Lecture 5: Conditional distribution, Markov chains, and martingales

Conditional distribution

For random vectors X and Y, is $P[X^{-1}(B)|Y=y]$ a probability measure for given y?
Problem: $P[X^{-1}(B)|Y=y]$ is defined a.s.

Theorem 1.7(i) (Existence of conditional distributions)

Let X be a random n-vector on a probability space $(\Omega,\mathscr{F},P)$ and $\mathscr{A}$ be a sub-σ-field of $\mathscr{F}$.
Then there exists a function $P(B,\omega)$ on $\mathscr{B}^n\times\Omega$ such that

(a) $P(B,\omega)=P[X^{-1}(B)|\mathscr{A}]$ a.s. for any fixed $B\in\mathscr{B}^n$, and
(b) $P(\cdot,\omega)$ is a probability measure on $(\mathscr{R}^n,\mathscr{B}^n)$ for any fixed $\omega\in\Omega$.

Let Y be measurable from $(\Omega,\mathscr{F},P)$ to $(\Lambda,\mathscr{G})$.
Then there exists $P_{X|Y}(B|y)$ such that

(a) $P_{X|Y}(B|y)=P[X^{-1}(B)|Y=y]$ a.s. P_Y for any fixed $B\in\mathscr{B}^n$, and
(b) $P_{X|Y}(\cdot|y)$ is a probability measure on $(\mathscr{R}^n,\mathscr{B}^n)$ for any fixed $y\in\Lambda$.

Theorem 1.7(ii)

Let $(\Lambda, \mathscr{G}, P_1)$ be a probability space.
Suppose that P_2 is a function from $\mathscr{B}^n \times \Lambda$ to $\mathscr{R}$ and satisfies

(a) $P_2(\cdot, y)$ is a probability measure on $(\mathscr{R}^n, \mathscr{B}^n)$ for any $y \in \Lambda$, and

(b) $P_2(B, \cdot)$ is Borel for any $B \in \mathscr{B}^n$.

Then there is a unique probability measure P on $(\mathscr{R}^n \times \Lambda, \sigma(\mathscr{B}^n \times \mathscr{G}))$ such that, for $B \in \mathscr{B}^n$ and $C \in \mathscr{G}$,

$$P(B \times C) = \int_C P_2(B, y) dP_1(y). \tag{1}$$

Furthermore, if $(\Lambda, \mathscr{G}) = (\mathscr{R}^m, \mathscr{B}^m)$, and $X(x, y) = x$ and $Y(x, y) = y$ define the coordinate random vectors, then $P_Y = P_1$, $P_{X|Y}(\cdot|y) = P_2(\cdot, y)$, and the probability measure in (1) is the joint distribution of (X, Y), which has the following joint c.d.f.:

$$F(x, y) = \int_{(-\infty, y]} P_{X|Y}((-\infty, x]|z) dP_Y(z), \quad x \in \mathscr{R}^n, y \in \mathscr{R}^m, \tag{2}$$

where $(-\infty, a]$ denotes $(-\infty, a_1] \times \cdots \times (-\infty, a_k]$ for $a = (a_1, ..., a_k)$.

Conditional distribution

For a fixed y, $P_{X|Y=y} = P_{X|Y}(\cdot|y)$ is called the conditional distribution of X given $Y = y$.

Two-stage experiment theorem

If $Y \in \mathscr{R}^m$ is selected in stage 1 of an experiment according to its marginal distribution $P_Y = P_1$, and X is chosen afterward according to a distribution $P_2(\cdot, y)$, then the combined two-stage experiment produces a jointly distributed pair (X, Y) with distribution $P_{(X,Y)}$ given by (1) and $P_{X|Y=y} = P_2(\cdot, y)$.
This provides a way of generating dependent random variables.

Example 1.23

A market survey is conducted to study whether a new product is preferred over the product currently available in the market.
Questionnaires are sent by mail along with the sample products (both new and old) to N customers randomly selected from a population.
Each customer is asked to fill out the questionnaire and return it.
Response is either 1 (new is better than old) or 0 (otherwise).

Some customers, however, do not return the questionnaires.

Let X be the number of ones in the returned questionnaires.
What is the distribution of X?

If every customer returns the questionnaire, then (from elementary probability) X has the binomial distribution $Bi(p,N)$ in Table 1.1 (assuming that the population is large enough so that customers respond independently), where $p \in (0,1)$ is the overall rate of customers who prefer the new product.

Now, let Y be the number of customers who respond, which is random.

Suppose that customers respond independently with the same probability $\pi \in (0,1)$.
Then P_Y is the binomial distribution $Bi(\pi,N)$.

Given $Y = y$ (an integer between 0 and N), $P_{X|Y=y}$ is the binomial distribution $Bi(p,y)$ if $y \geq 1$ and the point mass at 0 if $y = 0$.

Using (2) and the fact that binomial distributions have p.d.f.'s w.r.t. counting measure, we obtain that the joint c.d.f. of (X,Y) is

$$F(x,y) = \sum_{k=0}^{y} P_{X|Y=k}((-\infty, x]) \binom{N}{k} \pi^k (1-\pi)^{N-k}$$
$$= \sum_{k=0}^{y} \sum_{j=0}^{\min\{x,k\}} \binom{k}{j} p^j (1-p)^{k-j} \binom{N}{k} \pi^k (1-\pi)^{N-k}$$

for $x = 0, 1, ..., y$, $y = 0, 1, ..., N$.
The marginal c.d.f. $F_X(x) = F(x, \infty) = F(x, N)$.
The p.d.f. of X w.r.t. counting measure is

$$f_X(x) = \sum_{k=x}^{N} \binom{k}{x} p^x (1-p)^{k-x} \binom{N}{k} \pi^k (1-\pi)^{N-k}$$
$$= \binom{N}{x} (\pi p)^x (1-\pi p)^{N-x} \sum_{k=x}^{N} \binom{N-x}{k-x} \left(\frac{\pi - \pi p}{1 - \pi p} \right)^{k-x} \left(\frac{1-\pi}{1-\pi p} \right)^{N-k}$$
$$= \binom{N}{x} (\pi p)^x (1-\pi p)^{N-x}$$

for $x = 0, 1, ..., N$.
It turns out that the marginal distribution of X is the binomial $Bi(\pi p, N)$.

Markov chain

An important example of dependent sequence of random variables in statistical application

A sequence of random vectors $\{X_n : n = 1, 2, ...\}$ is a *Markov chain* or *Markov process* iff

$$P(B|X_1,...,X_n) = P(B|X_n) \text{ a.s.}, \quad B \in \sigma(X_{n+1}), \ n = 2, 3, ...$$

That is, given X_n, X_{n+1} and $(X_1,...,X_{n-1})$ are conditionally independent. We call the previous equation the "Markov property".

Remarks

- X_{n+1} (tomorrow) is conditionally independent of $(X_1,...,X_{n-1})$ (the past), given X_n (today).
- $(X_1,...,X_{n-1})$ is not necessarily independent of (X_n, X_{n+1}).
- A sequence of independent random vectors forms a trivial Markov chain

Example 1.24 (First-order autoregressive processes)

Let $\varepsilon_1, \varepsilon_2, ...$ be independent random variables defined on a probability space, $X_1 = \varepsilon_1$, and $X_{n+1} = \rho X_n + \varepsilon_{n+1}$, $n = 1, 2, ...$, where ρ is a constant in $\mathscr{R}$.

Then $\{X_n\}$ is called a first-order autoregressive process.

We now show that $\{X_n\}$ is a Markov chain

We need to show the Markov property, i.e., for any $B \in \mathscr{B}$ and $n = 1, 2, ...$,

$$P(X_{n+1} \in B | X_1, ..., X_n) = P_{\varepsilon_{n+1}}(B - \rho X_n) = P(X_{n+1} \in B | X_n) \text{ a.s.,}$$

where $B - y = \{x \in \mathscr{R} : x + y \in B\}$.

For any $y \in \mathscr{R}$,

$$P_{\varepsilon_{n+1}}(B - y) = P(\varepsilon_{n+1} + y \in B) = \int I_B(x + y) dP_{\varepsilon_{n+1}}(x)$$

and, by Fubini's theorem, $P_{\varepsilon_{n+1}}(B - y)$ is Borel.

Hence, $P_{\varepsilon_{n+1}}(B - \rho X_n)$ is Borel w.r.t. $\sigma(X_n)$ and, thus, is Borel w.r.t. $\sigma(X_1, ..., X_n)$.

Example 1.24 (continued)

Let $B_j \in \mathscr{B}$, $j = 1, ..., n$, and $A = \cap_{j=1}^n X_j^{-1}(B_j)$.
Since $\varepsilon_{n+1} + \rho X_n = X_{n+1}$ and ε_{n+1} is independent of $(X_1, ..., X_n)$, it follows from Theorem 1.2 and Fubini's theorem that

$$\begin{aligned}\int_A P_{\varepsilon_{n+1}}(B - \rho X_n) dP &= \int_{x_j \in B_j, j=1,...,n} \int_{t \in B - \rho x_n} dP_{\varepsilon_{n+1}}(t) dP_X(x) \\ &= \int_{x_j \in B_j, j=1,...,n, x_{n+1} \in B} dP_{(X, \varepsilon_{n+1})}(x, t) \\ &= P\left(A \cap X_{n+1}^{-1}(B)\right),\end{aligned}$$

where X and x denote $(X_1, ..., X_n)$ and $(x_1, ..., x_n)$, respectively, and x_{n+1} denotes $\rho x_n + t$.
Using this and the argument in the end of the proof for Proposition 1.11, we obtain $P(X_{n+1} \in B | X_1, ..., X_n) = P_{\varepsilon_{n+1}}(B - \rho X_n)$ a.s.
The proof for $P_{\varepsilon_{n+1}}(B - \rho X_n) = P(X_{n+1} \in B | X_n)$ a.s. is similar and simpler.

Proposition 1.12 (Characterizations of Markov chains)

A sequence of random vectors $\{X_n\}$ is a Markov chain if and only if one of the following three conditions holds.

(a) For any $n = 2, 3, ...$ and any integrable $h(X_{n+1})$ with a Borel function h,

$$E[h(X_{n+1})|X_1, ..., X_n] = E[h(X_{n+1})|X_n] \ \text{a.s.}$$

(b) For any $n = 1, 2, ...$ and $B \in \sigma(X_{n+1}, X_{n+2}, ...)$,

$$P(B|X_1, ..., X_n) = P(B|X_n) \ \text{a.s.}$$

(“the past and the future are conditionally independent given the present")

(c) For any $n = 2, 3, ...$, $A \in \sigma(X_1, ..., X_n)$, and $B \in \sigma(X_{n+1}, X_{n+2}, ...)$,

$$P(A \cap B|X_n) = P(A|X_n)P(B|X_n) \ \text{a.s.}$$

Proof

(i) The equivalence between (a) and the Markov property.

It is clear that (a) implies the Markov property.

If h is a simple function, then the Markov property and Proposition 1.10(iii) imply (a).

If h is nonnegative, then there are nonnegative simple functions $h_1 \le h_2 \le \cdots \le h$ such that $h_j \to h$.

Then the Markov property together with Proposition 1.10(iii) and (x) imply (a).

Since $h = h_+ - h_-$, we conclude that the Markov property implies (a).

(ii) The equivalence between (b) and the Markov property.

It is clear that (b) implies the Markov property.

Note that $\sigma(X_{n+1}, X_{n+2}, ...) = \sigma\left(\cup_{j=1}^{\infty} \sigma(X_{n+1}, ..., X_{n+j})\right)$ (Exercise 19).

Hence, to show that the Markov property implies (b), it suffices to show that $P(B|X_1, ..., X_n) = P(B|X_n)$ a.s. for $B \in \sigma(X_{n+1}, ..., X_{n+j})$ for any $j = 1, 2, ...$

We use induction.

The result for $j = 1$ follows from the Markov property.

Proof (continued)

Suppose that the result holds for any $B \in \sigma(X_{n+1}, ..., X_{n+j})$.
To show the result for any $B \in \sigma(X_{n+1}, ..., X_{n+j+1})$, it is enough (why?) to show that for any $B_1 \in \sigma(X_{n+j+1})$ and any $B_2 \in \sigma(X_{n+1}, ..., X_{n+j})$,
$P(B_1 \cap B_2 | X_1, ..., X_n) = P(B_1 \cap B_2 | X_n)$ a.s.
From the proof in (i), the induction assumption implies

$$E[h(X_{n+1}, ..., X_{n+j}) | X_1, ..., X_n] = E[h(X_{n+1}, ..., X_{n+j}) | X_n] \qquad (3)$$

for any Borel function h.
The result follows from

$$\begin{aligned}
E(I_{B_1} I_{B_2} | X_1, ..., X_n) &= E[E(I_{B_1} I_{B_2} | X_1, ..., X_{n+j}) | X_1, ..., X_n] \\
&= E[I_{B_2} E(I_{B_1} | X_1, ..., X_{n+j}) | X_1, ..., X_n] \\
&= E[I_{B_2} E(I_{B_1} | X_{n+j}) | X_1, ..., X_n] \\
&= E[I_{B_2} E(I_{B_1} | X_{n+j}) | X_n] \\
&= E[I_{B_2} E(I_{B_1} | X_n, ..., X_{n+j}) | X_n] \\
&= E[E(I_{B_1} I_{B_2} | X_n, ..., X_{n+j}) | X_n] \\
&= E(I_{B_1} I_{B_2} | X_n) \ \text{ a.s.,}
\end{aligned}$$

Proof (continued)

where the first and last equalities follow from Proposition 1.10(v), the second and sixth equalities follow from Proposition 1.10(vi), the third and fifth equalities follow from the Markov property, and the fourth equality follows from (3).

(iii) The equivalence between (b) and (c)

Let $A \in \sigma(X_1,...,X_n)$ and $B \in \sigma(X_{n+1},X_{n+2},...)$.
If (b) holds, then

$$\begin{aligned} E(I_A I_B|X_n) &= E[E(I_A I_B|X_1,...,X_n)|X_n] \\ &= E[I_A E(I_B|X_1,...,X_n)|X_n] \\ &= E[I_A E(I_B|X_n)|X_n] \\ &= E(I_A|X_n)E(I_B|X_n), \end{aligned}$$

which is (c).

Proof (continued)

Assume that (c) holds.
Let $A_1 \in \sigma(X_n)$, $A_2 \in \sigma(X_1,...,X_{n-1})$, and $B \in \sigma(X_{n+1},X_{n+2},...)$.
Then

$$\begin{aligned}
\int_{A_1\cap A_2} E(I_B|X_n)dP &= \int_{A_1} I_{A_2}E(I_B|X_n)dP \\
&= \int_{A_1} E[I_{A_2}E(I_B|X_n)|X_n]dP \\
&= \int_{A_1} E(I_{A_2}|X_n)E(I_B|X_n)dP \\
&= \int_{A_1} E(I_{A_2}I_B|X_n)dP \\
&= P(A_1\cap A_2\cap B).
\end{aligned}$$

Since disjoint unions of events of the form $A_1 \cap A_2$ as specified above generate $\sigma(X_1,...,X_n)$, this shows that $E(I_B|X_n) = E(I_B|X_1,...,X_n)$ a.s., which is (b).

Martingales

$\{X_n\}$: a sequence of integrable random variables on $(\Omega, \mathscr{F}, P)$
$\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}$: a sequence of σ-fields such that $\sigma(X_n) \subset \mathscr{F}_n$
$\{X_n, \mathscr{F}_n : n = 1, 2, ...\}$ or $\{X_n\}$ when $\mathscr{F}_n = \sigma(X_1, ..., X_n)$ is said to be a *martingale* if

$$E(X_{n+1}|\mathscr{F}_n) = X_n \text{ a.s., } n = 1, 2, ...$$

a *submartingale* or *supermartingale* if $=$ is replaced by $\geq$ or $\leq$
A simple property of a martingale (or a submartingale) $\{X_n, \mathscr{F}_n\}$ is that $E(X_{n+j}|\mathscr{F}_n) = X_n$ a.s. (or $E(X_{n+j}|\mathscr{F}_n) \geq X_n$ a.s.) and $EX_1 = EX_j$ (or $EX_1 \leq EX_2 \leq \cdots$) for any $j = 1, 2, ...$

Examples

- Y: an intgrable random variable, $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}$
  $\{E(Y|\mathscr{F}_n)\}$ is a martingale
- $X_n = \varepsilon_1 + \cdots + \varepsilon_n$, $n = 1, 2, ...$, ε_n's are independent

$$E(X_{n+1}|X_1, ..., X_n) = E(X_n + \varepsilon_{n+1}|X_1, ..., X_n) = X_n + E\varepsilon_{n+1} \text{ a.s.,}$$

  $\{X_n\}$ is a martingale or submartingale if $E\varepsilon_n = 0$ or ≥ 0 for all n

Proposition 1.13.

(i) If $\{X_n, \mathscr{F}_n\}$ is a martingale, φ is convex, and $\varphi(X_n)$ is integrable for all n, then $\{\varphi(X_n), \mathscr{F}_n\}$ is a submartingale.
(ii) If $\{X_n, \mathscr{F}_n\}$ is a submartingale, $\varphi(X_n)$ is integrable for all n, and φ is nondecreasing and convex, then $\{\varphi(X_n), \mathscr{F}_n\}$ is a submartingale.
Proof. (i) Note that $\varphi(X_n) = \varphi(E(X_{n+1}|\mathscr{F}_n)) \leq E[\varphi(X_{n+1}|\mathscr{F}_n)]$ a.s. by Jensen's inequality for conditional expectations (Exercise 89(c)).
(ii) Since φ is nondecreasing and $\{X_n, \mathscr{F}_n\}$ is a submartingale,
$\varphi(X_n) \leq \varphi(E(X_{n+1}|\mathscr{F}_n)) \leq E[\varphi(X_{n+1}|\mathscr{F}_n)]$ a.s.

Proposition 1.15.

Let $\{X_n, \mathscr{F}_n\}$ be a submartingale. If $c = \sup_n E|X_n| < \infty$, then $\lim_{n\to\infty} X_n = X$ a.s., where X is a random variable satisfying $E|X| \leq c$.

Example.

$Y_1, ..., Y_n$ are independent, $Y_n > 0$, and $EY_n = 1$
$\{X_n = Y_1 \cdots Y_n\}$ is a martingale
$E(X_{n+1}|X_1, ..., X_n) = E(Y_1 \cdots Y_{n+1}|Y_1, ..., Y_n) = Y_1 \cdots Y_n E(Y_{n+1}) = X_n$
$E|X_n| = 1$, hence $\lim_{n\to\infty} Y_1 \cdots Y_n = X$ a.s.