MATH 743
SPECTRAL THEORY OF ANDERSON TYPE HAMILTONIANS
Instructor: Vojkan Jaksic
LECTURE 4

Preliminaries 4: Probability theory

O.1 References

[Bur] Burill C.W.: *Measure, Integration, and Probability*, McGraw-Hill, 1972.

[Lo] Loeve M.: *Probability Theory*, Van Nostrand, 1963.

[Var] Varadhan S.R.S.: *Probability Theory*, Courant Lecture Notes 7.

[Si] Simon B.: *Functional Integration and Quantum Physics*, Academic Press 1979.

[AA] Arnold V.I., Avez A.: *Ergodic Problems in Classical Mechanics*, Addison-Wesley 1989.

1.1. Introduction

In this lecture we will review some basic notions of probability theory. The material of this lecture has been covered in depth last year in MATH 587 and 589. Since all the students have taken these courses, I will not prove much in this lecture (for the proofs you should consult [Bur] or your lecture notes from the last year). In Subsection 1.9 we will deal with a topic we have not discussed in MATH 587-589 (conditional probability distributions), and I will provide complete proofs.

The goal of the homework is to remind you of some material taught in MATH 587 and 589.

1.2. Basic notions

Throughout the lecture we will adopt the usual notation of probability theory. Let Ω be a set. We denote the elements of Ω by ω and call them *the elementary events*. Let $\mathcal{F}$ be a σ-algebra in Ω. The sets in $\mathcal{F}$ are called *events* (note then an elementary event may not be an event). Let P be a probability measure on $(\Omega, \mathcal{F})$. The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

A measurable function $X : \Omega \to \mathbb{R}$ is called a random variable. If X is integrable, $\mathbb{E}(X) := \int_\Omega X \mathrm{d}P$ is called the expectation of X. The variance of X is defined by

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

A measurable function $X : \Omega \to \mathbb{C}$ is called a *complex* random variable. A measurable function $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ is called an *extended* random variable.

Let X be a random variable. For $B \subset \mathbb{R}$ Borel let

$$\mu_X(B) := P(X^{-1}(B)).$$

μ_X is a Borel probability measure on $\mathbb{R}$ called the probability distribution of X. For any positive Borel function f,

$$\int_{\mathbb{R}} f \mathrm{d}\mu_X = \int_{\Omega} f(X) \mathrm{d}P.$$

In particular, $f(X)$ has expectation iff $f \in L^1(\mathbb{R}, \mathrm{d}\mu_X)$ and in this case $\mathbb{E}(f(X)) = \int_{\mathbb{R}} f \mathrm{d}\mu_X$.

For $x \in \mathbb{R}$ let $F_X(x) := P(X \leq x) = P(X^{-1}(-\infty, x])$. The function F_X is increasing, continuous from the right, $F_X(-\infty) = 0$ and $F_X(\infty) = 1$. F_X is called the probability distribution function of X and $\mathrm{d}\mu_X$ is equal to the Stieltjes measure $\mathrm{d}F_X$.

A random variable X is called absolutely continuous if $\mathrm{d}\mu_X = f(x)\mathrm{d}x$. The function f is called the density of X. The following three examples will be often mentioned.

(1) Let $[a, b]$ be a finite interval. A random variable X is uniformly distributed on $[a, b]$ if $\mathrm{d}\mu_X = (b-a)^{-1}\chi_{[a,b]}\mathrm{d}x$.

(2) Let $m \in \mathbb{R}$ and $\sigma > 0$. A random variable X has the normal $\mathcal{N}(m, \sigma^2)$ distribution if

$$\mathrm{d}\mu_X = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \mathrm{d}x.$$

Note that $\mathbb{E}(X) = m$, $\mathrm{Var}(X) = \sigma^2$.

(3) Let $a \in \mathbb{R}$, $b > 0$. A random variable X has the Cauchy $\mathcal{C}(a, b)$ distribution if

$$\mathrm{d}\mu_X = \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2} \mathrm{d}x.$$

Note that $\mathbb{E}(|X|) = \infty$.

The normal and Cauchy random variables have some special properties which are discussed in the homework.

A random variable X is called discrete if μ_X is a pure point measure. The set of atoms of μ_X is precisely the set of values of X and $P(X = x) = \mu_X(\{x\})$.

(4) Let n be a positive integer, $p \in [0, 1]$, $q = 1-p$. A random variable X has the binomial $B(n, p)$ distribution if

$$P(X = k) = \binom{n}{k} p^k q^{n-k},$$

for $k = 0, 1, \cdots, n$.

(5) Let $\lambda > 0$. A random variable X has the Poisson $\mathcal{P}(\lambda)$ distribution if

$$P(X = k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!},$$

for $k = 0, 1, \cdots$. Poisson random variables will play an important role in Lecture 11.

A measurable function $X : \Omega \to \mathbb{R}^n$ is called a *random vector.* Obviously, $X = (X_1, \cdots, X_n)$, where X_k are random variables. The probability distribution of X is

$$\mu_X(B) = P(X^{-1}(B)), \qquad B \subset \mathbb{R}^n \text{ Borel},$$

and for any positive Borel function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} f \mathrm{d}\mu_X = \int_\Omega f(X)\mathrm{d}P.$$

We will sometimes write $\mu_{X_1, \cdots, X_n}$ for μ_X and call it the joint probability distribution of $X_1, \cdots, X_n$. A random vector is called absolutely continuous if $\mathrm{d}\mu_X = f(x)\mathrm{d}x$. The function f is called the density of X. For $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ let

$$F_X(x) := P(X_1 \le x_1, \cdots, X_n \le x_n).$$

The function F_X is called the probability distribution function of X (or the joint probability distribution function of $X_1, \cdots, X_n$) and $\mathrm{d}\mu_X$ is equal to the Stieltjes measure $\mathrm{d}F_X$.
(6) Let $X = (X_1, \cdots, X_n)$ be a random vector, $m_k = \mathbb{E}(X_k)$, $m = (m_1, \cdots, m_n)$. Let $A = [a_{ij}]$ be a strictly positive matrix. X is called a normal random vector with variance A if

$$\mathrm{d}\mu_X(x) = (2\pi)^{-n/2}(\det A)^{-1/2}\mathrm{e}^{-\frac{1}{2}(x-m|A^{-1}(x-m))}.$$

The components $X_1, \cdots, X_n$ of X are called *jointly Gaussian* random variables.

Let $\{X_i\}_{i\in I}$ be a family of random variables. We will denote by $\sigma(\{X_i\}_{i\in I})$ the minimal σ-field w.r.t. which all the random variables X_i are measurable. $\sigma(\{X_i\}_{i\in I})$ is generated by the sets $\{X_i^{-1}(B) : i \in I, B \subset \mathbb{R} \text{ Borel}\}$.

1.3. Isomorphism of probability spaces

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Two events $E_1, E_2 \in \mathcal{F}$ are called equivalent iff $P((E_1 \setminus E_2) \cup (E_2 \setminus E_1)) = 0$. The set of the corresponding classes of equivalences, $\mathcal{C}$, inherits the notions of intersection and union.

An isomorphism of two probability spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$ is a bijection $T : \mathcal{C} \to \mathcal{C}'$ respecting countable intersections and unions such that $P' \circ T = P$. Two random variables X and X' correspond under T if $T(X^{-1}(B)) = X'^{-1}(B)$ for all Borel sets $B \subset \mathbb{R}$.

Let $\{X_k\}_{1\le k\le n}$ be random variables on Ω and $\mathcal{D} := \sigma(\{X_k\})$. The probability space $(\Omega, \mathcal{D}, P)$ encodes all information concerning $\{X_k\}$. Set $\Omega' = \mathbb{R}^n$, $\mathcal{F}' =$ Borel σ-algebra in $\mathbb{R}^n$, $P' = \mu_{X_1, \cdots, X_n}$. For $\omega' = (\omega'_1, \cdots, \omega'_k)$ let $X'_k(\omega) = \omega_k$. The map

$$T(X_1^{-1}(B_1) \cap \cdots \cap X_n^{-1}(B_n)) = B_1 \times \cdots \times B_n,$$

extends to an isomorphism of probability spaces $(\Omega, \mathcal{D}, P)$ and $(\Omega', \mathcal{F}', P')$ and X_k correspond to X_k' under this isomorphism.

A similar construction applies to any family of random variables. We will restrict ourselves to countable families. Let $\{X_n\}_{n\in\Gamma}$ be a countable family of random variables and $\mathcal{D} = \sigma(\{X_n\})_{n\in\Gamma}$. Let $\Omega' = \mathbb{R}^\Gamma$ and let $\mathcal{F}'$ be the Borel σ-algebra in Ω'. Recall that $\mathcal{F}'$ is generated by the cylinders

$$F = \{\omega' : (\omega'(i_1), \omega'(i_2), \cdots, \omega'(i_k)) \in B\} \tag{1}$$

where B is a Borel set in $\mathbb{R}^k$. Set

$$P'(F) := \mu_{X_{i_1}, \cdots, X_{i_k}}(B).$$

By the Daniell-Kolmogorov theorem (Theorem 5-6 C in [Bur]), P' is well-defined and uniquely extends to a probability measure on $(\Omega', \mathcal{F}')$. Set $X_n'(\omega') = \omega'(n)$. Let T be defined by

$$T(X_{i_1}^{-1}(B_1) \cap \cdots \cap X_{i_k}^{-1}(B)) = \{\omega' : (\omega'(i_1), \omega'(i_2), \cdots, \omega'(i_k)) \in B_1 \times \cdots \times B_k\}.$$

where B_j are Borel sets in $\mathbb{R}$. Then T extends to an isomorphism. X_n and X_n' correspond under T.

We will sometimes call P' the probability distribution of $\{X_n\}_{n\in\Gamma}$ and denote it by $\mu_{\{X_n\}_{n\in\Gamma}}$.

A sigma field $\mathcal{D} \subset \{\mathcal{F}\}$ is called *trivial* if $(\Omega, \mathcal{D}, P)$ is isomorphic to $(\Omega, \{\emptyset, \Omega\}, P)$. In other words, $\mathcal{D}$ is trivial iff for all $E \in \mathcal{D}$, $P(E) \in \{0, 1\}$.

Theorem 1.3.1. Let $\mathcal{D}$ be a trivial σ-field and X and extended random variable such that $\sigma(X) \subset \mathcal{D}$. Then $X = \text{const}$ P-a.s.

Proof. Let $L = \{x \in \mathbb{R} : P(X \le x) > 0\}$. Note that if $P(X \le x) > 0$, then $P(X \le x) = 1$. If $L = \emptyset$, then $X = \infty$ P-a.s. If L is nonempty and $\inf L = -\infty$, then $X = -\infty$ P-a.s.

Assume that $c := \inf L > -\infty$. Then for any $\epsilon > 0$, $P(X \le c - \epsilon) = 0$ and so

$$P(X < c) = P(\cup_{n=1}^\infty \{X \le c - 1/n\}) = 0.$$

Since for any $\epsilon > 0$,

$$P(X > c + \epsilon) = 1 - P(X \le c + \epsilon) = 0,$$

$$P(X > c) = P(\cup_{n=1}^\infty \{X > c + 1/n\}) = 0,$$

and so $X = c$ P-a.s. $\square$

1.4. Independence.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A collection of events $\mathcal{E} \subset \mathcal{F}$ is called independent if for each finite subcollection $\{E_1, \cdots, E_n\}$ of *distinct* events

$$P(E_1 \cap \cdots \cap E_n) = P(E_1) \cdots P(E_n).$$

A family of random variables $\{X_i\}_{i \in I}$ is called independent if any finite collection of the events

$$\{X_{i_1}^{-1}(B_1), \cdots, X_{i_n}^{-1}(B_{i_n}) \,:\, i_1, \cdots, i_n \,\text{distinct},\, B_1, \cdots, B_n \,\text{Borel}\}$$

is independent.

Theorem 1.4.1. The following statements are equivalent:
(1) The random variables $X_1, \cdots, X_n$ are independent.
(2) $F_{X_1, \cdots, X_n} = F_{X_1} \cdots F_{X_n}$.
(3) $\mu_{X_1, \cdots, X_n} = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$.

Theorem 1.4.2. Assume that the random vector $X = (X_1, \cdots, X_n)$ is absolutely continuous with density f_X. Then each X_k is an absolutely continuous random variable and $X_1, \cdots, X_n$ are independent iff $f_X = f_{X_1} \cdots f_{X_n}$.

Theorem 1.4.3. If $X_{m,k}$, $m = 1, \cdots, n$, $k = 1, \cdots, p_m$ is a collection of independent random variables and $g_m : \mathbb{R}^{p_m} \to \mathbb{R}$ are Borel functions, then the random variables

$$Y_m = g_m(X_{m,1}, \cdots, X_{m,p_m}), \qquad m = 1, \cdots, n,$$

are independent.

Theorem 1.4.4. Assume that Γ is countable. The following statements are equivalent:
(1) The random variables $\{X_n\}_{n \in \Gamma}$ are independent.
(2) $\mu_{\{X_n\}_{n \in \Gamma}} = \bigotimes_{n \in \Gamma} \mu_{X_n}$.

1.5. $0 - 1$ **laws**

Let E_n be a sequence of events. We define

$$\overline{\lim} E_n = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} E_n.$$

$\omega \in \overline{\lim} E_n$ iff ω belongs to infinitely many E_n's. Note also that all ω

$$\chi_{\overline{\lim} E_n}(\omega) = \overline{\lim}_{n \to \infty} \chi_{E_n}(\omega).$$

The simplest and the best known $0 - 1$ law is:

Theorem 1.5.1. (Borel $0-1$ law.) Let E_n be a sequence of independent events. Then

$$P(\overline{\lim} E_n) = \begin{cases} 0 & \text{if } \sum_n P(E_n) < \infty, \\ 1 & \text{if } \sum_n P(E_n) = \infty. \end{cases}$$

The independence condition can be relaxed. Let E_n be any sequence of events. If $\sum_n P(E_n) < \infty$, then $P(\overline{\lim} E_n) = 0$. On the other hand, if $\sum_n P(E_n) = \infty$,

$$|P(E_n \cap E_m) - P(E_n)P(E_m)| \leq \alpha_{nm} P(E_n)^{1/2} P(E_m)^{1/2},$$

and $\{\alpha_{nm}\}$ defines a bounded operator on $l^2(\mathbb{N})$, then $P(\overline{\lim} E_n) = 1$. It suffices, for example, that

$$\sup_n \sum_m \alpha_{nm} < \infty, \qquad \sup_m \sum_n \alpha_{nm} < \infty.$$

The Kolmogorov $0-1$ law will play an important role in these lectures. Let X_n be a sequence of random variable, $n \in \mathbb{N}$. The tail σ-field, τ, of this sequence is defined by

$$\tau(\{X_n\}) = \bigcap_{n=1}^{\infty} \sigma(\{X_k\}_{k \geq n}).$$

We are interested under what conditions $\tau(\{X_n\})$ is trivial. (In our study of Anderson type Hamiltonians we will use the triviality of τ and Theorem 1.3.1 to prove that certain functions are P-a.s. constant).

Theorem 1.5.2. (Kolmogorov's $0-1$ law.) Assume that the random variables $\{X_n\}_{n \in \mathbb{N}}$ are independent. The the tail σ-field $\tau(\{X_n\})$ is trivial.

The independence condition can be relaxed, see Subsection 1.8.

1.6. Ergodicity

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{C}$ the corresponding mod0 classes of equivalences. Let $T : \Omega \to \Omega$ be a measurable transformation. We assume that if $E \in \mathcal{F}$, then $T(E) \in \mathcal{F}$ and that T extends to a bijection $T : \mathcal{C} \to \mathcal{C}$. A set E is called T-invariant if $T\mathcal{C}_E = \mathcal{C}_E$, where $\mathcal{C}_E$ is the class of equivalences that contains E. In other words, E is T-invariant if $T(E) = E \bmod 0$, that is, if $P((T(E) \setminus E) \cup (E \setminus T(E)) = 0$.

The measure P is called T-invariant if $P(T(E)) = P(E)$ for all $E \in \mathcal{F}$.

The transformation T is called ergodic if P is T-invariant and

$$T(E) = E \bmod 0 \qquad \Rightarrow \qquad P(E) \in \{0, 1\}.$$

The ergodicity is another example of 0-1 law which we will use on several occasions.

Theorem 1.6.1. Assume that T is ergodic and let X be an extended random variable such that $X \circ T = X$ P-a.s. Then $X = \text{const}$ P-a.s.

Proof. The proof is very similar to the proof of Theorem 1.3.1. Let

$$L = \{x \in \mathbb{R} : P(X \geq x) > 0\}.$$

Note that since the set $\{X \geq x\}$ is T-invariant, $P(X \geq x) \in \{0, 1\}$.

If $L = \emptyset$, then $X = \infty$ P-a.s. If L is nonempty and $\inf L = -\infty$, then $X = -\infty$ P-a.s. If $c = \inf L > -\infty$, then for any $\epsilon > 0$, $P(X \leq x - \epsilon) = 0$ and $P(X > x + \epsilon) = 0$. Hence, $X = \text{const}$ P-a.s. $\square$

Theorem 1.6.2. (Birkhoff's ergodic theorem.) Assume that T is ergodic and that X is an integrable random variable. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X(T^k(\omega)) = \mathbb{E}(X) \qquad \text{for } P - a.e.\, \omega.$$

1.7. Conditional expectation and conditional probability

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{D} \subset \mathcal{F}$ a σ-algebra. Let X be an integrable random variable. For $D \in \mathcal{D}$ set

$$\mu(D) = \int_D X \mathrm{d}P.$$

Then μ is a signed measure on $(\Omega, \mathcal{D})$ and $P(D) = 0 \Rightarrow \mu(D) = 0$. Hence, there exists a $\mathcal{D}$-measurable function $\mathbb{E}(X|\mathcal{D})$ such that

$$\mu(D) = \int_D \mathbb{E}(X|\mathcal{D}) \mathrm{d}P.$$

The function $\mathbb{E}(X|\mathcal{D})$ is called the conditional expectation of X given $\mathcal{D}$. $\mathbb{E}(X|\mathcal{D})$ is uniquely specified (up to a set of measure zero, of course) by the following two properties:
(1) $\mathbb{E}(X|\mathcal{D})$ is $\mathcal{D}$-measurable.
(2) $\int_D X \mathrm{d}P = \int_D \mathbb{E}(X|\mathcal{D}) \mathrm{d}P$ for all $D \in \mathcal{D}$.

Consider the Hilbert space $\mathcal{H} := L^2(\Omega, \mathrm{d}P)$, and let $\mathcal{K}$ be the set of all $\mathcal{D}$-measurable functions in $\mathcal{H}$. $\mathcal{K}$ is a closed subspace of $\mathcal{H}$. Let Q be the orthogonal projection on $\mathcal{K}$. Then for $X \in \mathcal{H}$, $QX = \mathbb{E}(X|\mathcal{D})$.

Note that $\mathbb{E}(X|\mathcal{F}) = X$ and $\mathbb{E}(X|\{\emptyset, \Omega\}) = X$.

Theorem 1.7.1.
(1) If X is an integrable random variable, so is $\mathbb{E}(X|\mathcal{D})$ and $\mathbb{E}(\mathbb{E}(X|\mathcal{D})) = \mathbb{E}(X)$.

(2) If X and Y are integrable random variables and Y is $\mathcal{D}$-measurable, then $\mathbb{E}(YX|\mathcal{D}) = Y\mathbb{E}(X|\mathcal{D})$ a.s.

Theorem 1.7.2. (Properties of the conditional expectation.)
(1) If X_1, X_2 are integrable and $a_1, a_2 \in \mathbb{R}$, then

$$\mathbb{E}(a_1 X_1 + a_2 X_2|\mathcal{D}) = a_1\mathbb{E}(X_1|\mathcal{D}) + a_2\mathbb{E}(X_2|\mathcal{D}) \qquad P - a.s.$$

(2) If $X_1 \leq X_2$ a.s., then $\mathbb{E}(X_1|\mathcal{D}) \leq \mathbb{E}(X_2|\mathcal{D})$ a.s.
(3) If X_n is an increasing sequence of positive random variables, $X_n \to X$ and X is integrable, then $\mathbb{E}(X_n|\mathcal{D}) \to \mathbb{E}(X|\mathcal{D})$ a.s.
(4) If X_n is an increasing sequence of positive random variables and $X := \liminf_n X_n$ is integrable, then

$$\mathbb{E}(X|\mathcal{D}) \leq \liminf \mathbb{E}(X_n|\mathcal{D}) \qquad P - a.s.$$

(5) If X_n is a sequence of random variables such that $X_n \to X$ a.s. and $|X_n| \leq Y$ where Y is integrable, then

$$\lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{D}) = \mathbb{E}(X|\mathcal{D}) \qquad P - a.s.$$

(6) If f is convex and $X, f(X)$, are integrable, then

$$f(\mathbb{E}(X|\mathcal{D})) \leq \mathbb{E}(f(X)|\mathcal{D}) \qquad P - a.s.$$

Of course, you recognize (3) as the conditional monotone convergence theorem, (4) as the conditional Fatou's lemma, (5) as the conditional dominated convergence theorem, and (6) as the conditional Jensen inequality.

Theorem 1.7.3. (The martingale convergence theorem). Let $\{X_n\}_{n\in\mathbb{N}}$ be random variables and X an integrable random variable. Set

$$Y_n := \mathbb{E}(X|\sigma(\{X_k\}_{k\leq n})), \qquad Y_\infty := \mathbb{E}(X|\sigma(\{X_k\}_{k\in\mathbb{N}})).$$

Then $\lim_{n\to\infty} Y_n = Y_\infty$ P-a.s.

We now turn to the conditional probability. Let E be an event. The conditional probability of E given $\mathcal{D}$ is

$$P(E|\mathcal{D}) := \mathbb{E}(\chi_E|\mathcal{D}).$$

$P(E|\mathcal{D})$ is unique (up to a set of measure zero) $\mathcal{D}$-measurable function such that $\forall D \in \mathcal{D}$,

$$\int_D P(E|\mathcal{D})\mathrm{d}P = P(E \cap D).$$

Theorem 1.7.4. (Properties of the conditional probability.)
(1) $0 \le P(E|\mathcal{D}) \le 1$ a.s., $P(\emptyset|\mathcal{D}) = 0$ a.s., and $P(\Omega|\mathcal{D}) = 1$ a.s.
(2) If E_n is a sequence of disjoint events, then

$$P(\cup_{n=1}^{\infty} E_n|\mathcal{D}) = \sum_{n=1}^{\infty} P(E_n|\mathcal{D}) \qquad P-a.s.$$

(3) If $E_1 \subset E_2$, then $P(E_1|\mathcal{D}) \le P(E_2|\mathcal{D})$ a.s.
(4) If E_n is a decreasing sequence of events, $E_1 \supset E_2 \supset \cdots$, then

$$P(\cap_{n=1}^{\infty} E_n|\mathcal{D}) = \lim_{n\to\infty} P(E_n|\mathcal{D}) \qquad P-a.s.$$

Warning: On the basis of the above theorem one may think that for P-a.e. ω the map

$$\mathcal{F} \ni E \to P(E|\mathcal{D})(\omega), \tag{2}$$

is a measure. However, this may not be so, since the desired properties hold up to a set of measure zero. This is irrelevent if we deal with countable sequences of events, but it certainly becomes relevent if we consider *all* events simultaneously. In particular, the countable additivity of the function (2) may hold only a set of P-measure zero!

1.8. Kolmogorov's 0-1 law for dependent random variables

Let $X_n, n \in \mathbb{N}$, be a sequence of random variables and

$$\mathcal{D}_n = \sigma(\{X_k\}_{k\le n}), \qquad \mathcal{D}_{\overline{n}} = \sigma(\{X_k\}_{k\ge n}).$$

Theorem 1.8.1. The following statements are equivalent:
(1) The tail σ-field $\tau(\{X_n\}_{n\in\mathbb{N}})$ is trivial.
(2) For all n, all $A \in \mathcal{D}_n$, and all $E \in \tau(\{X_n\}_{n\in\mathbb{N}})$, $P(A \cap E) = P(A)P(E)$.

Proof. We only need to prove that (2) $\Rightarrow$ (1). Let $E \in \tau$. For $D \in \mathcal{D}_n$,

$$\int_D P(E|\mathcal{D}_n)\mathrm{d}P = P(E \cap D) = P(E)P(D),$$

and so $P(E|\mathcal{D}_n) = P(E)$ P-a.s. On the other hand, by the martingale convergence theorem,

$$\lim_{n\to\infty} P(E|\mathcal{D}_n) = P(E|\sigma(\{X_n\}_{n\in\mathbb{N}})) \qquad P-a.s.$$

Since $E \in \sigma(\{X_n\}_{n\in\mathbb{N}})$,

$$P(E|\sigma(\{X_n\}_{n\in\mathbb{N}})) = \mathbb{E}(\chi_E|\sigma(\{X_n\}_{n\in\mathbb{N}})) = \chi_E \qquad P-a.s.,$$

and so $\chi_E = P(E)$ P-a.s. Hence, $P(E) \in \{0,1\}$ and τ is trivial. $\square$

A sufficient condition for (2) in Theorem 1.8.1 is that for all n, all $A \in \mathcal{D}_n$ and every sequence $B_m \in \mathcal{D}_{\overline{m}}$,

$$\liminf_{m\to\infty} |P(A \cap B_m) - P(A)P(B_m)| = 0.$$

In fact, we can formulate a sufficient and necessary condition for the Kolmogorov 0-1 law in terms of vanishing of suitable correlations.

Theorem 1.8.2. The following statements are equivalent:
(1) The tail σ-field $\tau(\{X_n\}_{n\in\mathbb{N}})$ is trivial.
(2) For all n, all $A \in \mathcal{D}_n$, and every sequence if decreasing events $B_m \in \mathcal{D}_{\overline{m}}$,

$$\lim_{m\to\infty} |P(A \cap B_m) - P(A)P(B_m)| = 0.$$

Proof. The Theorem 1.8.1 yields that (2) $\Rightarrow$ (1) and we need to show that (1) $\Rightarrow$ (2).

Assume that (1) holds and (2) does not. Then there is n, $\epsilon > 0$, $A \in \mathcal{D}_n$, and a decreasing sequence of events $B_m \in \mathcal{D}_{\overline{m}}$ such that for all m,

$$|P(A \cap B_m) - P(A)P(B_m)| \geq \epsilon. \tag{3}$$

Let $\overline{B} = \cap B_m$. Then $\overline{B} \in \tau$, and since $P(\overline{B}) < 1$ (otherwise (3) cannot hold), $P(\overline{B}) = 0$. Since B_m is decreasing,

$$\lim_{m\to\infty} P(B_m) = P(\overline{B}) = 0,$$

and this contradicts (3). $\square$

1.9. Conditional probability distributions. Conditional Fubini's theorem.

In this subsection we will describe a technical device which will play an important role in our study of Anderson type Hamiltonians. In fact, the main technical conditions in Lectures 9 and 10 will be expressed in terms of this device. In telegraphic terms, this technical condition states that the sequence of random variables X_n (which defines the potential in the Anderson type Hamiltonian) is such that *for all n the conditional probability distribution of X_n, given any $\{X_m\}_{m\neq n}$, is absolutely continous.* This condition appeared in the work of Aizenman and Molchanov, and is a very natural condition for our purposes.

Our goal in this subsection is to give the mathematical meaning to the phrase *conditional probability distribution of X_n, given any $\{X_m\}_{m\neq n}$* and to prove related conditional Fubini's theorem. The conditional Fubini theorem will play a role in the proof of all major results in Lectures 9 and 10.

For the purposes of this section w.l.o.g. we assume that $\Omega = \mathbb{R}^\Gamma$, where $\Gamma = \{1, \cdots, N\}$ or $\Gamma = \mathbb{N}$. $\mathcal{F}$ is the Borel σ-field in Ω and $X_n(\omega) = \omega(n)$. Our distinguished random variable is X_1. Let $\mathcal{D} := \sigma(\{X_n\}_{n\geq 2})$. Let $\Gamma_2 = \Gamma \setminus \{1\}$, $\Omega_2 = \mathbb{R}^{\Gamma_2}$ and let $\mathcal{F}_2$ be the Borel σ-field in $\mathbb{R}^{\Gamma_2}$. $\mathcal{F}_2$ and $\mathcal{D}$ are identified in the obvious way—any $E \in \mathcal{D}$ has the form $E = \mathbb{R} \times E_2$ where $E_2 \in \mathcal{F}_2$ and we can identify E with E_2. We set $P_2 = P \restriction \mathcal{D}$. P_2 defines a measure on $(\Omega_2, \mathcal{F}_2)$ which we denote by the same letter.

Let $\Omega_1 = \mathbb{R}$ and let $\mathcal{F}_1$ be the usual Borel σ-field in Ω_1. Obviously,

$$\Omega = \Omega_1 \times \Omega_2, \qquad \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2.$$

We denote the elements of Ω_1 and Ω_2 by ω_1 and ω_2, so $\omega = (\omega_1, \omega_2)$.

Let X be a $\mathcal{D}$ measurable random variable. Then X does not depend on ω_1, that is, $X(\omega_1, \omega_2) = X(\omega_1', \omega_2)$ for all $\omega_1, \omega_1' \in \Omega_1$. To see this, let $X(\omega_1, \omega_2) = x$. Then $X^{-1}(\{x\})$ belongs to $\mathcal{D}$ and so has the form $\Omega_1 \times E$. If $X_2 : \Omega_2 \to \mathbb{R}$ is defined by $X_2(\omega_2) = X(\omega_1, \omega_2)$, then for all Borel sets $B \in \mathcal{F}_2$,

$$\int_{\mathbb{R}\times B} X \mathrm{d}P = \int_B X_2 \mathrm{d}P_2. \tag{4}$$

(In the sequel we will identify X and X_2 without further saying). The converse also holds—if X_2 is a random variable on Ω_2, then $X(\omega_1, \omega_2) = X_2(\omega_2)$ is a $\mathcal{D}$-measurable random variable on Ω and (4) holds.

For $x \in \mathbb{R}$ let $I_x = (-\infty, x]$. For any rational x set

$$F(x) := P(I_x \times \Omega_2 | \mathcal{D}).$$

More precisely, $F(x)$ is a specific $\mathcal{D}$-measurable function which satisfies

$$\int_D F(x) \mathrm{d}P = P(I_x \times D),$$

for all $D \in \mathcal{D}$. Since rationals are countable, there is a set $\tilde{\Omega} \in \mathcal{D}$ of P measure 1, such that all the properties listed in Theorem 1.7.4 hold for $\{I_x : x \in \mathbb{Q}\}$ and all $\omega \in \tilde{\Omega}$. Let $\tilde{\Omega} = \mathbb{R} \times \tilde{\Omega}_2$ where $\tilde{\Omega}_2 \in \mathcal{F}_2$ and $P_2(\tilde{\Omega}_2) = 1$. We denote also by $F(x)$ the random variable on Ω_2 associated to $F(x)$. Then for all $\omega_2 \in \tilde{\Omega}_2$, $F(x_1)(\omega_2) \leq F(x_2)(\omega_2)$ if $x_1 \leq x_2$, and

$$\lim_{x\to\infty, x\in\mathbb{Q}} F(x)(\omega_2) = 1, \qquad \lim_{x\to-\infty, x\in\mathbb{Q}} F(x)(\omega_2) = 0.$$

We now define for all $x \in \mathbb{R}$ and $\omega_2 \in \Omega_2$,

$$F(x)(\omega_2) = \inf\{F(r)(\omega_2) : r > x,\ r \in \mathbb{Q}\}.$$

By Theorem 9-7 A in [Bur], $x \to F(x)(\omega_2)$ is a probability distribution function, namely $F(x)(\omega_2)$ is monotone, right continuous and

$$\lim_{x\to\infty} F(x)(\omega_2) = 1, \qquad \lim_{x\to-\infty} F(x)(\omega_2) = 0.$$

We denote by $\mathrm{d}\mu^{\omega_2} = \mathrm{d}F(x)(\omega_2)$, $\omega_2 \in \tilde{\Omega}_2$, the corresponding Stieltjes measure. The measure μ^{ω_2} is called the probability distribution of X_1 given $\{X_n\}_{n\geq 2}$.

Theorem 1.9.2. For any Borel sets $B_1 \in \mathcal{F}_1$, $B_2 \in \mathcal{F}_2$,

$$P(B_1 \times B_2) = \int_{B_2} \mu^{\omega_2}(B_1)\mathrm{d}P_2(\omega_2). \tag{5}$$

Proof. It is not difficult to show that the function $\Omega_2 \ni \omega_2 \to \mu^{\omega_2}(B_1)$ is measurable (this is a homework problem).

Assume first that $B_1 = I_x = (-\infty, x]$, where x is rational.Then

$$\begin{aligned}\int_{B_2} \mu^{\omega_2}(B_1)\mathrm{d}P_2(\omega_2) &= \int_{\Omega_2} \chi_{B_2}(\omega_2)\mu^{\omega_2}(B_1)\mathrm{d}P_2(\omega_2) \\ &= \int_{\Omega} P(B_1 \times B_2|\mathcal{D})\mathrm{d}P = P(B_1 \times B_2).\end{aligned}$$

(We have used (4) and Theorem 1.7.1—make sure you understand this step.)

Note that for fixed B_2 the set functions

$$B_1 \to P(B_1 \times B_2) \qquad \text{and} \qquad B_1 \to \int_{B_2} \mu^{\omega_2}(B_1)\mathrm{d}P_2(\omega_2)$$

are Borel measures. These two measures coincide for $B_1 = I_x$, x rational, and so they coincide for all Borel B_1. □

Theorem 1.9.3. (The conditional Fubini theorem). Let X be a positive or an integrable random variable. Then

$$\int_{\Omega} X\mathrm{d}P = \int_{\Omega_2} \left[\int_{\Omega_1} X(\omega_1, \omega_2)\mathrm{d}\mu^{\omega_2}(\omega_1)\right] \mathrm{d}P_2(\omega_2).$$

If X is positive, then both sides are allowed to be infinite.

Proof. If X is a linear combination of the characteristic functions of the sets of the form $B_1 \times B_2$, then the result follows from Theorem 1.9.2. Since $\{B_1 \times B_2 : B_1 \in \mathcal{F}_1, B_2 \in \mathcal{F}_2\}$ generates $\mathcal{F}_1 \times \mathcal{F}_2 = \mathcal{F}$ the theorem holds for any characteristic function. Hence, it holds for any simple function, and by the monotone convergence, for any positive function. □

If the measure μ^{ω_2} is absolutely continuous w.r.t. the Lebesgue measure for P_2-a.e. ω_2, we say that the conditional probability distribution of X_1 given $\{X_n\}_{n\geq 1}$ is absolutely continuous.

Finally, we mention that if the variables $\{X_n\}$ are independent, that is, if $P = \bigotimes_{n\in\Gamma} \mu_{X_n}$, then μ^{ω_2} is equal to μ_{X_1} (and independent of ω_2). It is important to keep this special case in mind while reading the Lectures 9 and 10.

1.9. Homework.

[1] The material in Subsections 1.1,1.3–1.8, has been covered in MATH 587-589. Please read again the proofs of all the results described in these subsections.

[2] We have not discussed isomorphisms of probability spaces in MATH 587-589. Please think about the main points of Subsection 1.2. Let $\Omega = \mathbb{R}^{\mathbb{N}}$ and let $P = \prod_{n\in\mathbb{N}} \mu_n$ where all μ_n are equal, $\mu_n(\{1\}) = 1/2$, $\mu_n(\{-1\}) = 1/2$. Prove that (Ω, P) is isomorphic to $([0,1], \mathrm{d}x)$. What are the random variables corresponded to $X_n(\omega) = \omega(n)$?

Pleased read page 8 in [Si] and make sure that you understand the content of the third paragraph.

Please read Chapter 2 and Appendix 6 in [AA].

[3] Let $X_1, \cdots, X_n$ be independent and $X = \alpha_1 X_1 + \cdots + \alpha_n X_n$, where α_j are real.
(1) Assume that X_j is $\mathcal{N}(m_j, \sigma_j^2)$. Prove that X is $\mathcal{N}(m, \sigma)$ and compute m and σ.
(2) Assume that X_j is $\mathcal{C}(a_j, b_j)$. Prove that X is $\mathcal{C}(a, b)$ and compute a and b.

[4] Supplement the proofs of Theorems 1.9.2 and 1.9.3 with necessary measurability arguments.

[5] Define the conditional probability distribution of any finite subcollection $\{X_m\}_{1\le m\le n}$ given $\{X_m\}_{m>n}$. Prove the analogs of Theorems 1.9.2 and 1.9.3. Show that the conditional probability distribution of $\{X_m\}_{1\le m\le n}$ is absolutely continuous iff the conditional probability distribution of X_j given $\{X_m\}_{m\neq j}$ is absolutely continuous for $j = 1, \cdots, n$.