Lecture 27. Using σ-algebras to characterize dependence.

We know what it means that σ-algebras $\mathcal{A}$ and $\mathcal{B} \subseteq \mathcal{F}$ are dependent: this means that there exist such events $A \in \mathcal{A}$, $B \in \mathcal{B}$ that $P(A \cap B) \neq P(A) \cdot P(B)$. It seems to be interesting to introduce some number-valued coefficient characterizing the degree of this dependence: this would enable us, e. g., to say meaningfully such things as "the σ-algebras $\mathcal{A}$ and $\mathcal{B}$ are more dependent than the σ-algebras $\mathcal{C}$ and $\mathcal{D}$".

Many different "dependence coefficients" have been introduced. We are going to consider only one of them, introduced by M.Rosenblatt.

We define

$$\begin{aligned}\alpha^*(\mathcal{A}, \mathcal{B}) = \sup\{|\mathrm{Cov}(\eta, \zeta)|\colon\ &\eta,\ \zeta \text{ are random variables, } |\eta|,\ |\zeta| \le 1,\\ &\text{measurable: } \eta \text{ with respect to } \mathcal{A},\ \zeta \text{ with respect to } \mathcal{B}\}.\end{aligned} \tag{27.1}$$

It is clear that if $\mathcal{A}$ and $\mathcal{B}$ are independent, the random variables η and ζ are also independent, the covariance is equal to 0, and $\alpha^*(\mathcal{A}, \mathcal{B}) = 0$.

To be precise, M.Rosenblatt introduced a little different coefficient of dependence:

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|P(A \cap B) - P(A) \cdot P(B)|\colon\ A \in \mathcal{A}, B \in \mathcal{B}\}; \tag{27.2}$$

but it is almost the same: it can be proved that

$$\alpha(\mathcal{A}, \mathcal{B}) \le \alpha^*(\mathcal{A}, \mathcal{B}) \le 4\alpha(\mathcal{A}, \mathcal{B}); \tag{27.3}$$

and the coefficient α^* is more convenient for me.

In Lecture 11 we considered the zero-one law for independent random variables. It turns out that the same law holds for sequences of dependent random variables provided that the dependence vanishes in the limit:

Theorem 27.1. *Let* ξ_0, ξ_1, ξ_2, ..., ξ_n ... *be a sequence of* (dependent, in general) *random variables*; *and let the dependence coefficient*

$$\alpha^*(\mathcal{F}_{\le n}, \mathcal{F}_{\ge m}) \to 0 \qquad (m \to \infty). \tag{27.4}$$

Then every event in the tail σ*-algebra has either probability* 1, *or* 0:

$$A \in \mathcal{F}_{\ge \infty} \ \Rightarrow\ P(A) = 0 \text{ or } P(A) = 1. \tag{27.5}$$

The **proof** is essentially the same as that of Theorem 11.2: Let $A \in \mathcal{F}_{\ge \infty}$; then, as for every event belonging to $\mathcal{F}_{\ge 0}$, we can for every $\varepsilon > 0$ find a natural n and an event $A_\varepsilon \in \mathcal{F}_{\le n}$ such that the probability of the symmetric difference $P(A \Delta A_\varepsilon) < \varepsilon$. Now, for every $m \ge n$ the event $A \in \mathcal{F}_{\ge m}$, and we have:

$$|P(A_\varepsilon \cap A) - P(A_\varepsilon) \cdot P(A)| = |\mathrm{Cov}(I_{A_\varepsilon}, I_A)| \le \alpha^*(\mathcal{F}_{\le n}, \mathcal{F}_{\ge m}). \tag{27.6}$$

Letting $m \to \infty$, we get that the left-hand side of (27.7) is equal to 0, and

$$P(A_\varepsilon \cap A) = P(A_\varepsilon) \cdot P(A); \tag{27.7}$$

after this the proof follows that of Theorem 11.2.

For ergodic Markov chains the dependence coefficient $\alpha^*(\mathcal{F}_{\le n}, \mathcal{F}_{\ge m}) \to 0$ as $m \to \infty$; and even

$$\alpha^*(\mathcal{F}_{\le n}, \mathcal{F}_{\ge m}) \le \beta(m - n), \qquad \text{where } \beta(k) \to 0 \quad (k \to \infty). \tag{27.8}$$

We are going to prove this for *finite* ergodic chains (the proof for infinite chains, i. e., for countably infinite space X in which the random variables ξ_i take values, is more complicated).

Theorem 27.2. *Let $\xi_0, \xi_1, \xi_2, \ldots, \xi_n, \ldots$ be a finite ergodic Markov chain. Then* (27.8) *holds.*

Proof. Let η, ζ be random variables, $|\eta|, |\zeta| \le 1$, measurable, respectively, with respect to $\mathcal{F}_{\le n}$ and to $\mathcal{F}_{\ge m}$, $m > n$. For η this means that

$$\eta = f(\xi_0, \xi_1, \ldots, \xi_n), \tag{27.9}$$

where f is a function on X^{n+1}, $|f(x_0, x_1, \ldots, x_n)| \le 1$.

We have:

$$\operatorname{Cov}(\eta, \zeta) = E(\eta\zeta) - E\eta \cdot E\zeta. \tag{27.10}$$

The random variable η takes finitely many values, and

$$E\eta = \sum_{x_0, x_1, \ldots, x_n \in X} q_{x_0} \cdot p_{x_0 x_1} \cdot \ldots \cdot p_{x_{n-1}, x_n} \cdot f(x_0, x_1, \ldots, x_n). \tag{27.11}$$

To $E\zeta$ we apply the total expectation formula (23 – 24.11) with $i = y \in X$ and $A_i = \{\xi_m = y\}$:

$$E\zeta = \sum_{y \in X} P\{\xi_m = y\} \cdot E\{\zeta | \xi_m = y\}. \tag{27.12}$$

Introducing the notation

$$g(y) = E\{\zeta | \xi_m = y\} \tag{27.13}$$

(of course this function depends on what random variable ζ we are considering, and on m), we can write:

$$E\zeta = \sum_{y \in X} P\{\xi_m = y\} \cdot g(y) = Eg(\xi_m). \tag{27.14}$$

Of course, the function $g(y)$ is bounded by the same constant 1 as the random variable ζ.

To $E(\eta\zeta)$ we also apply the total expectation formula, with $(x_0, x_1, \ldots, x_n, y) \in X^{n+2}$ instead of i and $\{\xi_0 = x_0, \xi_1 = x_1, \ldots, \xi_n = x_n, \xi_m = y\}$ instead of A_z:

$$\begin{aligned} E(\eta\zeta) = \sum_{x_0, x_1, \ldots, x_n, y \in X} & P\{\xi_0 = x_0, \xi_1 = x_1, \ldots, \xi_n = x_n, \xi_m = y\} \times \\ & \times E\{\eta\zeta | \xi_0 = x_0, \xi_1 = x_1, \ldots, \xi_n = x_n, \xi_m = y\}. \end{aligned} \tag{27.15}$$

Since $\eta = f(\xi_0, \xi_1, ..., \xi_n)$, and we have $\xi_0 = x_0, \xi_1 = x_1, ..., \xi_n = x_n$ in the condition, we can change η in the conditional expectation to $f(x_0, x_1, ..., x_n)$:

$$\begin{aligned} E\{\eta\zeta | \xi_0 = x_0,\, \xi_1 = x_1, ..., \xi_n = x_n,\, \xi_m = y\} \\ &= E\{f(x_0, x_1, ..., x_n) \cdot \zeta | \xi_0 = x_0,\, \xi_1 = x_1, ..., \xi_n = x_n,\, \xi_m = y\} \\ &= f(x_0,\, x_1, ..., x_n) \cdot E\{\zeta | \xi_0 = x_0,\, \xi_1 = x_1, ..., \xi_n = x_n,\, \xi_m = y\}, \end{aligned} \tag{27.16}$$

because $f(x_0, x_1, ..., x_n)$ in this conditional expectation is equal to a constant (does not depend on ω).

Now, according to Theorem 25.8, we have for every $A \in \mathcal{F}_{\geq m}$, every set $C \in X^m$, and every $y \in X$ (we change in formula (25.30) B to A, A to $\{(\xi_0, ..., \xi_{m-1}) \in C\}$, n to m and x to y):

$$P\{A | (\xi_0,\, \xi_1,\, ...,\, \xi_{m-1}) \in C,\, \xi_m = y\} = P\{A | \xi_m = y\}. \tag{27.17}$$

Since these two probability measures, considered only on the σ-algebra $\mathcal{F}_{\geq m}$, coincide, we have for every $\mathcal{F}_{\geq m}$-measurable random variable ζ:

$$E\{\zeta | (\xi_0,\, \xi_1,\, ...,\, \xi_{m-1}) \in C,\, \xi_m = y\} = E\{\zeta | \xi_m = y\} = g(y). \tag{27.18}$$

Now we take $C = \{(x_0, x_1, ..., x_n, x_{n+1}, ..., x_{m-1})$, where $x_0, x_1, ..., x_n$ are the same as in (27.16), and $x_{n+1}, ..., x_{m-1}$ run over the whole space $X\}$; and we can write:

$$E\{\eta\zeta | \xi_0 = x_0,\, \xi_1 = x_1, ..., \xi_n = x_n,\, \xi_m = y\} = f(x_0,\, x_1, ..., x_n) \cdot g(y). \tag{27.19}$$

Formula (27.15) becomes

$$\begin{aligned} E(\eta\zeta) = \sum_{x_0,\, x_1,\, ...,\, x_n,\, y \in X} P\{\xi_0 = x_0,\, \xi_1 = x_1, ..., \xi_n = x_n,\, \xi_m = y\} \cdot f(x_0,\, x_1, ..., x_n)\, g(y) \\ = E\big(f(\xi_0,\, \xi_1,\, ...,\, \xi_n) \cdot g(\xi_m)\big). \end{aligned} \tag{27.20}$$

So we have discovered that $\alpha^*(\mathcal{F}_{\leq n}, \mathcal{F}_{\geq m})$ can be expressed as the supremum in which we consider only random variables ζ of the form $g(\xi_m)$.

Note how we passed from expectations $E\zeta$, $E(\eta\zeta)$ to formulas containing sums, and then back to formulas containing only expectations (formulas (27.14), (27.20)). We'll use this device in a more systematic way when we learn about *conditional expectations with respect to σ-algebras* (in the lectures to follow).

Now let us express $E\zeta = Eg(\xi_m)$, $E(\eta\zeta) = E\big(\eta\, g(\xi_m)\big)$ using the transition probabilities.

Using once more the total expectation formula (with the events $\{\xi_0 = x\}$ as A_i), we obtain:

$$E\zeta = Eg(\xi_m) = \sum_{x \in X} P\{\xi_0 = x\} \cdot E\{g(\xi_m) | \xi_0 = x\} = \sum_{x \in X} \big[q_x \cdot \sum_{y \in X} p_{xy}^{(m)}\, g(y)\big]. \tag{27.21}$$

To $E\big(\eta\, g(\xi_m)\big)$ we apply the same formula with the events $\{\xi_0 = x_0,\, \xi_1 = x_1, ..., \xi_n = x_n\}$ as A_i:

$$\begin{aligned} E\big(f(\xi_0, \xi_1, ..., \xi_n)\zeta\big) &= E\big(f(\xi_0, \xi_1, ..., \xi_n) g(\xi_m)\big) \\ &= \sum_{x_0, x_1, ..., x_n \in X} P\{\xi_0 = x_0, \xi_1 = x_1, ..., \xi_n = x_n\} \times \\ &\qquad \times E\{f(\xi_0, \xi_1, ..., \xi_n) g(\xi_m) | \xi_0 = x_0, \xi_1 = x_1, ..., \xi_n = x_n\} \\ &= \sum_{x_0, x_1, ..., x_n \in X} P\{\xi_0 = x_0, \xi_1 = x_1, ..., \xi_n = x_n\} \times \\ &\qquad \times f(x_0, x_1, ..., \xi_n) \cdot E\{g(\xi_m) | \xi_0 = x_0, \xi_1 = x_1, ..., \xi_n = x_n\} \\ &= \sum_{x_0, x_1, ..., x_n \in X} P\{\xi_0 = x_0, \xi_1 = x_1, ..., \xi_n = x_n\} \cdot f(x_0, x_1, ..., x_n) \cdot E\{g(\xi_m) | \xi_n = x_n\} \\ &= \sum_{x_0, x_1, ..., x_n \in X} P\{\xi_0 = x_0, \xi_1 = x_1, ..., \xi_n = x_n\} \cdot \big[f(x_0, x_1, ..., x_n) \cdot \sum_{y \in X} p_{x_m, y}^{(m-n)}\, g(y)\big]. \end{aligned} \tag{27.22}$$

Now let $\gamma(k) = \max_{x,\, y \in X} |p_{xy}^{(k)} - p_y|$ (in the case of a finite ergodic chain, $\gamma(k) \to 0$ as $k \to \infty$). It is clear that for $0 \le l < k$ we have $\gamma(k) \le \gamma(l)$. Indeed, we have

$$p_{xy}^{(k)} - p_y = \sum_{z \in X} p_{xz}^{(k-l)} [p_{xy}^{(l)} - p_y]; \tag{27.23}$$

the difference in the brackets is between $\mp\gamma(l)$, the sum $\sum_{z \in X} p_{xz}^{(k-l)} = 1$ for every x, and we have $-\gamma(l) \le p_{xy}^{(k)} - p_y \le \gamma(l)$ for every $x, y \in X$, from which $\gamma(k) \le \gamma(l)$.

Using the inequalities $p_y - \gamma(m-n) \le p_{x_m, n}^{(m-n)} \le p_y + \gamma(m-n)$, $p_y \cdot g(y) - \gamma(m-n) \le p_{x_m, n}^{(m-n)} \cdot g(y) \le p_y \cdot g(y) + \gamma(m-n)$ (remember, $|g(y)| \le 1$), we obtain from (27.22):

$$\begin{aligned} Ef(\xi_0, \xi_1, ..., \xi_n) \cdot \sum_{y \in X} p_y \cdot g(y) - \gamma(m-n) &\le E\big(f(\xi_0, \xi_1, ..., \xi_n)\zeta\big) \cdot \sum_{y \in X} p_y \cdot g(y) \\ &\le Ef(\xi_0, \xi_1, ..., \xi_n) \cdot \sum_{y \in X} p_y \cdot g(y) + \gamma(m-n) \end{aligned} \tag{27.24}$$

(again, remember that $|f| \le 1$).

Similarly, from (27.21):

$$\sum_{y \in X} p_y \cdot g(y) - \gamma(m) \le E\zeta \le \sum_{y \in X} p_y \cdot g(y) - \gamma(m). \tag{27.25}$$

Since $\gamma(m) \le \gamma(m-n)$, we can replace $\mp\gamma(m)$ here with $\mp\gamma(m-n)$.

Finally, we get from these inequalities and (27.24):

$$|E(\eta\zeta) - E\eta \cdot E\zeta| \le 2\gamma(m-n), \tag{27.26}$$

which goes to 0 as $m - n \to \infty$.

Theorems 27.1, 27.2 together yield

Theorem 27.3. *For a finite ergodic Markov chain all events of the tail* σ*-algebra* $\mathcal{F}_{\geq\infty}$ *have either probability* 0 *or* 1.

In the next lecture we'll talk of the law of large numbers for dependent random variables.

By the way, it follows from formula (27.1) that for random variables η, ζ with $|\eta| \leq C_1$, $|\zeta| \leq C_2$, measurable with respect to the σ-algebras $\mathcal{A}$, $\mathcal{B}$ we have:

$$|E(\eta\zeta)| \leq |E\eta| \cdot |E\zeta| + C_1 C_2 \cdot \alpha^*(\mathcal{A}, \mathcal{B}). \tag{27.27}$$

We'll use this in the next lecture.