
&= \int_{\mathbb{R}} \int_X 1_{g(x) > y \geq 0} d\mu dy \\
&= \int_{\mathbb{R}} \mu(\{x : g(x) > y \geq 0\}) dy \\
&= \int_0^{\infty} \mu(\{x : g(x) > y\}) dy
\end{aligned}$$

---

**1.7.3.** Let  $F, G$  be Stieltjes measure functions and let  $\mu, \nu$  be the corresponding measures on  $(\mathbb{R}, \mathcal{R})$ . Show that

- (i)  $\int_{(a,b]} \{F(y) - F(a)\} dG(y) = (\mu \times \nu)(\{(x, y) : a < x \leq y \leq b\})$
- (ii)  $\int_{(a,b]} F(y) dG(y) + \int_{(a,b]} G(y) dF(y) = F(b)G(b) - F(a)G(a) + \sum_{x \in (a,b]} \mu(\{x\})\nu(\{x\})$
- (iii) If  $F = G$  is continuous then  $\int_{(a,b]} 2F(y) dF(y) = F^2(b) - F^2(a)$ .

To see the second term in (ii) is needed, let  $F(x) = G(x) = 1_{[0,\infty)}(x)$  and  $a < 0 < b$ .

sol.

(i)

$$\begin{aligned} \int_{(a,b]} \{F(y) - F(a)\} dG(y) &= \int_{(a,b]} \int_{(a,y]} 1 dF(x) dG(y) \\ &= \int_{(a,y] \times (a,b]} 1 d(\mu \times \nu) \\ &= (\mu \times \nu)(\{(x, y) : a < x \leq y \leq b\}) \end{aligned}$$

(ii) By (i),

$$\begin{aligned} \int_{(a,b]} F(y) dG(y) &= (\mu \times \nu)(\{(x, y) : a < x \leq y \leq b\}) + F(a)(G(b) - G(a)) \\ \int_{(a,b]} G(y) dF(y) &= (\mu \times \nu)(\{(x, y) : a < y \leq x \leq b\}) + G(a)(F(b) - F(a)) \\ \Rightarrow \int_{(a,b]} F(y) dG(y) + \int_{(a,b]} G(y) dF(y) &= (\mu \times \nu)(\{(x, y) : a < x, y \leq b\}) + F(a)(G(b) - G(a)) + G(a)(F(b) - F(a)) \\ &= F(b)G(b) - F(a)G(a) + \sum_{x \in (a,b]} \mu(\{x\})\nu(\{x\}) \end{aligned}$$

(iii) By (ii),

$$\int_{(a,b]} 2F(y) dF(y) = F^2(b) - F^2(a) + \sum_{x \in (a,b]} \mu(\{x\})^2$$

Since  $F$  is continuous,  $\mu(\{x\}) = 0$  for all  $x \in (a, b]$ .

□

**1.7.4.** Let  $\mu$  be a finite measure on  $\mathbb{R}$  and  $F(X) = \mu((-\infty, x])$ . Show that

$$\int (F(x+c) - F(x)) dx = c\mu(\mathbb{R})$$

sol.

$$\begin{aligned} \int (F(x+c) - F(x)) dx &= \int \mu((x, x+c]) dx \\ &= \int \int 1_{(x, x+c]}(y) d\mu dx \\ &= \int \int 1_{[y-c, y)}(x) dx d\mu \\ &= \int c d\mu = c\mu(\mathbb{R}) \end{aligned}$$

□



---

**1.7.5.** Show that  $e^{-xy} \sin x$  is integrable in the strip  $0 < x < a$ ,  $0 < y$ . Perform the double integral in the two orders to get:

$$\int_0^a \frac{\sin x}{x} dx = \arctan(a) - (\cos a) \int_0^\infty \frac{e^{-ay}}{1+y^2} dy - (\sin a) \int_0^\infty \frac{ye^{-ay}}{1+y^2} dy$$

and replace  $1+y^2$  by 1 to conclude  $|\int_0^a (\sin x)/x dx - \arctan(a)| \leq 2/a$  for  $a \geq 1$ .

## Chapter 2

# Law of Large Numbers

### 2.1 Independence

**2.1.1.** Suppose  $(X_1, \dots, X_n)$  has density  $f(x_1, x_2, \dots, x_n)$ , that is

$$P((X_1, X_2, \dots, X_n) \in A) = \int_A f(x) dx \text{ for } A \in \mathcal{R}^n$$

If  $f(x)$  can be written as  $g_1(x_1) \cdots g_n(x_n)$  where the  $g_m \geq 0$  are measurable, then  $X_1, X_2, \dots, X_n$  are independent. Note that the  $g_m$  are not assumed to be probability densities.

*sol.* Since  $f$  is integrable,  $g_i$ s are integrable. If we set  $f_i(x) := g_i(x) / \int g_i(x) dx$ , then  $f_i$ s are probability densities. Thus,  $X_1, X_2, \dots, X_n$  are independent.  $\square$

**2.1.2.** Suppose  $X_1, \dots, X_n$  are random variables that take values in countable sets  $S_1, \dots, S_n$ . Then in order for  $X_1, \dots, X_n$  to be independent, it is sufficient that whenever  $x_i \in S_i$

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$$

*sol.* We can make  $\sigma$ -algebra by singletons. Thus, it is sufficient to show that above equality holds.  $\square$

**2.1.3.** Let  $\rho(x, y)$  be a metric. (i) Suppose  $h$  is differentiable with  $h(0) = 0$ ,  $h'(x) > 0$  for  $x > 0$  and  $h'(x)$  decreasing on  $[0, \infty)$ . Then  $h(\rho(x, y))$  is a metric. (ii)  $h(x) = x/(x+1)$  satisfies the hypotheses in (i).

*sol.*

(i)  $h(\rho(x, x)) = h(0) = 0$ .

$$h(\rho(x, y)) = h(\rho(y, x))$$

Suppose,  $\rho(x, y) \leq \rho(x, z) + \rho(y, z)$ .

$$\begin{aligned} h(\rho(y, z) + \rho(x, z)) - h(\rho(x, z)) &= \int_{\rho(x, z)}^{\rho(y, z) + \rho(x, z)} h'(x) dx \\ &< \int_0^{\rho(y, z)} h'(x) dx = h(\rho(y, z)) \\ h(\rho(y, z) + \rho(x, z)) &< h(\rho(y, z)) + h(\rho(x, z)) \end{aligned}$$

Thus,

$$h(\rho(x, y)) < h(\rho(y, z)) + h(\rho(x, z))$$

(ii)  $h(0) = 0$

$h'(x) = 1/(x+1)^2 > 0$ . Thus  $h'(x)$  is decreasing.

□

**2.1.4.** Let  $\Omega = (0, 1)$ ,  $\mathcal{F} = \text{Borel sets}$ ,  $P = \text{Lebesgue measure}$ .  $X_n(\omega) = \sin(2\pi n\omega)$ ,  $n = 1, 2, \dots$  are uncorrelated but not independent.

*sol.*

1. Uncorrelated For  $n < m$ ,

$$\begin{aligned} EX_n X_m &= \int X_n X_m dP = \int_0^1 \sin(2\pi n x) \sin(2\pi m x) dx \\ &= \frac{1}{2} \left( \frac{\sin(2\pi(m-n)x)}{m-n} - \frac{\sin(2\pi(m+n)x)}{m+n} \right) \Big|_0^1 \\ &= 0 \end{aligned}$$

$$EX_n = \int_0^1 \sin(2\pi n x) dx = 0$$

Thus,  $EX_n X_m - EX_n EX_m = 0$  for all  $n \neq m$ .

2. Not Independent

$$P(X_4 > 0 | X_2 > 0) = 1 \neq P(X_4 > 0) = 1/2$$

□

**2.1.5.** (i) Show that if  $X$  and  $Y$  are independent with distributions  $\mu$  and  $\nu$  then

$$P(X + Y = 0) = \sum_y \mu(\{-y\}) \nu(\{y\})$$

(ii) Conclude that if  $X$  has continuous distribution  $P(X = Y) = 0$ .

*sol.*

(i)

$$\begin{aligned} P(X + Y = 0) &= P(\cup_y (\{X = -y\} \cap \{Y = y\})) \\ &= \sum_y P(\{X = -y\} \cap \{Y = y\}) \\ &= \sum_y P(\{X = -y\}) P(\{Y = y\}) = \sum_y \mu(\{-y\}) \nu(\{y\}) \end{aligned}$$

(ii) Since  $X$  is continuous, finite sets has measure 0. Thus,

$$P(X + Y = 0) = \sum_y \mu(\{-y\}) \nu(\{y\}) = 0$$

□

**2.1.6.** Prove directly from the definition that if  $X$  and  $Y$  are independent and  $f$  and  $g$  are measurable functions then  $f(X)$  and  $g(Y)$  are independent.

*sol.* Choose  $A, B \in \mathcal{R}$ .

$$\begin{aligned} P(f(X) \in A, g(Y) \in B) &= P(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &= P(X \in f^{-1}(A)) P(Y \in g^{-1}(B)) \\ &= P(f(X) \in A) P(g(Y) \in B) \end{aligned}$$

Since  $f, g$  are measurable,  $f^{-1}(A) \in \sigma(X), g^{-1}(B) \in \sigma(Y)$

□

**2.1.7.** Let  $K \geq 3$  be a prime and let  $X$  and  $Y$  be independent random variables that are uniformly distributed on  $\{0, 1, \dots, K-1\}$ . For  $0 \leq n < K$ , let  $Z_n = X + nY \pmod K$ . Show that  $Z_0, Z_1, \dots, Z_{K-1}$  are pairwise independent, i.e., each pair is independent. They are not independent because if we know the values of two of the variables then we know the values of all the variables.

*sol.*

$$\begin{aligned} P(Z_n(\omega_1) = A, Z_m(\omega_2) = B) &= K^{-2} \sum_i \sum_j 1_{Z_n(\omega_1)=i} 1_{Z_m(\omega_2)=j} \\ &= K^{-2} 1_{\{Z_n(\omega_1)=A\}} 1_{\{Z_m(\omega_2)=B\}} \\ &= K^{-1} 1_{\{Z_n(\omega_1)=A\}} K^{-1} 1_{\{Z_m(\omega_2)=B\}} \end{aligned}$$

□

**2.1.8.** Find four random variables taking values in  $\{-1, 1\}$  so that any three are independent but all four are not.

*sol.* Let  $X_1, X_2, X_3, X_4$  be independent random variables with  $P(X_i = 1) = P(X_i = -1) = 1/2$ . Define  $Y_1 = X_1 X_2, Y_2 = X_2 X_3, Y_3 = X_3 X_4, Y_4 = X_4 X_1$ . Then  $Y_1 Y_2 Y_3 Y_4 = 1$ . Thus all four random variables are not independent. However, all three random variables are independent. □

**2.1.9.** Let  $\Omega = \{1, 2, 3, 4\}$ ,  $\mathcal{F} =$  all subsets of  $\Omega$ , and  $P(\{i\}) = 1/4$ . Give an example of two collections of sets  $\mathcal{A}_1$  and  $\mathcal{A}_2$  that are independent but whose generated  $\sigma$ -fields are not

*sol.* Let  $\mathcal{A}_1 = \{\{1, 2\}\}, \mathcal{A}_2 = \{\{1, 3\}, \{1, 4\}\}$  □

**2.1.10.** Show that if  $X$  and  $Y$  are independent, integer-valued random variables, then

$$P(X + Y = n) = \sum_m P(X = m) P(Y = n - m)$$

*sol.* We can apply ex 2.1.5. (i). □

**2.1.11.** In Example 1.6.13, we introduced the Poisson distribution with parameter  $\lambda$ , which is given by  $P(Z = k) = e^{-\lambda} \lambda^k / k!$  for  $k = 0, 1, 2, \dots$ . Use the previous exercise to show that if  $X = \text{Poisson}(\lambda)$  and  $Y = \text{Poisson}(\mu)$  are independent then  $X + Y = \text{Poisson}(\lambda + \mu)$ .

*sol.* Let  $Z = X + Y$ .

$$\begin{aligned} P(Z = n) &= \sum_m P(X = m) P(Y = n - m) \\ &= \sum_m e^{-\lambda} \lambda^m / m! e^{-\mu} \mu^{n-m} / (n-m)! \\ &= e^{-\lambda-\mu} \mu^n \sum_m \frac{\lambda^m \mu^{-m}}{m! (n-m)!} \\ &= e^{-\lambda-\mu} \mu^n \left(1 + \frac{\lambda}{\mu}\right)^n / n! \\ &= e^{-\lambda-\mu} (\lambda + \mu)^n / n! \end{aligned}$$

Thus  $X + Y = \text{Poisson}(\lambda + \mu)$ . □

**2.1.12.**  $X$  is said to have a Binomial( $n, p$ ) distribution if

$$P(X = m) = \binom{n}{m} p^m (1-p)^{n-m}$$

(i) Show that if  $X = \text{Binomial}(n, p)$  and  $Y = \text{Binomial}(m, p)$  are independent then  $X + Y = \text{Binomial}(n + m, p)$ . (ii) Look at Example 1.6.12 and use induction to conclude that the sum of  $n$  independent Bernoulli( $p$ ) random variables is Binomial( $n, p$ ).

sol.

(i) Let  $Z = X + Y$ .

$$\begin{aligned}
 P(Z = k) &= \sum_i P(X = i)P(Y = k - i) \\
 &= \sum_i \binom{n}{i} p^n (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m+i-k} \\
 &= p^k (1-p)^{n+m-k} \sum_i \binom{n}{i} \binom{m}{k-i} \\
 &= \binom{n+m}{k} p^k (1-p)^{n+m-k}
 \end{aligned}$$

Thus  $X + Y = \text{Binomial}(n + m, p)$ .

(ii) (a)  $\text{Bernoulli}(p) \equiv \text{Binomial}(1, p)$

(b) Suppose sum of  $n$  independent  $\text{Bernoulli}(p)$  random variables is  $\text{Binomial}(n, p)$ .

By (i) and (ii)-(a), then sum of  $n+1$  independent  $\text{Bernoulli}(p)$  random variables is  $\text{Binomial}(n+1, p)$ .

□

**2.1.13.** It should not be surprising that the distribution of  $X + Y$  can be  $F * G$  without the random variables being independent. Suppose  $X, Y \in \{0, 1, 2\}$  and take each value with probability  $1/3$ . (a) Find the distribution of  $X + Y$  assuming  $X$  and  $Y$  are independent. (b) Find all the joint distributions  $(X, Y)$  so that the distribution of  $X + Y$  is the same as the answer to (a).

sol.

(a)

$$\begin{aligned}
 P(X + Y = 0) &= P(\{X = 0\} \cap \{Y = 0\}) = P(\{X = 0\})P(\{Y = 0\}) = 1/9 \\
 P(X + Y = 1) &= \sum_{i \in \{0,1\}} P(\{X = i\} \cap \{Y = 1 - i\}) = \sum_i P(\{X = i\})P(\{Y = 1 - i\}) = 2/9 \\
 P(X + Y = 2) &= \sum_{i \in \{0,1,2\}} P(\{X = i\} \cap \{Y = 2 - i\}) = \sum_i P(\{X = i\})P(\{Y = 2 - i\}) = 3/9 \\
 P(X + Y = 3) &= \sum_{i \in \{1,2\}} P(\{X = i\} \cap \{Y = 3 - i\}) = \sum_i P(\{X = i\})P(\{Y = 3 - i\}) = 2/9 \\
 P(X + Y = 4) &= P(\{X = 2\} \cap \{Y = 2\})P(\{X = 2\})P(\{Y = 2\}) = 1/9
 \end{aligned}$$

(b) Let  $P(X = 0, Y = 1) := p$ . Use sum of each column and row is  $1/3$ .

$X, Y$	0	1	2
0	1/9	p	2/9 - p
1	2/9 - p	1/9	p
2	p	2/9 - p	1/9

□

**2.1.14.** Let  $X, Y \geq 0$  be independent with distribution functions  $F$  and  $G$ . Find the distribution function of  $XY$ .

*sol.* Let  $H$  be the distribution function of  $XY$ .

$$\begin{aligned} H(z) &= P(XY \leq z) \\ &= \int_0^\infty \int_0^\infty 1_{\{XY \leq z\}} dF dG \\ &= \int_0^\infty \int_0^{z/y} 1 dF dG \\ &= \int_0^\infty F(z/y) dG \end{aligned}$$

□

**2.1.15.** If we want an infinite sequence of coin tossings, we do not have to use Kolmogorov's theorem. Let  $\Omega$  be the unit interval  $(0, 1)$  equipped with the Borel sets  $\mathcal{F}$  and Lebesgue measure  $P$ . Let  $Y_n(\omega) = 1$  if  $[2^n \omega]$  is odd and 0 if  $[2^n \omega]$  is even. Show that  $Y_1, Y_2, \dots$  are independent with  $P(Y_k = 0) = P(Y_k = 1) = 1/2$ .

*sol.* Let  $i_n \in \{0, 1\}, n \in \mathbb{N}$ . Define  $x = \sum_{j=1}^m i_j 2^{-j}$ . It is trivial that for each  $j$ ,  $P(Y_j = 1) = P(Y_j = 0) = 1/2$ . For arbitrary  $m \in \mathbb{N}$

$$P(Y_1 = i_1, \dots, Y_m = i_m) = P(\omega \in [x, x + 2^{-m})) = 2^{-m}$$

Therefore,  $Y_1, Y_2, \dots$  are independent

□

## 2.2 Weak Laws of Large Numbers

**2.2.1.** Let  $X_1, X_2, \dots$  be uncorrelated with  $EX_i = \mu_i$  and  $\text{var}(X_i)/i \rightarrow 0$  as  $i \rightarrow \infty$ . Let  $S_n = X_1 + \dots + X_n$  and  $\nu_n = ES_n/n$  then as  $n \rightarrow \infty$ ,  $S_n/n - \nu_n \rightarrow 0$  in  $L^2$  and in probability.

*sol.* Choose a positive number  $\epsilon$ . For given  $\epsilon$ , there exists  $\delta_i$  such that

$$\text{Var}(X_i) < \delta_i + i\epsilon$$

Let  $\delta = \max\{\delta_1, \dots, \delta_n\}$ . Then

$$\sum \text{Var}(X_i) < \sum \delta_i + \epsilon \sum i < n\delta + n^2\epsilon$$

$$\begin{aligned} E(S_n/n - \nu_n)^2 &= \frac{1}{n^2} \sum \text{Var}(X_i) \\ &< \frac{1}{n} \delta + \epsilon \end{aligned}$$

Thus,  $E(S_n/n - \nu_n)^2 \rightarrow 0$  as  $n \rightarrow \infty$  since  $\epsilon$  is arbitrary positive.

$$\begin{aligned} E(S_n/n - \nu_n)^2 \rightarrow 0 &\Rightarrow P((S_n/n - \nu_n)^2 > \epsilon^2) \rightarrow 0 \\ &\Rightarrow P(|(S_n/n - \nu_n)| > \epsilon) \rightarrow 0 \end{aligned}$$

□

**2.2.2.** The  $L^2$  weak law generalizes immediately to certain dependent sequences. Suppose  $EX_n = 0$  and  $EX_n X_m \leq r(n - m)$  for  $m \leq n$  (no absolute value on the left-hand side!) with  $r(k) \rightarrow \infty$  as  $k \rightarrow \infty$ . Show that  $(X_1 + \dots + X_n)/n \rightarrow 0$  in probability.

*sol.*  $L^2$  convergence implies convergence in probability. Thus, it is sufficient to show that the sequences converges in  $L^2$ .

By Cauchy-Schwarz,

$$EX_n X_m \leq (EX_n^2 EX_m^2)^{0.5} = EX_n^2 = r(0)$$

Thus, for all  $i$ ,  $r(i) \leq r(0)$ . Choose a positive number  $\epsilon$ . For given  $\epsilon$ , there exists  $K$  such that for all  $k > K$ ,  $|r(k)| < \epsilon$ . For  $n > K$ ,

$$E((X_1 + \dots + X_n)/n)^2 \leq \frac{1}{n^2} n(2K + 1)r(0) + n^2\epsilon \leq \frac{2K + 1}{n} r(0) + \epsilon$$

Since  $\epsilon$  is arbitrary positive, the sequence converges to 0 in  $L^2$ .

□

---

**2.2.3** (Monte Carlo integration). (i)  $f$  be a measurable function on  $[0, 1]$  with  $\int_0^1 |f(x)| dx < \infty$ . Let  $U_1, U_2, \dots$  be independent and uniformly distributed on  $[0, 1]$ , and let

$$I_n = n^{-1}(f(U_1) + \dots + f(U_n))$$

Show that  $I_n \rightarrow I \equiv \int_0^1 f dx$  in probability. (ii) Suppose  $\int_0^1 |f(x)|^2 dx < \infty$ . Use Chebyshev's inequality to estimate  $P(|I_n - I| > a/n^{1/2})$ .

*sol.*

(i) By assumption,  $Ef(U_i)$  exists. Thus,

$$\sum f(U_i)/n - \int_0^1 f dx \xrightarrow{p} 0$$

Thus, the sum converges to  $I \equiv \int_0^1 f dx$  in probability.

(ii) By Chebyshev's inequality

$$\begin{aligned} P(|I_n - I| > a/n^{1/2}) &\leq n \text{Var}(I_n)/a^2 = \text{Var}(f(U_i))/a^2 \\ &= \frac{1}{a^2} \left[ \int_0^1 f^2 dx - \left( \int_0^1 f dx \right)^2 \right] \end{aligned}$$

□

**2.2.4.** Let  $X_1, X_2, \dots$  be i.i.d. with  $P(X_i = (-1)^k k) = C/k^2 \log k$  for  $k \geq 2$  where  $C$  is chosen to make the sum of the probabilities = 1. Show that  $E|X_i| = \infty$ , but there is a finite constant  $\mu$  so that  $S_n/n \rightarrow \mu$  in probability.

*sol.*  $E|X_i| = \infty$

$$E|X_i| = \sum_{k=2}^{\infty} C/k \log k$$

Thus, we can check convergence by whether the below integral converges

$$\int_2^{\infty} C/x \log x dx = C \log \log x|_2^{\infty} = \infty$$

Thus,  $E|X_i| = \infty$

$$\begin{aligned} P(|X| > n) &= \sum_{k=n+1}^{\infty} C/k^2 \log k \leq C/\log n \int_n^{\infty} 1/n^2 = C/n^2 \log n \\ &\Rightarrow nP(|X| > n) \leq C/\log n \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Thus,  $S_n/n$  converges  $\lim EX_i 1_{\{X_i \leq n\}}$ .

$$\begin{aligned} \lim EX_i 1_{\{X_i \leq n\}} &= \lim \sum_{k=2}^n (-1)^k C/k \log k \\ &= \sum_{k=2}^{\infty} (-1)^k C/k \log k \end{aligned}$$

The series converges since  $C/k \log k$  is decreasing and converges to 0. □

**2.2.5.** Let  $X_1, X_2, \dots$  be i.i.d with  $P(X_i > x) = e/x \log x$  for  $x \geq e$ . Show that  $E|X_i| = \infty$ , but there is a sequece of constants  $\mu_n \rightarrow \infty$  so that  $S_n/n - \mu_n \rightarrow 0$  in probability.

*sol.*  $X_i$  is a non-negative random variable.

$$E|X_i| = EX_i = \int_e^\infty e/x \log x dx = e \log \log x|_e^\infty = \infty$$

$$xP(X_i > x) = e/\log x \rightarrow 0 \text{ as } x \rightarrow \infty$$

Thus,

$$\begin{aligned} \mu_n &= EX_i 1_{\{X_i \leq n\}} \\ &= \int_e^n e/x \log x dx = e \log \log n \end{aligned}$$

□

**2.2.6.** (i) Show that if  $X \geq 0$  is integer valued  $EX = \sum_{n \geq 1} P(X \geq n)$ . (ii) Find a similar expression for  $EX^2$ .

*sol.*  $X$  is integer valued variable.

$$X = \sum_{i=1}^{\infty} I(X \geq i) \Rightarrow EX = \sum_{i=1}^{\infty} P(X \geq i)$$

$$X^2 = \sum_{i=1}^{\infty} (2i-1)I(X \geq i) \Rightarrow EX^2 = \sum_{i=1}^{\infty} (2i-1)P(X \geq i)$$

□

**2.2.7.** Generalize Lemma 2.2.13 to conclude that if  $H(x) = \int_{(-\infty, x]} h(y) dy$  with  $h(y) \geq 0$ , then

$$EH(X) = \int_{-\infty}^{\infty} h(y) P(X \geq y) dy$$

An important special case is  $H(x) = \exp(\theta x)$  with  $\theta > 0$ .

*sol.*

$$\begin{aligned} EH(x) &= \int H(X) dP = \int \int_{(-\infty, x]} h(y) dy dP \\ &= \int \int_{[y, \infty)} h(y) dP dy \\ &= \int h(y) P(X \geq y) dy \end{aligned}$$

□

**2.2.8** (An unfair “fair game”). Let  $p_k = 1/2^k k(k+1)$ ,  $k = 1, 2, \dots$  and  $p_0 = 1 - \sum_{k \geq 1} p_k$ .

$$\sum_{k=1}^{\infty} 2^k p_k = (1 - 1/2) + (1/2 - 1/3) + \dots = 1$$

so if we let  $X = 1, X_2, \dots$  be i.i.d. with  $P(X_n = -1) = p_0$  and

$$P(X_n = 2^k - 1) = p_k \text{ for } k \geq 1$$

then  $EX_n = 0$ . Let  $S_n = X_1 + \dots + X_n$ . Use Theorem 2.2.11 with  $b_n = 2^{m(n)}$  where  $m(n) = \min\{m : 2^{-m} m^{-3/2} \leq n^{-1}\}$  to conclude that

$$S_n/(n/\log_2 n) \rightarrow -1 \text{ in probability}$$

*sol.*

□

**2.2.9** (Weak law for positive variables). Suppose  $X_1, X_2, \dots$  are i.i.d.,  $P(0 \leq X_i < \infty) = 1$  and  $P(X_i > x) > 0$  for all  $x$ . Let  $\mu(s) = \int_0^s x dF(x)$  and  $\nu(s) = \mu(s)/s(1-F(s))$ . It is known that there exist constants  $a_n$  so that  $S_n/a_n \rightarrow 1$  in probability, if and only if  $\nu(s) \rightarrow \infty$  as  $s \rightarrow \infty$ . Pick  $b_n \geq 1$  so that  $n\mu(b_n) = b_n$  (this works for large  $n$ ), and use Theorem 2.2.11 to prove that the condition is sufficient.

*sol.*

□



## 2.3 Borel-Cantelli Lemmas

**2.3.1.** Prove that  $P(\limsup A_n) \geq \limsup P(A_n)$  and  $P(\liminf A_n) \leq \liminf P(A_n)$ .

*sol.*  $\liminf A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k$  and  $\limsup A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$

$$\begin{aligned} P(\liminf A_n) &= \lim P(\bigcap_{k \geq n} A_k) = \liminf P(\bigcap_{k \geq n} A_k) \leq \liminf P(A_n) \\ P(\limsup A_n) &= \lim P(\bigcup_{k \geq n} A_k) = \limsup P(\bigcup_{k \geq n} A_k) \geq \limsup P(A_n) \end{aligned}$$

□

**2.3.2.** Prove the first result in Theorem 2.3.4 directly from the definition.

*sol.* For arbitrary  $\varepsilon > 0$ , there exists  $\delta > 0$  which implies

$$|X_n - X| < \delta \Rightarrow |f(X_n) - f(X)| < \varepsilon$$

Thus,  $X_n \rightarrow X$  in probability implies  $f(X_n) \rightarrow f(X)$  in probability. □

**2.3.3.** Let  $\ell_n$  be the length of the head run at time  $n$ . See Example 8.3.1. for the precise definition. Show that  $\limsup_{n \rightarrow \infty} \ell_n / \log_2 n = 1$ ,  $\liminf_{n \rightarrow \infty} \ell_n = 0$  a.s.

**2.3.4** (Fatou's lemma). Suppose  $X_n \geq 0$  and  $X_n \rightarrow X$  in probability. Show that  $\liminf_{n \rightarrow \infty} EX_n \geq EX$ .

*sol.* Pick a subsequence  $X_{n(m)}$  which satisfies that

$$\lim_{m \rightarrow \infty} EX_{n(m)} = \liminf_{n \rightarrow \infty} EX_n$$

Since  $X_n$  converges in probability, we can find a further subsequence  $X_{n(m_k)}$  such that  $X_{n(m_k)} \rightarrow X$  a.s. If we apply Fatou's lemma,

$$EX \leq \liminf_{k \rightarrow \infty} EX_{n(m_k)} \leq \liminf_{n \rightarrow \infty} EX_n$$

□

**2.3.5** (Dominated convergence). Suppose  $X_n \rightarrow X$  in probability and (a)  $|X_n| \leq Y$  with  $EY < \infty$  or (b) there is a continuous function  $g$  with  $g(x) > 0$  for large  $x$  with  $|x|/g(x) \rightarrow 0$  as  $|x| \rightarrow \infty$  so that  $Eg(X_n) \leq C < \infty$  for all  $n$ . Show that  $EX_n \rightarrow EX$ .

*sol.* Convergence in probability implies pointwise convergence. Thus, we need a dominating function to apply Dominated Convergence Theorem.

(a)  $Y$  dominates  $X_n$ .

(b) Let  $Z = g(X)$ . Then,  $Z$  dominates  $X_n$ .

Thus, we can apply DCT. □

**2.3.6** (Metric for convergence in probability). Show (a) that  $d(X, Y) = E(|X - Y|/(1 + |X - Y|))$  defines a metric on the set of random variables, i.e., (i)  $d(X, Y) = 0$  if and only if  $X = Y$  a.s., (ii)  $d(X, Y) = d(Y, X)$ , (iii)  $d(X, Z) \leq d(X, Y) + d(Y, Z)$  and (b) that  $d(X_n, X) \rightarrow 0$  as  $n \rightarrow \infty$  if and only if  $X_n \rightarrow X$  in probability.

*sol.*

(a) (i)  $X = Y$  a.s.  $\Leftrightarrow d(X, Y) = 0$

(ii)  $d(X, Y) = E(|X - Y|/(1 + |X - Y|)) = E(|Y - X|/(1 + |Y - X|)) = d(Y, X)$

(iii)  $|X - Z| = |X - Y + Y - Z| \leq |X - Y| + |Y - Z|$ . By result of 2.1.3,  $d(X, Z) \leq d(X, Y) + d(Y, Z)$ .

(b) ( $\Rightarrow$ ) For given  $\epsilon > 0$ ,

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(|X_n - X|/(1 + |X_n - X|) > \epsilon/(1 + \epsilon)) \\ &< d(X_n, X)(1 + \epsilon)/\epsilon \rightarrow 0 \quad \because \text{Chebychev} \end{aligned}$$

Thus,  $X_n \rightarrow X$  in probability.

( $\Leftarrow$ )  $X_n \rightarrow X$  in probability and  $|d(\cdot, \cdot)| \leq 1$ . Thus,  $d(X_n, X) \rightarrow d(X, X) = 0$  by DCT. □

**2.3.7.** Show that random variables are a complete space under the metric defined in the previous exercise, i.e., if  $d(X_m, X_n) \rightarrow 0$  whenever  $m, n \rightarrow \infty$  then there is a r.v.  $X_\infty$  so that  $X_n \rightarrow X_\infty$  in probability.

*sol.* Choose  $N_k$  so that if  $n, m \geq N_k$ , then  $d(X_n, X_m) \leq 2^{-k}$ . Pick a subsequence  $\{X_{n_k}\}$  which satisfies  $n_k > N_k$ . Then

$$P(|X_{n_{k+1}} - X_{n_k}| > k^{-2}) \leq d(X_{n_{k+1}}, X_{n_k}) \frac{1 + k^{-2}}{k^{-2}} \leq 2^{-k}(1 + k^2)$$

The last term is summable so by the Borel Cantelli Lemma,  $\limsup P(|X_{n_{k+1}} - X_{n_k}| > k^{-2}) = 0$ . Thus, there exists  $X$  such that  $X_{n_k} \rightarrow X$  in probability. Let  $X'$  be the limit of another subsequence. By the Cauchy assumption,  $d(X_{n_k}, X_{m_k}) \rightarrow 0$ . The metric is defined as expectation. By exercise 2.3.4,  $d(X, X') = 0$ . Thus the random variable exists. □

**2.3.8.** Let  $A_n$  be a sequence of independent events with  $P(A_n) < 1$  for all  $n$ . Show that  $P(\cup A_n) = 1$  implies  $\sum_n P(A_n) = \infty$  and hence  $P(A_n \text{ i.o.}) = 1$ .

*sol.* By second Borel Cantelli Lemma, it is sufficient to show that  $P(\cup A_n) = 1$  implies  $\sum_n P(A_n) = \infty$ . Suppose  $\sum_n P(A_i)$  is finite. Thus, for some  $N$ ,  $P(A_n) = 0$  if  $n > N$ .

$$\begin{aligned} P(\cup A_n) = 1 &\Rightarrow P(\cap A_n^c) = 0 \\ &\Rightarrow \prod_n [1 - P(A_n)] = 0 \end{aligned}$$

However,  $\prod_n [1 - P(A_n)]$  converges to  $\prod_{n=1}^N [1 - P(A_n)] > 0$ . Thus, there is contradiction. □

**2.3.9.** (i) If  $P(A_n) \rightarrow 0$  and  $\sum_{n=1}^\infty P(A_n^c \cap A_{n+1}) < \infty$  then  $P(A_n \text{ i.o.}) = 0$ . (ii) Find an example of a sequence  $A_n$  to which the result in (i) can be applied but the Borel - Cantelli lemma cannot.

*sol.* (i) Let  $B_n = A_n^c \cap A_{n+1}$ . The partial sum of  $P(B_n)$  goes to 0.

$$\begin{aligned} P\left(\bigcup_{n=m}^\infty A_i\right) &\leq P(A_m) + \sum_{n=m}^\infty P(B_i) \rightarrow 0 \text{ as } n \rightarrow \infty \\ P(\limsup A_n) &= \inf P\left(\bigcup_{n=m}^\infty A_i\right) = 0 \end{aligned}$$

(ii) Let  $A_n := [0, n^{-1}]$ . Then the lemma can not be applied since  $\sum P(A_n) = \infty$ . However,  $P(A_n) \rightarrow 0$ ,  $P(A_n^c \cap A_{n+1}) \rightarrow 0$ . □

**2.3.10** (Kochen-Stone lemma). Suppose  $\sum P(A_k) = \infty$ . Use Exercises 1.6.6 and 2.3.1 to show that if

$$\limsup_{n \rightarrow \infty} \left( \sum_{k=1}^n P(A_k) \right)^2 / \left( \sum_{1 \leq j, k \leq n} P(A_j \cap A_k) \right) = \alpha > 0$$

then  $P(A_n \text{ i.o.}) \geq \alpha$ . The case  $\alpha = 1$  contains Theorem 2.3.7

*sol.* Define  $X_n := \sum_{i=1}^n 1\{A_i\}$ . Then,  $EX_n = \sum_{i=1}^n P(A_i)$ ,  $EX_n^2 = \sum_{1 \leq i, j \leq n} P(A_i \cap A_j)$ . By exercise 1.6.6.

$$P(X_n \geq 0) \geq (EX_n)^2 / EX_n^2 \Rightarrow \limsup P(X_n \geq 0) \geq \limsup (EX_n)^2 / EX_n^2 = \alpha$$

□

**2.3.11.** Let  $X_1, X_2, \dots$  be independent with  $P(X_n = 1) = p_n$  and  $P(X_n = 0) = 1 - p_n$ . Show that (i)  $X_n \rightarrow 0$  in probability if and only if  $p_n \rightarrow 0$ , and (ii)  $X_n \rightarrow 0$  a.s. if and only if  $\sum p_n < \infty$ .

*sol.*

(i) For given  $0 < \epsilon < 1$ ,

$$P(|X_n| > \epsilon) = P(X_n > \epsilon) = P(X_n = 1) = p_n$$

Thus,  $X_n \rightarrow 0 \Leftrightarrow p_n \rightarrow 0$ .

(ii)  $(\Rightarrow)$   $X_n \rightarrow 0$  a.s. implies, for some  $N$ ,  $X_n = 0$  if  $n > N$ . Thus,  $\sum_n p_n = \sum_{n=1}^N p_n < \infty$ .

$(\Leftarrow)$   $\sum p_n < \infty \Rightarrow$  for some  $N$ ,  $p_n = 0$  if  $n > N$ . Thus  $X_n \rightarrow 0$  a.s. □

**2.3.12.** Let  $X_1, X_2, \dots$  be a sequence of r.v.'s on  $(\Omega, \mathcal{F}, P)$  where  $\Omega$  is a countable set and  $\mathcal{F}$  consists of all subsets of  $\Omega$ . Show that  $X_n \rightarrow X$  in probability implies  $X_n \rightarrow X$  a.s.

*sol.* Since  $\Omega$  is countable, the image of a random variable is countable. And all subsets are measurable by assumption. For every  $\epsilon > 0$ ,

$$\lim P(|X_n - X| < \epsilon) = 1$$

Thus, for any  $\omega \in \Omega$ ,  $|X_n(\omega) - X(\omega)| < \epsilon$ . This implies  $X_n \rightarrow X$  a.s. □

**2.3.13.** If  $X_n$  is any sequence of random variables, there are constants  $c_n \rightarrow \infty$  so that  $X_n/c_n \rightarrow 0$  a.s.

*sol.* For every  $\epsilon > 0$ , choose  $m$  s.t.  $\epsilon < 2^{-m}$ . Choose  $c_n$  satisfies that  $P(X_n > c_n 2^{-n}) \leq 2^{-n}$ . Then  $\sum P(X_n/c_n > \epsilon) < \sum P(X_n/c_n > 2^{-m}) < \infty$ . Thus,  $P(\limsup X_n/c_n > 0) = 0$ . □

**2.3.14.** Let  $X_1, X_2, \dots$  be independent. Show that  $\sup X_n < \infty$  a.s. if and only if  $\sum_n P(X_n > A) < \infty$  for some  $A$ .

*sol.*

$(\Rightarrow)$  Let  $A > \sup X_n$ . Then  $\sum P(X_n > A) = 0$ .

$(\Leftarrow)$   $\sum P(X_n > A) < \infty \Rightarrow P(\limsup X_n > A) = 0$ . Thus  $\sup X_n < \infty$  a.s. □

**2.3.15.** Let  $X_1, X_2, \dots$  be i.i.d. with  $P(X_i > x) = e^{-x}$ , let  $M_n = \max_{1 \leq m \leq n} X_m$ . Show that (i)  $\limsup_{n \rightarrow \infty} X_n / \log n = 1$  a.s. and (ii)  $M_n / \log n \rightarrow 1$  a.s.

*sol.*

(i)  $P(X_n \geq \log n) = 1/n \Rightarrow \sum P(X_n \geq \log n) = \infty$ . Thus  $P(X_n / \log n \geq 1 \text{ i.o.}) = 1$ . However, for any  $\epsilon > 0$ ,  $P(X_n \geq (1 + \epsilon) \log n) = n^{-1-\epsilon}$ . Thus the sum of probability is 0. Then  $P(X_n / \log n \geq 1 + \epsilon \text{ i.o.}) = 0$ .

(ii) By (i),  $M_n \leq (1 + \epsilon) \log n \Rightarrow \limsup M_n / \log n \leq 1$ .

$$\begin{aligned} P(M_n < (1 - \epsilon) \log n) &= (1 - n^{-1+\epsilon})^n \\ &\leq \exp(-n^\epsilon) \end{aligned}$$

Thus, by Borel Cantelli,  $P(M_n < (1 - \epsilon) \log n \text{ i.o.}) = 0 \Rightarrow \liminf M_n / \log n \geq 1$ . □

**2.3.16.** Let  $X_1, X_2, \dots$  be i.i.d. with distribution  $F$ , let  $\lambda_n \uparrow \infty$ , and let  $A_n = \{\max_{1 \leq m \leq n} X_m \lambda_n\}$ . Show that  $P(A_n \text{ i.o.}) = 0$  or 1 according as  $\sum_{n \geq 1} (1 - F(\lambda_n)) < \infty$  or  $\infty$ .

*Proof.* By Borel Cantelli lemma,  $\sum_{n \geq 1} (1 - F(\lambda_n)) < \infty$  implies  $P(X_n > \lambda_n \text{ i.o.}) = 0$ . Thus,  $P(A_n \text{ i.o.}) = 0$ . Otherwise,  $\sum_{n \geq 1} (1 - F(\lambda_n)) = \infty$  implies  $P(X_n > \lambda_n \text{ i.o.}) = 1$  by second Borel Cantelli lemma. Thus,  $P(A_n \text{ i.o.}) = 1$ . □

---

**2.3.17.** Let  $Y_1, Y_2, \dots$  be i.i.d. Find necessary and sufficient conditions for

- (i)  $Y_n/n \rightarrow 0$  almost surely
- (ii)  $(\max_{m \leq n} Y_m)/n \rightarrow 0$  almost surely
- (iii)  $(\max_{m \leq n} Y_m)/n \rightarrow 0$  in probability
- (iv)  $Y_n/n \rightarrow 0$  in probability

**2.3.18.** Let  $0 \leq X_1 \leq X_2 \leq \dots$  be random variables with  $EX_n \sim an^\alpha$  with  $a, \alpha > 0$ , and  $\text{var}(X_n) \leq Bn^\beta$  with  $\beta < 2\alpha$ . Show that  $X_n/n^\alpha \rightarrow a$  a.s.

*sol.* Let  $Y_n := X_n/n^\alpha$ . Then  $EY_n \sim \alpha$  and  $\text{var}(Y_n) \leq Bn^{\beta-2\alpha}$ . By Chebyshev,

$$P(|Y_n - \alpha| > \epsilon) \leq \text{Var}(Y_n)/\epsilon^2$$

Since the right term is summable,  $P(|Y_n - \alpha| > \epsilon, \text{ i.o.}) = 0$ . Thus  $X_n/n^\alpha \rightarrow a$  a.s.  $\square$

**2.3.19.** Let  $X_n$  be independent Poisson r.v.'s with  $EX_n = \lambda_n$ , and let  $S_n = X_1 + \dots + X_n$ . Show that if  $\sum \lambda_n = \infty$  then  $S_n/ES_n \rightarrow 1$  a.s.

*sol.*  $\text{Var}(S_n) = \sum \text{Var}(X_n) = \sum \lambda_n$ .

$$P(|S_n/ES_n - 1| > \epsilon) \leq \epsilon^{-2} / \sum \lambda_n$$

$\square$

**2.3.20.** Show that if  $X_n$  is the outcome of the  $n$ th play of the St. Petersburg game (Example 2.2.16) then  $\limsup_{n \rightarrow \infty} X_n/(n \log_2 n) = \infty$  a.s. and hence the same result holds for  $S_n$ . This shows that the convergence  $S_n/(n \log_2 n) \rightarrow 1$  in probability proved in Section 2.2. does not occur a.s.

## 2.4 Strong Law of Large Numbers

**2.4.1** (Lazy janitor). Suppose the  $i$ th light bulb burns for an amount of time  $X_i$  and then remains burned out for time  $Y_i$  before being replaced. Suppose the  $X_i, Y_i$  are positive and independent with the  $X$ 's having distribution  $F$  and the  $Y$ 's having distribution  $G$ , both of which have finite mean. Let  $R_t$  be the amount of time in  $[0, t]$  that we have a working light bulb. Show that  $R_t/t \rightarrow EX_i/(EX_i + EY_i)$  almost surely.

*sol.* Let  $\underline{R}_t = \sum_{n \leq N(t)} X_n$ ,  $\bar{R}_t = \sum_{n \leq N(t)+1} X_n$ . Then,  $\underline{R}_t \leq R_t \leq \bar{R}_t$ . SLLN implies

$$\lim \underline{R}_t/N(t) = \lim \bar{R}_t/(N(t) + 1) = EX_1 \text{ a.s.}$$

Also

$$\lim N(t)/t = \lim (N(t) + 1)/t = 1/(EX_1 + EY_1) \text{ a.s.}$$

Thus,

$$\lim \underline{R}_t/t = \lim \bar{R}_t/t = EX_1/(EX_1 + EY_1) \text{ a.s.}$$

Finally

$$R_t/t \rightarrow EX_1/(EX_1 + EY_1) \text{ a.s.}$$

$\square$

**2.4.2.** Let  $X_0 = (1, 0)$  and define  $X_n \in \mathbb{R}^2$  inductively by declaring that  $X_{n+1}$  is chosen at random from the ball of radius  $|X_n|$  centered at the origin, i.e.,  $X_{n+1}/|X_n|$  is uniformly distributed on the ball of radius 1 and independent of  $X_1, \dots, X_n$ . Prove that  $n^{-1} \log |X_n| \rightarrow c$  a.s. and compute  $c$

*sol.* By assumption  $|X_1|, |X_2|/|X_1|, |X_3|/|X_2|, \dots$  have are i.i.d. with distribution  $U$  which  $P(U < u) = u^2$  for  $u \in [0, 1]$ . Thus,  $|X_n| = U_1 U_2 \dots U_n$ . Then

$$\frac{\log |X_n|}{n} = \frac{\sum_{m=1}^n \log U_m}{n} \rightarrow E \log U_m$$

$E \log U_m = \int_0^1 2u \log u du = -1/2$ . Thus,  $c = -1/2$ .  $\square$

**2.4.3** (Investment problem). We assume that at the beginning of each year totu can buy bonds for \$1 that are worth \$  $a$  at the end of the year or stocks that are worth a random amount  $V \geq 0$ . If you always invest a fixed proportion  $p$  of your wealth in bonds, then your wealth at the end of year  $n+1$  is  $W_{n+1} = (ap + (1-p)V_n)W_n$ . Suppose  $V_1, V_2, \dots$  are i.i.d. with  $EV_n^2 < \infty$  and  $E(V_n^{-2}) < \infty$ . (i) Show that  $n^{-1} \log W_n \rightarrow c(p)$  a.s. (ii) Show that  $c(p)$  is concave. (iii) By investigating  $c'(0)$  and  $c'(1)$ , give conditions on  $V$  that guarantee that the optimal choice of  $p$  is in  $(0,1)$ . (iv) Suppose  $P(V=1) = P(V=4) = 1/2$ . Find the optimal  $p$  as a function of  $a$ .

sol.

$$\begin{aligned} \log W_{n+1} &= \log W_n + \log(ap + (1-p)V_n) \\ &= \sum_{i=1}^n \log(ap + (1-p)V_i) \end{aligned}$$

By SLLN

$$\frac{\log W_n}{n} \rightarrow E \log(ap + (1-p)V_1) \equiv c(p)$$

Concavity of  $c(p)$

$$\begin{aligned} c'(p) &= E \left( \frac{a - V_1}{ap + (1-p)V_1} \right) \\ c''(p) &= -E \left( \frac{((a - V_1)^2)}{(ap + (1-p)V_1)^2} \right) \end{aligned}$$

$c(p)$  is concave since  $c$  is twice differentiable and second derivative is negative.

To have a maximum in  $(0,1)$ ,  $c'(0) > 0$  and  $c'(1) < 0$ . Thus,

$$aEV_1^{-1} > 1, \quad EV_1 > a$$

Using above condition, optimal  $p$  lies in  $(0,1)$  when  $8/5 < a < 5/2$ . Optimal  $p$  satisfies  $c'(p) = 0$ . Thus, we can obtain that

$$p^* = \frac{5a - 8}{-2a^2 + 6a - 8}$$

If  $a \leq 8/5$ , then  $p^* = 0$  and if  $a \geq 5/2$ , then  $p^* = 1$  □

## 2.5 Convergence of Random Series\*

**2.5.1.** Suppose  $X_1, X_2, \dots$  are i.i.d. with  $EX_i = 0$ ,  $\text{var}(X_i) = C < \infty$ . Use Theorem 2.5.5 with  $n = m^\alpha$  where  $\alpha(2p-1) > 1$  to conclude that if  $S_n = X_1 + \dots + X_n$  and  $p > 1/2$  then  $S_n/n^p \rightarrow 0$  almost surely.

**2.5.2.**

**2.5.3.**

**2.5.4.**

**2.5.5.**

**2.5.6.**

**2.5.7.**

**2.5.8.**

**2.5.9.**

**2.5.10.**

**2.5.11.**

**2.5.12.**