Math 831 : Theory of Probability

1. Introduction: The basic model and repeated trials

- Experiments
- Repeated trials and intuitive probability
- Numerical observations and "random variables"
- The distribution of a random variable
- The law of large numbers and expectations

Experiments

Probability models *experiments* in which repeated *trials* typically result in different *outcomes*.

As a part of mathematics, Kolmolgorov's axioms [3] for experiments determine probability in the same sense that Euclid's axioms determine geometry.

As a means of understanding the "real world," probability identifies surprising regularities in highly irregular phenomena.

Anticipated regularity

If we roll a die 100 times we anticipate that about a sixth of the time the roll is 5.

If that doesn't happen, we suspect that something is wrong with the die or the way it was rolled.

Probabilities of events

Events are statements about the outcome of the experiment: $\{$the roll is 6$\}$, $\{$the rat died$\}$, $\{$the television set is defective$\}$

The anticipated regularity is that

$$P(A) \approx \frac{\#\text{times } A \text{ occurs}}{\#\text{of trials}}$$

This presumption is called the *relative frequency* interpretation of probability.

“Definition” of probability

The probability of an event A should be

$$P(A) = \lim_{n\to\infty} \frac{\#\text{times } A \text{ occurs in first } n \text{ trials}}{n}$$

The mathematical problem: Make sense out of this.

The real world relationship: Probabilities are predictions about the future.

Random variables

In performing an experiment numerical measurements or observations are made. Call these *random variables* since they vary randomly.

Give the quantity a name: X

$\{X = a\}$ and $\{a < X < b\}$ are statements about the outcome of the experiment, that is, are events

The distribution of a random variable

If X_k is the value of X observed on the kth trial, then we should have

$$P\{X = a\} = \lim_{n\to\infty} \frac{\#\{k \le n : X_k = a\}}{n}$$

If X has only finitely many possible values, then

$$\sum_{a\in\mathcal{R}(X)} P\{X = a\} = 1.$$

This collection of probabilities determine the *distribution* of X.

Distribution function

More generally,

$$P\{X \leq x\} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{(-\infty,x]}(X_k)$$

$F_X(x) \equiv P\{X \leq x\}$ is the *distribution function* for X.

The law of averages

If $\mathcal{R}(X) = \{a_1, \ldots, a_m\}$ is finite, then

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \lim_{n\to\infty} \sum_{l=1}^{m} a_l \frac{\#\{k \leq n : X_k = a_l\}}{n} = \sum_{l=1}^{m} a_l P\{X = a_l\}$$

More generally, if $\mathcal{R}(X) \subset [c, d]$, $-\infty < c < d < \infty$, then

$$\begin{aligned}
\sum_l x_l P\{x_l < X \leq x_{l+1}\} &= \lim_{n\to\infty} \sum_{l=1}^m x_l \frac{\#\{k \leq n : x_l < X_k \leq x_{l+1}\}}{n} \\
&\leq \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} \\
&\leq \lim_{n\to\infty} \sum_{l=1}^m x_{l+1} \frac{\#\{k \leq n : x_l < X_k \leq x_{l+1}\}}{n} \\
&= \sum_l x_{l+1} P\{x_l < X \leq x_{l+1}\} \\
&= \sum_l x_{l+1} (F_X(x_{l+1}) - F_X(x_l)) \\
&\to \int_c^d x dF_X(x)
\end{aligned}$$

The expectation as a Stieltjes integral

If $\mathcal{R}(X) \subset [c, d]$, define

$$E[X] = \int_c^d x dF_X(x).$$

If the relative frequency interpretation is valid, then

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = E[X].$$

A random variable without an expectation

Example 1.1 *Suppose*

$$P\{X \leq x\} = \frac{x}{1+x}, \qquad x \geq 0.$$

Then

$$\begin{aligned} \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} &\geq \sum_{l=0}^{m} l P\{l < X \leq l+1\} \\ &= \sum_{l=0}^{m} l\left(\frac{l+1}{l+2} - \frac{l}{l+1}\right) \\ &= \sum_{l=0}^{m} \frac{l}{(l+2)(l+1)} \to \infty \quad \text{as } m \to \infty \end{aligned}$$

2. The Kolmogorov axioms

- The sample space and events
- Probability measures
- Random variables

The Kolmogorov axioms: The sample space

The possible outcomes of the experiment form a *set* Ω called the *sample space*.

Each *event* (statement about the outcome) can be identified with the subset of the sample space for which the statement is true.

The Kolmogorov axioms: The collection of events

If

$$\begin{aligned} A &= \{\omega \in \Omega : \text{ statement I is true for } \omega\} \\ B &= \{\omega \in \Omega : \text{ statement II is true for } \omega\} \end{aligned}$$

Then

$$\begin{aligned} A \cap B &= \{\omega \in \Omega : \text{ statement I and statement II are true for } \omega\} \\ A \cup B &= \{\omega \in \Omega : \text{ statement I or statement II is true for } \omega\} \\ A^c &= \{\omega \in \Omega : \text{ statement I is not true for } \omega\} \end{aligned}$$

Let $\mathcal{F}$ be the collection of events. Then $A, B \in \mathcal{F}$ should imply that $A \cap B$, $A \cup B$, and A^c are all in $\mathcal{F}$. $\mathcal{F}$ is an *algebra* of subsets of Ω.

In fact, we assume that $\mathcal{F}$ is a σ-algebra (closed under countable unions and complements).

The Kolmogorov axioms: The probability measure

Each event $A \in \mathcal{F}$ is assigned a probability $P(A) \geq 0$.

From the relative frequency interpretation, we must have

$$P(A \cup B) = P(A) + P(B)$$

for disjoint events A and B and by induction, if $A_1, \ldots, A_m$ are disjoint

$$P(\cup_{k=1}^{m} A_k) = \sum_{k=1}^{m} P(A_k) \qquad \textit{finite additivity}$$

In fact, we assume *countable additivity*: If $A_1, A_2, \ldots$ are disjoint events, then

$$P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k).$$

$P(\Omega) = 1$.

A probability space is a measure space

A *measure space* $(M, \mathcal{M}, \mu)$ consists of a set M, a σ-algebra of subsets $\mathcal{M}$, and a nonnegative function μ defined on $\mathcal{M}$ that satisfies $\mu(\emptyset) = 0$ and countable additivity.

A *probability space* is a measure space $(\Omega, \mathcal{F}, P)$ satsifying $P(\Omega) = 1$.

Random variables

If X is a random variable, then we must know the value of X if we know that outcome $\omega \in \Omega$ of the experiment. Consequently, X is a function defined on Ω.

The statement $\{X \leq c\}$ must be an event, so

$$\{X \leq c\} = \{\omega : X(\omega) \leq c\} \in \mathcal{F}.$$

In other words, X is a *measurable function* on $(\Omega, \mathcal{F}, P)$.

Distributions

Definition 2.1 *The* Borel subsets $\mathcal{B}(\mathbb{R})$ *is the smallest* σ*-algebra of subsets of* $\mathbb{R}$ *containing* $(-\infty, c]$ *for all* $c \in \mathbb{R}$.

See Problem 1.

If X is a random variable, then $\{B : \{X \in B\} \in \mathcal{F}\}$ is a σ-algebra containing $\mathcal{B}(\mathbb{R})$. See Problem 2.

Definition 2.2 *The* distribution *of a* $\mathbb{R}$*-valued random variable* X *is the Borel measure defined by* $\mu_X(B) = P\{X \in B\}$, $B \in \mathcal{B}(\mathbb{R})$.

μ_X is called the measure *induced* by the function X.

3. Modeling information

- Information and σ-algebras
- Information from observing random variables

Information and σ-algebras

If I know whether or not $\omega \in A$, then I know whether or not $\omega \in A^c$.

If in addition, I know whether or not $\omega \in B$, then I know whether or not $\omega \in A \cup B$ and whether or not $\omega \in A \cap B$.

Consequently, we will assume that "available information" corresponds to a σ-algebra of events.

Information obtained by observing random variables

For example, if X is a random variable, $\sigma(X)$ will denote the smallest σ-algebra containing $\{X \leq c\}$, $c \in \mathbb{R}$. $\sigma(X)$ is called the σ-algebra *generated* by X.

Lemma 3.1 *Let X be a random variable. Then*

$$\sigma(X) = \{\{X \in B\} : B \in \mathcal{B}(\mathbb{R})\}.$$

If $X_1, \ldots, X_m$ are random variables, then $\sigma(X_1, \ldots, X_m)$ denotes the smallest σ-algebra containing $\{X_i \leq c\}$, $c \in \mathbb{R}$, $i = 1, \ldots, m$.

Lemma 3.2 *Let $X_1, \ldots, X_m$ be random variables. Then*

$$\sigma(X_1, \ldots, X_m) = \{\{(X_1, \ldots, X_m) \in B\} : B \in \mathcal{B}(\mathbb{R}^m)\}$$

where $\mathcal{B}(\mathbb{R}^m)$ is the Borel σ-algebra in $\mathbb{R}^m$.

Compositions

Let $f : \mathbb{R}^m \to \mathbb{R}$ and $Y = f(X_1, \ldots, X_m)$. Let $f^{-1}(B) = \{x \in \mathbb{R}^m : f(x) \in B\}$. Then

$$\{Y \in B\} = \{(X_1, \ldots, X_m) \in f^{-1}(B)\}$$

Definition 3.3 $f : \mathbb{R}^m \to \mathbb{R}$ *is* Borel measurable *if and only if* $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^m)$ *for each* $B \in \mathcal{B}(\mathbb{R})$.

Note that $\{B \subset \mathbb{R} : f^{-1}(B) \in \mathcal{B}(\mathbb{R}^m)\}$ is a σ-algebra.

Lemma 3.4 *If* $X_1, \ldots, X_m$ *are random variables and* f *is Borel measurable, then* $Y = f(X_1, \ldots, X_m)$ *is a random variable.*

Note that every continuous function $f : \mathbb{R}^m \to \mathbb{R}$ is Borel measurable.

4. Measure and integration

- Closure properties of the collection of random variables
- Almost sure convergence
- Some properties of measures
- Integrals and expectations
- Convergence theorems
- When are two measures equal: The Dynkin-class theorem
- Distributions and expectations
- Markov and Chebychev inequalities
- Convergence of series

Closure properties of collection random variables

Lemma 4.1 *Suppose* $\{X_n\}$ *are* $\mathcal{D}$*-measurable,* $[-\infty, \infty]$*-valued random variables. Then*

$$\sup_n X_n, \quad \inf_n X_n, \quad \limsup_{n\to\infty} X_n, \quad \liminf_{n\to\infty} X_n$$

are $\mathcal{D}$*-measurable,* $[-\infty, \infty]$*-valued random variables*

Proof. Let $Y = \sup_n X_n$. Then $\{Y \leq c\} = \cap_n \{X_n \leq c\} \in \mathcal{D}$. Let $Z = \inf_n X_n$. Then $\{Z \geq c\} = \cap_n \{X_n \geq c\} \in \mathcal{D}$.

Note that $\liminf_{n\to\infty} X_n = \sup_n \inf_{m\geq n} X_m$. □

Almost sure convergence

Definition 4.2 *A sequence of random variables* $\{X_n\}$ converges almost surely *(a.s.) if* $P\{\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n\} = 1.$

We write $Z = \lim_{n\to\infty} X_n$ a.s. if

$$P\{Z = \limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n\} = 1$$

Properties of measures

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

If $A, B \in \mathcal{F}$ and $A \subset B$, then $P(A) \leq P(B)$. $(B = A \cup (A^c \cap B))$

If $\{A_k\} \subset \mathcal{F}$, then

$$P(\cup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k)$$

Define $B_1 = A_1$, $B_k = A_k \cap (A_1 \cup \cdots \cup A_{k-1})^c \subset A_k$, and note that $\{B_k\}$ are disjoint and $\cup_{k=1}^{\infty} B_k = \cup_{k=1}^{\infty} A_k$

If $A_1 \subset A_2 \subset A_3 \subset \cdots$, then $P(\cup_{k=1}^{\infty} A_k) = \lim_{n\to\infty} P(A_n) = \sum_{k=1}^{\infty} P(A_k \cap A_{k-1}^c)$

If $A_1 \supset A_2 \supset A_3 \supset \cdots$, then $P(\cap_{k=1}^{\infty} A_k) = \lim_{n\to\infty} P(A_n) = P(A_1) - \sum_{k=1}^{\infty} P(A_k^c \cap A_{k-1})$

Expectations and integration

Simple functions/Discrete random variables

Suppose X assumes finitely many values $\{a_1, \ldots, a_m\}$. Then

$$E[X] = \sum_{k=1}^{m} a_k P\{X = a_k\}$$

If

$$X = \sum_{k=1}^{m} a_k \mathbf{1}_{A_k},$$

then

$$\int_\Omega X dP = \sum_{k=1}^{m} a_k P(A_k)$$

Nonnegative random variables

If $P\{0 \leq X < \infty\} = 1$, then

$$E[X] = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} P\{\frac{k}{2^n} < X \leq \frac{k+1}{2^n}\} = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k+1}{2^n} P\{\frac{k}{2^n} < X \leq \frac{k+1}{2^n}$$

If $P\{0 \leq X \leq \infty\} = 1$, then $\int_\Omega X dP$ is defined by

$$\int_\Omega X dP = \sup\{\int_\Omega Y dP : Y \leq X, Y \text{ simple}\}$$

If $P\{X = \infty\} > 0$, then $\int_\Omega X dP = \infty$. If $P\{0 \leq X < \infty\} = 1$,

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} P\{\frac{k}{2^n} < X \leq \frac{k+1}{2^n}\} \leq \int_\Omega X dP$$

$$\leq \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k+1}{2^n} P\{\frac{k}{2^n} < X \leq \frac{k+1}{2^n}\}$$

General expecation/integral

Let $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$. If

$$E[|X|] = E[X^+] + E[X^-] < \infty,$$

then

$$\int_\Omega X dP = E[X] \equiv E[X^+] - E[X^-]$$

Properties

Linearity: $E[aX + bY] = aE[X] + bE[Y]$

Monotonicity: $P\{X \leq Y\} = 1$ implies $E[X] \leq E[Y]$

Approximation

Lemma 4.3 *Let* $X \geq 0$. *Then* $\lim_{c\to\infty} E[X \wedge c] = E[X]$.

Proof. By monotonicity, $\lim_{c\to\infty} E[X \wedge c] \leq E[X]$. If Y is a simple random variable and $Y \leq X$, then for $c \geq \max_\omega Y(\omega)$, $Y \leq X \wedge c$ and $E[X \wedge c] \geq E[Y]$. Consequently,

$$\lim_{c\to\infty} E[X \wedge c] \geq \sup_{\{Y \text{ simple}: Y \leq X\}} E[Y] = E[X]$$

□

The monotone convergence theorem

Theorem 4.4 *Let* $0 \leq X_1 \leq X_2 \leq \cdots$ *be random variables and define* $X = \lim_{n\to\infty} X_n$. *Then* $\lim_{n\to\infty} E[X_n] = E[X]$.

Proof. For $\epsilon > 0$, let $A_n = \{X_n \leq X - \epsilon\}$. Then $A_1 \supset A_2 \supset \cdots$ and $\cap A_n = \emptyset$. For $c > 0$, $X \wedge c \leq \mathbf{1}_{A_n^c}(X_n \wedge c + \epsilon) + c\mathbf{1}_{A_n}$, so

$$E[X \wedge c] \leq \epsilon + E[X_n \wedge c] + cP(A_n).$$

Consequently,

$$E[X] = \lim_{c\to\infty} E[X \wedge c] = \lim_{c\to\infty} \lim_{n\to\infty} E[X_n \wedge c] \leq \lim_{n\to\infty} E[X_n]$$

$\square$

Fatou's lemma

Lemma 4.5 *Let* $X_n \geq 0$. *Then*

$$\liminf_{n\to\infty} E[X_n] \geq E[\liminf_{n\to\infty} X_n]$$

Proof.

$$\begin{aligned}
\liminf_{n\to\infty} E[X_n] &\geq \lim_{n\to\infty} E[\inf_{m\geq n} X_m] \\
&= E[\lim_{n\to\infty} \inf_{m\geq n} X_n] \\
&= E[\liminf_{n\to\infty} X_n]
\end{aligned}$$

□

Dominated converence theorem

Theorem 4.6 *Suppose* $|X_n| \leq Y_n$, $\lim_{n\to\infty} X_n = X$ *a.s.*, $\lim_{n\to\infty} Y_n = Y$ *a.s., and* $\lim_{n\to\infty} E[Y_n] = E[Y] < \infty$. *Then* $\lim_{n\to\infty} E[X_n] = E[X]$.

Proof.

$$\liminf E[Y_n + X_n] \geq E[Y + X],$$

so $\liminf E[X_n] \geq E[X]$. Similarly,

$$\liminf E[Y_n - X_n] \geq E[Y - X],$$

so $\limsup E[X_n] = -\liminf E[-X_n] \leq E[X]$. □

The Dynkin-class theorem

A collection $\mathcal{D}$ of subsets of Ω is a *Dynkin class* if $\Omega \in \mathcal{D}$, $A, B \in \mathcal{D}$ and $A \subset B$ imply $B - A \in \mathcal{D}$, and $\{A_n\} \subset \mathcal{D}$ with $A_1 \subset A_2 \subset \cdots$ implies $\cup_n A_n \in \mathcal{D}$.

Theorem 4.7 *Let $\mathcal{S}$ be a collection of subsets of Ω such that $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$. If $\mathcal{D}$ is a Dynkin class with $\mathcal{S} \subset \mathcal{D}$, then $\sigma(\mathcal{S}) \subset \mathcal{D}$.*

$\sigma(\mathcal{S})$ denotes the smallest σ-algebra containing $\mathcal{S}$.

Example 4.8 *If Q_1 and Q_2 are probability measures on Ω, then $\{B : Q_1(B) = Q_2(B)\}$ is a Dynkin class.*

Proof. Let $D(\mathcal{S})$ be the smallest Dynkin-class containing $\mathcal{S}$.

If $A, B \in \mathcal{S}$, then $A^c = \Omega - A$, $B^c = \Omega - B$, and $A^c \cup B^c = \Omega - A \cap B$ are in $D(\mathcal{S})$.

Consequently, $A^c \cup B^c - A^c = A \cap B^c$, $A^c \cup B = \Omega - A \cap B^c$, $A^c \cap B^c = A^c \cup B - B$, and $A \cup B = \Omega - A^c \cap B^c$ are in $D(\mathcal{S})$.

For $A \in \mathcal{S}$, $\{B : A \cup B \in D(\mathcal{S})\}$ is a Dynkin class containing $\mathcal{S}$, and hence $D(\mathcal{S})$.

Consequently, for $A \in D(\mathcal{S})$, $\{B : A \cup B \in D(\mathcal{S})\}$ is a Dynkin class containing $\mathcal{S}$ and hence $D(\mathcal{S})$.

It follows that $A, B \in D(\mathcal{S})$ implies $A \cup B \in D(\mathcal{S})$. But if $D(\mathcal{S})$ is closed under finite unions it is closed under countable unions. $\square$

Equality of two measures

Lemma 4.9 *Let μ and ν be measures on $(M, \mathcal{M})$. Let $\mathcal{S} \subset \mathcal{M}$ be closed under finite intersections. Suppose that $\mu(M) = \nu(M)$ and $\mu(B) = \nu(B)$ for each $B \in \mathcal{S}$. Then $\mu(B) = \nu(B)$ for each $B \in \sigma(\mathcal{S})$.*

Proof. Since $\mu(M) = \nu(M)$, $\{B : \mu(B) = \nu(B)\}$ is a Dynkin-class containing $\mathcal{S}$ and hence contains $\sigma(\mathcal{S})$. $\square$

For example: $M = \mathbb{R}^d$, $\mathcal{S} = \{\prod_{i=1}^d (-\infty, c_i] : c_i \in \mathbb{R}\}$. If

$$P\{X_1 \leq c_1, \ldots, X_d \leq c_d\} = P\{Y_1 \leq c_1, \ldots, Y_d \leq c_d\}, \quad c_1, \ldots, c_d \in \mathbb{R},$$

then

$$P\{(X_1, \ldots, X_d) \in B\} = P\{(Y_1, \ldots, Y_d) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^d).$$

The distribution and expectation of a random variable

We defined

$$\mu_X(B) = P\{X \in B\}, \quad B \in \mathcal{B}(\mathbb{R}).$$

If X is simple,

$$X = \sum a_k \mathbf{1}_{\{X=a_k\}},$$

then

$$\int_\Omega XdP = E[X] = \sum_k a_k P\{X = a_k\} = \sum a_k \mu_X\{a_k\} = \int_\mathbb{R} x\mu_X(dx)$$

and for positive X,

$$\begin{aligned} E[X] = \int_\Omega XdP &= \lim_{n\to\infty} \int_\Omega \sum_l \frac{l}{n} \mathbf{1}_{\{\frac{l}{n}<X\leq\frac{l+1}{n}\}} dP \\ &= \lim_{n\to\infty} \sum_l \frac{l}{n} \mu_X(\frac{l}{n}, \frac{l+1}{n}] = \int_\mathbb{R} x\mu_X(dx). \end{aligned}$$

Expectation of a function of a random varialbe

Lemma 4.10 *Assume that $g : \mathbb{R} \to [0, \infty)$ is a Borel measurable function, and let $Y = g(X)$. Then*

$$E[Y] = \int_\Omega Y dP = \int_\mathbb{R} g(x)\mu_X(dx) \tag{4.1}$$

More generally, Y is integrable with respect to P if and only if g is integrable with respect to μ_X and (4.1) holds.

Proof. Let $Y_n = \sum \frac{l}{n}\mathbf{1}_{\{\frac{l}{n}<Y\leq\frac{l+1}{n}\}}$ and $g_n = \sum_l \frac{l}{n}\mathbf{1}_{g^{-1}((\frac{l}{n},\frac{l+1}{n}])}$. Then

$$\begin{aligned} E[Y_n] &= \sum_l \frac{l}{n}P\{\frac{l}{n} < Y \leq \frac{l+1}{n}\} \\ &= \sum_l \frac{l}{n}P\{X \in g^{-1}((\frac{l}{n},\frac{l+1}{n}])\} \\ &= \sum_l \frac{l}{n}\mu_X(g^{-1}((\frac{l}{n},\frac{l+1}{n}])) \\ &= \int_{\mathbb{R}} g_n(x)\mu_X(dx), \end{aligned}$$

and the lemma follows by the monotone convergence theorem.

The last assertion follow by the fact that $Y^+ = g^+(X)$ and $Y^- = g^-(X)$. $\square$

Lebesgue measure

Lebesgue measure L is the measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying $L(a, b] = b - a$. Note that uniqueness follows by the Dynkin class theorem.

If g is Riemann integrable, then

$$\int_{\mathbb{R}} g(x) L(dx) = \int_{\mathbb{R}} g(x) dx,$$

so one usually writes $\int_{\mathbb{R}} g(x)dx$ rather than $\int_{\mathbb{R}} g(x)L(dx)$. Note that there are many functions that are Lebesgue integrable but not Riemann integrable.

Distributions with a (Lebesgue) density

A random variable X has a Lebesgue density f_X if

$$P\{X \in B\} = \mu_X(B) = \int_{\mathbb{R}} \mathbf{1}_B(x) f_X(x)dx \equiv \int_B f_X(x)dx.$$

By an argument similar to the proof of Lemma 4.10,

$$E[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)dx.$$

Markov inequality

Lemma 4.11 *Let Y be a nonnegative random variable. Then for $c > 0$,*

$$P\{Y \geq c\} \leq \frac{E[Y]}{c}.$$

Proof. Observing that

$$c\mathbf{1}_{\{Y \geq c\}} \leq Y,$$

the inequality follows by monotonicity. □

Chebychev inequality

If X is integrable and $Y = X - E[X]$, then $E[Y] = 0$. (Y is X "centered at its expectation.") The variance of X is defined by

$$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

Then

$$P\{|X - E[X]| \geq \epsilon\} = P\{(X - E[X])^2 \geq \epsilon^2\} \leq \frac{Var(X)}{\epsilon^2}.$$

Convergence of series

A series $\sum_{k=1}^{\infty} a_k$ *converges* if $\lim_{n\to\infty} \sum_{k=1}^{n} a_k$ exists and is finite.

A series *converges absolutely* if $\lim_{n\to\infty} \sum_{k=1}^{n} |a_k| < \infty$.

Lemma 4.12 *If a series converges absolutely, it converges.*

Series of random variables

Let $X_1, X_2, \ldots$ be random variables. The series $\sum_{k=1}^{\infty} X_k$ *converges almost surely*, if

$$P\{\omega : \lim_{n\to\infty} \sum_{k=1}^{n} X_k(\omega) \text{ exists and is finite}\} = 1.$$

The series *converges absolutely almost surely*, if

$$P\{\omega : \sum_{k=1}^{\infty} |X_k(\omega)| < \infty\} = 1.$$

Lemma 4.13 *If a series of random variables converges absolutely almost surely, it converges almost surely.*

If $\sum_{k=1}^{\infty} E[|X_k|] < \infty$*, then* $\sum_{k=1}^{\infty} X_k$ *converges absolutely almost surely.*

Proof.

If $\sum_{k=1}^{\infty}|X_k(\omega)| = \lim_{n\to\infty}\sum_{k=1}^{n}|X_k(\omega)| < \infty$, then $n < m$,

$$|\sum_{k=1}^{m} X_k(\omega) - \sum_{k=1}^{n} X_k(\omega)| \le \sum_{k=n+1}^{\infty} |X_k(\omega)| = \sum_{k=1}^{\infty}|X_k(\omega)| - \sum_{k=1}^{n}|X_k(\omega)|$$

By the monotone convergence theorem

$$E[\sum_{k=1}^{\infty}|X_k|] = \lim_{n\to\infty} E[\sum_{k=1}^{n}|X_k|] = \sum_{k=1}^{\infty} E[|X_k|] < \infty,$$

which implies $\sum_{k=1}^{\infty}|X_k| < \infty$ almost surely. □

5. Discrete and combinatorial probability

- Discrete probability spaces
- Probability spaces with equally likely outcomes
- Elementary combinatorics
- Binomial distribution

Discrete probability spaces

If Ω is countable and $\{\omega\} \in \mathcal{F}$ for each $\omega \in \Omega$, then $\mathcal{F}$ is the collection of all subsets of Ω and for each $A \subset \Omega$,

$$P(A) = \sum_{\omega \in A} P\{\omega\}.$$

Similarly, if $X \geq 0$,

$$E[X] = \sum_{\omega \in \Omega} X(\omega) P\{\omega\}$$

Probability spaces with equally likely outcomes

If Ω is finite and all elements in Ω are "equally likely," that is, $P\{\omega\} = P\{\omega'\}$ for $\omega, \omega' \in \Omega$, then for $A \subset \Omega$,

$$P(A) = \frac{\#A}{\#\Omega}.$$

Calculating probabilities becomes a counting problem.

Ordered sampling without replacement

An urn U contains n balls. m balls are selected randomly one at a time without replacement:

$$\Omega_o = \{(a_1, \ldots, a_m) : a_i \in U, a_i \neq a_j\}$$

Then

$$\#\Omega_o = n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!}.$$

Unordered samples

A urn U contains n balls. m are selected randomly.

$$\Omega_u = (\alpha \subset U : \#\alpha = m\}$$

Each $\alpha = \{a_1, \ldots, a_m\} \in \Omega_u$ can be ordered in $m!$ different ways, so

$$\#\Omega_o = \#\Omega_u \times m!.$$

Therefore,

$$\#\Omega_u = \frac{n!}{(n-m)!m!} = \binom{n}{m}.$$

Numbered balls

Suppose the balls are numbered 1 through n. Assuming ordered sampling, let X_k be the number on the kth ball drawn, $k = 1, \ldots, m$.

$\Omega = \{(a_1, \ldots, a_m) : a_i \in \{1, \ldots, n\}, a_i \neq a_j \text{ for } i \neq j\}$

$\#\Omega = \frac{n!}{(n-m)!}$

$X_k(a_1, \ldots, a_m) = a_k$.

$$E[X_k] = \frac{1}{n} \sum_{l=1}^{n} l = \frac{n+1}{2}$$

Flip a fair coin 6 times

$\Omega = \{(a_1, \ldots, a_6) : a_i \in \{H, T\}\} \qquad \#\Omega = 2^6$

$$X_k(a_1, \ldots, a_6) = \begin{cases} 1 & k\text{th flip is heads} \\ 0 & k\text{th flip is tails} \end{cases}$$

$$X_k(a_1, \ldots, a_6) = \begin{cases} 1 & \text{if } a_k = H \\ 0 & \text{if } a_k = T \end{cases}$$

Let

$$S_6 = \sum_{k=1}^{6} X_k$$

$$P\{S_6 = l\} = \frac{\#\{S_6 = l\}}{\#\Omega} = \binom{6}{l} \frac{1}{2^6}$$

Lopsided coins

We want to model n flips of a coin for which the probability of heads is p. Let

$$\Omega = \{a_1 \cdots a_n : a_i = H \text{ or } T\}.$$

Let $S_n(a_1 \cdots a_n)$ be the number of indices i such that $a_i = H$, and define

$$P(a_1 \cdots a_n) = p^{S_n}(1-p)^{n-S_n}.$$

Note that with this definition of P, S_n has a binomial distribution

$$P\{S_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Moments

S_n is binomially distributed with parameters n and p if

$$P\{S_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Then

$$E[S_n] = np$$

and for $m < n$,

$$E[S_n(S_n-1)\cdots(S_n-m)] = n(n-1)\cdots(n-m)p^{m+1}.$$

Consequently,

$$E[S_n^2] = n(n-1)p^2 + np, \quad Var(S_n) = np(1-p)$$

6. Product measures and repeated trials

- Product spaces
- Product measure
- Tonelli's theorem
- Fubini's theorem
- Infinite product measures
- Relative frequency

Product spaces

Let $(M_1, \mathcal{M}_1, \mu_1)$ and $(M_2, \mathcal{M}_2, \mu_2)$ be measure spaces. Define

$$M_1 \times M_2 = \{(z_1, z_2) : z_1 \in M_1, z_2 \in M_2\}$$

For example: $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$

Let

$$\mathcal{M}_1 \times \mathcal{M}_2 = \sigma\{A_1 \times A_2 : A_1 \in \mathcal{M}_1, A_2 \in \mathcal{M}_2\}$$

For example: $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$.

Measurability of section

Lemma 6.1 *If* $A \in \mathcal{M}_1 \times \mathcal{M}_2$ *and* $z_1 \in M_1$*, then*

$$A_{z_1} = \{z_2 : (z_1, z_2) \in A\} \in \mathcal{M}_2.$$

Proof. Let $\Gamma_{z_1} = \{A \in \mathcal{M}_1 \times \mathcal{M}_2 : A_{z_1} \in \mathcal{M}_2\}$. Note that $A_1 \times A_2 \in \Gamma_{z_1}$ for $A_1 \in \mathcal{M}_1$ and $A_2 \in \mathcal{M}_2$. Check that Γ_{z_1} is a σ-algebra. □

A measurability lemma

Lemma 6.2 *If* $A \in \mathcal{M}_1 \times \mathcal{M}_2$*, then*

$$f_A(z_1) = \int_{M_2} \mathbf{1}_A(z_1, z_2)\mu_2(dz_2)$$

is measurable.

Proof. Check that the collection of A satisfying the conclusion of the lemma is a Dynkin class containing $A_1 \times A_2$, for $A_1 \in \mathcal{M}_1$ and $A_2 \in \mathcal{M}_2$. □

Product measure

Lemma 6.3 *For $A \in \mathcal{M}_1 \times \mathcal{M}_2$, define*

$$\mu_1 \times \mu_2(A) = \int_{M_1} \int_{M_2} \mathbf{1}_A(z_1, z_2)\mu_2(dz_2)\mu_1(dz_1). \tag{6.1}$$

Then $\mu_1 \times \mu_2$ is a measure satisfying

$$\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \quad A_1 \in \mathcal{M}_1, A_2 \in \mathcal{M}_2. \tag{6.2}$$

There is a most one measure on $\mathcal{M}_1 \times \mathcal{M}_2$ satisfying (6.2), so

$$\int_{M_1} \int_{M_2} \mathbf{1}_A(z_1, z_2)\mu_2(dz_2)\mu_1(dz_1) = \int_{M_2} \int_{M_1} \mathbf{1}_A(z_1, z_2)\mu_1(dz_1)\mu_2(dz_2). \tag{6.3}$$

Proof. $\mu_1 \times \mu_2$ is countably additive by the linearity of the integral and the monotone convergence theorem.

Uniqueness follows by Lemma 4.9. □

Tonelli's theorem

Theorem 6.4 *If f is nonnegative $\mathcal{M}_1 \times \mathcal{M}_2$-measurable function, then*

$$\begin{aligned}\int_{M_1}\int_{M_2} f(z_1,z_2)\mu_2(dz_2)\mu_1(dz_1) &= \int_{M_2}\int_{M_1} f(z_1,z_2)\mu_1(dz_1)\mu_2(dz_2) \\ &= \int_{M_1\times M_2} f(z_1,z_2)\mu_1\times\mu_2(dz_1\times dz_2)\end{aligned}$$

Proof. The result holds for simple functions by (6.3) and the definition of $\mu_1 \times \mu_2$, and in general, by the monotone convergence theorem.
□

Example

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $([0,\infty), \mathcal{B}[0,\infty), dx)$, be the measure space corresponding to Lebesgue measure on the half line. Then if $P\{X \geq 0\} = 1$,

$$E[X] = \int_\Omega \int_0^\infty \mathbf{1}_{[0,X(\omega))}(x) dx P(d\omega) = \int_0^\infty P\{X > x\} dx$$

Fubini's theorem

Theorem 6.5 *If f is $\mathcal{M}_1 \times \mathcal{M}_2$-measurable and*

$$\int_{M_1} \int_{M_2} |f(z_1, z_2)| \mu_2(dz_2) \mu_1(dz_1) < \infty,$$

then f is $\mu_1 \times \mu_2$ integrable and

$$\begin{aligned} \int_{M_1} \int_{M_2} f(z_1, z_2) \mu_2(dz_2) \mu_1(dz_1) &= \int_{M_2} \int_{M_1} f(z_1, z_2) \mu_1(dz_1) \mu_2(dz_2) \\ &= \int_{M_1 \times M_2} f(z_1, z_2) \mu_1 \times \mu_2(dz_1 \times dz_2) \end{aligned}$$

Infinite product spaces

The extension to $(M_1 \times \cdots \times M_m, \mathcal{M}_1 \times \cdots \times \mathcal{M}_m, \mu_1 \times \cdots \times \mu_m)$ is immediate. The extension to infinite product spaces is needed to capture the idea of an infinite sequence of repeated trials of an experiment. Let $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, 2, \ldots$ be probability spaces (possibly all copies of the same probability space). Let

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots = \{(\omega_1, \omega_2, \ldots) : \omega_i \in \Omega_i, i = 1, 2, \ldots\}$$

and

$$\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots = \sigma(A_1 \times \cdots \times A_m \times \Omega_{m+1} \times \Omega_{m+2} \times \cdots : A_i \in \mathcal{F}_i, m = 1, 2, \ldots)$$

We want a probability measure satisfying

$$P(A_1 \times \cdots \times A_m \times \Omega_{m+1} \times \Omega_{m+2} \times \cdots) = \prod_{i=1}^{m} P_i(A_i). \tag{6.4}$$

Construction of P

For $A \in \mathcal{F}$, let

$$Q_n(A, \omega_{n+1}, \omega_{n+2}, \ldots) = \int \mathbf{1}_A(\omega_1, \omega_2, \ldots) P_1(d\omega_1) \cdots P_n(d\omega_n).$$

The necessary measurability follows as above. Let $\mathcal{F}^m = \sigma(A_1 \times \cdots \times A_m \times \Omega_{m+1} \times \Omega_{m+2} \times \cdots : A_i \in \mathcal{F}_i, i = 1, \ldots, m)$. Then clearly

$$P(A) \equiv \lim_{n \to \infty} Q_n(A, \omega_{n+1}, \omega_{n+2}, \ldots)$$

exists for each $A \in \cup_m \mathcal{F}^m$. It follows, as in Problems 5 and 6 that P is a countably additive set function on $\cup_m \mathcal{F}^m$.

Caratheodary extension theorem

Theorem 6.6 *Let M be a set, and let $\mathcal{A}$ be an algebra of subsets of M. If μ is a σ-finite measure (countably additive set function) on $\mathcal{A}$, then there exists a unique extension of μ to a measure on $\sigma(\mathcal{A})$.*

Consequently, P defined above extends to a measure on $\vee_m \mathcal{F}^m$. The uniqueness of P satisfying (6.4) follows by the Dynkin class theorem.

Expectation of the product of component (independent) random variables

Lemma 6.7 *Suppose* X_k *is integrable and* $X_k(\omega) = Y_k(\omega_k)$ *for* $k = 1, 2, \ldots$. *Then*

$$E[\prod_{k=1}^{m} X_k] = \prod_{k=1}^{m} \int_{\Omega_k} Y_k dP_k = \prod_{k=1}^{m} E[X_k]$$

Relative frequency

For a probability space $(\Omega, \mathcal{F}, P)$, let $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$ denote the infinite product space with each factor given by $(\Omega, \mathcal{F}, P)$. Fix $A \in \mathcal{F}$, and let

$$S_n(\omega_1, \omega_2, \ldots) = \sum_{k=1}^{n} \mathbf{1}_A(\omega_k).$$

Then S_n is binomially distributed with parameters n and $p = P(A)$.

$$E[S_n] = nP(A)$$

and

$$P^\infty\{|n^{-1}S_n - P(A)| \geq \epsilon\} \leq \frac{E[(S_n - nP(A))^2]}{n^2\epsilon^2} = \frac{P(A)(1 - P(A))}{n\epsilon^2},$$

Almost sure convergence of relative frequency

Letting $X_k(\omega) = \mathbf{1}_A(\omega_k)$, by the Markov inequality

$$P^\infty\{|n^{-1}S_n - P(A)| \geq \epsilon\} \leq \frac{E[(S_n - nP(A))^4]}{n^4\epsilon^4}$$

$$= \frac{E[(X_k - P(A))^4]}{n^3\epsilon^4} + \frac{3(n-1)E[(X_k - P(A))^2]^2}{n^3\epsilon^4}$$

and

$$\sum_n P^\infty\{|n^{-1}S_n - P(A)| \geq \epsilon\} < \infty.$$

Therefore

$$P^\infty\{\limsup_{n\to\infty} |n^{-1}S_n - P(A)| > \epsilon\} \leq \sum_{n=m}^{\infty} P^\infty\{|n^{-1}S_n - P(A)| \geq \epsilon\} \to 0$$

7. Independence

- Independence of σ-algebras and random variables
- Independence of generated σ-algebras
- Bernoulli sequences and the law of large numbers
- Tail events and the Kolmogorov zero-one law

Independence

Definition 7.1 *σ-algebras $\mathcal{D}_i \subset \mathcal{F}$, $i = 1, \ldots, m$, are* independent *if and only if*

$$P(D_1 \cap \cdots \cap D_m) = \prod_{i=1}^{m} P(D_i), \quad D_i \in \mathcal{D}_i$$

Random variable $X_1, \ldots, X_m$ are independent *if and only if $\sigma(X_1), \ldots, \sigma(X_m)$ are independent.*

An infinite collection of σ-algebras/random variables is independent if every finite subcollection is independent.

Lemma 7.2 *If $\mathcal{D}_1, \ldots, \mathcal{D}_m$ are independent σ-algebras, X_k is $\mathcal{D}_k$-measurable, $k = 1, \ldots, m$, then $X_1, \ldots, X_m$ are independent.*

Independence of generated σ-algebras

For a collection of σ-algebras $\{\mathcal{G}_\alpha, \alpha \in \mathcal{A}\}$, let $\vee_{\alpha\in\mathcal{A}}\mathcal{G}_\alpha$ denote the smallest σ-algebra containing $\cup_{\alpha\in\mathcal{A}}\mathcal{G}_\alpha$.

Lemma 7.3 *Suppose $\{\mathcal{D}_\alpha, \alpha \in \mathcal{A}\}$ are independent. Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ and $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$. Then $\vee_{\alpha\in\mathcal{A}_1}\mathcal{D}_\alpha$ and $\vee_{\alpha\in\mathcal{A}_2}\mathcal{D}_\alpha$ are independent. (But see Problem 7.)*

Proof. Let

$$\mathcal{S}_i = \{A_1 \cap \cdots \cap A_m : A_k \in \mathcal{D}_{\alpha_k}, \alpha_1, \ldots, \alpha_m \in \mathcal{A}_i, \alpha_k \neq \alpha_l \text{ for } k \neq m\}$$

Let $A \in \mathcal{S}_1$, and let $\mathcal{G}_2^A$ be the collection of $B \in \mathcal{F}$ such that $P(A \cap B) = P(A)P(B)$. Then $\mathcal{G}_2^A$ is a Dynkin class containing $\mathcal{S}_2$ and hence containing $\vee_{\alpha\in\mathcal{A}_2}\mathcal{D}_\alpha$. Similarly, let $B \in \vee_{\alpha\in\mathcal{A}_2}\mathcal{D}_\alpha$, and let $\mathcal{G}_1^B$ bet the collection of $A \in \mathcal{F}$ such that $P(A \cap B) = P(A)P(B)$. Again, $\mathcal{G}_1^B$ is a Dynkin class containing $\mathcal{S}_1$ and hence $\vee_{\alpha\in\mathcal{A}_1}\mathcal{D}_\alpha$. $\square$

Consequences of independence

Lemma 7.4 *If* $X_1, \ldots, X_m$ *are independent,* $g_1 : \mathbb{R}^k \rightarrow \mathbb{R}$ *and* $g_2 : \mathbb{R}^{m-k} \rightarrow \mathbb{R}$ *are Borel measurable,* $Y_1 = g_1(X_1, \ldots, X_k)$*, and* $Y_2 = g_2(X_{k+1}, \ldots, X_m)$*, then* Y_1 *and* Y_2 *are independent.*

Lemma 7.5 *If* $X_1, \ldots, X_m$ *are independent and integrable, then* $E[\prod_{k=1}^m X_k] = \prod_{k=1}^m E[X_k]$.

Proof. Check first for simple random variables and then approximate. □

Bernoulli trials

Definition 7.6 *A sequence of random variables* $\{X_i\}$ *is* Bernoulli *if the random variables are independent and* $P\{X_i = 1\} = 1 - P\{X_i = 0\} = p$ *for some* $0 \leq p \leq 1$.

If $\{X_i\}$ is Bernoulli, then

$$S_n = \sum_{i=1}^{n} X_i$$

is binomially distributed and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = p \quad a.s.$$

Law of large numbers for bounded random variables

Theorem 7.7 *Let* $\{Y_i\}$ *be independent and identically distributed random variables with* $P\{|Y| \leq c\} = 1$ *for some* $0 < c < \infty$. *Then* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = E[Y]$ *a.s.*

Proof.(See the "law of averages.") For each m,

$$\begin{aligned}
\sum_{l} \frac{l}{m}\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\frac{l}{m}<Y_i\leq\frac{l+1}{m}\}} &\leq \frac{1}{n}\sum_{i=1}^{n} Y_i \\
&\leq \sum_{l} \frac{l+1}{m}\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\frac{l}{m}<Y_i\leq\frac{l+1}{m}\}} \\
&\to \sum_{l} \frac{l+1}{m} P\{\frac{l}{m} < Y \leq \frac{l+1}{m}\}
\end{aligned}$$

□

Tail events and the Kolmogorov zero-one law

Lemma 7.8 *Let $\mathcal{D}_1, \mathcal{D}_2, \ldots$ be independent σ-algebras, and define*

$$\mathcal{T} = \cap_m \vee_{n \geq m} \mathcal{D}_n,$$

where $\vee_{n\geq m}\mathcal{D}_n$ denotes the smallest σ-algebra containing $\cup_{n\geq m}\mathcal{D}_n$. If $A \in \mathcal{T}$, then $P(A) = 0$ or $P(A) = 1$.

Proof. Note that for $m > k$, $\vee_{n\geq m}\mathcal{D}_n$ is independent of $\vee_{l\leq k}\mathcal{D}_l$. Consequently, for all k, $\vee_{l\leq k}\mathcal{D}_l$ is independent of $\mathcal{T}$ which implies

$$P(A \cap B) = P(A)P(B) \quad A \in \cup_{k=1}^{\infty} \vee_{l\leq k} \mathcal{D}_l, B \in \mathcal{T}.$$

But the collection of A satisfying this identity is a monotone class containing $\cup_{k=1}^{\infty} \vee_{l\leq k} \mathcal{D}_l$ and hence contains $\sigma(\cup_{k=1}^{\infty} \vee_{l\leq k} \mathcal{D}_l) \supset \mathcal{T}$. Therefore $P(B) = P(B \cap B) = P(B)^2$, which imples $P(B) = 0$ or 1. □

Borel-Cantelli lemma

Let $A_1, A_2, \ldots$ be events and define

$B = \cap_m \cup_{n \geq m} A_n = \{\omega : \omega \in A_n \text{ for infinitely many } n\} \equiv \{A_n \text{ occurs i.o.}\}$

Note that

$$P(B) \leq \sum_{n=m}^{\infty} P(A_n). \tag{7.5}$$

Lemma 7.9 *If* $\sum_{n=1}^{\infty} P(A_n) < \infty$, *then* $P\{A_n \text{ occurs i.o.}\} = 0$.

If $A_1, A_2, \ldots$ *are independent and* $\sum_{n=1}^{\infty} P(A_n) = \infty$, *then* $P\{A_n \text{ occurs i.o.}\} = 1$.

See Problem 9.

Proof.

The first part follows by (7.5).

Noting that $B^c = \cup_m \cap_{n\geq m} A_n^c$,

$$P(B^c) \leq \sum_m P(\cap_{n\geq m} A_n^c) = \sum_m \prod_{n\geq m} P(A_n^c) \leq \sum_m e^{-\sum_{n=m}^{\infty} P(A_n)} = 0.$$

□

8. L^p spaces

- Metric spaces
- A metric on the space of random variables
- Normed linear spaces
- L^p spaces
- Projections in L^2

Metric spaces

$d : S \times S \to [0, \infty)$ is a *metric* on S if and only if

- $d(x, y) = d(y, x)$, $x, y \in S$
- $d(x, y) = 0$ if and only if $x = y$
- $d(x, y) \leq d(x, z) + d(z, y)$, $x, y, z \in S$, (triangle inequality)

If d is a metric then $d \wedge 1$ is a metric.

Examples

- $\mathbb{R}^m$ $\quad d(x, y) = |x - y|$
- $C[0, 1]$ $\quad d(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|$

Convergence

Let (S, d) be a metric space.

Definition 8.1 *A sequence* $\{x_n\} \subset S$ converges *if there exists* $x \in S$ *such that* $\lim_{n\to\infty} d(x_n, x) = 0$. x *is called the* limit *of* $\{x_n\}$ *and we write* $\lim_{n\to\infty} x_n = x$.

Lemma 8.2 *A sequence* $\{x_n\} \subset S$ *has at most one limit.*

Proof. Suppose $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} x_n = y$. Then

$$d(x, y) \leq d(x, x_n) + d(y, x_n) \to 0.$$

Consequently, $d(x, y) = 0$ and $x = y$. □

Open and closed sets

Let (S, d) be a metric space. For $\epsilon > 0$ and $x \in S$, let $B_\epsilon(x) = \{y \in S : d(x, y) < \epsilon\}$. $B_\epsilon(x)$ is called the *open ball of radius ϵ centered at x.*

Definition 8.3 *A set $G \subset S$ is* open *if and only if for each $x \in G$, there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset G$. A set $F \subset S$ is* closed *if and only if F^c is open. The* closure *$\bar{H}$ of a set $H \subset S$ is the smallest closed set containing H.*

Lemma 8.4 *A set $F \subset S$ is closed if and only if for each convergent $\{x_n\} \subset F$, $\lim_{n\to\infty} x_n \in F$.*

If $H \subset S$, then $\bar{H} = \{x : \exists\{x_n\} \subset H, \lim_{n\to\infty} x_n = x\}$.

Completeness

Definition 8.5 *A sequence* $\{x_n\} \subset S$ *is* Cauchy *if and only if*

$$\lim_{n,m\to\infty} d(x_n, x_m) = 0,$$

that is, for each $\epsilon > 0$*, there exists* n_ϵ *such that*

$$\sup_{n,m\geq n_\epsilon} d(x_n, x_m) \leq \epsilon.$$

Definition 8.6 *A metric space* (S, d) *is* complete *if and only if every Cauchy sequence has a limit.*

Recall that the space of real numbers can be defined to be the completion of the rational numbers under the usual metric.

Completeness is a metric property

Two metrics generate the same *topology* if the collection of open sets is the same for both metrics. In particular, the collection of convergent sequences is the same.

Completeness depends on the metric, not the topology: For example

$$r(x, y) = |\frac{x}{1 + |x|} - \frac{y}{1 + |y|}|$$

is a metric giving the usual topology on the real line, but $\mathbb{R}$ is not complete under this metric.

Separability

Definition 8.7 *A set $D \subset S$ is* dense *in S if for every $x \in S$, there exists $\{x_n\} \subset D$ such that* $\lim_{n\to\infty} x_n = x$.

S is separable *if there is a countable set $D \subset S$ that is dense in S.*

Lemma 8.8 *S is separable if and only if there is $\{x_n\} \subset S$ such that for each $\epsilon > 0$, $S \subset \cup_{n=1}^{\infty} B_\epsilon(x_n)$.*

Note that $\mathbb{Q}$ is dense in $\mathbb{R}$, so $\mathbb{R}$ is separable.

Continuity of a metric

Lemma 8.9 *If* $\lim_{n\to\infty} x_n = x$ *and* $\lim_{n\to\infty} y_n = y$*, then* $\lim_{n\to\infty} d(x_n, y_n) = d(x, y)$.

Proof. By the triangle inequality

$$d(x,y) \leq d(x,x_n) + d(x_n,y_n) + d(y,y_n)$$

and

$$d(x_n,y_n) \leq d(x,x_n) + d(x,y) + d(y,y_n)$$

□

Equivalence relations

Definition 8.10 *Let S be a set, and $E \subset S \times S$. If $(a, b) \in E$, write $a \sim b$. E is an* equivalence relation *on S if*

- *Reflexivity: $a \sim a$*
- *Symmetry: If $a \sim b$ then $b \sim a$*
- *Transitivity: If $a \sim b$ and $b \sim c$ then $a \sim c$.*

$G \subset S$ is an equivalence class *if $a, b \in G$ implies $a \sim b$ and $a \in G$ and $b \sim a$ implies $b \in G$.*

Lemma 8.11 *If G_1 and G_2 are equivalence classes, then either $G_1 = G_2$ or $G_1 \cap G_2 = \emptyset$.*

Each $a \in S$ is in some equivalence class.

Equivalence classes of random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let S be the collection of random variables on $(\Omega, \mathcal{F}, P)$. Then $X \sim Y$ if and only if $X = Y$ a.s. defines an equivalence relation on S. L^0 will denote the collection of equivalence classes of random variables.

In practice, we will write X for the random variable and for the equivalence class of all random variables equivalent to X. For example, we will talk about X being *the* almost sure limit of $\{X_n\}$ even though any other random variable satisfying $Y = X$ a.s. would also be the almost sure limit of $\{X_n\}$.

Convergence in probability

Definition 8.12 *A sequence of random variables* $\{X_n\}$ converges in probability *to* X *(*$X_n \xrightarrow{P} X$*) if and only if for each* $\epsilon > 0$*,*

$$\lim_{n\to\infty} P\{|X - X_n| > \epsilon\} = 0.$$

Define

$$d(X,Y) = E[1 \wedge |X - Y|].$$

Lemma 8.13 *d is a metric on L^0 and* $\lim_{n\to\infty} d(X_n, X) = 0$ *if and only if* $X_n \xrightarrow{P} X$*.*

Proof. Note that $1 \wedge |x-y|$ defines a metric on $\mathbb{R}$, so

$$E[1 \wedge |X-Y|] \leq E[1 \wedge |X-Z|] + E[1 \wedge |Z-Y|].$$

Since for $0 < \epsilon \leq 1$,

$$P\{|X-X_n| > \epsilon\} \leq \frac{d(X,X_n)}{\epsilon},$$

$\lim_{n\to\infty} d(X_n, X) = 0$ implies if $X_n \xrightarrow{P} X$. Observing that

$$1 \wedge |X-X_n| \leq \mathbf{1}_{\{|X-X_n|>\epsilon\}} + \epsilon,$$

$X_n \xrightarrow{P} X$ implies

$$\limsup E[1 \wedge |X-X_n|] \leq \epsilon.$$

□

Linear spaces

Definition 8.14 *A set L is a (real)* linear space *if there is a notion of* addition $+ : L \times L \to L$ *and* scalar multiplication $\cdot : \mathbb{R} \times L \to L$ *satisfying*

- *For all $u, v, w \in L$, $u + (v + w) = (u + v) + w$*
- *For all $v, w \in L$, $v + w = w + v$.*
- *There exists an element $0 \in L$ such that $v + 0 = v$ for all $v \in L$.*
- *For all $v \in L$, there exists $w \in L$ $(-v)$ such that $v + w = 0$.*
- *For all $a \in \mathbb{R}$ and $v, w \in L$, $a(v + w) = av + aw$.*
- *For all $a, b \in \mathbb{R}$ and $v \in L$, $(a + b)v = av + bv$.*
- *For all $a, b \in \mathbb{R}$ and $v \in \mathbb{R}$, $a(bv) = (ab)v$.*
- *For all $v \in L$, $1v = v$.*

Norms

Definition 8.15 *If L is a linear space, then $\| \cdot \| : L \to [0, \infty)$ defines a* norm *on L if*

- $\|u\| = 0$ *if and only if* $u = 0$.
- *For $a \in \mathbb{R}$ and $u \in L$,* $\|au\| = |a|\|u\|$.
- *For $u, v \in L$,* $\|u + v\| \leq \|u\| + \|v\|$.

Note that $d(u, v) = \|u - v\|$ defines a metric on L.

L^p spaces

Fix $(\Omega, \mathcal{F}, P)$. For $p \geq 1$, let L^p be the collection of (equivalence classes of) random variables satisfying $E[|X|^p] < \infty$.

Lemma 8.16 *L^p is a linear space.*

Proof. Note that $|a+b|^p \leq 2^{p-1}(|a|^p + |b|^p)$. □

A geometric inequality

Let $p, q \geq 1$ satisfy $p^{-1} + q^{-1} = 1$, and note that $\frac{1}{p-1} = \frac{p}{p-1} - 1 = q - 1$. If $f(x) = x^{p-1}$, then $f^{-1}(y) = y^{\frac{1}{p-1}} = y^{q-1}$, and hence for $a, b \geq 0$,

$$ab \leq \int_0^a x^{p-1}dx + \int_0^b y^{q-1}dy = \frac{1}{p}a^p + \frac{1}{q}b^q.$$

Consequently, if $E[|X|^p] = E[|Y|^q] = 1$, then

$$E[|XY|] \leq 1.$$

Hölder inequality

Lemma 8.17 *Let* $p > 1$ *and* $\frac{1}{p} + \frac{1}{q} = 1$. *Suppose* $X \in L^p$ *and* $Y \in L^q$, *then*

$$E[XY] \leq E[|XY|] \leq E[|X|^p]^{1/p} E[|Y|^q]^{1/q}$$

Proof. Define

$$\tilde{X} = \frac{X}{E[|X|^p]^{1/p}} \qquad \tilde{Y} = \frac{Y}{E[|Y|^q]^{1/q}}.$$

Then

$$E[|\tilde{X}\tilde{Y}|] \leq 1$$

and the lemma follows. □

Minkowski inequality

Lemma 8.18 *Suppose* $X, Y \in L^p$, $p \geq 1$. *Then*

$$E[|X+Y|^p]^{1/p} \leq E[|X|^p]^{1/p} + E[|Y|^p]^{1/p}$$

Proof. Since $|X+Y| \leq |X| + |Y|$ and $\frac{p-1}{p} + \frac{1}{p} = 1$,

$$\begin{aligned} E[|X+Y|^p] &\leq E[|X||X+Y|^{p-1}] + E[|Y||X+Y|^{p-1}] \\ &\leq E[|X|^p]^{1/p} E[|X+Y|^p]^{\frac{p-1}{p}} + E[|X|^p]^{1/p} E[|X+Y|^p]^{\frac{p-1}{p}}. \end{aligned}$$

□

The L^p norm

For $p \geq 1$, define

$$\|X\|_p = E[|X|^p]^{1/p}.$$

Then, by the Minkowski inequality, $\|X\|_p$ is a norm on L^p. Note that the Hölder inequality becomes

$$E[XY] \leq \|X\|_p \|Y\|_q.$$

The L^∞ norm

Define

$$\|X\|_\infty = \inf\{c > 0 : P\{|X| \geq c\} = 0\},$$

and let L^∞ be the collection of (equivalence classes) of random variables X such that $\|X\|_\infty < \infty$.

$\|\cdot\|_\infty$ is a norm and

$$E[XY] \leq \|X\|_\infty \|Y\|_1.$$

Completeness

Lemma 8.19 *Suppose* $\{X_n\}$ *is a Cauchy sequence in* L^p. *Then there exists* $X \in L^p$ *such that* $\lim_{n\to\infty} \|X - X_n\|_p = 0$.

Proof. Select n_k such that $n, m \geq n_k$ implies $\|X_n - X_m\|_p \leq 2^{-k}$. Assume that $n_{k+1} \geq n_k$. Then

$$\sum_k E[|X_{n_{k+1}} - X_{n_k}|] \leq \sum_k \|X_{n_{k+1}} - X_{n_k}\|_p < \infty,$$

so the series $\sum_k (X_{n_{k+1}} - X_{n_k})$ converges a.s. to a random variable X. Fatou's lemma implies $X \in L^p$ and that

$$E[|X - X_{n_l}|^p] \leq \lim_{k\to\infty} E[|X_{n_k} - X_{n_l}|^p] \leq 2^{-pl}.$$

Therefore

$$\lim_{n\to\infty} \|X - X_n\|_p \leq \limsup_{n\to\infty} (\|X - X_{n_k}\|_p + \|X_{n_k} - X_n\|_p) \leq 2^{-k}.$$

□

Best L^2 approximation

Lemma 8.20 *Let M be a closed linear subspace of L^2, and let $X \in L^2$. Then there exists a unique $Y \in M$ such that $E[(X-Y)^2] = \inf_{Z \in M} E[(X-Z)^2]$.*

Proof. Let $\rho = \inf_{Z \in M} E[(X-Z)^2]$, and let $Y_n \in M$ satisfy $\lim_{n\to\infty} E[(X-Y_n)^2] = \rho$. Then noting that

$$E[(Y_n - Y_m)^2] = E[(X-Y_n)^2] + E[(X-Y_m)^2] - 2E[(X-Y_n)(X-Y_m)]$$

we have

$$\begin{aligned} 4\rho &\leq E[(2X - (Y_n + Y_m))^2] \\ &= E[(X-Y_n)^2] + E[(X-Y_m)^2] + 2E[(X-Y_n)(X-Y_m)] \\ &= 2E[(X-Y_n)^2] + 2E[(X-Y_m)^2] - E[(Y_n - Y_m)^2], \end{aligned}$$

and it follows that $\{Y_n\}$ is Cauchy in L^2. By completeness, there exists Y such that $Y = \lim_{n\to\infty} Y_n$, and $\rho = E[(X-Y)^2]$.

Note that uniqueness also follows from the inequality. □

Orthogonality

The case $p = 2$ (which implies $q = 2$) has special properties. The first being the idea of *orthogonality.*

Definition 8.21 *Let $X, Y \in L^2$. Then X and Y are* orthogonal *($X \perp Y$) if and only if $E[XY] = 0$.*

Lemma 8.22 *Let M be a closed linear subspace of L^2, and let $X \in L^2$. Then the best approximation constructed in Lemma 8.20 is the unique $Y \in M$ such that $(X - Y) \perp Z$ for every $Z \in M$.*

Proof. Suppose $Z \in M$. Then

$$\begin{aligned} E[(X-Y)^2] &\leq E[(X-(Y+aZ))^2] \\ &= E[(X-Y)^2] - 2aE[Z(X-Y)] + a^2E[Z^2]. \end{aligned}$$

Since a may be either positive or negative, we must have

$$E[Z(X-Y)] = 0.$$

Uniqueness follows from the fact that $E[Z(X-Y_1)] = 0$ and $E[Z(X-Y_2)] = 0$ for all $Z \in M$ implies

$$E[(Y_1-Y_2)^2] = E[(Y_1-Y_2)(X-Y_2)] - E[(Y_1-Y_2)(X-Y_1)] = 0.$$

□

Projections in L^2

Lemma 8.23 *Let M be a closed linear subspace of L^2, and for $X \in L^2$, denote the Y from Lemma 8.20 by $P_M X$. Then P_M is a linear operator on L^2, that is,*

$$P_M(a_1 X_1 + a_2 X_2) = a_1 P_M X_1 + a_2 P_M X_2.$$

Proof. Since

$$E[Z(a_1 X_1 + a_2 X_2 - (a_1 P_M X_1 + a_2 P_M X_2)]$$
$$= a_1 E[Z(X_1 - P_M X_1)] + a_2 E[Z(X_2 - P_M X_2)]$$

the conclusion follows by the uniqueness in Lemma 8.22. □

Best linear approximation

Let $Y \in L^2$ and $M = \{aY + b : a, b \in \mathbb{R}\}$, and let $X \in L^2$. Then

$$P_M X = a_X Y + b_X$$

where

$$b_X = E[X - a_X Y] \qquad a_X = \frac{E[XY] - E[X]E[Y]}{E[Y^2] - E[Y]^2} = \frac{Cov(X,Y)}{Var(Y)}$$

Compute

$$\inf_{a,b} E[(X - (aY + b))^2]$$

9. Conditional expectations

- Definition
- Relation to elementary definitions
- Properties of conditional expectation
- Jensen's inequality
- Definition of $E[X|\mathcal{D}]$ for arbitrary nonnegative random variables
- Convergence theorems
- Conditional probability
- Regular conditional probabilities/distributions

Best approximation using available information

Let $\mathcal{D} \subset \mathcal{F}$ be a sub-σ-algebra representing the available information. Let X be a random variable, not necessarily $\mathcal{D}$-measurable. We want to approximate X using the available information.

Definition of conditional expectation

Let $\mathcal{D} \subset \mathcal{F}$ be a sub-σ-algebra, and let $L^2(\mathcal{D})$ be the linear space of $\mathcal{D}$-measurable random variables in L^2. Define

$$E[X|\mathcal{D}] = P_{L^2(\mathcal{D})}X.$$

Then by orthogonality (Lemma 8.22),

$$E[X\mathbf{1}_D] = E[E[X|\mathcal{D}]\mathbf{1}_D], \quad D \in \mathcal{D}.$$

We extend the defintion to L^1.

Definition 9.1 *Let $X \in L^1$. Then $E[X|\mathcal{D}]$ is the unique $\mathcal{D}$-measurable random variable satisfying*

$$E[X\mathbf{1}_D] = E[E[X|\mathcal{D}]\mathbf{1}_D], \quad D \in \mathcal{D}.$$

Monotonicity

Lemma 9.2 *Let* $X_1, X_2 \in L^1$*,* $X_1 \geq X_2$ *a.s., and suppose* $Y_1 = E[X_1|\mathcal{D}]$ *and* $Y_2 = E[X_2|\mathcal{D}]$*. Then* $Y_1 \geq Y_2$ *a.s.*

Proof. Let $D = \{Y_2 > Y_1\}$. Then

$$0 \leq E[(X_1 - X_2)\mathbf{1}_D] = E[(Y_1 - Y_2)\mathbf{1}_D] \leq 0.$$

□

Existence for L^1

Lemma 9.3 *Let* $X \in L^1$, $X \geq 0$. *Then*

$$E[X|\mathcal{D}] = \lim_{c\to\infty} E[X \wedge c|\mathcal{D}] \tag{9.1}$$

Proof. Note that the right side of (9.1) (call it Y) is $\mathcal{D}$-measurable and for $D \in \mathcal{D}$,

$$E[X\mathbf{1}_D] = \lim_{c\to\infty} E[(X \wedge c)\mathbf{1}_D] = \lim_{c\to\infty} E[E[X \wedge c|\mathcal{D}]\mathbf{1}_D] = E[Y\mathbf{1}_D],$$

where the first and last equalities hold by the monotone convergence theorem and the middle equality holds by definition. □

Verifying that a random variable is a conditional expectation

To show that $Y = E[X|\mathcal{D}]$, one must verify

1. Y is $\mathcal{D}$-measurable
2. $$E[Y\mathbf{1}_D] = E[X\mathbf{1}_D], \quad D \in \mathcal{D}. \tag{9.2}$$

Assuming that $X, Y \in L^1$, if $\mathcal{S} \subset \mathcal{D}$ is closed under intersections, $\Omega \in \mathcal{S}$, and $\sigma(\mathcal{S}) = \mathcal{D}$, then to verify (9.2), it is enough to show that $E[Y\mathbf{1}_D] = E[X\mathbf{1}_D]$ for $D \in \mathcal{S}$.

Relation to elementary definitions

Suppose that $\{D_k\} \subset \mathcal{F}$ is a partition of Ω and $\mathcal{D} = \sigma\{D_k\}$. Then

$$E[X|\mathcal{D}] = \sum_k \frac{E[X\mathbf{1}_{D_k}]}{P(D_k)}\mathbf{1}_{D_k}.$$

Suppose the (X, Y) have a joint density $f_{XY}(x, y)$. Define

$$g(y) = \frac{\int_{-\infty}^{\infty} x f_{XY}(x, y)dx}{f_Y(y)}.$$

Then

$$E[X|Y] \equiv E[X|\sigma(Y)] = g(Y).$$

Properties of conditional expectation

- Linearity: Assume that $X, Y \in L^1$.

$$E[aX + bY|\mathcal{D}] = aE[X|\mathcal{D}] + bE[Y|\mathcal{D}]$$

- Monotonicity/positivity: If $X, Y \in L^1$ and $X \geq Y$ a.s., then

$$E[X|\mathcal{D}] \geq E[Y|\mathcal{D}]$$

- Iteration: If $\mathcal{D}_1 \subset \mathcal{D}_2$ and $X \in L^1$, then

$$E[X|\mathcal{D}_1] = E[E[X|\mathcal{D}_2]|\mathcal{D}_1] \tag{9.3}$$

- Factoring: If $X, XY \in L^1$ and Y is $\mathcal{D}$-measurable, then

$$E[XY|\mathcal{D}] = YE[X|\mathcal{D}].$$

  In particular, if Y is $\mathcal{D}$-measurable, $E[Y|\mathcal{D}] = Y$.

- Independence: If $\mathcal{H}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$, then

$$E[X|\sigma(\mathcal{G}, \mathcal{H})] = E[X|\mathcal{G}]$$

$G \in \mathcal{G}$, $H \in \mathcal{H}$,

$$\begin{aligned} E[E[X|\mathcal{G}]\mathbf{1}_{G\cap H}] &= E[E[X|\mathcal{G}]\mathbf{1}_G\mathbf{1}_H] \\ &= E[E[X|\mathcal{G}]\mathbf{1}_G]E[\mathbf{1}_H] \\ &= E[X\mathbf{1}_G]E[\mathbf{1}_H] \\ &= E[X\mathbf{1}_G\mathbf{1}_H] \end{aligned}$$

In particular, if X is independent of $\mathcal{H}$,

$$E[X|\mathcal{H}] = E[X].$$

Jensen's inequality

Lemma 9.4 *If* $\varphi : \mathbb{R} \to \mathbb{R}$ *is convex and* $X, \varphi(X) \in L^1$*, then*

$$E[\varphi(X)|\mathcal{D}] \geq \varphi(E[X|\mathcal{D}]).$$

Proof. Let

$$\varphi^+(x) = \lim_{h \to 0+} \frac{\varphi(x+h) - \varphi(x)}{h}.$$

Then

$$\varphi(x) - \varphi(y) \geq \varphi^+(y)(x-y).$$

Consequently,

$$\begin{aligned} E[\varphi(X) - \varphi(E[X|\mathcal{D}])|\mathcal{D}] &\geq E[\varphi^+(E[X|\mathcal{D}])(X - E[X|\mathcal{D}])|\mathcal{D}] \\ &= \varphi^+(E[X|\mathcal{D}])E[X - E[X|\mathcal{D}]|\mathcal{D}] \\ &= 0 \end{aligned}$$

□

Definition of $E[X|\mathcal{D}]$ for arbitrary nonnegative random variables

Let $X \geq 0$ a.s. Then

$$Y \equiv \lim_{c\to\infty} E[X \wedge c|\mathcal{D}]$$

is $\mathcal{D}$-measurable and satisfied

$$E[Y\mathbf{1}_D] = E[X\mathbf{1}_D]$$

(allowing $\infty = \infty$). Consequently, we can extend the definition of conditional expectation to all nonnegative random variables.

Monotone convergence theorem

Lemma 9.5 *Let* $0 \leq X_1 \leq X_2 \leq \cdots$. *Then*

$$\lim_{n\to\infty} E[X_n|\mathcal{D}] = E[\lim_{n\to\infty} X_n|\mathcal{D}] \quad a.s.$$

Proof. Let $Y = \lim_{n\to\infty} E[X_n|\mathcal{D}]$. Then Y is $\mathcal{D}$-measurable and for $D \in \mathcal{D}$,

$$E[Y\mathbf{1}_D] = \lim_{n\to\infty} E[E[X_n|\mathcal{D}]\mathbf{1}_D] = \lim_{n\to\infty} E[X_n\mathbf{1}_D] = E[(\lim_{n\to\infty} X_n)\mathbf{1}_D].$$

□

Functions of independent random variables

Lemma 9.6 *Suppose X is independent of $\mathcal{D}$ and Y is $\mathcal{D}$-measurable, If $\varphi : \mathbb{R}^2 \to [0,\infty)$ and $g(y) = \int_{\mathbb{R}} \varphi(x,y)\mu_X(dx)$ then*

$$E[\varphi(X,Y)|\mathcal{D}] = g(Y), \tag{9.4}$$

and hence, (9.4) holds for all φ such that $E[|\varphi(X,Y)|] < \infty$.

Proof. Let $A, B \in \mathcal{B}(\mathbb{R})$. Then setting $g(y) = \mu_X(A)\mathbf{1}_B(y)$,

$$E[\mathbf{1}_A(X)\mathbf{1}_B(Y)|\mathcal{D}] = \mu_X(A)\mathbf{1}_B(Y).$$

Let $C \in \mathcal{B}(\mathbb{R}^2)$ and $g_C(y) = \int_{\mathbb{R}} \mathbf{1}_C(x,y)\mu_X(dx)$. The collection of C such that

$$E[\mathbf{1}_C(X,Y)|\mathcal{D}] = g_C(Y)$$

is a Dynkin class and consequently, contains all of $\mathcal{B}(\mathbb{R}^2)$. By linearity, (9.4) holds for all simple functions and extends to all nonnegative φ be the monotone convergence theorem. □

Functions of known and unknown random variables

Lemma 9.7 *Let X be a random variable and $\mathcal{D} \subset \mathcal{F}$ a sub-σ-algebra. Let Ξ be the collection of $\varphi : \mathbb{R}^2 \to [0, \infty]$ such that φ is Borel measurable and there exists $\mathcal{B}(\mathbb{R}) \times \mathcal{D}$-measurable $\psi : \mathbb{R} \times \Omega \to [0, \infty]$ satisfying*

$$E[\varphi(X, Y)|\mathcal{D}](\omega) = \psi(Y(\omega), \omega) \quad \text{(almost surely)}, \tag{9.5}$$

for all $\mathcal{D}$-measurable random variables, Y. Then Ξ is closed under positive linear combinations and under convergence of increasing sequences.

Since Ξ contains all functions of the form $\gamma_1(x)\gamma_2(y)$, for Borel measurable $\gamma_i : \mathbb{R} \to [0, \infty]$, Ξ is the collection of all Borel measurable $\varphi : \mathbb{R}^2 \to [0, \infty]$.

Proof. Linearity follows from the linearity of the conditional expectation.

Suppose $\{\varphi_n\} \subset \Xi$ and $\varphi_1 \leq \varphi_2 \leq \cdots$, and let $\varphi = \lim_{n\to\infty} \varphi_n$. Then by the monotonicity of the conditional expectation, the corresponding ψ_n must satisfy $\psi_n(Y(\omega),\omega) \geq \psi_{n-1}(Y(\omega),\omega)$ almost surely for each $\mathcal{D}$-measurable Y. Consequently, $\hat{\psi}_n = \psi_1 \vee \cdots \vee \psi_n$ must satisfy $\hat{\psi}_n(Y(\omega),\omega) = \psi_n(Y(\omega),\omega)$ almost surely for each $\mathcal{D}$-measurable Y, and hence

$$E[\varphi_n(X,Y)|\mathcal{D}](\omega) = \hat{\psi}_n(Y(\omega),\omega) \quad a.s.$$

for each $\mathcal{D}$-measurable Y. Defining $\psi(x,y) = \lim_{n\to\infty} \hat{\psi}_n(x,y)$, (9.5) holds. □

Fatou's lemma

Lemma 9.8 *Suppose* $X_n \geq 0$ *a.s. Then*

$$\liminf_{n\to\infty} E[X_n|\mathcal{D}] \geq E[\liminf_{n\to\infty} X_n|\mathcal{D}] \quad a.s.$$

Proof. Since

$$E[X_n|\mathcal{D}] \geq E[\inf_{m\geq n} X_m|\mathcal{D}],$$

the lemma follows by the monotone convergence theorem. □

Dominated convergence theorem

Lemma 9.9 *Suppose* $|X_n| \leq Y_n$, $\lim_{n\to\infty} X_n = X$ *a.s.*, $\lim_{n\to\infty} Y_n = Y$ *a.s. with* $E[Y] < \infty$, *and* $\lim_{n\to\infty} E[Y_n|\mathcal{D}] = E[Y|\mathcal{D}]$ *a.s. Then*

$$\lim_{n\to\infty} E[X_n|\mathcal{D}] = E[X|\mathcal{D}] \quad a.s.$$

Proof. The proof is the same as for expectations. For example

$$\begin{aligned} \liminf E[Y_n - X_n|\mathcal{D}] &\geq E[\liminf_{n\to\infty}(Y_n - X_n)|\mathcal{D}] \\ &= E[Y - X|\mathcal{D}], \end{aligned}$$

so

$$\limsup_{n\to\infty} E[X_n|\mathcal{D}] \leq E[X|\mathcal{D}].$$

□

Conditional probability

Conditional probability is simply defined as

$$P(A|\mathcal{D}) = E[\mathbf{1}_A|\mathcal{D}]$$

and if $\{A_k\}$ are disjoint, the monotone convergence theorem implies

$$P(\cup_{k=1}^{\infty} A_k|\mathcal{D}) = E[\mathbf{1}_{\cup A_k}|\mathcal{D}] = \sum_{k=1}^{\infty} E[\mathbf{1}_{A_k}|\mathcal{D}] = \sum_{k=1}^{\infty} P(A_k|\mathcal{D}).$$

BUT, we need to remember that conditional expectations are only unique in the equivalence class sense, so the above identy only asserts that $P(\cup_{k=1}^{\infty} A_k|\mathcal{D})$ and $\sum_{k=1}^{\infty} P(A_k|\mathcal{D})$ are equal almost surely. That does not guarantee tha $A \to P(A|\mathcal{D})(\omega)$ is a probability measure for any fixed $\omega \in \Omega$.

Random measures

Definition 9.10 *Let* $(S, \mathcal{S})$ *be a measurable space, and let* $\mathcal{M}(S)$ *be the space of* σ*-finite measures on* $(S, \mathcal{S})$*. Let* Ξ *be the smallest* σ*-algebra of subsets of* $\mathcal{M}(S)$ *containing sets of the form* $G_{A,c} = \{\mu \in \mathcal{M}(S) : \mu(A) \leq c\}$*,* $A \in \mathcal{S}$*,* $c \in \mathbb{R}$*. A* random measure ξ *on* $(S, \mathcal{S})$ *is a measurable mapping form* $(\Omega, \mathcal{F}, P)$ *to* $(\mathcal{M}(S), \Xi)$*.* ξ *is a* random probability measure *if* $\xi(S) \equiv 1$

Note that since $\{\xi(A) \leq c\} = \xi^{-1}(G_{A,c})$, $\xi(A)$ is a random variable for each $A \in \mathcal{S}$.

Regular conditional probabilities/distributions

Definition 9.11 *Let $\mathcal{D} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. A* regular conditional probability *given $\mathcal{D}$ is a random probability measure ξ on $(\Omega, \mathcal{F})$ such that $\xi(A) = P(A|\mathcal{D})$ for each $A \in \mathcal{F}$.*

Definition 9.12 *Let (S, d) be a complete, separable metric space, and let Z be an S-valued random variable. (Z is a measurable mapping from $(\Omega, \mathcal{F}, P)$ to $(S, \mathcal{B}(S))$.) Then ξ is a* regular conditional distribution *for Z given $\mathcal{D}$ if ξ is a random probability measure on $(S, \mathcal{B}(S))$ and $\xi(B) = P(Z \in B|\mathcal{D})$, $B \in \mathcal{B}(S)$.*

Existence of regular conditional distributions

Lemma 9.13 *Let (S, d) be a complete, separable metric space, let Z be an S-valued random variable, and let $\mathcal{D} \subset \mathcal{F}$ be a sub-σ-algebra. Then there exists a regular conditional distribution for Z given $\mathcal{D}$.*

Proof. If $S = \mathbb{R}$, the construction is simple. By monotonicity, and the countability of the rationals, we can construcut $F(x, \omega)$, $x \in \mathbb{Q}$, such that $x \in \mathbb{Q} \to F(x, \omega) \in [0, 1]$ is nondecreasing for each $\omega \in \Omega$ and $F(x) = P\{Z \leq x|\mathcal{D}\}$, $x \in \mathbb{Q}$. Then for $x \in \mathbb{R}$, define

$$\bar{F}(x, \omega) = \inf_{y>x.y\in\mathbb{Q}} F(y, \omega) = \lim_{y\in\mathbb{Q}\to x+} F(x, \omega).$$

Then the monotone convergence theorem implies

$$\bar{F}(x) = P\{Z \leq x|\mathcal{D}\}$$

and for each ω, $\bar{F}(\cdot, \omega)$ is a cumulative distribution function. Defining $\xi((-\infty, x], \omega) = \bar{F}(x, \omega)$, ξ extends to a random probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Finally, note that the collection of $B \in \mathcal{B}(\mathbb{R})$ such that $\xi(B) = P\{Z \in B|\mathcal{D}\}$ is a Dynkin class so contains $\mathcal{B}(\mathbb{R})$. □

10. Change of measure

- Absolute continuity and the Radon Nikodym theorem
- Applications of absolute continuity
- Bayes formula

Absolute continuity and the Radon-Nikodym theorem

Definition 10.1 *Let P and Q be probability measures on $(\Omega, \mathcal{F})$. Then P is* absolutely continuous *with respect to Q ($P << Q$) if and only if $Q(A) = 0$ implies $P(A) = 0$.*

Theorem 10.2 *If $P << Q$, then there exists a random variable $L \geq 0$ such that*

$$P(A) = E^Q[\mathbf{1}_A L] = \int_A L dQ, \quad A \in \mathcal{F}.$$

Consequently, Z is P-integrable if and only if ZL is Q-integrable, and

$$E^P[Z] = E^Q[ZL].$$

Standard notation: $\frac{dP}{dQ} = L$.

Maximum likelihood estimation

Suppose for each $\alpha \in \mathcal{A}$,

$$P_\alpha(\Gamma) = \int_\Gamma L_\alpha dQ$$

and

$$L_\alpha = H(\alpha, X_1, X_2, \ldots X_n)$$

for random variables $X_1, \ldots, X_n$. The maximum likelihood estimate $\hat{\alpha}$ for the "true" parameter $\alpha_0 \in \mathcal{A}$ based on observations of the random variables $X_1, \ldots, X_n$ is the value of α that maximizes

$$H(\alpha, X_1, X_2, \ldots X_n).$$

Sufficiency

If $dP_\alpha = L_\alpha dQ$ where

$$L_\alpha(X,Y) = H_\alpha(X)G(X,Y),$$

then X is a *sufficient statistic* for α. Without loss of generality, we can assume $E^Q[G(X,Y)] = 1$ and hence $d\hat{Q} = G(X,Y)dQ$ defines a probability measure.

Example 10.3 *If $(X_1, \ldots, X_n)$ are iid $N(\mu, \sigma^2)$ under $P_{(\mu,\sigma)}$ and $Q = P_{(0,1)}$, then*

$$L_{(\mu,\sigma)} = \frac{1}{\sigma^n} \exp\left\{ -\frac{1-\sigma^2}{2\sigma^2} \sum_{i=1}^n X_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n X_i - \frac{\mu^2}{\sigma^2} \right\}$$

so $(\sum_{i=1}^n X_i^2, \sum_i^n X_i)$ is a sufficient statistic for (μ, σ).

Parameter estimates and sufficiency

Theorem 10.4 *If $\hat{\theta}(X,Y)$ is an estimator of $\theta(\alpha)$ and φ is convex, then*

$$E^{P_\alpha}[\varphi(\theta(\alpha)-\hat{\theta}(X,Y))] \geq E^{P_\alpha}[\varphi(\theta(\alpha)-E^{\hat{Q}}[\hat{\theta}(X,Y)|X])]$$

Proof.

$$\begin{aligned}
E^{P_\alpha}[\varphi(\theta(\alpha)-\hat{\theta}(X,Y))] &= E^{\hat{Q}}[\varphi(\theta(\alpha)-\hat{\theta}(X,Y))H_\alpha(X)] \\
&= E^{\hat{Q}}[E^{\hat{Q}}[\varphi(\theta(\alpha)-\hat{\theta}(X,Y))|X]H_\alpha(X)] \\
&\geq E^{\hat{Q}}[\varphi(\theta(\alpha)-E^{\hat{Q}}[\hat{\theta}(X,Y)|X])H_\alpha(X)] \\
&= E^{P_\alpha}[\varphi(\theta(\alpha)-E^{\hat{Q}}[\hat{\theta}(X,Y)|X])]
\end{aligned}$$

□

Other applications

Finance: Asset pricing models depend on finding a change of measure under which the price process becomes a martingale.

Stochastic Control: For a controlled diffusion process

$$X(t) = X(0) + \int_0^t \sigma(X(s))dW(s) + \int_0^t b(X(s), u(s))ds$$

where the control only enters the drift coefficient, the controlled process can be obtained from an uncontrolled process satisfying via a change of measure.

Bayes Formula

Lemma 10.5 *(Bayes Formula) If $dP = LdQ$, then*

$$E^P[Z|\mathcal{D}] = \frac{E^Q[ZL|\mathcal{D}]}{E^Q[L|\mathcal{D}].} \tag{10.1}$$

Proof. Clearly the right side of (10.1) is $\mathcal{D}$-measurable. Let $D \in \mathcal{D}$. Then

$$\begin{aligned}
\int_D \frac{E^Q[ZL|\mathcal{D}]}{E^Q[L|\mathcal{D}].} dP &= \int_D \frac{E^Q[ZL|\mathcal{D}]}{E^Q[L|\mathcal{D}]} LdQ \\
&= \int_D \frac{E^Q[ZL|\mathcal{D}]}{E^Q[L|\mathcal{D}]} E^Q[L|\mathcal{D}]dQ \\
&= \int_D E^Q[ZL|\mathcal{D}]dQ = \int_D ZLdQ = \int_D ZdP
\end{aligned}$$

which verifies the identity. □

Example

For general random variables, suppose X and Y are independent on $(\Omega, \mathcal{F}, Q)$. Let $L = H(X,Y) \geq 0$, and $E[H(X,Y)] = 1$. Define

$$\begin{aligned} \nu_Y(\Gamma) &= Q\{Y \in \Gamma\} \\ dP &= H(X,Y)dQ. \end{aligned}$$

Bayes formula becomes

$$E^P[g(Y)|X] = \frac{E^Q[g(Y)H(X,Y)|X]}{E^Q[H(X,Y)|X]} = \frac{\int g(y)H(X,y)\nu_Y(dy)}{\int H(X,y)\nu_Y(dy)}$$

11. Filtrations and martingales

- Discrete time stochastic processes
- Filtrations and adapted processes
- Markov chains
- Martingales
- Stopping times
- Optional sampling theorem
- Doob's inequalities
- Martingales and finance

Discrete time stochastic processes

Let (E, r) be a complete, separable metric space. A sequence of E-valued random variables $\{X_n, n = 0, 1, \ldots\}$ will be called a *discrete time stochastic process* with *state space* E.

Filtrations and adapted processes

Definition 11.1 *A* filtration *is a sequence of* σ*-algebras* $\{\mathcal{F}_n, n = 0, 1, 2, \ldots\}$ *satisfying* $\mathcal{F}_n \subset \mathcal{F}$*, and* $\mathcal{F}_n \subset \mathcal{F}_{n+1}$*,* $n = 0, 1, \ldots$*.*

A stochastic process $\{X_n\}$ *is* adapted *to a filtration* $\{\mathcal{F}_n\}$ *if* X_n *is* $\mathcal{F}_n$*-measurable for each* $n = 0, 1, \ldots$*.*

If $\{X_n\}$ *is a stochastic process, then the* natural *filtration* $\{\mathcal{F}_n^X\}$ *for* X *is the filtration given by* $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$*.*

Markov chains

Definition 11.2 *A stochastic process* $\{X_n\}$ *adapted to a filtration* $\{\mathcal{F}_n\}$ *is* $\{\mathcal{F}_n\}$-Markov *if* $E[f(X_{n+1})|\mathcal{F}_n] = E[f(X_{n+1})|X_n]$ *for each* $n = 0, 1, \ldots$ *and each* $f \in B(E)$. *(*$B(E)$ *denotes the bounded, Borel measurable functions on* E*.)*

Lemma 11.3 *Let* Y *be an* E*-valued random variable and* $Z \in L^1$*. Then there exists a Borel measurable function* g*, such that*

$$E[Z|Y] = g(Y) \quad a.s.$$

Lemma 11.4 *If* $\{X_n\}$ *is* $\{\mathcal{F}_n\}$*-Markov, then for each* $k \geq 1$*,* $n \geq 0$*, and* $f \in B(E)$*,* $E[f(X_{n+k})|\mathcal{F}_n] = E[f(X_{n+k})|X_n]$*.*

Proof. Proceeding by induction, the assertion is true for $k = 1$. Suppose it holds for k_0. Then there exists $g \in B(E)$ such that

$$
\begin{aligned}
E[f(X_{n+k_0+1})|\mathcal{F}_n] &= E[E[f(X_{n+k_0+1})|\mathcal{F}_{n+k_0}]|\mathcal{F}_n] \\
&= E[E[f(X_{n+k_0+1})|X_{n+k_0}]|\mathcal{F}_n] \\
&= E[g(X_{n+k_0})|\mathcal{F}_n] \\
&= E[g(X_{n+k_0})|X_n].
\end{aligned}
$$

□

Martingales

Definition 11.5 *A* $\mathbb{R}$*-valued stochastic process* $\{M_n\} \subset L^1$ *adapted to a filtration* $\{\mathcal{F}_n\}$ *is a* $\{\mathcal{F}_n\}$*-martingale if* $E[M_{n+1}|\mathcal{F}_n] = M_n$ *for each* $n = 0, 1, \ldots$.

Lemma 11.6 *If* $\{M_n\}$ *is a* $\{\mathcal{F}_n\}$-martingale, *then* $E[M_{n+k}|\mathcal{F}_n] = M_n$, $n = 0, 1, \ldots$, $k \geq 1$.

Stopping times

Definition 11.7 *A random variable τ with values in $\{0, 1, \ldots, \infty\}$ is a* $\{\mathcal{F}_n\}$-stopping time *if $\{\tau = k\} \in \mathcal{F}_k$ for each $k = 0, 1, \ldots$. A stopping time is* finite, *if $P\{\tau < \infty\} = 1$.*

Lemma 11.8 *If τ is a $\{\mathcal{F}_n\}$-stopping time, then $\{\tau = \infty\} \in \vee_k \mathcal{F}_k$.*

Lemma 11.9 *A random variable τ with values in $\{0, 1, \ldots, \infty\}$ is a $\{\mathcal{F}_n\}$-* stopping time *if and only if $\{\tau \leq k\} \in \mathcal{F}_k$ for each $k = 0, 1, \ldots$.*

Proof. If τ is a stopping time, then $\{\tau \leq k\} = \cup_{l=0}^{k} \{\tau = l\} \in \mathcal{F}_k$. If $\{\tau \leq k\} \in \mathcal{F}_k$, $k \in \mathbb{N}$, then $\{\tau = k\} = \{\tau \leq k\} \cap \{\tau > k-1\} \in \mathcal{F}_k$. □

Hitting times

Lemma 11.10 *Let* $\{X_n\}$ *be an* E*-valued stochastic process adapted to* $\{\mathcal{F}_n\}$*. Let* $B \in \mathcal{B}(E)$*, and define* $\tau_B = \min\{n : X_n \in B\}$ *with* $\tau_B = \infty$ *if* $\{n, X_n \in B\}$ *is empty. Then* τ_B *is a* $\{\mathcal{F}_n\}$*-stopping time.*

Proof. Note that $\{\tau_B = k\} = \{X_k \in B\} \cap \cap_{l=0}^{k-1}\{X_l \in B^c\}$. □

Closure properties of the collection of stopping times

Lemma 11.11 *Suppose that* $\tau, \tau_1, \tau_2, \ldots$ *are* $\{\mathcal{F}_n\}$*-stopping times, and that* $c \in \mathbb{N} = \{0, 1, \ldots\}$. *Then*

- $\max_k \tau_k$ *and* $\min_k \tau_k$ $\{\mathcal{F}_n\}$*-stopping times.*
- $\tau \wedge c$ *and* $\tau \vee c$ *are* $\{\mathcal{F}_n\}$*-stopping times.*
- $\tau + c$ *is a* $\{\mathcal{F}_n\}$*-stopping time.*
- *If* $\{X_n\}$ *is* $\{\mathcal{F}_n\}$*-adapted with state space* E *and* $B \in \mathcal{B}(E)$, *then* $\gamma = \min\{\tau + n : n \geq 0, X_{\tau+n} \in B\}$ *is a* $\{\mathcal{F}_n\}$*-stopping time.*

Proof. For example, $\{\max_k \tau_k \leq n\} = \cap_k \{\tau_k \leq n\} \in \mathcal{F}_n$, and

$$\{\gamma = m\} = \cup_{l=0}^{m} \{\tau = l\} \cap \{X_m \in B\} \cap \cap_{k=l}^{m-1} \{X_k \in B^c\} \in \mathcal{F}_m$$

□

Stopped processes

Lemma 11.12 *Suppose* $\{X_n\}$ *is* $\{\mathcal{F}_n\}$*-adapted and* τ *is a* $\{\mathcal{F}_n\}$*-stopping time. Then* $\{X_{\tau\wedge n}\}$ *is* $\{\mathcal{F}_n\}$*-adapted.*

Proof. We have

$$\{X_{\tau\wedge n} \in B\} = (\cup_{l=0}^{n-1}\{\tau = l\}\cap\{X_l \in B\})\cup(\{\tau \geq n\}\cap\{X_n \in B\}) \in \mathcal{F}_n.$$

□

Information available at time τ

Definition 11.13

$$\begin{aligned}\mathcal{F}_\tau &= \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n, n = 0, 1, \ldots\} \\ &= \{A \in \mathcal{F} : A \cap \{\tau \le n\} \in \mathcal{F}_n, n = 0, 1, \ldots\}\end{aligned}$$

Lemma 11.14 *$\mathcal{F}_\tau$ is a σ-algebra.*

Stopped processes

Lemma 11.15 *If $\{X_n\}$ is $\{\mathcal{F}_n\}$-adapted, then for each $m \in \mathbb{N}$, $X_{m\wedge\tau}$ is $\mathcal{F}_\tau$-measurable.*

If $\mathcal{F}_n = \mathcal{F}_n^X$, $n \in \mathbb{N}$ and τ is finite for all $\omega \in \Omega$, then $\mathcal{F}_\tau^X = \sigma(X_{m\wedge\tau}, m \in \mathbb{N})$.

Proof. Note that so $\{X_{m\wedge\tau} \in B\} \in \mathcal{F}_\tau$, since

$$\{X_{m\wedge\tau} \in B\} \cap \{\tau = n\} = \{X_{m\wedge n} \in B\} \cap \{\tau = n\} \in \mathcal{F}_n, \tag{11.1}$$

If $\{\mathcal{F}_n\} = \{\mathcal{F}_n^X\}$, then by (11.1), $\sigma(X_{m\wedge\tau}, m \in \mathbb{N}) \subset \mathcal{F}_\tau^X$. If $A \in \mathcal{F}_\tau^X$, then $A \cap \{\tau = n\} \in \mathcal{F}_n^X$, so

$$A \cap \{\tau = n\} = \{(X_0, \ldots, X_n) \in B_n\} = \{(X_0, X_{1\wedge\tau}, \ldots, X_{n\wedge\tau}) \in B_n\}$$

and $A = \cup_n \{(X_0, X_{1\wedge\tau}, \ldots, X_{n\wedge\tau}) \in B_n\} \in \sigma(X_{m\wedge\tau}, m \in \mathbb{N})$. □

Monotonicity of information

Lemma 11.16 *If τ and σ are $\{\mathcal{F}_n\}$-stopping times and $\sigma \le \tau$ for all $\omega \in \Omega$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.*

Proof. If $A \in \mathcal{F}_\sigma$, then

$$A \cap \{\tau \le n\} = A \cap \{\sigma \le n\} \cap \{\tau \le n\} \in \mathcal{F}_n,$$

so $A \in \mathcal{F}_\tau$. □

Conditional expectations given $\mathcal{F}_\tau$

Lemma 11.17 *Let $Z \in L^1$, and let τ be a finite $\{\mathcal{F}_n\}$-stopping time. Then*

$$E[Z|\mathcal{F}_\tau] = \sum_{n=0}^{\infty} E[Z|\mathcal{F}_n]\mathbf{1}_{\{\tau=n\}}.$$

Proof. Problem 12 □

Sub- and supermartingales

Definition 11.18 *Let* $\{X_n\} \subset L^1$ *be a stochastic process adapted to* $\{\mathcal{F}_n\}$*. Then* $\{X_n\}$ *is a* submartingale *if and only if*

$$E[X_{n+1}|\mathcal{F}_n] \geq X_n, \quad n = 0, 1, \ldots$$

and $\{X_n\}$ *is a* supermartingale *if and only if*

$$E[X_{n+1}|\mathcal{F}_n] \leq X_n, \quad n = 0, 1, \ldots$$

Martingales and Jensen's inequality

Lemma 11.19 *If φ is convex and X is a martingale with $E[|\varphi(X_n)|] < \infty$, then $Y_n = \varphi(X_n)$ is a submartingale.*

If φ is convex and nondecreasing and X is a submartingale, with $E[|\varphi(X_n)|] < \infty$, then $Y_n = \varphi(X_n)$ is a submartingale.

Stopped submartingales

Lemma 11.20 *Suppose that X is a $\{\mathcal{F}_n\}$-submartingale and τ is a $\{\mathcal{F}_n\}$-stopping time. Then*

$$E[X_{\tau\wedge n}|\mathcal{F}_{n-1}] \geq X_{\tau\wedge(n-1)},$$

and hence $\{X_{\tau\wedge n}\}$ is a $\{\mathcal{F}_n\}$-submartingale.

Proof.

$$\begin{aligned} E[X_{\tau\wedge n}|\mathcal{F}_{n-1}] &= E[X_n \mathbf{1}_{\{\tau>n-1\}}|\mathcal{F}_{n-1}] + E[X_{\tau\wedge(n-1)}\mathbf{1}_{\{\tau\leq n-1\}}|\mathcal{F}_{n-1}] \\ &\geq X_{n-1}\mathbf{1}_{\{\tau>n-1\}} + X_{\tau\wedge(n-1)}\mathbf{1}_{\{\tau\leq n-1\}} \\ &= X_{\tau\wedge(n-1)} \end{aligned}$$

□

By iteration, for $m \leq n$ $E[X_{\tau\wedge n}|\mathcal{F}_m] \geq X_{\tau\wedge m}$.

Optional sampling theorem

Theorem 11.21 *Let X be a $\{\mathcal{F}_n\}$-submartingale and τ_1 and τ_2 be $\{\mathcal{F}_n\}$-stopping times. Then*

$$E[X_{\tau_2\wedge n}|\mathcal{F}_{\tau_1}] \geq X_{\tau_1\wedge\tau_2\wedge n}$$

Proof.

$$\begin{aligned} E[X_{\tau_2\wedge n}|\mathcal{F}_{\tau_1}] &= \sum_{m=0}^{\infty} E[X_{\tau_2\wedge n}|\mathcal{F}_m]\mathbf{1}_{\{\tau_1=m\}} \\ &\geq X_{\tau_2\wedge n}\mathbf{1}_{\{\tau_1>n\}} + \sum_{m=0}^{n} X_{\tau_2\wedge m}\mathbf{1}_{\{\tau_1=m\}} \\ &= X_{\tau_1\wedge\tau_2\wedge n} \end{aligned}$$

□

Corollary 11.22 *If in addition,* $\tau_2 < \infty$ *a.s.,* $E[|X_{\tau_2}|] < \infty$*, and*

$$\lim_{n\to\infty} E[|X_n|\mathbf{1}_{\{\tau_2 \geq n\}}] = 0,$$

then

$$E[X_{\tau_2}|\mathcal{F}_{\tau_1}] \geq X_{\tau_1 \wedge \tau_2}$$

Doob's inequalities

Theorem 11.23 *Let* $\{X_n\}$ *be a nonnegative submartingale. Then*

$$P\{\max_{m\leq n} X_m \geq x\} \leq \frac{E[X_n]}{x}$$

Proof. Let $\tau = \min\{m : X_m \geq x\}$. Then

$$E[X_n] \geq E[X_{n\wedge\tau}] \geq xP\{\tau \leq n\}.$$

□

Kolmogorov inequality

Lemma 11.24 *Let* $\{\xi_i\}$ *be independent random variables with* $E[\xi_i] = 0$ *and* $Var(\xi_i) < \infty$*. Then*

$$P\{\sup_{m\leq n} |\sum_{i=1}^{m} \xi_i| \geq r\} \leq \frac{1}{r^2} \sum_{i=1}^{n} E[\xi_i^2]$$

Proof. Since $M_n = \sum_{i=1}^{n} \xi_i$ is a martingale, M_n^2 is a submartingale, and $E[M_n^2] = \sum_{i=1}^{n} E[\xi_i^2]$. □

Doob's inequalities

Theorem 11.25 *Let* $\{X_n\}$ *be a nonnegative submartingale. Then for* $p > 1$,

$$E[\max_{m \leq n} X_m^p] \leq \left(\frac{p}{p-1}\right)^p E[X_n^p]$$

Corollary 11.26 *If* M *is a square integrable martingale, then*

$$E[\max_{m \leq n} M_m^2] \leq 4E[M_n^2].$$

Proof. Let $\tau_z = \min\{n : X_n \geq z\}$ and $Z = \max_{m \leq n} X_m$. Then

$$\begin{aligned} E[X^p_{n\wedge\tau_z}] &\leq E[X^p_n] \\ zP\{Z \geq z\} &\leq E[X_{\tau_z}\mathbf{1}_{\{\tau_z \leq n\}}] \leq E[X_n \mathbf{1}_{\{\tau_z \leq n\}}] \end{aligned}$$

and

$$\begin{aligned} E[\varphi(Z\wedge\beta)] &= \int_0^\beta \varphi'(z)P\{Z \geq z\}dz \\ &\leq \int_0^\beta \frac{\varphi'(z)}{z} E[X_n \mathbf{1}_{\{Z \geq z\}}]dz = E[X_n \psi(Z\wedge\beta)], \end{aligned}$$

where $\psi(z) = \int_0^z x^{-1}\varphi'(x)dx$. If $\varphi(z) = z^p$ and $\frac{1}{p} + \frac{1}{q} = 1$, $q = \frac{p}{p-1}$, then $\psi(z) = \frac{p}{p-1}z^{p-1}$ and

$$E[X_n\psi(Z\wedge\beta)] = \frac{p}{p-1}E[X_n(Z\wedge\beta)^{p-1}] \leq \frac{p}{p-1}E[X_n^p]^{1/p}E[(Z\wedge\beta)^p]^{1-1/p}.$$

Stopping condition for a martingale

Lemma 11.27 *Let* $\{X_n\}$ *be adapted to* $\{\mathcal{F}_n\}$*. Then* $\{X_n\}$ *is an* $\{\mathcal{F}_n\}$*-martingale if and only if*

$$E[X_{\tau\wedge n}] = E[X_0]$$

for every $\{\mathcal{F}_n\}$*-stopping time* τ *and each* $n = 0, 1, \ldots$*.*

Proof. Problem 14. □

Martingale differences

Definition 11.28 $\{\xi_n\}$ *are* martingale differences *for* $\{\mathcal{F}_n\}$ *if for each* n, ξ_n *is* $\mathcal{F}_n$*-measurable and* $E[\xi_n|\mathcal{F}_{n-1}] = 0$.

□

Lemma 11.29 *Let* $\{\xi_n\}$ *be martingale differences and* $\{Y_n\}$ *be adapted, with* $\xi_n, Y_n \in L^2$, $n = 0, 1, \ldots$. *Then*

$$M_n = \sum_{k=1}^{n} Y_{k-1}\xi_k$$

is a martingale.

Model of a market

Consider financial activity over a time interval $[0, T]$ modeled by a probability space $(\Omega, \mathcal{F}, P)$.

Assume that there is a "fair casino" or market which is *complete* in the sense that at time 0, for each event $A \in \mathcal{F}$, a price $Q(A) \geq 0$ is fixed for a bet or a contract that pays one dollar at time T if and only if A occurs.

Assume that the market is *frictionless* in that an investor can either buy or sell the contract at the same price and that it is *liquid* in that there is always a buyer or seller available. Also assume that $Q(\Omega) < \infty$.

An investor can construct a *portfolio* by buying or selling a variety of contracts (possibly countably many) in arbitrary multiples.

No arbitrage condition

If a_i is the "quantity" of a contract for A_i ($a_i < 0$ corresponds to selling the contract), then the payoff at time T is

$$\sum_i a_i \mathbf{1}_{A_i}.$$

Require $\sum_i |a_i| Q(A_i) < \infty$ (only a finite amount of money changes hands) so that the initial cost of the portfolio is (unambiguously)

$$\sum_i a_i Q(A_i).$$

The market has *no arbitrage* if no combination (buying and selling) of countably many policies with a net cost of zero results in a positive profit at no risk.

That is, if $\sum |a_i|Q(A_i) < \infty$,

$$\sum_i a_i Q(A_i) = 0, \text{ and } \sum_i a_i \mathbf{1}_{A_i} \geq 0 \quad a.s.,$$

then

$$\sum_i a_i \mathbf{1}_{A_i} = 0 \quad a.s.$$

Consequences of the no arbitrage condition

Lemma 11.30 *Assume that there is no arbitrage. If $P(A) = 0$, then $Q(A) = 0$. If $Q(A) = 0$, then $P(A) = 0$.*

Proof. Suppose $P(A) = 0$ and $Q(A) > 0$. Buy one unit of Ω and sell $Q(\Omega)/Q(A)$ units of A.

$$\text{Cost} = Q(\Omega) - \frac{Q(\Omega)}{Q(A)}Q(A) = 0$$

$$\text{Payoff} = 1 - \frac{Q(\Omega)}{Q(A)}\mathbf{1}_A = 1 \quad a.s.$$

which contradicts the no arbitrage assumption.

Now suppose $Q(A) = 0$. Buy one unit of A. The cost of the portfolio is $Q(A) = 0$ and the payoff is $\mathbf{1}_A \geq 0$. So by the no arbitrage assumption, $\mathbf{1}_A = 0$ a.s., that is, $P(A) = 0$. □

Price monotonicity

Lemma 11.31 *If there is no arbitrage and* $A \subset B$, *then* $Q(A) \leq Q(B)$, *with strict inequality if* $P(A) < P(B)$.

Proof. Suppose $P(B) > 0$ (otherwise $Q(A) = Q(B) = 0$) and $Q(B) \leq Q(A)$. Buy one unit of B and sell $Q(B)/Q(A)$ units of A.

$$\text{Cost} = Q(B) - \frac{Q(B)}{Q(A)}Q(A) = 0$$

$$\text{Payoff} = \mathbf{1}_B - \frac{Q(B)}{Q(A)}\mathbf{1}_A = \mathbf{1}_{B-A} + (1 - \frac{Q(B)}{Q(A)})\mathbf{1}_A \geq 0,$$

Payoff $= 0$ a.s. implies $Q(B) = Q(A)$ and $P(B - A) = 0$. □

Q must be a measure

Theorem 11.32 *If there is no arbitrage, Q must be a measure on $\mathcal{F}$.*

Proof. $A_1, A_2, \ldots$ disjoint and $A = \cup_{i=1}^{\infty} A_i$. Assume $P(A_i) > 0$ for some i. (Otherwise, $Q(A) = Q(A_i) = 0$.)

Let $\rho \equiv \sum_i Q(A_i)$, and buy one unit of A and sell $Q(A)/\rho$ units of A_i for each i.

$$\text{Cost} = Q(A) - \frac{Q(A)}{\rho} \sum_i Q(A_i) = 0$$

$$\text{Payoff} = \mathbf{1}_A - \frac{Q(A)}{\rho} \sum_i \mathbf{1}_{A_i} = (1 - \frac{Q(A)}{\rho})\mathbf{1}_A.$$

If $Q(A) \leq \rho$, then $Q(A) = \rho$.

If $Q(A) \geq \rho$, sell one unit of A and buy $Q(A)/\rho$ units of A_i. □

Theorem 11.33 *If there is no arbitrage, $Q << P$ and $P << Q$. (P and Q are equivalent measures.)*

Proof. The result follows from Lemma 11.30. □

Pricing general payoffs

If X and Y are random variables satisfying $X \leq Y$ a.s., then no arbitrage should mean

$$Q(X) \leq Q(Y).$$

It follows that for any Q-integrable X, the price of X is

$$Q(X) = \int X dQ$$

By the Radon-Nikodym theorm, $dQ = LdP$, for some nonnegative, integrable random variable L, and

$$Q(X) = E^P[XL]$$

Assets that can be traded at intermediate times

$\{\mathcal{F}_n\}$ represents the information available at time n.

B_n is the price at time n of a bond that is worth \$1 at time T (e.g. $B_n = \frac{1}{(1+r)^{T-n}}$), that is, at any time $0 \leq n \leq T$, B_n is the price of a contract that pays exactly \$1 at time T.

Note that $B_0 = Q(\Omega)$

Define $\hat{Q}(A) = Q(A)/B_0$.

Martingale properties of tradable assets

Let X_n be the price at time n of another tradable asset, that is, X_n is the buying or selling price at time n of an asset that will be worth X_T at time T. $\{X_n\}$ must be $\{\mathcal{F}_n\}$-adapted.

For any stopping time $\tau \leq T$, we can buy one unit of the asset at time 0, sell the asset at time τ and use the money received (X_τ) to buy X_τ/B_τ units of the bond. Since the payoff for this strategy is X_τ/B_τ (the value of the bonds at time T), we must have

$$X_0 = \int \frac{X_\tau}{B_\tau} dQ = \int \frac{B_0 X_\tau}{B_\tau} d\hat{Q}.$$

Theorem 11.34 *If X is the price of a tradable asset, then X/B is a martingale on $(\Omega, \mathcal{F}, \hat{Q})$.*

Equivalent martingale measures

Consider a simple model on $(\Omega, \mathcal{F}, P)$ of a financial market consisting of one tradable asset (stock) $\{X_n, 0 \leq n \leq T\}$, a bond $\{B_n, 0 \leq n \leq T\}$, and information filtration $\mathcal{F}_n = \sigma(X_k, B_k, 0 \leq k \leq n)$. Assume that X_0 and B_0 are almost surely constant and that the market is complete in the sense that every payoff of the form $Z = F(B_0, X_0, \ldots, B_T, X_T)$ for some bounded function F has a price at which it can be bought or sold at time zero. Then if there is no arbitrage, there must by a probability measure Q that is equivalent to P such that the price of Z is given by

$$B_0 E^Q[F(B_0, X_0, \ldots, B_T, X_T)]$$

and X/B is a martingale on $(\Omega, \mathcal{F}, Q)$. (Note that we have dropped the hat on $\hat{Q}$ to simplify notation)

Self-financing trading strategies

A trading strategy is an adapted process $\{(\alpha_n, \beta_n)\}$, where α_n gives the number of shares of the stock owned at time n and β_n, the number of units of the bond. The trading strategy is *self-financing* if

$$\alpha_{n-1}X_n + \beta_{n-1}B_n = \alpha_n X_n + \beta_n B_n, \quad n > 0.$$

Note that if α_{n-1} shares of stock are owned at time $n-1$ and β_{n-1} units of the bond, then at time n, the value of the portfolio is $\alpha_{n-1}X_n + \beta_{n-1}B_n$, and "self-financing" simply means that money may be transfered from the stock to the bond or vice versus, but no money is taken out and no money is added.

Binomial model

Assume that $B_n = (1+r)^{-(T-n)}$, $0 < P\{X_{n+1} = (1+u)X_n\} < 1$, and

$$P\{X_{n+1} = (1-d)X_n\} = 1 - P\{X_{n+1} = (1+u)X_n\},$$

for some $-d < r < u$, so that we can write $X_{n+1} = (1+\xi_{n+1})X_n$, where $\mathcal{R}(\xi_{n+1}) = \{-d, u\}$. Since $E^Q[X_{n+1}(1+r)^{T-(n+1)}|\mathcal{F}_n] = X_n(1+r)^{T-n}$,

$$E^Q[X_{n+1}|\mathcal{F}_n] = X_n(1+r), \qquad E^Q[\xi_{n+1}|\mathcal{F}_n] = r,$$

and hence

$$Q\{\xi_{n+1} = u|\mathcal{F}_n\} = \frac{r+d}{u+d},$$

so that under Q, the $\{\xi_n\}$ are iid with

$$Q\{\xi_n = u\} = 1 - Q\{\xi_n = -d\} = \frac{r+d}{u+d}.$$

In particular, there is only one possible choice of Q defined on $\mathcal{F}_T$.

Hedging

Theorem 11.35 *For the binomial model, for each* $Z = F(X_0, \ldots, X_T)$*, there exists a self-financing trading strategy such that*

$$\alpha_{T-1}X_T + \beta_{T-1} = F(X_0, \ldots, X_T). \tag{11.2}$$

Proof. Note that the self-financing requirement becomes

$$\alpha_{n-1}X_n + \beta_{n-1}(1+r)^{-(T-n)} = \alpha_n X_n + \beta_n(1+r)^{-(T-n)}, \quad n > 0,$$

and (11.2) and the martingale property for X/B would imply

$$\begin{aligned} E^Q[F(X_0,\ldots,X_T)|\mathcal{F}_{T-1}] &= \alpha_{T-1}(1+r)X_{T-1} + \beta_{T-1} \\ &= (1+r)(\alpha_{T-2}X_{T-1} + \beta_{T-2}B_{T-1}) \\ E^Q[F(X_0,\ldots,X_T)|\mathcal{F}_n] &= (1+r)^{T-n}(\alpha_n X_n + \beta_n B_n) \\ &= (1+r)^{T-n}(\alpha_{n-1}X_n + \beta_{n-1}B_n). \end{aligned}$$

Let

$$H_n(X_0, \ldots, X_n) = E^Q[F(X_0, \ldots, X_T)|\mathcal{F}_n]$$

We can solve

$$\begin{aligned}
\alpha_{T-1}X_{T-1}(1+u) + \beta_{T-1} &= F(X_0, \ldots, X_{T-1}, X_{T-1}(1+u)) \\
\alpha_{T-1}X_{T-1}(1-d) + \beta_{T-1} &= F(X_0, \ldots, X_{T-1}, X_{T-1}(1-d))
\end{aligned}$$

and

$$\begin{aligned}
(1+r)^{T-n}(\alpha_{n-1}X_{n-1}(1+u) + \beta_{n-1}B_n) &= H_n(X_0, \ldots, X_{n-1}, X_{n-1}(1+u)) \\
(1+r)^{T-n}(\alpha_{n-1}X_{n-1}(1-d) + \beta_{n-1}B_n) &= H_n(X_0, \ldots, X_{n-1}, X_{n-1}(1-d))
\end{aligned}$$

Note that the solution will be adapted and

$$H_n(X_0, \ldots, X_n) = (1+r)^{T-n}(\alpha_{n-1}X_n + \beta_{n-1}B_n)$$

Since

$$
\begin{aligned}
&H_n(X_0,\ldots,X_n)\\
&\quad= E^Q[H_{n+1}(X_0,\ldots,X_{n+1})|\mathcal{F}_n]\\
&\quad= \frac{r+d}{u+d}H_{n+1}(X_0,\ldots,X_n,X_n(1+u))\\
&\qquad\qquad +\frac{u-r}{u+d}H_{n+1}(X_0,\ldots,X_n,X_n(1-d))\\
&\quad= \frac{r+d}{u+d}(1+r)^{T-n-1}(\alpha_n X_n(1+u)+\beta_n B_{n+1})\\
&\qquad\qquad +\frac{u-r}{u+d}(1+r)^{T-n-1}(\alpha_n X_n(1-d)+\beta_n B_{n+1})\\
&\quad= (1+r)^{T-n}(\alpha_n X_n+\beta_n B_n),
\end{aligned}
$$

the solution is self-financing. □

Corollary 11.36 *For the binomial model, if all self-financing strategies are allowed, then the market is complete.*

12. Martingale convergence

- Properties of convergent sequences
- Upcrossing inequality
- Martingale convergence theorem
- Uniform integrability
- Reverse martingales
- Martingales with bounded increments
- Extended Borel-Cantelli lemma
- Radon-Nikodym theorem
- Law of large numbers for martingales

Properties of convergent sequences

Lemma 12.1 *Let* $\{x_n\} \subset \mathbb{R}$. *Suppose that for each* $a < b$, *the sequence crosses the interval* $[a, b]$ *only finitely often. Then either* $\lim_{n\to\infty} x_n = \infty$, $\lim_{n\to\infty} x_n = -\infty$, *or* $\lim_{n\to\infty} x_n = x$ *for some* $x \in \mathbb{R}$. *(Note that it is sufficient to consider rational* a *and* b.*)*

Proof. For each $a < b$, either $\limsup_{n\to\infty} x_n \leq b$ or $\liminf_{n\to\infty} x_n \geq a$. Suppose there exists $b_0 \in \mathbb{R}$ such that $\limsup_{n\to\infty} x_n \leq b_0$, and let $\bar{b} = \inf\{b : \limsup_{n\to\infty} x_n \leq b\}$. If $\bar{b} = -\infty$, then $\lim_{n\to\infty} x_n = -\infty$. Otherwise, for each $\epsilon > 0$, $\liminf_{n\to\infty} x_n \geq \bar{b} - \epsilon$, and hence, $\lim_{n\to\infty} x_n = \bar{b}$. □

(sub)-martingale transforms

Let $\{H_n\}$ and $\{X_n\}$ be $\{\mathcal{F}_n\}$-adapted, and define

$$H \cdot X_n = \sum_{k=1}^{n} H_{k-1}(X_k - X_{k-1}).$$

Lemma 12.2 *If X is a submartingale (supermartingale) and H is a non-negative, adapted sequence, then $H \cdot X$ is a submartingale (supermartingale)*

Proof.

$$\begin{aligned} E[H \cdot X_{n+1}|\mathcal{F}_n] &= H \cdot X_n + E[H_n(X_{n+1} - X_n)|\mathcal{F}_n] \\ &= H \cdot X_n + H_n E[X_{n+1} - X_n|\mathcal{F}_n] \\ &\geq H \cdot X_n \end{aligned}$$

□

Upcrossing inequality

For $a \leq b$, let $U_n(a,b)$ be the number of times the sequence $\{X_k\}$ crosses from below a to above b by time n.

Lemma 12.3 *Let $\{X_n\}$ be a submartingale. Then for $a < b$,*

$$\frac{E[(X_n - a)^+]}{b - a} \geq E[U_n(a,b)]$$

Proof. Define

$$\begin{aligned}\sigma_1 &= \min\{n : X_n \le a\}\\ \tau_i &= \min\{n > \sigma_i : X_n \ge b\}\\ \sigma_{i+1} &= \min\{n > \tau_i : X_n \le a\}\end{aligned}$$

and

$$H_k = \sum_i \mathbf{1}_{\{\tau_i \le k < \sigma_{i+1}\}}.$$

Then $H \cdot X_n = \sum_i (X_{n\wedge\sigma_{i+1}} - X_{n\wedge\tau_i})$ and $U_n(a,b) = \max\{i : \tau_i \le n\}$. Then since if $\tau_i < \infty$, $X_{\tau_i} \ge b$ and if $\sigma_i < \infty$, $X_{\sigma_i} \le a$,

$$\begin{aligned}-H \cdot X_n &\ge (b-a)U_n(a,b) - \sum_i (X_{n\wedge\sigma_{i+1}} - a)1_{\{\tau_i \le n < \tau_{i+1}\}}\\ &\ge (b-a)U_n(a,b) - (X_n - a)^+\end{aligned}$$

and hence

$$0 \ge (b-a)E[U_n(a,b)] - E[(X_n - a)^+]$$

□

Martingale convergence theorem

Theorem 12.4 *Let* $\{X_n\}$ *be a submartingale with* $\sup_n E[X_n^+] < \infty$. *Then* $\lim_{n\to\infty} X_n$ *exists a.s.*

Corollary 12.5 *If* $\{X_n\}$ *is a nonnegative supermartingale, then* $X_\infty = \lim_{n\to\infty} X_n$ *exists a.s. and* $E[X_0] \geq E[X_\infty]$

Proof. For each $a < b$,

$$E[\lim_{n\to\infty} U_n(a,b)] \leq \sup_n \frac{E[(X_n - a)^+]}{b-a} \leq \frac{\sup_n E[X_n^+] + |a|}{b-a} < \infty.$$

Therefore, with probability one, $U_\infty(a,b) < \infty$ for all rational a, b. Consequently, there exists $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$ such that for $\omega \in \Omega_0$, either $\lim_{n\to\infty} X_n(\omega) = \infty$, $\lim_{n\to\infty} X_n(\omega) = -\infty$, or $\lim_{n\to\infty} X_n(\omega) = X_\infty(\omega)$ for some $X_\infty(\omega) \in \mathbb{R}$.

Since $E[X_n] \geq E[X_0]$,

$$E[|X_n|] = 2E[X_n^+] - E[X_n] \leq 2E[X_n^+] - E[X_0].$$

Consequently,

$$E[\liminf_{n\to\infty} |X_n|] \leq \liminf_{n\to\infty} E[|X_n|] < \infty,$$

so $P\{\lim_{n\to\infty} |X_n| = \infty\} = 0$ and $\lim_{n\to\infty} X_n \in \mathbb{R}$ with probability one. □

Examples

$X_n = \prod_{k=1}^n \xi_k$ ξ_k iid with $\xi_k \geq 0$ a.s. and $E[\xi_k] = 1$. $\{X_n\}$ is a nonnegative martingale. Hence, $\lim_{n\to\infty} X_n$ exists. What is it?

$S_n = 1 + \sum_{k=1}^n \eta_k$, η_k iid, integer-valued, nontrivial, with $E[\eta_k] = 0$, $\eta_k \geq -1$ a.s.

Let $\tau = \inf\{n : S_n = 0\}$. Then for $X_n = S_{n\wedge\tau}$, $\lim_{n\to\infty} X_n$ must be zero.

Properties of integrable random variables

Lemma 12.6 *If X is integrable, then for $\epsilon > 0$ there exists a $K > 0$ such that*

$$\int_{\{|X|>K\}} |X| dP < \epsilon.$$

Proof. $\lim_{K\to\infty} |X|\mathbf{1}_{\{|X|>K\}} = 0$ a.s. □

Lemma 12.7 *If X is integrable, then for $\epsilon > 0$ there exists a $\delta > 0$ such that $P(F) < \delta$ implies $\int_F |X| dP < \epsilon$.*

Proof. Let $F_n = \{|X| \geq n\}$. Then $nP(F_n) \leq E[|X|\mathbf{1}_{F_n}] \to 0$. Select n so that $E[|X|\mathbf{1}_{F_n}] \leq \epsilon/2$, and let $\delta = \frac{\epsilon}{2n}$. Then $P(F) < \delta$ implies

$$\int_F |X| dP \leq \int_{F_n \cap F} |X| dP + \int_{F_n^c \cap F} |X| dP < \frac{\epsilon}{2} + n\delta = \epsilon$$

□

Uniform integrability

Theorem 12.8 *Let $\{X_\alpha\}$ be a collection of integrable random variables. The following are equivalent:*

a) $\sup E[|X_\alpha|] < \infty$ *and for* $\epsilon > 0$ *there exists* $\delta > 0$ *such that* $P(F) < \delta$ *implies* $\sup_\alpha \int_F |X_\alpha| dP < \epsilon$.

b) $\lim_{K\to\infty} \sup_\alpha E[|X_\alpha| \mathbf{1}_{\{|X_\alpha|>K\}}] = 0$.

c) $\lim_{K\to\infty} \sup_\alpha E[|X_\alpha| - |X_\alpha| \wedge K] = 0$

d) There exists a (strictly) convex function φ *on* $[0,\infty)$ *with* $\lim_{r\to\infty} \frac{\varphi(r)}{r} = \infty$ *such that* $\sup_\alpha E[\varphi(|X_\alpha|)] < \infty$.

Proof. a) implies b) follows by

$$P\{|X_\alpha| > K\} \leq \frac{E[|X_\alpha|]}{K}$$

b) implies d): Let $N_1 = 0$, and for $k > 1$, select N_k such that

$$\sum_{k=1}^\infty k \sup_\alpha E[\mathbf{1}_{\{|X_\alpha|>N_k\}}|X_\alpha|] < \infty$$

Define $\varphi(0) = 0$ and

$$\varphi'(x) = k, \qquad N_k \leq x < N_{k+1}.$$

Recall that $E[\varphi(|X|)] = \int_0^\infty \varphi'(x)P\{|X| > x\}dx$, so

$$E[\varphi(|X_\alpha|)] = \sum_{k=1}^\infty k \int_{N_k}^{N_{k+1}} P\{|X_\alpha| > x\}dx \leq \sum_{k=1}^\infty k \sup_\alpha E[\mathbf{1}_{\{|X_\alpha|>N_k\}}|X_\alpha|].$$

To obtain strictly convex $\tilde{\varphi}$, define

$$\tilde{\varphi}'(x) = k - \frac{1}{N_{k+1} - N_k}(N_{k+1} - x) \le \varphi'(x).$$

d) implies b): Assume for simplicity that $\varphi(0) = 0$ and φ is increasing. Then $r^{-1}\varphi(r)$ is increasing and $|X_\alpha|^{-1}\varphi(|X_\alpha|)\mathbf{1}_{\{|X_\alpha|>K\}} \ge K^{-1}\varphi(K)$, so $E[\mathbf{1}_{\{|X_\alpha|>K\}}|X_\alpha|] \le \frac{K}{\varphi(K)}E[\varphi(|X_\alpha|)]$

b) implies a): $\int_F |X_\alpha| dP \le P(F)K + E[\mathbf{1}_{\{|X_\alpha|>K\}}|X_\alpha|]$.

To see that (b) is equivalent to (c), observe that

$$E[|X_\alpha| - |X_\alpha| \wedge K] \le E[|X_\alpha|\mathbf{1}_{\{|X_\alpha|>K\}}] \le 2E[|X_\alpha| - |X_\alpha| \wedge \frac{K}{2}]$$

$\square$

Uniformly integrable families

- For X integrable, $\Gamma = \{E[X|\mathcal{D}] : \mathcal{D} \subset \mathcal{F}\}$
- For $X_1, X_2, \ldots$ integrable and identically distributed

$$\Gamma = \{\frac{X_1 + \cdots + X_n}{n} : n = 1, 2, \ldots\}$$

- For $Y \geq 0$ integrable, $\Gamma = \{X : |X| \leq Y\}$.

Uniform integrability and L^1 convergence

Theorem 12.9 *$X_n \to X$ in L^1 iff $X_n \to X$ in probability and $\{X_n\}$ is uniformly integrable.*

Proof. If $X_n \to X$ in L^1, then

$$\lim_{n\to\infty} E[|X_n| - |X_n| \wedge K] = E[|X| - |X| \wedge K]$$

and Condition (c) of Theorem 12.8 follows from the fact that

$$\lim_{K\to\infty} E[|X| - |X| \wedge K] = \lim_{K\to\infty} E[|X_n| - |X_n| \wedge K] = 0.$$

Conversely, let $f_K(x) = ((-K) \vee x) \wedge K$, and note that $|x - f_K(x)| = |x| - K \wedge |x|$. Since

$$|X_n - X| \leq |X_n - f_K(X_n)| + |f_K(X_n) - f_K(X)| + |X - f_K(X)|,$$

$$\limsup_{n\to\infty} E[|X_n - X|] \leq 2 \sup_n E[|X_n - f_K(X_n)|] \stackrel{K\to\infty}{\to} 0.$$

□

Convergence of conditional expectations

Theorem 12.10 *Let $\{\mathcal{F}_n\}$ be a filtration and $Z \in L^1$. Then $M_n = E[Z|\mathcal{F}_n]$ is a $\{\mathcal{F}_n\}$-martingale and $\lim_{n\to\infty} M_n$ exists a.s. and in L^1. If Z is $\vee_n\mathcal{F}_n$-measurable, then $Z = \lim_{n\to\infty} M_n$.*

Proof. Since $E[|M_n|] \leq E[|Z|]$, almost sure convergence follows by the martingale convergence theorem and L^1-convergence from the uniform integrability of $\{M_n\}$.

Suppose Z is $\vee_n\mathcal{F}_n$-measurable, and let $Y = \lim_{n\to\infty} M_n$. Then Y is $\vee_n\mathcal{F}_n$-measurable, and for $A \in \cup_n\mathcal{F}_n$,

$$E[\mathbf{1}_A Z] = \lim_{n\to\infty} E[\mathbf{1}_A M_n] = E[\mathbf{1}_A Y].$$

Therefore $E[\mathbf{1}_A Z] = E[\mathbf{1}_A Y]$ for all $A \in \vee_n\mathcal{F}_n$. Taking $A = \{Y > Z\}$ and $\{Y < Z\}$ gives the last statement of the theorem. □

Null sets and complete probability spaces

Let $(\Omega, \mathcal{F}, P)$ be a probability space. The collection of *null sets* $\mathcal{N}$ is the collection of all events $A \in \mathcal{F}$ such that $P(A) = 0$. $(\Omega, \mathcal{F}, P)$ is *complete*, if $A \in \mathcal{N}$ and $B \subset A$ implies $B \in \mathcal{N} \subset \mathcal{F}$.

Lemma 12.11 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let*

$$\bar{\mathcal{F}} = \{A \subset \Omega : \exists B \in \mathcal{F}, C \in \mathcal{N} \ni A \triangle B \subset C\}.$$

Then $\bar{\mathcal{F}}$ *is a* σ*-algebra and* P *extends to a measure* $\bar{P}$ *on* $\bar{\mathcal{F}}$. $(\Omega, \bar{\mathcal{F}}, \bar{P})$ *is called the* completion *of* $(\Omega, \mathcal{F}, P)$.

If $(\Omega, \mathcal{F}, P)$ is complete and $\mathcal{D} \subset \mathcal{F}$ is a σ-algebra, then the completion of $\mathcal{D}$ is $\bar{\mathcal{D}} = \sigma(\mathcal{D} \cup \mathcal{N})$.

Extension of Kolmogorov zero-one law

Corollary 12.12 *Let $\{\mathcal{D}_n\}$ and $\mathcal{G}$ be independent σ-algebras, and let*

$$\mathcal{T} = \cap_n \mathcal{G} \vee \vee_{m \geq n} \mathcal{D}_m.$$

Then $\bar{\mathcal{T}} = \bar{\mathcal{G}}$, where $\bar{\mathcal{T}}$ is the completion of $\mathcal{T}$ and $\bar{\mathcal{G}}$ is the completion of $\mathcal{G}$.

Proof. Clearly, $\mathcal{G} \subset \mathcal{T}$. Let $\mathcal{F}_n = \mathcal{G} \vee \vee_{k=1}^n \mathcal{D}_k$. Then for $A \in \mathcal{T}$, by Problem 11,

$$E[\mathbf{1}_A|\mathcal{F}_n] = E[\mathbf{1}_A|\mathcal{G}].$$

But $\mathbf{1}_A = \lim_{n\to\infty} E[\mathbf{1}_A|\mathcal{F}_n]$ a.s., so $A \in \bar{\mathcal{G}}$. □

Reverse martingale convergence theorem

Theorem 12.13 *Let* $\{\mathcal{G}_n\}$ *be* σ*-algebras in* $\mathcal{F}$ *satisfying* $\mathcal{G}_n \supset \mathcal{G}_{n+1}$ *and let* $Z \in L^1$. *Then* $\lim_{n\to\infty} E[Z|\mathcal{G}_n]$ *exists a.s. and in* L^1.

Proof. Let $Y_k^N = E[Z|\mathcal{G}_{N-k}]$, $0 \leq k \leq N$. Then $\{Y_k^N\}$ is a martingale, and the upcrossing inequality for $\{Y_k^N\}$ gives a "downcrossing" inequality for $\{E[Z|\mathcal{G}_n]\}$. □

A proof of the law of large numbers

Let $\{\xi_i\}$ be iid random variables with $E[|\xi_i|] < \infty$. Define

$$X_n = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

and $\mathcal{G}_n = \sigma(X_n, \xi_{n+1}, \xi_{n+2}, \ldots)$. Then $\mathcal{G}_n \supset \mathcal{G}_{n+1}$ and

$$E[\xi_1|\mathcal{G}_n] = X_n$$

is a reverse martingale, so X_n converges a.s. and in L^1. By the Kolmogorov zero-one law, the limit must be a constant and hence $E[\xi_i]$

Asymptotic behavior of a martingale with bounded increments

Theorem 12.14 *Let* $\{M_n\}$ *be a martingale and suppose that*

$$E[\sup_n |M_{n+1} - M_n|] < \infty$$

Let $H_1 = \{\lim_{n\to\infty} M_n \text{ exists}\}$ *and*

$$H_2 = \{\limsup_{n\to\infty} M_n = \infty, \liminf_{n\to\infty} M_n = -\infty\}.$$

Then $P(H_1 \cup H_2) = 1.$

Proof. Let $H_2^+ = \{\limsup_{n\to\infty} M_n = \infty\}$ and $H_2^- = \{\liminf_{n\to\infty} M_n = -\infty\}$. For $c > 0$, let $\tau_c = \inf\{n : M_n > c\}$. Then $\{M_n^{\tau_c}\}$ is a martingale satisfying

$$E[|M_k^{\tau_c}|] \leq 2E[\sup_n |M_{n+1} - M_n|] + 2c - E[M_k^{\tau_c}].$$

Consequently, $Y_c = \lim_{n\to\infty} M_n^{\tau_c}$ exists almost surely. Then $H_1 = \cup_c \{Y_c < c\}$ and $H_2^+ \supset \cap_c \{Y_c \geq c\}$. Consequently, $P(H_1 \cup H_2^+) = 1$. Similarly, $P(H_1 \cup H_2^-) = 1$, and hence $P(H_1 \cup (H_2^+ \cap H_2^-)) = 1$. □

Extended Borel-Cantelli lemma

Recalling the Borel-Cantelli lemma, we have the following corollary:

Corollary 12.15 *For* $n = 1, 2, \ldots$, *let* $A_n \in \mathcal{F}_n$. *Then*

$$G_1 \equiv \{\sum_{n=1}^{\infty} \mathbf{1}_{A_n} = \infty\} = \{\sum_{n=1}^{\infty} P(A_n|\mathcal{F}_{n-1}) = \infty\} \equiv G_2 \quad a.s.$$

Proof.

$$M_n = \sum_{i=1}^{n} (\mathbf{1}_{A_i} - P(A_i|\mathcal{F}_{i-1}))$$

is a martingale satisfying $\sup_n |M_{n+1} - M_n| \leq 1$. Consequently, with H_1 and H_2 defined as in Theorem 12.14, $P(H_1 \cup H_2) = 1$.

Clearly, $H_2 \subset G_1$ and $H_2 \subset G_2$. For $\omega \in H_1$, $\lim_{n\to\infty} M_n(\omega)$ exists, so either both $\sum_{n=1}^{\infty} \mathbf{1}_{A_n} < \infty$ and $\sum_{n=1}^{\infty} P(A_n|\mathcal{F}_{n-1}) < \infty$ or $\omega \in G_1$ and $\omega \in G_2$. $\square$

Jensen's inequality revisited

Lemma 12.16 *If φ is strictly convex and increasing on $[0,\infty)$ and $X \geq 0$, then*

$$E[\varphi(X)|\mathcal{D}] = \varphi(E[X|\mathcal{D}]) < \infty \quad a.s.$$

implies that $X = E[X|\mathcal{D}]$ a.s.

Proof. Strict convexity implies that for $x \neq y$,

$$\varphi(x) - \varphi(y) > \varphi^+(y)(x-y).$$

Consequently,

$$E[\varphi(X) - \varphi(E[X|\mathcal{D}]) - \varphi^+(E[X|\mathcal{D}])(X - E[X|\mathcal{D}])|\mathcal{D}] = 0$$

implies $X = E[X|\mathcal{D}]$ a.s. □

Radon-Nikodym theorem

Theorem 12.17 *Let ν and μ be finite measures on $(S, \mathcal{S})$. Suppose that for each $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that $\mu(A) < \delta_\epsilon$ implies $\nu(A) < \epsilon$. Then there exists a nonnegative $\mathcal{S}$-measurable function g such that*

$$\nu(A) = \int_A g d\mu, \quad A \in \mathcal{S}.$$

Proof. Without loss of generality, assume that μ is a probability measure. For a partition $\{B_k\} \subset \mathcal{S}$, define

$$X^{\{B_k\}} = \sum_k \frac{\nu(B_k)}{\mu(B_k)} \mathbf{1}_{B_k},$$

which will be a random variable on the probability space $(S, \mathcal{S}, \mu)$.

Then$E[X^{\{B_k\}}] = \nu(S) < \infty$,

$$\mu\{X^{\{B_k\}} \geq K\} \leq \frac{\nu(S)}{K}$$

and for $K > \frac{\nu(S)}{\delta_\epsilon}$,

$$\int_{\{X^{\{B_k\}} \geq K\}} X^{\{B_k\}} d\mu = \nu(X^{\{B_k\}} \geq K) < \epsilon.$$

It follows that $\{X^{\{B_k\}}\}$ is uniformly integrable. Therefore there is a strictly convex, increasing function φ such that

$$\sup_{\{B_k\}} E[\varphi(X^{\{B_k\}})] < \infty$$

Let $\mathcal{D}^{\{B_k\}} \subset \mathcal{S}$ be the σ-algebra generated by $\{B_k\}$. If $\{C_l\}$ is a refinement of $\{B_k\}$, then

$$E[X^{\{C_l\}} | \mathcal{D}^{\{B_k\}}] = X^{\{B_k\}}.$$

Let $\{A_n\} \subset \mathcal{S}$, and let $\mathcal{F}_n = \sigma(A_1, \ldots, A_n)$. Then there exists a finite partition $\{B_k^n\}$ such that $\mathcal{F}_n = \sigma(\{B_k^n\})$. Let

$$M_n = X^{\{B_k^n\}} = \sum_k \frac{\nu(B_k^n)}{\mu(B_k^n)} \mathbf{1}_{B_k^n}.$$

Then $\{M_n\}$ is a $\{\mathcal{F}_n\}$-martingale, and $M_n \to M^{\{A_n\}}$ a.s. and in L^1.

Let

$$\gamma = \sup_{\{A_n\} \subset \mathcal{S}} E[\varphi(M^{\{A_n\}}] = \lim_{m \to \infty} E[\varphi(M^{\{A_n^m\}})].$$

Let $\{\hat{A}_n\} = \cup_m \{A_n^m\}$. Then

$$E[M^{\{\hat{A}_n\}} | \sigma(\{A_n^m\})] = M^{\{A_n^m\}}$$

and $E[\varphi(M^{\{\hat{A}_n\}})] = \gamma$.

For each $A \in \mathcal{S}$, we must have $E[\varphi(M^{\{\hat{A}_n\}\cup\{A\}})] = \gamma$, and hence

$$E[\varphi(M^{\{\hat{A}_n\}\cup\{A\}}) - \varphi(M^{\{\hat{A}_n\}}) - \varphi^+(M^{\{\hat{A}_n\}})(M^{\{\hat{A}_n\}\cup\{A\}} - M^{\{\hat{A}_n\}})] = 0,$$

which implies $M^{\{\hat{A}_n\}\cup\{A\}} = M^{\{\hat{A}_n\}}$ a.s. and

$$\nu(A) = E[\mathbf{1}_A M^{\{\hat{A}_n\}\cup\{A\}}] = E[\mathbf{1}_A M^{\{\hat{A}_n\}}] = \int_A M^{\{\hat{A}_n\}} d\mu.$$

□

Kronecker's lemma

Lemma 12.18 *Let* $\{A_n\}$ *and* $\{Y_n\}$ *be sequences of random variables where* $A_0 > 0$ *and* $A_{n+1} \geq A_n$*,* $n = 0, 1, 2, \ldots$*. Define* $R_n = \sum_{k=1}^n \frac{1}{A_{k-1}}(Y_k - Y_{k-1})$*. and suppose that* $\lim_{n\to\infty} A_n = \infty$ *and that* $\lim_{n\to\infty} R_n$ *exists a.s. Then,* $\lim_{n\to\infty} \frac{Y_n}{A_n} = 0 \quad a.s.$

Proof.

$$\begin{aligned} A_n R_n &= \sum_{k=1}^n (A_k R_k - A_{k-1} R_{k-1}) = \sum_{k=1}^n R_{k-1}(A_k - A_{k-1}) + \sum_{k=1}^n A_k (R_k - R_{k-1}) \\ &= Y_n - Y_0 + \sum_{k=1}^n R_{k-1}(A_k - A_{k-1}) + \sum_{k=1}^n \frac{1}{A_{k-1}}(Y_k - Y_{k-1})(A_k - A_{k-1}) \end{aligned}$$

and

$$\frac{Y_n}{A_n} = \frac{Y_0}{A_n} + R_n - \frac{1}{A_n}\sum_{k=1}^n R_{k-1}(A_k - A_{k-1}) - \frac{1}{A_n}\sum_{k=1}^n \frac{1}{A_{k-1}}(Y_k - Y_{k-1})(A_k - A_{k-1})$$

□

Law of large numbers for martingales

Lemma 12.19 *Suppose* $\{A_n\}$ *is as in Lemma 12.18 and is adapted to* $\{\mathcal{F}_n\}$*, and suppose* $\{M_n\}$ *is a* $\{\mathcal{F}_n\}$*-martingale such that for each* $\{\mathcal{F}_n\}$*-stopping time* τ*,* $E[A_{\tau-1}^{-2}(M_\tau - M_{\tau-1})^2\mathbf{1}_{\{\tau<\infty\}}] < \infty$*. If*

$$\sum_{k=1}^{\infty} \frac{1}{A_{k-1}^2}(M_k - M_{k-1})^2 < \infty \quad a.s.,$$

then $\lim_{n\to\infty} \frac{M_n}{A_n} = 0$ *a.s.*

Proof. Without loss of generality, we can assume that $A_n \geq 1$. Let

$$\tau_c = \min\{n : \sum_{k=1}^{n} \frac{1}{A_{k-1}^2}(M_k - M_{k-1})^2 \geq c\}.$$

Then

$$\sum_{k=1}^{\infty} \frac{1}{A_{k-1}^2}(M_{k\wedge\tau_c} - M_{(k-1)\wedge\tau_c})^2 \leq c + \frac{1}{A_{\tau_c-1}^2}(M_{\tau_c} - M_{\tau_c-1})^2 \mathbf{1}_{\{\tau_c<\infty\}}.$$

Defining $R_n^c = \sum_{k=1}^{n} \frac{1}{A_{k-1}}(M_{k\wedge\tau_c} - M_{(k-1)\wedge\tau_c})$, $\sup_n E[(R_n^c)^2] < \infty$, and hence, $\{R_n^c\}$ converges a.s. Consequently, by Lemma 12.18, $\lim_{n\to\infty} \frac{M_{n\wedge\tau_c}}{A_n} = 0$ a.s. Since

$$\{\lim_{n\to\infty} \frac{M_n}{A_n} = 0\} \supset \cup_c(\{\lim_{n\to\infty} \frac{M_{n\wedge\tau_c}}{A_n} = 0\} \cap \{\tau_c = \infty\}),$$

$P\{\lim_{n\to\infty} \frac{M_n}{A_n} = 0\} = 1$. □

Three series theorem

Theorem 12.20 *Let* $\{\xi_n\}$ *be* $\{\mathcal{F}_n\}$*-adapted and define* $\eta_n = \xi_n \mathbf{1}_{\{|\xi_n| \leq b\}}$*. If*

$$\sum_{n=1}^{\infty} P\{|\xi_{n+1}| > b | \mathcal{F}_n\} < \infty \; a.s., \quad \sum_{n=1}^{\infty} E[\eta_{n+1} | \mathcal{F}_n] \text{ converges } a.s.,$$

and

$$\sum_{k=1}^{\infty} E[(\eta_k - E[\eta_k | \mathcal{F}_{k-1}])^2 | \mathcal{F}_{k-1}] < \infty,$$

then $\sum_{n=1}^{\infty} \xi_n$ *converges a.s.*

Proof. Let $\tau_c = \inf\{n : \sum_{k=1}^n E[(\eta_k - E[\eta_k|\mathcal{F}_{k-1}])^2|\mathcal{F}_{k-1}] > c\}$, and note that $\{\tau_c = n\} \in \mathcal{F}_{n-1}$. Then $M_n = \sum_{k=1}^n (\eta_k - E[\eta_k|\mathcal{F}_{k-1}])$ is a martingale with bounded increments as is $M_{n\wedge\tau_c}$. Since

$$E[(M_{n\wedge\tau_c})^2] = \sum_{k=1}^n E[E[(\eta_k - E[\eta_k|\mathcal{F}_{k-1}])^2 \mathbf{1}_{\{\tau_c \geq k\}}] \leq c + 4b^2,$$

$\lim_{n\to\infty} M_{n\wedge\tau_c}$ exists a.s. Since $\lim_{c\to\infty} P\{\tau_c = \infty\} = 1$, $\lim_{n\to\infty} M_n$ exists a.s. Since the extended Borel-Cantelli lemma implies $\sum_{n=1}^\infty \mathbf{1}_{\{|\xi_n|>b\}} < \infty$ a.s., $\sum_{n=1}^\infty \xi_n$ converges a.s. □

Geometric convergence

Lemma 12.21 *Let* $\{M_n\}$ *be a martingale with* $|M_{n+1} - M_n| \leq c$ *a.s. for each* n *and* $M_0 = 0$*. Then for each* $\epsilon > 0$*, there exist* C *and* η *such that*

$$P\{\frac{1}{n}|M_n| \geq \epsilon\} \leq Ce^{-n\eta}.$$

Proof. Let $\hat{\varphi}(x) = e^{-x} + e^x$ and $\varphi(x) = e^x - 1 - x$. Then, setting $X_k = M_k - M_{k-1}$

$$\begin{aligned} E[\hat{\varphi}(aM_n)] &= 2 + \sum_{k=1}^{n} E[\hat{\varphi}(aM_k) - \hat{\varphi}(aM_{k-1})] \\ &= 2 + \sum_{k=1}^{n} E[\exp\{aM_{k-1}\}\varphi(aX_k) + \exp\{-aM_{k-1}\}\varphi(-aX_k)] \\ &\leq 2 + \sum_{k=1}^{n} \varphi(ac)E[\hat{\varphi}(aM_{k-1})], \end{aligned}$$

and hence

$$E[\hat{\varphi}(aM_n)] \leq 2e^{n\varphi(ac)}.$$

Consequently,

$$P\{\sup_{k\leq n} \frac{1}{n}|M_k| \geq \epsilon\} \leq \frac{E[\hat{\varphi}(aM_n)]}{\hat{\varphi}(an\epsilon)} \leq 2e^{n(\varphi(ac)-a\epsilon)}.$$

Then $\eta = \sup_a(a\epsilon - \varphi(ac)) > 0$, and the lemma follows. □

Truncation

In the usual formulations of the law of large numbers, $A_n = n$ or equivalently, $A_n = n+1$, so we would like to know

$$E[\tau^{-2}(M_\tau - M_{\tau-1})^2 \mathbf{1}_{\{\tau<\infty\}}] < \infty \qquad (12.1)$$

and

$$\sum_{k=1}^{\infty} \frac{1}{k^2}(M_k - M_{k-1})^2 < \infty \quad a.s.$$

Define $\rho_k(x) = ((-k) \vee x) \wedge k$ and

$$\xi_n = \rho_n(M_n - M_{n-1}) - E[\rho_n(M_n - M_{n-1})|\mathcal{F}_{n-1}].$$

Then $\hat{M}_n = \sum_{k=1}^n \xi_n$ is a martingale satisfying (12.1).

13. Characteristic functions and Gaussian distributions

- Definition of characteristic function
- Inversion formula
- Characteristic functions in $\mathbb{R}^d$
- Characteristic functions and independence
- Examples
- Existence of a density
- Gaussian distributions
- Conditions for independence

Characteristic functions

Definition 13.1 *Let X be a $\mathbb{R}$-valued random variable. Then the* characteristic function *for X is*

$$\varphi_X(\theta) = E[e^{i\theta X}] = \int_{\mathbb{R}} e^{i\theta x}\mu_X(dx).$$

The characteristic function is the Fourier transform *of μ_X.*

Lemma 13.2 *φ_X is uniformly continuous.*

Proof.

$$|\varphi_X(\theta + h) - \varphi_X(\theta)| \leq E[|e^{i(\theta+h)X} - e^{i\theta X}|] = E[|e^{ihX} - 1|]$$

□

Lemma 13.3 *If X and Y are independent and $Z = X + Y$, then $\varphi_Z = \varphi_X \varphi_Y$.*

Inversion formula

Theorem 13.4

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-i\theta a}-e^{-i\theta b}}{i\theta}\varphi_X(\theta)d\theta=\frac{1}{2}\mu_X\{a\}+\mu_X(a,b)+\frac{1}{2}\mu_X\{b\}$$

Proof.

$$\int_{-T}^{T} \frac{\sin \theta z}{\theta} d\theta = \mathrm{sgn}(z) 2 \int_0^{T|z|} \frac{\sin u}{u} du \equiv R(z,T)$$

$$\begin{aligned}
\int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta} \varphi_X(\theta) d\theta &= \int_{\mathbb{R}} \int_{-T}^{T} \frac{e^{i\theta(x-a)} - e^{i\theta(x-b)}}{i\theta} d\theta \mu_X(dx) \\
&= \int_{\mathbb{R}} \int_{-T}^{T} \frac{\sin \theta(x-a) - \sin \theta(x-b)}{\theta} d\theta \mu_X(dx) \\
&= \int_{\mathbb{R}} (R(x-a,T) - R(x-b,T)) \mu_X(dx)
\end{aligned}$$

The theorem follows from the fact that

$$\lim_{T \to \infty} (R(x-a,T) - R(x-b,T)) = \begin{cases} 2\pi & a < x < b \\ \pi & x \in \{a,b\} \\ 0 & \text{otherwise} \end{cases}$$

□

Characteristic functions for $\mathbb{R}^d$

Let X be a $\mathbb{R}^d$-valued random variable and for $\theta \in \mathbb{R}^d$, define

$$\varphi_X(\theta) = E[e^{i\theta\cdot X}].$$

Define

$$I_{a,b}(x) = \begin{cases} 1 & a < x < b \\ \frac{1}{2} & x \in \{a, b\} \\ 0 & \text{otherwise} \end{cases}$$

Corollary 13.5

$$\lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{[-T,T]^d} \prod_{l=1}^{d} \frac{e^{-i\theta_l a_l} - e^{-i\theta_l b_l}}{i\theta_l} \varphi_X(\theta) d\theta = \int \prod_{l=1}^{d} I_{a_l,b_l}(x_l) \mu_X(dx)$$

Independence

Lemma 13.6 $X_1, \ldots, X_d$ *are independent if and only if*

$$E[e^{i\sum_{k=1}^d \theta_k X_k}] = \prod_{k=1}^d \varphi_{X_k}(\theta_k)$$

Proof. Let $\tilde{X}_k$, $k = 1, \ldots, d$ be independent with $\mu_{\tilde{X}_k} = \mu_{X_k}$. Then The characteristic function of $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ is the same as the characteristic function of $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_d)$ so the distributions are the same. $\square$

Examples

- Poisson

$$\varphi_X(\theta) = \sum_{k=0}^{\infty} e^{i\theta k} e^{-\lambda} \frac{\lambda^k}{k!} = \exp\{\lambda(e^{i\theta} - 1)\}$$

- Normal

$$\varphi_X(\theta) = \int_{-\infty}^{\infty} e^{i\theta x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \exp\{i\theta\mu - \frac{\theta^2\sigma^2}{2}\}$$

- Uniform

$$\varphi_X(\theta) = \int_a^b e^{i\theta x} \frac{1}{b-a} dx = \frac{e^{i\theta b} - e^{i\theta a}}{i\theta(b-a)}$$

- Binomial

$$\varphi_X(\theta) = \sum_{k=0}^{n} e^{i\theta k} \binom{n}{k} p^k (1-p)^{n-k} = (pe^{i\theta} + (1-p))^n$$

- Exponential

$$\varphi_X(\theta) = \int_0^\infty e^{i\theta x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - i\theta}$$

Sufficient conditions for existence of density

Lemma 13.7 *L^1 characteristic function: If $\int |\varphi_X(\theta)| d\theta < \infty$, then X has a continuous density*

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

Proof.

$$\int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta} \varphi_X(\theta) d\theta = \int_a^b \int_{-T}^{T} e^{-i\theta x} \varphi_X(\theta) d\theta dx$$

□

Computation of moments

Lemma 13.8 *If* $m \in \mathbb{N}^+$ *and* $E[|X|^m] < \infty$, *then*

$$E[X^m] = (-i)^m \frac{d^m}{d\theta^m} \varphi_X(\theta)|_{\theta=0}.$$

Gaussian distributions

Definition 13.9 $X = (X_1, \ldots, X_d)$ *is* jointly Gaussian *if and only if* $a \cdot X = \sum_{k=1}^{d} a_k X_k$ *is Gaussian for each* $a \in \mathbb{R}^d$.

Lemma 13.10 *Let* $X = (X_1, \ldots, X_d)$*,* $X_k \in L^2$*, and define* $\mu_k = E[X_k]$*,* $\sigma_{kl} = Cov(X_k, X_l)$*, and* $\Sigma = ((\sigma_{kl}))$*. Then* X *is Gaussian if and only if*

$$\varphi_X(\theta) = \exp\{i\mu \cdot \theta - \frac{1}{2}\theta^T \Sigma \theta\}. \tag{13.1}$$

Proof. Suppose (13.1) holds, and let $Z = \sum_{k=1}^{d} a_k X_k$. Then $\varphi_Z(\theta) = \exp\{i\theta\mu \cdot a - \frac{1}{2}\theta^2 a^T \Sigma a\}$, so Z is Gaussian.

If X is Gaussian, then $\theta \cdot X$ is Gaussian with mean $\mu \cdot \theta$ and $Var(\theta \cdot X) = \theta^T \Sigma \theta$, so (13.1) follows. □

Independence of jointly Gaussian random variables

Lemma 13.11 *Let $X = (X_1, \ldots, X_d)$ be Gaussian. Then the X_k are independent if and only if $Cov(X_k, X_l) = 0$ for all $k \neq l$.*

Proof. Of course independence implies the covariances are zero. If the covariances are zero, then

$$\varphi_X(\theta) = \exp\{i\mu \cdot \theta - \frac{1}{2}\sum_{k=1}^{d} \theta_k^2 \sigma_{kk}\} = \prod_{k=1}^{d} e^{i\mu_k\theta_k - \frac{1}{2}\theta_k^2\sigma_{kk}},$$

and independence follows by Lemma 13.6. □

Linear transformations

Lemma 13.12 *Suppose that X is Gaussian in $\mathbb{R}^d$ and that A is a $m \times d$ matrix. Then $Y = AX$ is Gaussian in $\mathbb{R}^m$.*

Proof. Since $\sum_{j=1}^m b_j Y_j = \sum_{k=1}^d \sum_{j=1}^m b_j a_{jk} X_k$, and linear combination of the $\{Y_j\}$ has a Gaussian distribution. $\square$

Representation as a linear combination of independent Gaussians

Let (X_1, X_2) be Gaussian and define $Y_1 = X_1$ and $Y_2 = X_2 - \frac{Cov(X_1,X_2)}{Var(X_1)} X_1$. Then $Cov(Y_1, Y_2) = 0$ and hence Y_1 and Y_2 are independent. Note that $X_2 = Y_2 + \frac{Cov(X_1,X_2)}{Var(X_1)} Y_1$.

More generally, for $(X_1, \ldots, X_d)$ Gaussian, define $Y_1 = X_1$ and recursively, define $Y_k = X_k + \sum_{l=1}^{k-1} b_{kl} X_l$ so that

$$Cov(Y_k, X_m) = Cov(X_k, X_m) + \sum_{l=1}^{k-1} b_{kl} Cov(X_l, X_m) = 0, \quad m = 1, \ldots, k-1.$$

Then, Y_k is independent of $X_1, \ldots, X_{k-1}$ and hence of $Y_1, \ldots, Y_{k-1}$.

Conditional expectations

Lemma 13.13 *Let* $(X_1, \ldots, X_d)$ *be Gaussian. Then there exist constants* $c_0, c_1, \ldots, c_{d-1}$ *such that*

$$E[X_d | X_1, \ldots, X_{d-1}] = c_0 + \sum_{k=1}^{d-1} c_k X_k. \tag{13.2}$$

Proof. By the previous discussion, it is possible to define

$$Y_d = X_d - \sum_{k=1}^{d-1} c_k X_k$$

so that $Cov(Y_d, X_k) = 0$, $k = 1, \ldots, d-1$. Then Y_d is independent of $(X_1, \ldots, X_{d-1})$ and (13.2) holds with $c_0 = E[Y_d]$.

□

14. Convergence in distribution

- Definitions
- Separating and convergence determining sets
- First proof of the central limit theorem
- Tightness and Helly's theorem
- Convergence based on characteristic functions
- Continuous mapping theorem
- Convergence in $\mathbb{R}^d$

Convergence in distribution: Classical definition in $\mathbb{R}$

Definition 14.1 *A sequence of $\mathbb{R}$-valued random variables $\{X_n\}$ converges in distribution to a random variable X (denoted $X_n \Rightarrow X$) if and only if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at each point of continuity x of F_X.

Lemma 14.2 $F_X(x) - F_X(x-) = P\{X = x\}$*, so* $P\{X = x\} = 0$ *implies that x is a point of continuity of F_X. F_X has at most countably many discontinuities.*

Lemma 14.3 *If $F_{X_n}(x) \to F_X(x)$ for x in a dense set, then $X_n \Rightarrow X$.*

Weak convergence of measures

Definition 14.4 *Let* (E, r) *be a complete, separable metric space. A sequence of probability measures* $\{\mu_n\} \subset \mathcal{P}(E)$ *converges weakly to* $\mu \in \mathcal{P}(E)$ *(denoted* $\mu_n \Rightarrow \mu$*) if and only if*

$$\int_E g d\mu_n \to \int_E g d\mu, \quad \text{for every } g \in C_b(E).$$

In particular, $\mu_{X_n} \Rightarrow \mu_X$ *if and only if*

$$E[g(X_n)] \to E[g(X)], \quad \text{for every } g \in C_b(E).$$

We then say X_n converges in distribution *to* X*.*

Equivalence of definitions in $\mathbb{R}$

Lemma 14.5 *If $E = \mathbb{R}$, then $X_n \Rightarrow X$ if and only if $\mu_{X_n} \Rightarrow \mu_X$.*

Proof. For $\epsilon > 0$ and $z \in \mathbb{R}$, let

$$f_z'(x) = -\epsilon^{-1}\mathbf{1}_{(z,z+\epsilon)}(x), \qquad f_z(z) = 1.$$

Then

$$\mathbf{1}_{(-\infty,z]}(x) \leq f_z(x) \leq \mathbf{1}_{(-\infty,z+\epsilon]}(x).$$

Then $\mu_{X_n} \Rightarrow \mu_X$ implies

$$\limsup_{n\to\infty} F_{X_n}(z) \leq F_X(z+\epsilon) \leq \liminf_{n\to\infty} F_{X_n}(z+2\epsilon),$$

so $\limsup_{n\to\infty} F_{X_n}(z) \leq F_X(z)$ and $\liminf_{n\to\infty} F_{X_n}(z) \geq F_X(z-)$.

Conversely, if $g \in C_c^1(\mathbb{R})$,

$$\begin{aligned} E[g(X_n)] &= g(0) + \int_{[0,\infty)} \int_0^y g'(x)dx\mu_{X_n}(dy) - \int_{(-\infty,0)} \int_y^0 g'(x)dx\mu_{X_n}(dy) \\ &= g(0) + \int_{[0,\infty)} g'(x)\mu_{X_n}[x,\infty)dx - \int_{(-\infty,0)} g'(x)\mu_{X_n}(-\infty,x]dx \end{aligned}$$

□

Separating and convergence determining sets

Definition 14.6 *A collection of functions* $H \subset C_b(E)$ *is* separating *if*

$$\int_E f d\mu = \int_E f d\nu, \quad \text{for every } f \in H,$$

implies $\mu = \nu$.

A collection of functions $H \subset C_b(E)$ *is* convergence determining *if*

$$\lim_{n\to\infty} \int_E f d\mu_n = \int_E f d\mu, \quad \text{for every } f \in H,$$

implies $\mu_n \Rightarrow \mu$.

C_c is convergence determining

Lemma 14.7 *$C_c(\mathbb{R})$, the space of continuous functions with compact support, is convergence determine.*

Proof. Assume $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ for all $f \in C_c(\mathbb{R})$. Let $f_K(x) \in C_c(\mathbb{R})$ satisfy $0 \le f_K(x) \le 1$, $f_K(x) = 1$, $|x| \le K$, and $f_K(x) = 0$, $|x| \ge K+1$. Then $E[f_K(X_n)] \to E[f_K(X)]$ implies

$$\limsup_{n\to\infty} P\{|X_n| \ge K+1\} \le P\{|X| \ge K\},$$

and for $g \in C_b(\mathbb{R})$,

$$\limsup_{n\to\infty} |E[g(X_n)] - E[g(X_n)f_K(X_n)]| \le \|g\| E[1 - f_K(X)].$$

Since $\lim_{K\to\infty} E[1 - f_K(X)] = 0$, by Problem 19,

$$\lim_{n\to\infty} E[g(X_n)] = \lim_{K\to\infty} \lim_{n\to\infty} E[g(X_n)f_K(X_n)] = \lim_{K\to\infty} E[g(X)f_K(X)] = E[g(X)$$

□

C_c^∞ is convergence determining

Let

$$\rho(x) = \begin{cases} c\exp\{-\frac{1}{(x+1)(1-x)}\} & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where c is selected so that $\int_{\mathbb{R}} \rho(x)dx = 1$. Then ρ is in C_c^∞ and for $f \in C_c(\mathbb{R})$,

$$f_\epsilon(x) = \int_{\mathbb{R}} f(y)\epsilon^{-1}\rho(\epsilon^{-1}(x-y))dy,$$

$f_\epsilon \in C_c^\infty(\mathbb{R})$ and $\lim_{\epsilon\to\infty} \sup_x |f(x) - f_\epsilon(x)| = 0$.

Note that $C_u(\mathbb{R})$, the collection of uniformly continuous functions is also convergence determining since it contains $C_c(\mathbb{R})$.

The central limit theorem: First proof

Theorem 14.8 *Let* $X_1, X_2, \ldots$ *be iid with* $E[X_k] = \mu$ *and* $Var(X_k) = \sigma^2 < \infty$, *and define*

$$Z_n = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}}{\sigma}(\frac{1}{n}\sum_{k=1}^n X_k - \mu) = \frac{1}{\sqrt{n}}\sum_{k=1}^n \frac{X_k - \mu}{\sigma}.$$

Then $Z_n \Rightarrow Z$, *where*

$$P\{Z \le z\} = \Phi(z) \equiv \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Remark 14.9 *If* X *satisfies* $E[X] = \mu$ *and* $Var(X) = \sigma^2 < \infty$, *then* $Y = \frac{X-\mu}{\sigma}$ *has expectation* 0 *and variance* 1. *Note that the conversion of* X *to* Y *is essentially a change of units. (Think conversion of Fahrenheit to Celsius.)* Y *is the* standardized *version of* X. *Distributions of standardized random variables have the same location (balance point)* 0 *and the same degree of "spread" as measured by their standard deviations.*

Sums of independent Gaussian random variables are Gaussian

Lemma 14.10 *If* $X_1, X_2, \dots$ *are independent Gaussian random variables with* $E[X_k] = 0$ *and* $Var(X_k) = \sigma^2$*, then for each* $n \geq 1$*,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k$$

is Gaussian with expectation zero and variance σ^2*.*

Proof. Without loss of generality, we can assume that $E[X_k] = 0$ and $Var(X_k) = 1$. Let $\xi_1, \xi_2, \ldots$ be iid Gaussian (normal) random variables with $E[\xi_k] = 0$ and $Var(\xi_k) = 1$. For $0 \le m \le n$, define

$$Z_n^{(m)} = \sum_{k=1}^{m} X_k + \sum_{k=m+1}^{n} \xi_k, \qquad \hat{Z}_n^{(m)} = \sum_{k=1}^{m-1} X_k + \sum_{k=m+1}^{n} \xi_k$$

Then for $f \in C_c^\infty(\mathbb{R})$,

$$\begin{aligned} f(\frac{1}{\sqrt{n}}Z_n^{(n)}) - f(\frac{1}{\sqrt{n}}Z_n^{(0)}) &= \sum_{m=1}^{n}(f(\frac{1}{\sqrt{n}}Z_n^{(m)}) - f(\frac{1}{\sqrt{n}}Z_n^{(m-1)})) \\ &= \sum_{m=1}^{n}\Big(f'(\frac{1}{\sqrt{n}}\hat{Z}_n^{(m)})\frac{1}{\sqrt{n}}(X_m - \xi_m) + \frac{1}{2}f''(\frac{1}{\sqrt{n}}\hat{Z}_n^{(m)})\frac{1}{n}(X_m^2 - \xi_m^2) \\ &\qquad + R(\frac{1}{\sqrt{n}}\hat{Z}_n^{(m)}, \frac{1}{\sqrt{n}}X_m) - R(\frac{1}{\sqrt{n}}\hat{Z}_n^{(m)}, \frac{1}{\sqrt{n}}\xi_m)\Big), \end{aligned}$$

where

$$R(z,h) = \int_z^{z+h}\int_z^y (f''(u) - f''(z))dudy$$

and

$$|R(z,h)| \leq \frac{1}{2}h^2 \sup_{|u-v|\leq h} |f''(u) - f''(v)|.$$

Consequently,

$$\begin{aligned}
&|E[f(\frac{1}{\sqrt{n}}Z_n^{(n)})] - E[f(\frac{1}{\sqrt{n}}Z_n^{(0)})]| \\
&\quad \leq E[X_1^2 \sup_{|u-v|\leq \frac{1}{\sqrt{n}}|X_1|} |f''(u) - f''(v)|] \\
&\qquad + E[\xi_1^2 \sup_{|u-v|\leq \frac{1}{\sqrt{n}}|\xi_1|} |f''(u) - f''(v)|] \to 0
\end{aligned}$$

by the dominated convergence theorem. □

Helly's theorem

Theorem 14.11 *Let* $\{X_n\}$ *be a sequence of* $\mathbb{R}$*-valued random variables Suppose that for each* $\epsilon > 0$*, there exists a* $K_\epsilon > 0$ *such that*

$$\sup_n F_{X_n}(-K_\epsilon)+1-F_{X_n}(K_\epsilon) = \sup_n(P\{X_n \leq -K_\epsilon\}+P\{X_n > K_\epsilon\}) < \epsilon.$$

Then there exists a subsequence $\{n_m\}$ *and a random variable* X *such that* $X_{n_m} \Rightarrow X$*.*

Proof. Select a subsequence of $\{F_{X_n}\}$ such that $F_{X_{n_m}}(y)$ converges for each rational y. Call the limit $F^0(y)$ and define

$$F_X(x) = \inf_{y\in\mathbb{Q},y>x} F^0(y) \geq \sup_{y\in\mathbb{Q},y<x} F^0(y)$$

Then F_X is a cdf, and by monotonicity, $F_{X_{n_m}}(x) \to F_X(x)$ for each continuity point x. $\square$

Tightness

Definition 14.12 *A sequence of random variables* $\{X_n\}$ *is* tight, *if for each* $\epsilon > 0$ *there exists* $K_\epsilon > 0$ *such that* $P\{|X_n| > K_\epsilon\} \leq \epsilon$.

If $\{X_n\}$ is tight, then Helly's states that there exists a subsequence that converges in distribution. Note that the original sequence converges if there is only one possible limit distribution.

Lemma 14.13 *Suppose* $\psi \geq 0$ *and* $\lim_{r\to\infty} \psi(r) = \infty$. *If* $\sup_n E[\psi(X_n)] < \infty$, *then* $\{X_n\}$ *is tight.*

Lévy's convergence theorem

Theorem 14.14 *If* $\lim \varphi_{X_n}(\theta) = g(\theta)$ *for every* θ *and* g *is continuous at* 0*, then* g *is the characteristic function for a random variable* X *and* $X_n \Rightarrow X$.

Proof. Assume tightness. Then convergence follows from the inversion formula.

Proof of tightness:

$$\begin{aligned}
\delta^{-1}\int_0^\delta (2-\varphi_{X_n}(\theta)-\varphi_{X_n}(-\theta))\,d\theta &= \delta^{-1}\int_{-\delta}^{\delta}\int_{\mathbb{R}}(1-e^{i\theta x})\mu_{X_n}(dx)d\theta \\
&= \int_{\mathbb{R}}\delta^{-1}\int_{-\delta}^{\delta}(1-e^{i\theta x})d\theta\mu_{X_n}(dx) \\
&= \int_{\mathbb{R}}2(1-\tfrac{\sin\delta x}{\delta x})\mu(dx) \\
&\geq \mu_{X_n}\{x:|x|>2\delta^{-1}\}
\end{aligned}$$

□

The central limit theorem: Second proof

Proof. Let Z_n be as before. Then assuming $E[X_k] = 0$ and $Var(X_k) = 1$,

$$\begin{aligned}\varphi_{Z_n}(\theta) &= E[e^{i\theta\frac{1}{\sqrt{n}}X}]^n = \varphi_X(\frac{1}{\sqrt{n}}\theta)^n \\ &= (E[e^{i\frac{1}{\sqrt{n}}\theta X} - 1 - i\frac{\theta}{\sqrt{n}}X + \frac{\theta^2}{2n}X^2] + 1 - \frac{\theta^2}{2n})^n.\end{aligned}$$

Claim:

$$\begin{aligned}&\lim_{n\to\infty} nE[e^{i\frac{1}{\sqrt{n}}\theta X} - 1 - i\frac{\theta}{\sqrt{n}}X + \frac{\theta^2}{2n}X^2] \\ &\quad = -\lim_{n\to\infty} E[X^2\int_0^\theta\int_0^v (e^{i\frac{1}{\sqrt{n}}uX} - 1)dudv = 0,\end{aligned}$$

so $\varphi_{Z_n}(\theta) \to e^{-\frac{1}{2}\theta^2}$ and Z_n converges in distribution to a *standard normal* random variable. □

Triangular arrays

Definition 14.15 *A collection of random variables* $\{X_{nk}, 1 \leq k \leq N_n, n = 1, 2, \ldots\}$ *is refered to as a* triangular array. *The triangular array is a* null array *(or* uniformly asymptotically negligible*) if*

$$\lim_{n\to\infty} \sup_k E[|X_{nk}| \wedge 1] = 0$$

Lindeberg conditions

Theorem 14.16 *Let* $\{X_{nk}, 1 \leq k \leq N_n, n = 1, 2, \ldots\}$ *be a triangular array of independent, mean zero random variables, and let* Z *be standard normal. Suppose that* $\lim_{n\to\infty} \sum_k E[X_{nk}^2] \to 1$*. Then*

$$\sum_k X_{nk} \Rightarrow Z \text{ and } \sup_k E[X_{nk}^2] \to 0$$

if and only if for each $\epsilon > 0$*,*

$$\lim_{n\to\infty} \sum_k E[X_{nk}^2 \mathbf{1}_{\{|X_{nk}|>\epsilon\}}] = 0.$$

Proof. Let $\{\xi_{nk}\}$ be independent Gaussian (normal) random variables with $E[\xi_{nk}] = 0$ and $Var(\xi_{nk}) = Var(X_{nk})$. For $0 \leq m \leq m_n$, define

$$Z_n^{(m)} = \sum_{k=1}^{m} X_{nk} + \sum_{k=m+1}^{m_n} \xi_{nk}, \qquad \hat{Z}_n^{(m)} = \sum_{k=1}^{m-1} X_{nk} + \sum_{k=m+1}^{m_n} \xi_{nk}$$

Then for $f \in C_c^\infty(\mathbb{R})$,

$$f(Z_n^{(n)}) - f(Z_n^{(0)}) = \sum_{m=1}^{m_n} (f(Z_n^{(m)}) - f(Z_n^{(m-1)}))$$

$$= \sum_{m=1}^{m_n} \Big(f'(\hat{Z}_n^{(m)})(X_{nm} - \xi_{nm}) + \frac{1}{2} f''(\hat{Z}_n^{(m)})(X_{nm}^2 - \xi_{nm}^2)$$

$$+ R(\hat{Z}_n^{(m)}, X_{nm}) - R(\hat{Z}_n^{(m)}, \xi_{nm}) \Big),$$

Consequently,

$$
\begin{aligned}
|E[f(Z_n^{(n)})] - E[f(Z_n^{(0)})]| \\
&\leq \sum_{m=1}^{m_n} E[X_{nm}^2 \sup_{|u-v|\leq|X_{nm}|} |f''(u) - f''(v)|] \\
&\qquad + \sum_{m=1}^{m_n} E[\xi_{nm}^2 \sup_{|u-v|\leq|\xi_{nm}|} |f''(u) - f''(v)|] \\
&\leq 2\|f''\| \sum_{m=1}^{m_n} E[X_{nm}^2 \mathbf{1}_{\{|X_{nm}|>\epsilon\}}] \\
&\qquad + \sum_{m=1}^{m_n} E[X_{nm}^2] \sup_{|u-v|\leq\epsilon} |f''(u) - f''(v)| \\
&\qquad + \sum_{m=1}^{m_n} E[\xi_{nm}^2 \sup_{|u-v|\leq|\xi_{nm}|} |f''(u) - f''(v)|] \to 0.
\end{aligned}
$$

For the converse, see Theorem 5.15 in Kallenberg. □

Types of convergence

Consider

a) $X_n \to X$ almost surely

b) $X_n \to X$ in probability

c) $X_n \Rightarrow X$ (X_n converges to X is distribution)

Lemma 14.17 *$X_n \to X$ in probability if and only if $E[|X_n - X| \wedge 1] \to 0$.*

Lemma 14.18 *Almost sure convergence implies convergence in probability. Convergence in probability implies convergence in distribution.*

Proof. If $X_n \to X$ almost surely, the $E[|X_n - X| \wedge 1] \to 0$ by the bounded convergence theorem.

If $g \in C_c^\infty(\mathbb{R})$, then $|g(x)-g(y)| \le (\|g'\||x-y|)\wedge(2\|g\|)$. Consequently, if $X_n \to X$ in probability,

$$|E[g(X_n)] - E[g(X)]| \le (2\|g\|) \vee \|g'\| E[|X_n - X| \wedge 1] \to 0.$$

□

Skorohod representation theorem

Theorem 14.19 *If $X_n \Rightarrow X$, then there exists a probability space and random variable $\tilde{X}_n$, $\tilde{X}$ such that $\mu_{\tilde{X}_n} = \mu_{X_n}$, $\mu_{\tilde{X}} = \mu_X$, and $\tilde{X}_n \to \tilde{X}$ almost surely.*

Proof. Define

$$G_n(y) = \inf\{x : P\{X_n \leq x\} \geq y\}, \quad G(y) = \inf\{x : P\{X \leq x\} \geq y\}.$$

Let ξ by uniform $[0,1]$. Then $G(\xi) \leq x$ if and only if $P\{X \leq x\} \geq \xi$, so

$$P\{G(\xi) \leq x\} = P\{P\{X \leq x\} \geq \xi\} = P\{X \leq x\}.$$

$F_{X_n}(x) \to F_X(x)$ for all but countably many x implies $G_n(y) \to G(y)$ for all but countably many y. □

Continuous mapping theorem

Theorem 14.20 *Let* $H : \mathbb{R} \to \mathbb{R}$*, and let* $C_H = \{x : H \text{ is continuous at } x\}$*. If* $X_n \Rightarrow X$ *and* $P\{X \in C_H\} = 1$*, then* $H(X_n) \Rightarrow H(X)$*.*

Proof. The result follows immediately from the Skorohod representation theorem. □

Convergence in distribution in $\mathbb{R}^d$

Definition 14.21 $\{\mu_n\} \subset \mathcal{P}(\mathbb{R}^d)$ *is* tight *if and only if for each* $\epsilon > 0$ *there exists a* $K_\epsilon > 0$ *such that* $\sup_n \mu_n(B_{K_\epsilon}(0)^c) \leq \epsilon$.

Definition 14.22 $\{X^n\}$ *in* $\mathbb{R}^d$ *is* tight *if and only if for each* $\epsilon > 0$ *there exists a* $K_\epsilon > 0$ *such that* $\sup_n P\{|X_n| > K_\epsilon\} \leq \epsilon$.

Lemma 14.23 *Let* $X^n = (X_1^n, \ldots, X_d^n)$. *Then* $\{X_n\}$ *is tight if and only if* $\{X_k^n\}$ *is tight for each* k.

Proof. Note that

$$P\{|X^n| \geq K\} \leq \sum_{k=1}^{d} P\{|X_k^n| \geq d^{-1}K\}$$

□

Tightness implies relative compactness

Lemma 14.24 *If $\{X_n\} \subset \mathbb{R}^d$ is tight, then there exists a subsequence $\{n_k\}$ and a random variable X such that $X^{n_k} \Rightarrow X$.*

Proof. Since $\{X_n\}$ is tight, for $\epsilon > 0$, there exists a $K > 0$ such that

$$\begin{aligned} |\varphi_{X^n}(\theta_1) - \varphi_{X^n}(\theta_2)| &= K|\theta_1 - \theta_2|P\{|X| \leq K\} + 2P\{|X| > K\} \\ &\leq K|\theta_1 - \theta_2| + 2\epsilon, \end{aligned}$$

which implies that $\{\varphi_{X^n}(\theta)\}$ is uniformly equicontinuous. Selecting a subsequence along which $\varphi_{X^n}(\theta)$ converges for every θ with rational components, and by the equicontinuity, for every θ. Equicontinuity also implies that the limit is continuous, so the limit is the characteristic function of a probability distribution. □

Convergence determining sets in $\mathbb{R}^d$

Lemma 14.25 *$C_c(\mathbb{R}^d)$ is convergence determining.*

Proof. Assume $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ for all $f \in C_c(\mathbb{R}^d)$. Let $f_K(x) \in C_c(\mathbb{R}^d)$ satisfy $0 \le f_K(x) \le 1$, $f_K(x) = 1$, $|x| \le K$, and $f_K(x) = 0$, $|x| \ge K+1$. Then $E[f_K(X_n)] \to E[f_K(X)]$ implies

$$\limsup_{n\to\infty} P\{|X_n| \ge K+1\} \le P\{|X| \ge K\},$$

and for $g \in C_b(\mathbb{R}^d)$,

$$\limsup_{n\to\infty} |E[g(X_n)] - E[g(X_n)f_K(X_n)]| \le \|g\| E[1 - f_K(X)].$$

Since $\lim_{K\to\infty} E[1 - f_K(X)] = 0$, by Problem 19,

$$\lim_{n\to\infty} E[g(X_n)] = \lim_{K\to\infty}\lim_{n\to\infty} E[g(X_n)f_K(X_n)] = \lim_{K\to\infty} E[g(X)f_K(X)] = E[g(X)$$

$\square$

Convergence by approximation

Lemma 14.26 *Suppose that for each* $\epsilon > 0$*, there exists* $X^{n,\epsilon}$ *such that*

$$P\{|X^n - X^{n,\epsilon}| > \epsilon\} \leq \epsilon,$$

and that $X^{n,\epsilon} \Rightarrow X^\epsilon$*. Then there exist* X *such that* $X^\epsilon \stackrel{\epsilon\to 0}{\Rightarrow} X$ *and* $X^n \Rightarrow X$*.*

Proof. Since $|e^{i\theta\cdot x} - e^{i\theta\cdot y}| \leq |\theta||x-y|$,

$$|\varphi_{X^n}(\theta) - \varphi_{X^{n,\epsilon}}(\theta)| \leq |\theta|\epsilon(1-\epsilon) + 2\epsilon.$$

By Problem 19,

$$\lim_{n\to\infty} \varphi_{X^n}(\theta) = \lim_{\epsilon\to\infty} \varphi_{X^{n,\epsilon}}(\theta)$$

□

Convergence in distribution of independent random variables

Lemma 14.27 *For each* n*, suppose that* $\{X_1^n, \ldots, X_d^n\}$ *are independent, and assume that* $X_k^n \Rightarrow X_k$*,* $k = 1, \ldots, d$*. Then* $(X_1^n, \ldots, X_d^n) \Rightarrow (X_1, \ldots, X_d)$*, where* $X_1, \ldots, X_d$ *are independent.*

Continuous mapping theorem

Theorem 14.28 *Suppose* $\{X_n\}$ *in* $\mathbb{R}^d$ *satisfies* $X_n \Rightarrow X$ *and* $F : \mathbb{R}^d \to \mathbb{R}^m$ *is continuous. Then* $F(X_n) \Rightarrow F(X)$.

Proof. Let $g \in C_b(\mathbb{R}^m)$. Then $g \circ F \in C_b(\mathbb{R}^d)$ and $E[g(F(X_n))] \to E[g(F(X))]$. $\square$

Convergence in $\mathbb{R}^\infty$

$\mathbb{R}^\infty$ is a metric space with metric

$$d(x,y) = \sum_k 2^{-k} |x^{(k)} - y^{(k)}| \wedge 1$$

Note that $x_n \to x$ in $\mathbb{R}^\infty$ if and only if $x_n^{(k)} \to x^{(k)}$ for each k.

Lemma 14.29 *$X_n \Rightarrow X$ in $\mathbb{R}^\infty$ if and only if $(X_n^{(1)}, \ldots, X_n^{(d)}) \Rightarrow (X^{(1)}, \ldots, X^{(d)})$ for each d.*

15. Poisson convergence and Poisson processes

- Poisson approximation of the binomial distribution
- The Chen-Stein method
- Poisson processes
- Marked Poisson processes
- Poisson random measures
- Compound Poisson distributions

Poisson approximation of the binomial distribution

Theorem 15.1 *Let S_n be binomially distributed with parameters n and p_n, and suppose that $\lim_{n\to\infty} np_n = \lambda$. Then $\{S_n\}$ converges in distribution to a Poisson random variable with parameter λ.*

Proof. Check that

$$\lim_{n\to\infty} P\{S_n = k\} = \lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

or note that

$$E[e^{i\theta S_n}] = ((1-p_n) + p_n e^{i\theta})^n \to e^{\lambda(e^{i\theta}-1)}.$$

□

A characterization of the Poisson distribution

Lemma 15.2 *A nonnegative, integer-valued random variable Z is Poisson distributed with parameter λ if and only if*

$$E[\lambda g(Z+1) - Zg(Z)] = 0 \tag{15.1}$$

for all bounded g.

Proof. Let $g_k(j) = \delta_{jk}$. Then (15.1) implies

$$\lambda P\{Z = k-1\} - kP\{Z = k\} = 0$$

and hence

$$P\{Z = k\} = \frac{\lambda^k}{k!} P\{Z = 0\}.$$

□

The Chen-Stein equation

Let Z_λ denote a Poisson distributed random variable with $E[Z_\lambda] = \lambda$.

Lemma 15.3 *Let h be bounded and $E[h(Z_\lambda)] = 0$. Then there exists a bounded function g such that*

$$\lambda g(k+1) - kg(k) = h(k), \quad k \in \mathbb{N}.$$

Proof. Let $g(0) = 0$ and define recursively

$$g(k+1) = \frac{1}{\lambda}(h(k) + kg(k)).$$

$\gamma(k) = \frac{\lambda^k g(k)}{(k-1)!}$. Then

$$\gamma(k+1) = \gamma(k) + \frac{\lambda^k h(k)}{k!} = \sum_{l=0}^{k} \frac{\lambda^l}{l!} h(l) = -\sum_{l=k+1}^{\infty} \frac{\lambda^l}{l!} h(l)$$

and

$$g(k+1) = \frac{k!}{\lambda^{k+1}} \sum_{l=0}^{k} \frac{\lambda^l}{l!} h(l) = -\frac{k!}{\lambda^{k+1}} \sum_{l=k+1}^{\infty} \frac{\lambda^l}{l!} h(l),$$

and hence, for $k+2 > \lambda$

$$|g(k+1)| \leq \|h\| \sum_{j=0}^{\infty} \frac{\lambda^j}{(k+1+j)!/k!} \leq \frac{\|h\|}{(k+1)} \frac{k+2}{k+2-\lambda}$$

□

Poisson error estimates

Lemma 15.4 *Let W be a nonnegative, integer-valued random variable. Then*

$$P\{W \in A\} - P\{Z_\lambda \in A\} = E[\lambda g_{\lambda,A}(W+1) - W g_{\lambda,A}(W)]$$

where $g_{\lambda,A}$ is the solution of

$$\lambda g_{\lambda,A}(k+1) - k g_{\lambda,A}(k) = \mathbf{1}_A(k) - P\{Z_\lambda \in A\}$$

Sums of independent indicators

Let $\{X_i\}$ be independent with $P\{X_i = 1\} = 1 - P\{X_i = 0\} = p_i$, and define $W = \sum_i$, and $W_k = W - X_k$. Then

$$E[Wg(W)] = \sum_i E[X_i g(W_i + 1)] = \sum_i p_i E[g(W_i + 1)]$$

and setting $\lambda = \sum_i p_i$,

$$\begin{aligned} E[\lambda g(W+1) - Wg(W)] &= \sum_i p_i E[g(W+1) - g(W_i+1)] \\ &= \sum_i p_i E[X_i(g(W_i+2) - g(W_i+1)] \\ &= \sum_i p_i^2 E[(g(W_i+2) - g(W_i+1)] \end{aligned}$$

and hence

$$\begin{aligned} &|P\{W \in A\} - P\{Z_\lambda \in A\}| \\ &\quad \leq \sum_i p_i^2 \max(\sup_{k\geq 1}(g_{\lambda,A}(k+1) - g_{\lambda,A}(k)), \sup_{k\geq 1}(g_{\lambda,A^c}(k+1) - g_{\lambda,A^c}(k))). \end{aligned}$$

Estimate on g

Lemma 15.5 *For every $A \subset \mathbb{N}$,*

$$\sup_{k \geq 1}(g_{\lambda,A}(k+1) - g_{\lambda,A}(k)) \leq \frac{1 - e^{-\lambda}}{\lambda}$$

Proof. See Lemma 1.1.1 of [1]. □

Bernoulli processes

For each $n = 1, 2, \ldots$, let $\{\xi_k^n\}$ be a sequence of Bernoulli trials with $P\{\xi_k^n = 1\} = p_n$, and assume that $np_n \to \lambda$. Define

$$\begin{aligned} N_n(t) &= \textstyle\sum_{k=1}^{[nt]} \xi_k^n \\ \tau_l^n &= \inf\{t : N_n(t) = l\} \ . \\ \gamma_l^n &= \tau_l^n - \tau_{l-1}^n. \end{aligned}$$

Lemma 15.6 *For $t_0 = 0 < t_1 < \cdots < t_m$, then $N_n(t_k) - N_n(t_{k-1})$, $k = 1, \ldots, m$, are independent and converge in distribution to independent Poisson random variables.*

Interarrival times

Lemma 15.7 $\{\gamma_l^n\}$ *are independent and identically distributed.*

Proof. To simplify notation, let $n = 1$. Define $\mathcal{F}_k = \sigma(\xi_i, i \leq k)$. Compute

$$\begin{aligned} P\{\gamma_{l+1} > m|\mathcal{F}_{\tau_l}\} &= \sum_k E[\mathbf{1}_{\{\xi_{\tau_l+1}=\cdots=\xi_{\tau_l+m}=0\}}|\mathcal{F}_k]\mathbf{1}_{\{\tau_l=k\}} \\ &= \sum_k E[\mathbf{1}_{\{\xi_{k+1}=\cdots=\xi_{k+m}=0\}}|\mathcal{F}_k]\mathbf{1}_{\{\tau_l=k\}} \\ &= (1-p_1)^m \end{aligned}$$

□

Convergence of the interarrival times

Lemma 15.8 $(\gamma_1^n, \gamma_2^n, \ldots) \Rightarrow (\gamma_1, \gamma_2, \ldots)$ *where the* γ_k *are independent exponentials.*

Proof. By Lemmas 14.27 and 14.29, it is enough to show the convergence of $\{\gamma_k^n\}$ for each k. Note that

$$P\{\gamma_k^n > s\} = P\{n\gamma_k^n > [ns]\} = (1 - p_n)^{[ns]} \to e^{-\lambda s}.$$

□

Continuous time stochastic processes

Definition 15.9 *A family of σ-algebras $\{\mathcal{F}_t\} = \{\mathcal{F}_t, t \geq 0\}$ is a* filtration*, if $s < t$ implies $\mathcal{F}_s \subset \mathcal{F}_t$.*

A stochastic process $X = \{X(t), t \geq 0\}$ is adapted *to $\{\mathcal{F}_t\}$ if $X(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$.*

A nonnegative random variable τ is a $\{\mathcal{F}_t\}$-stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \geq 0$.

*A stochastic process X is a $\{\mathcal{F}_t\}$-*martingale (submartingale, supermartingale)*) if X is $\{\mathcal{F}_t\}$-adapted and*

$$E[X(s)|\mathcal{F}_t] = (\geq, \leq) X(t), \quad \forall t < s.$$

A stochastic process is cadlag *(continue à droite limite à gauche), if for each (or almost every) $\omega \in \Omega$, $t \to X(t, \omega)$ is right continuous and has a left limit at each $t > 0$.*

The Poisson process

The convergence in distribution of the increments and interarrival times suggest convergence (in some sense) of the Bernoulli process to a process N with independent, Poisson distributed increments. Convergence of the interarrival times suggests defining

$$N(t) = \max\{l : \sum_{k=1}^{l} \gamma_k \leq t\},$$

so that N is a cadlag, piecewise constant process.

Defined this way, the *Poisson process* is an example of a *renewal process*.

Setting $\mathcal{F}_t^N = \sigma(N(s) : s \leq t)$, the jump times

$$\tau_l = \inf\{t : N(t) \geq l\}$$

are $\{\mathcal{F}_t^N\}$-stopping times.

Relationship between N and $\{\tau_l\}$

Note that

$$P\{\tau_l > t\} = P\{N(t) < l\} = \sum_{k=0}^{l-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

and differentiating

$$f_{\tau_l}(t) = \frac{\lambda^l t^{l-1}}{(l-1)!} e^{-\lambda t}.$$

Martingale properties

If N is a Poisson process with parameter λ, then

$$M(t) = N(t) - \lambda t$$

is a martingale.

Theorem 15.10 *(Watanabe) If N is a counting process and*

$$M(t) = N(t) - \lambda t$$

is a martingale, N is a Poisson process with parameter λ.

Proof.

$$
\begin{aligned}
&E[e^{i\theta(N(t+r)-N(t))}|\mathcal{F}_t^N] \\
&\quad = 1+\sum_{k=0}^{n-1} E[(e^{i\theta(N(s_{k+1})-N(s_k)}-1-(e^{i\theta}-1)(N(s_{k+1})-N(s_k))e^{i\theta(N(s_k)-N(t))}|\mathcal{F}_t^N] \\
&\qquad +\sum_{k=0}^{n-1}\lambda(s_{k+1}-s_k)(e^{i\theta}-1)E[e^{i\theta(N(s_k)-N(t))}|\mathcal{F}_t^N]
\end{aligned}
$$

The first term converges to zero by the dominated convergence theorem, so we have

$$
E[e^{i\theta(N(t+r)-N(t))}|\mathcal{F}_t^N]=1+\lambda(e^{i\theta}-1)\int_0^r E[e^{i\theta(N(t+s)-N(t))}|\mathcal{F}_t^N]ds
$$

and $E[e^{i\theta(N(t+r)-N(t))}|\mathcal{F}_t^N]=e^{\lambda(e^{i\theta}-1)t}$. Since $\{e^{i\theta x}:\theta\in\mathbb{R}\}$ is separating, $N(t+r)-N(t)$ is independent of $\mathcal{F}_t^N$. □

Thinning Poisson processes

Theorem 15.11 *Let N be a Poisson process with parameter λ, and let $\{\xi_k\}$ be a Bernoulli sequence with $P\{\xi_k = 1\} = 1 - P\{\xi_k = 0\} = p$. Define*

$$N_1(t) = \sum_{k=1}^{N(t)} \xi_k, \quad N_2(t) = \sum_{k=1}^{N(t)} (1 - \xi_k).$$

Then N_1 and N_2 are independent Poisson processes with parameter λp and $\lambda(1-p)$ respectively.

Proof. Consider

$$\begin{aligned} E[e^{i(\theta_1 N_1(t)+\theta_2 N_2(t))}] &= E[\exp\{i\sum_{k=1}^{N(t)}((\theta_1-\theta_2)\xi_k+\theta_2)\}] \\ &= E[(e^{i\theta_1}p+e^{i\theta_2}(1-p))^{N(t)}] \\ &= \exp\{\lambda((e^{i\theta_1}p+e^{i\theta_2}(1-p))-1) \\ &= e^{\lambda p(e^{i\theta_1}-1)}e^{\lambda(1-p)(e^{i\theta_2}-1)}. \end{aligned}$$

By a similar argument, for $0=t_0<\cdots<t_m$, $N_1(t_k)-N_1(t_{k-1})$, $N_2(t_k)-N_2(t_{k-1})$, $k=1,\ldots,m$ are independent Poisson distributed.

$$\sigma(N_i)=\sigma(N_i(s),s\geq 0)=\vee_n\sigma(N_i(2^{-n}),N_i(2\times 2^{-n}),N_i(3\times 2^{-n}\ldots),$$

so independence of $\sigma(N_1(2^{-n}),N_1(2\times 2^{-n}),N_1(3\times 2^{-n}\ldots)$ and $\sigma(N_2(2^{-n}),N_2(2\times 2^{-n}),N_2(3\times 2^{-n}\ldots)$ implies independence of $\sigma(N_1)$ and $\sigma(N_2)$. □

Sums of independent Poisson processes

Lemma 15.12 *If* N_k, $k = 1, 2, \ldots$ *are independent Poisson processes with parameters* λ_k *satisfying* $\lambda = \sum_k \lambda_k < \infty$, *then* $N = \sum_k N_k$ *is a Poisson process with parameter* λ.

Marked Poisson processes

Let N be a Poisson process with parameter λ, and let $\{\eta_k\}$ be independent and identically distributed $\mathbb{R}^d$-valued random variable.

Assign η_k to the kth arrival time for N. η_k is sometimes refered to as the *mark* associated with the kth arrival time.

Note that for $A \in \mathcal{B}(\mathbb{R}^d)$

$$N(A,t) = \#\{k : \tau_k \leq t, \eta_k \in A\} = \sum_{k=1}^{N(t)} \mathbf{1}_A(\eta_k)$$

is a Poisson process with parameter $\lambda\mu_\eta(A)$, and that for disjoint $A_1, A_2, \ldots$, $N(A_i, \cdot)$ are independent.

Space-time Poisson random measures

Theorem 15.13 *Let ν be a σ-finite measure on $\mathbb{R}^d$. Then there exists a stochastic process $\{N(A,t) : A \in \mathcal{B}(\mathbb{R}^d), t \geq 0\}$, such that for each A satisfying $\nu(A) < \infty$, $N(A,\cdot)$ is a Poisson process with parameter $\nu(A)$ and for $A_1, A_2, \ldots$ disjoint with $\nu(A_i) < \infty$, $N(A_i \cdot)$ are independent.*

Proof. Let $\{D_m\}$ be disjoint with $\cup_m D_m = \mathbb{R}^d$ and $\nu(D_m) < \infty$, let $\{N_m\}$ be independent Poisson processes with parameters $\nu(D_m)$, and let $\{\eta_k^m\}$ be independent random variables with

$$P\{\eta_k^m \in A\} = \frac{\nu(A \cap D_m)}{\nu(D_m)}.$$

Then for each $A \in \mathcal{B}(\mathbb{R}^d)$ with $\nu(A) < \infty$, set

$$N(A,t) = \sum_m \sum_{k=1}^{N_m(t)} \mathbf{1}_A(\eta_k^m) = \sum_m \sum_{k=1}^{N_m(t)} \mathbf{1}_{A\cap D_m}(\eta_k^m).$$

Note that $\sum_{k=1}^{N_m(t)} \mathbf{1}_{A\cap D_m}(\eta_k^m)$ is a Poisson process with parameter $\nu(D_m) \times \frac{\nu(A\cap D_m)}{\nu(D_m)} = \nu(A \cap D_m)$. $\square$

Poisson approximation to multinomial

Theorem 15.14 *Let* $\{\eta_k^n\}$ *be independent with values in* $\{0,1,\ldots,m\}$*, and let* $p_{kl}^n = P\{\eta_{kl}^n = l\}$*. Suppose that* $\sup_k P\{\eta_{kl}^n > 0\} \to 0$ *and* $\sum_k P\{\eta_{kl}^n = l\} \to \lambda_l$ *for* $l > 0$*. Define* $N_l^n = \#\{k : \eta_{kl}^n = l\}$*,* $l = 1,\ldots,m$*. Then* $(N_1^n,\ldots,N_m^n) \Rightarrow (N_1,\ldots,N_m)$*, where* $\{N_l\}$ *are independent Poisson distributed random variables with* $E[N_l] = \lambda_l$*.*

Compound Poisson distributions

Let ν be a finite measure on $\mathbb{R}$ and let N be the Poisson random measure satisfying $N(A)$ Poisson distributed with parameter $\nu(A)$. Then writing

$$N = \sum_{k=1}^{N(\mathbb{R})} \delta_{X_k}$$

where $N(\mathbb{R})$ is Poisson distributed with parameter $\nu(\mathbb{R})$ and the $\{X_k\}$ are independent with distribution $\mu(A) = \frac{\nu(A)}{\nu(\mathbb{R})}$,

$$Y = \int_{\mathbb{R}} xN(dx) = \sum_{k=1}^{N(\mathbb{R})} X_k$$

has distribution satisfying

$$\varphi_Y(\theta) = E[\varphi_X(\theta)^{N(\mathbb{R})}] = e^{\int_{\mathbb{R}}(e^{i\theta x}-1)\nu(dx)} \tag{15.2}$$

16. Infinitely divisible distributions

- Other conditions for normal convergence
- More general limits
- Infinitely divisible distributions
- Stable distributions

Conditions for normal convergence

Theorem 16.1 *Let $\{\xi_{nk}\}$ be a null array (uniformly asymptotically negligible). Then $Z_n = \sum_k \xi_{nk}$ converges in distribution to Gaussian random variable Z with $E[Z\} = \mu$ and $Var(Z) = \sigma^2$ if and only if the following conditions hold:*

a) For each $\epsilon > 0$, $\sum_k P\{|\xi_{nk}| > \epsilon\} \to 0$.

b) $\sum_k E[\xi_{nk}\mathbf{1}_{\{|\xi_{nk}|\leq 1\}}] \to \mu$.

c) $\sum_k Var(\xi_{nk}\mathbf{1}_{\{|\xi_{nk}|\leq 1\}}) \to \sigma^2$.

Proof. Let $\hat{Z}_n = \sum_k \xi_{nk} \mathbf{1}_{\{|\xi_{nk}| \leq 1\}}$. Then

$$P\{Z_n \neq \hat{Z}_n\} \leq \sum_k P\{|\xi_{nk}| > 1\} \to 0,$$

so it is enough to show $\hat{Z}_n \Rightarrow Z$. Let

$$\zeta_{nk} = \xi_{nk} \mathbf{1}_{\{|\xi_{nk}| \leq 1\}} - E[\xi_{nk} \mathbf{1}_{\{|\xi_{nk}| \leq 1\}}].$$

Then noting that $\eta_n = \max_k |E[\xi_{nk} \mathbf{1}_{\{|\xi_{nk}| \leq 1\}}]| \to 0$,

$$\sum_k E[\zeta_{nk}^2 \mathbf{1}_{\{|\zeta_{nk}| > \epsilon\}}] \leq (1 + \eta_n)^2 \sum_k P\{|\xi_{nk}| > \epsilon - \eta_n\} \to 0,$$

Theorem 14.16 implies $\hat{Z}_n \Rightarrow Z$. □

The iid case

Theorem 16.2 *Let* ξ_k *be iid and let* $a_n \to \infty$. *Define* $\mu_n = E[\xi \mathbf{1}_{\{|\xi| \leq a_n\}}]$. *Suppose that for each* $\epsilon > 0$,

$$\lim_{n\to\infty} nP\{|\xi| > a_n \epsilon\} = 0$$

and that

$$\lim_{n\to\infty} \frac{n}{a_n^2}(E[\xi^2 \mathbf{1}_{\{|\xi|\leq a_n\}}] - \mu_n^2) = \sigma^2.$$

Then

$$\frac{\sum_{k=1}^n \xi_k - n\mu_n}{a_n} \Rightarrow Z$$

where Z *is normal with* $E[Z] = 0$ *and* $Var(Z) = \sigma^2$.

An example of normal convergence with infinite variance

Example 16.3 *Let*

$$f_\xi(x) = \begin{cases} 2x^{-3} & x \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\frac{n}{a_n^2} E[\xi^2 \mathbf{1}_{\{|\xi| \leq a_n\}}] = 2\frac{n}{a_n^2} \log n \to 2, \quad \mu_n = 2(1 - a_n^{-1}) \to 2,$$

and taking $a_n = \sqrt{n \log n}$*,*

$$nP\{|\xi| > a_n \epsilon\} = n \frac{1}{(a_n \epsilon)^2} \to 0.$$

Consequently, for Z *normal with mean zero and variance 2,*

$$\frac{\sum_{k=1}^n \xi_k - 2n(1 - a_n^{-1})}{a_n} \Rightarrow Z.$$

More general limits

Let $\{\xi_{nk}\}$ be a null array, but suppose that Condition (a) of Theorem 16.1 fails. In particular, suppose

$$\lim_{n\to\infty}\sum_k P\{\xi_{nk} > z\} = H_+(z), \quad \lim_{n\to\infty}\sum_k P\{\xi_{nk} \le -z\} = H_-(z)$$

for all but countably many $z > 0$. (Let D be the exceptional set.)

For $a_i, b_i \notin D$ and $0 < a_i < b_i$ or $a_i < b_i < 0$, define $N_n(a_i, b_i] = \#\{k : \xi_{nk} \in (a_i, b_i]\}$. Then

$$(N_n(a_1, b_1], N_n(a_2, b_2], \ldots) \Rightarrow (N(a_1, b_1], N(a_2, b_2], \ldots) \tag{16.1}$$

where $N(a, b]$ is Poisson distributed with expectation $H_+(a) - H_+(b)$ if $0 < a < b$ and expectation $H_-(b) - H_-(a)$ if $a < b < 0$ and $N(a_1, b_1], \ldots, N(a_m, b_m]$ are independent if $(a_1, b_1], \ldots, (a_m, b_m]$ are disjoint. (See Theorem 15.14.)

Compound Poisson part

Lemma 16.4 *Assume that for all but countably many $z > 0$,*

$$\lim_{n\to\infty}\sum_k P\{\xi_{nk} > z\} = H_+(z), \quad \lim_{n\to\infty}\sum_k P\{\xi_{nk} \leq -z\} = H_-(z), \tag{16.2}$$

and let C_H be the collection of z such that H_+ and H_- are continuous at z. Let ν be the measure on $\mathbb{R}$ satisfying $\nu\{0\} = 0$ and $\nu(z,\infty) = H_+(z)$ and $\nu(-\infty,-z) = H_-(z)$ for all $z \in C_H$. Then for each $\epsilon > 0$, $\epsilon \in C_H$,

$$Y_n^\epsilon = \sum_k \xi_{nk}\mathbf{1}_{\{|\xi_{nk}|>\epsilon\}} \Rightarrow Y^\epsilon \tag{16.3}$$

where Y^ϵ is compound Poisson with distribution determined by ν restricted to $(-\infty,-\epsilon)| \cup (\epsilon,\infty)$, that is

$$\varphi_{Y^\epsilon}(\theta) = e^{\int_{[-\epsilon,\epsilon]^c}(e^{i\theta x}-1)\nu(dx)} \tag{16.4}$$

Proof. Let
$$N_n^\epsilon(a,b] = \#\{k : \xi_{nk} \in (a,b] \cap [-\epsilon,\epsilon]^c\}.$$
Then
$$\sum a_j N_n^\epsilon(a_j,a_{j+1}] \leq \sum_k \xi_{nk}\mathbf{1}_{\{|\xi_{nk}|>\epsilon\}} \leq \sum a_{j+1} N_n^\epsilon(a_j,a_{j+1}].$$
Assuming that $a_j \in C_H$, by (16.1)
$$E[e^{i\theta \sum_j a_j N_n^\epsilon(a_j,a_{j+1}]}] \to \prod \varphi_{N^\epsilon(a_j,a_{j+1}]}(a_j\theta) = e^{\sum_j \nu_\epsilon(a_j,a_{j+1}](e^{i\theta a_j}-1)}.$$
Taking a limit as $\max(a_{j+1}-a_j) \to 0$ gives the rightside of (16.4). The convergence in (16.3) follows by Problem 24. □

Gaussian part

Lemma 16.5 *Suppose*

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\sum_k Var(\xi_{nk}\mathbf{1}_{\{|\xi_{nk}|\leq\epsilon\}})=\sigma^2, \tag{16.5}$$

then there exist $\epsilon_n\to 0$ *such that*

$$\sum_k \xi_{nk}\mathbf{1}_{\{|\xi_{nk}|\leq\epsilon_n\}}-\sum_k E[\xi_{nk}\mathbf{1}_{\{|\xi_{nk}|\leq\epsilon_n\}}]\Rightarrow Z,$$

where Z *is normal with* $E[Z]=0$ *and* $Var(Z)=\sigma^2$*.*

General limit theorem

Theorem 16.6 *Suppose that* $\{\xi_{nk}\}$ *is a null array satisfying (16.2) and (16.5). Then for* $\tau \in C_H$,

$$Z_n = \sum_k (\xi_{nk} - E[\xi_{nk} \mathbf{1}_{\{|\xi_{nk}| \leq \tau\}}])$$

converges in distribution to a random variable Z *with*

$$\varphi_Z(\theta) = \exp\{-\frac{\sigma^2}{2}\theta^2 + \int_{\mathbb{R}} (e^{i\theta z} - 1 - \mathbf{1}_{[-\tau,\tau]}(z) z i\theta)\nu(dz)\}$$

Proof. Let

$$\begin{aligned} Z_n &= \sum_k \xi_{nk}\mathbf{1}_{\{|\xi_{nk}|\le\epsilon\}} - \sum_k E[\xi_{nk}\mathbf{1}_{\{|\xi_{nk}|\le\epsilon\}}] \\ &\qquad + \sum_k (\xi_{nk}\mathbf{1}_{\{|\xi_{nk}|>\epsilon\}} - E[\xi_{nk}\mathbf{1}_{\{\epsilon<|\xi_{nk}|\le\tau\}}]) \\ &= Z_n^\epsilon + Y_n^\epsilon - A_n^\epsilon \end{aligned}$$

Then

$$A_n^\epsilon = \sum_k E[\xi_{nk}\mathbf{1}_{\{\epsilon<|\xi_{nk}|\le\tau\}}] \to \int_{[-\tau,-\epsilon)\cup(\epsilon,\tau]} z\nu(dz)$$

and

$$E[e^{i\theta(Y_n^\epsilon - A_n^\epsilon)}] \to \exp\{\int_{[-\epsilon,\epsilon]^c} (e^{i\theta z} - 1 - \mathbf{1}_{[-\tau,\tau]}(z)zi\theta)\nu(dz)\}.$$

In addition,

$$\begin{aligned} &|E[e^{i\theta Z_n}] - E[e^{i\theta Z_n^\epsilon}]E[e^{i\theta(Y_n^\epsilon - A_n^\epsilon)}]| \\ &\qquad \le \end{aligned}$$

Infinitely divisible distributions

Lemma 16.7 *Let* $\sigma^2, \tau > 0$, $a \in \mathbb{R}$, *and* ν *a measure on* $\mathbb{R} - \{0\}$ *satisfying* $\int_{\mathbb{R}} |z|^2 \wedge 1 \nu(dz) < \infty$. *Then*

$$\varphi_Z(\theta) = \exp\{\frac{\sigma^2}{2}\theta^2 + ia\theta + \int_{\mathbb{R}} (e^{i\theta z} - 1 - i\theta z \mathbf{1}_{[-\tau,\tau]}(z))\nu(dz)\} \quad (16.6)$$

is the characteristic function of a random variable satisfying

$$Z = \sigma Z_0 + a + \int_{[-\tau,\tau]} z\tilde{\xi}(dz) + \int_{[-\tau,\tau]^c} z\xi(dz)$$

where Z_0 *is standard normal,* ξ *is a Poisson random measure with* $E[\xi(A)] = \nu(A)$ *independent of* Z_0, *and* $\tilde{\xi} = \xi - \nu$. *The first integral is defined by*

$$\int_{[-\tau,\tau]} z\tilde{\xi}(dz) = \lim_{n\to\infty} (\int_{[-\tau,-\epsilon_n]} z\tilde{\xi}(dz) + \int_{[\delta_n,\tau]} z\tilde{\xi}(dz)) \quad (16.7)$$

for any sequences ϵ_n, δ_n *that decrease to zero.*

Proof. Note that

$$M_n^+ = \int_{[\delta_n,\tau]} z\tilde{\xi}(dz)), \quad M_n^- = \int_{[-\tau,-\epsilon_n]} z\tilde{\xi}(dz)$$

are martingales satisfying

$$E[(M_n^+)^2] = \int_{[\delta_n,\tau]} z^2\nu(dz)), \quad E[(M_n^-)^2] = \int_{[-\tau,-\epsilon_n]} z^2\nu(dz)),$$

and the limit in (16.7) exists by the martingale convergence theorem. The form of the characteristic function then follows by (15.2). □

Property of infinite divisibility

Lemma 16.8 *Let φ_Z be given by (16.6), and define*

$$\varphi_{Z_n}(\theta) = \exp\{\frac{\sigma^2}{n2}\theta^2 + i\frac{a}{n}\theta + \frac{1}{n}\int_{\mathbb{R}}(e^{i\theta z} - 1 - i\theta z\mathbf{1}_{[-\tau,\tau]}(z))\nu(dz)\}. \quad (16.8)$$

Then (16.8) defines a characteristic functions and if $Z_n^{(k)}$ are iid with that distribution, then

$$\sum_{k=1}^{n} Z_n^{(k)}$$

has the same distribution as Z.

Regular variation

Lemma 16.9 *Let* U *be a positive, monotone function on* $(0,\infty)$*. Suppose that*

$$\lim_{t\to\infty}\frac{U(tx)}{U(t)} = \psi(x) \le \infty$$

for x *in a dense set of points* D*. Then*

$$\psi(x) = x^\rho,$$

for some $-\infty \le \rho \le \infty$*.*

Proof. Since

$$\frac{U(tx_1x_2)}{U(t)} = \frac{U(tx_1x_2)}{U(tx_1)}\frac{U(tx_1)}{U(t)},$$

if $\psi(x_1)$ and $\psi(x_2)$ are finite and positive, then so is

$$\psi(x_1x_2) = \psi(x_1)\psi(x_2). \tag{16.9}$$

If $\psi(x_1) = \infty$, then $\psi(x_1^n) = \infty$ and $\psi(x_1^{-n}) = 0$ for all $n = 1, 2, \ldots$. By monotonicity, either $\psi(x) = x^\infty$ or $x^{-\infty}$. If $0 < \psi(x) < \infty$, for some $x \in D$, then by monotonicity $0 < \psi(x) < \infty$ for all $x \in D$. Extending ψ to a right continuous function, (16.9) holds for all x_1, x_2. Setting $\gamma(y) = \log\psi(e^y)$, we have $\gamma(y_1 + y_2) = \gamma(y_1) + \gamma(y_2)$, monotonicity implies $\gamma(y) = \rho y$ for some ρ, and hence, $\psi(x) = x^\rho$ for some ρ. $\square$

Renormalized sums of iid random variables

Let $X_1, X_2, \ldots$ be iid with cdf F, and consider

$$Z_n = \frac{1}{a_n} \sum_{k=1}^{n} (X_k - b_n),$$

where $0 < a_n \to \infty$ and $b_n \in \mathbb{R}$. Setting

$$\xi_{nk} = \frac{X_k - b_n}{a_n},$$

$$\sum_k P\{\xi_{nk} > z\} = n(1 - F(a_n z + b_n))$$

and

$$\sum_k P\{\xi_{nk} \leq -z\} = nF(-a_n z + b_n)$$

A convergence lemma

Lemma 16.10 *Suppose that* F *is a cdf and that for a dense set* D *of* $z > 0$,

$$\lim_{n\to\infty} n(1-F(a_n z+b_n)) = V^+(z) \geq 0, \qquad \lim_{n\to\infty} nF(-a_n z+b_n)) = V^-(z) \geq 0$$

where $V^+(z), V^-(z) < \infty$, $\lim_{z\to\infty} V^+(z) = \lim_{z\to\infty} V^-(z) = 0$, *and there exists* $\epsilon > 0$ *such that for*

$$\mu_n^\epsilon = \int_{-a_n\epsilon+b_n}^{a_n\epsilon+b_n} (z-b_n)dF(z),$$

$$\limsup_{n\to\infty} na_n^{-2} \int_{-a_n\epsilon+b_n}^{a_n\epsilon+b_n} (z-b_n-\mu_n^\epsilon)^2 dF(z) < \infty.$$

Then $\lim_{n\to\infty} a_n^{-1} b_n = 0$, *and if* $V^+(z) > 0$ *for some* $z > 0$, $V^+(z) = \lambda^+ z^{-\alpha}$, $0 < \alpha < 2$, *and similarly for* V^-.

Proof. For $z > 0$, we must have

$$\lim_{n\to\infty} a_n z + b_n = \infty, \qquad \lim_{n\to\infty} -a_n z + b_n = -\infty$$

which implies $\limsup |a_n^{-1} b_n| \leq z$. Since z can be arbitrarily small, $\lim_{n\to\infty} a_n^{-1} b_n = 0$.

If $V^+(z) > 0$, then there exists $\hat{z} > 0$ such that $V^+(\hat{z} - \delta) > V^+(\hat{z} + \delta)$ for all $\delta > 0$. For each $z \in D$, $z < \hat{z}$, we must have $\limsup \frac{a_{n+1}}{a_n} z < \hat{z} + \delta$, $\delta > 0$. Consequently, $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 1$. Let $N(t) = n$, if $a_n \leq t < a_{n+1}$. Then

$$\lim_{t\to\infty} \frac{1 - F(xt)}{1 - F(\hat{x}t)} = \lim_{t\to\infty} \frac{1 - F(a_{N(t)}(x\frac{t}{a_{N(t)}} - \frac{b_{N(t)}}{a_{N(t)}}) + b_{N(t)})}{1 - F(a_{N(t)}(\hat{x}\frac{t}{a_{N(t)}} - \frac{b_{N(t)}}{a_{N(t)}}) + b_{N(t)})} = \frac{V^+(x)}{V^+(\hat{x})}$$

for each $x, \hat{x}$ that are points of continuity of the right continuous extension of V^+. It follows that $V^+(x) = \alpha^{-1}\lambda^+ x^{-\alpha}$ for some $-\alpha = \rho < 0$.

To see the $\alpha < 2$, assume for simplicity that F is symmetric so that $b_n = \mu_n^\epsilon = 0$ and $V^+ = V^-$. Then by Fatou's lemma,

$$\begin{aligned}
\limsup_{n\to\infty} n a_n^{-2} \int_{-a_n\epsilon}^{a_n\epsilon} z^2 dF(z) &= \limsup 4n \int_0^\epsilon u(F(a_n\epsilon) - F(a_n u))du \\
&\geq 4 \int_0^\epsilon u(V^+(u) - V^+(\epsilon))du \\
&= 4 \int_0^\epsilon u\lambda^+(u^{-\alpha} - \epsilon^{-\alpha})du,
\end{aligned}$$

and we must have $\alpha < 2$. □

Stable distributions

Let

$$\varphi_Z(\theta) = \exp\{ia\theta + \int_0^\infty (e^{i\theta z} - 1 - i\theta z \mathbf{1}_{[-\tau,\tau]}(z)) \frac{\lambda^+}{z^{\alpha+1}} dz + \int_{-\infty}^0 (e^{i\theta z} - 1 - i\theta z \mathbf{1}_{[-\tau,\tau]}(z)) \frac{\lambda^-}{|z|^{\alpha+1}} dz\}$$

Then Z is *stable* in the sense that if Z_1 and Z_2 are independent copies of Z, then there exist a, b such that

$$\hat{Z} = \frac{Z_1 + Z_2 - b}{c}$$

has the same distribution as Z.

$$
\begin{aligned}
\varphi_{\hat{Z}}(\theta) &= \exp\{i\frac{b}{c}\theta + ia\theta + 2\int_0^\infty (e^{i\theta c^{-1}z} - 1 - i\theta c^{-1} z\mathbf{1}_{[-\tau,\tau]}(z))\frac{\lambda^+}{z^{\alpha+1}}dz \\
&\qquad + 2\int_{-\infty}^0 (e^{i\theta c^{-1}z} - 1 - i\theta c^{-1} z\mathbf{1}_{[-\tau,\tau]}(z))\frac{\lambda^-}{|z|^{\alpha+1}}dz\} \\
&= \exp\{i\frac{b}{c}\theta + ia\theta + 2\int_0^\infty (e^{i\theta z} - 1 - i\theta z\mathbf{1}_{[-\tau,\tau]}(cz))\frac{\lambda^+}{c^{\alpha+1}z^{\alpha+1}}cdz \\
&\qquad + 2\int_{-\infty}^0 (e^{i\theta z} - 1 - i\theta z\mathbf{1}_{[-\tau,\tau]}(cz))\frac{\lambda^-}{c^{\alpha+1}|z|^{\alpha+1}}cdz\}
\end{aligned}
$$

so $c^\alpha = 2$ $(c = 2^{\frac{1}{\alpha}})$ and

$$
\begin{aligned}
\frac{b}{c} &= \int_0^\infty z(\mathbf{1}_{[-\tau,\tau]}(cz)) - \mathbf{1}_{[-\tau,\tau]}(z))\frac{\lambda^+}{z^{\alpha+1}}dz \\
&\qquad + \int_{-\infty}^0 z(\mathbf{1}_{[-\tau,\tau]}(cz) - \mathbf{1}_{[-\tau,\tau]}(z))\frac{\lambda^-}{|z|^{\alpha+1}}dz
\end{aligned}
$$

17. Martingale central limit theorem

- A convergence lemma
- Martingale central limit theorem
- Martingales associated with Markov chains
- Central limit theorem for Markov chains

A convergence lemma

The proof of the martingale central limit theorem given here follows Sethuraman [4].

Lemma 17.1 *Suppose*

1. $U_n \to a$ *in probability*
2. $\{T_n\}$ *and* $\{|T_nU_n|\}$ *are uniformly integrable*
3. $E[T_n] \to 1$

Then $E[T_nU_n] \to a$.

Proof. The sum of uniformly integrable random variables is uniformly integrable and $T_n(U_n - a) \to 0$ in probability, so
$E[T_nU_n] = E[T_n(U_n - a)] + E[aT_n] \to a$. □

Martingale central limit theorem

Definition 17.2 $\{\xi_k\}$ *is a martingale difference array with respect to* $\{\mathcal{F}_k\}$ *if* $\{\xi_k\}$ *is* $\{\mathcal{F}_k\}$ *adapted and* $E[\xi_{k+1}|\mathcal{F}_k] = 0$ *for each* $k = 0, 1, \ldots$.

Theorem 17.3 *For each* n *let* $\{\mathcal{F}_k^n\}$ *be a filtration and* $\{\xi_k^n\}$ *be a martingale difference array with respect to* $\{\mathcal{F}_k^n\}$*, that is,* $X_k^n = \sum_{j=1}^k \xi_j^n$ *is an* $\{\mathcal{F}_k^n\}$*-martingale. Suppose that* $E[\max_j |\xi_j^n|] \to 0$ *and* $\sum_j (\xi_j^n)^2 \to \sigma^2$ *in probability. Then*

$$Z^n = \sum_j \xi_j^n \Rightarrow Z$$

where Z *is* $N(0, \sigma^2)$.

Proof. Assume that $\sigma^2 = 1$. Let $\eta_1^n = \xi_1^n$ and $\eta_j^n = \xi_j^n \mathbf{1}_{\{\sum_{1\le i<j}(\xi_i^n)^2\le 2\}}$. Then $\{\eta_j^n\}$ is also a martingale difference array, and $P\{\sum_j \eta_j^n \neq \sum_j \xi_j^n\} \to 0$.

Since

$$\log(1+ix) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(ix)^k = ix + \frac{x^2}{2} + \sum_{k=3}^{\infty} \frac{(-1)^{k-1}}{k}(ix)^k,$$

setting

$$r(x) = \sum_{l=2}^{\infty} \frac{(-1)^l}{2l} x^{2l} - i \sum_{l=1}^{\infty} \frac{(-1)^l}{2l+1} x^{2l+1},$$

$$\exp\{ix\} = (1+ix)\exp\{-\frac{x^2}{2} + r(x)\}$$

where $|r(x)| \le C|x|^3$ for $|x| \le .5$.

Let $T_n = \prod_j (1+i\theta\eta_j^n)$ and $U_n = \exp\{-\frac{\theta^2}{2}\sum_j(\eta_j^n)^2 + \sum_j r(\theta\eta_j^n)\}$ Clearly, $\{T_nU_n\}$ is uniformly integrable, $E[T_n] = 1$, and $U_n \to e^{-\theta^2/2}$. We also claim that $\{T_n\}$ is uniformly integrable.

$$|T_n| = \sqrt{\prod_j (1+\theta^2(\eta_j^n)^2)} \leq \sqrt{e^{2\theta^2}(1+\theta^2 \max_j |\xi_j^n|^2)}.$$

Consequently, by Lemma 17.1,

$$E[e^{i\theta\sum_j \eta_j^n}] = E[T_nU_n] \to e^{-\frac{\theta^2}{2}}.$$

□

Markov chains

Let

$$X_{k+1} = F(X_k, Z_{k+1}, \beta_0)$$

where the $\{Z_k\}$ are iid and X_0 is independent of the $\{Z_k\}$

Lemma 17.4 $\{X_k\}$ *is a Markov chain.*

Martingales associated with Markov chains

Let μ_Z be the distribution of Z_k and define

$$H(x,\beta) = \int F(x,z,\beta)\mu_Z(dz) \tag{17.1}$$

Then

$$M_n = \sum_{k=1}^{n} X_k - H(X_{k-1},\beta_0)$$

is a martingale. Define

$$P_\beta f(x) = \int f(F(x,z,\beta))\mu_Z(dz)$$

Then

$$M_n^f = \sum_{k=1}^{n} f(X_k) - P_{\beta_0} f(X_{k-1})$$

is a martingale and by Lemma 12.19, $\lim_{n\to\infty} \frac{1}{n} M_n^f = 0 \quad a.s.$

Stationary distributions

π is a stationary distribution for a Markov chain if $\mu_{X_0} = \pi$ implies $\mu_{X_k} = \pi$ for all $k = 1, 2, \ldots$

Lemma 17.5 *π is a stationary distribution for the Markov chain if and only if*

$$\int f d\pi = \int P_{\beta_0} f d\pi$$

Ergodicity for Markov chains

Definition 17.6 *A Markov chain is* ergodic *if and only if there is a unique stationary distribution for the chain.*

If $\{X_k\}$ is ergodic and $\mu_{X_0} = \pi$, then

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to \int f d\pi \quad a.s. \text{ and in } L^1$$

for all f satisfying $\int |f| d\pi < \infty$. (This will be proved next semester.)

Let

$$Q_\beta h(y) = \int h(F(y,z,\beta),y)\mu_Z(dz)$$

Then

$$\tilde{M}_n^h = \sum_{k=1}^n h(X_k, X_{k-1}) - Q_{\beta_0} h(X_{k-1})$$

is a martingale. If the chain is ergodic and $\mu_{X_0} = \pi$, then for h satisfying $\int Q_{\beta_0}|h|(x)\pi(dx) < \infty$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n h(X_k, X_{k-1}) = \int Q_{\beta_0} h(x)\pi(dx) \quad a.s.$$

Central limit theorem for Markov chains

Theorem 17.7 *Let* $\{X_k\}$ *be a stationary, ergodic Markov chain. Then for* f *satisfying* $\int f^2 d\pi < \infty$

$$\frac{1}{\sqrt{n}} M_n^f \Rightarrow Y^f$$

where Y^f *is normal with mean zero and variance* $\int f^2 d\pi - \int (P_{\beta_0} f)^2 d\pi$.

Proof.

$$\frac{1}{n} \sum_{k=1}^{n} (f(X_k) - P_{\beta_0} f(X_{k-1}))^2 \to \int f^2 d\pi - \int (P_{\beta_0} f)^2 d\pi,$$

and the theorem follows. □

A parameter estimation problem

Recalling the definition of H (17.1),

$$E[\sum_{k=1}^{n} X_k - H(X_{k-1}, \beta_0)] = 0$$

and

$$\sum_{k=1}^{n} X_k - H(X_{k-1}, \beta) = 0$$

is an *unbiased estimating equation* for β_0. A solution $\hat{\beta}_n$ is called a *martingale estimator* for β_0.

Asymptotic normality

$$\begin{aligned} M_n &= \sum_{k=1}^{n} H(X_{k-1},\hat{\beta}_n) - H(X_{k-1},\beta_0) \\ &= \sum_{k=1}^{n} H'(X_{k-1},\beta_0)(\hat{\beta}_n-\beta_0) + \sum_{k=1}^{n} \frac{1}{2} H''(X_{k-1},\tilde{\beta}_n)(\hat{\beta}_n-\beta_0)^2 \end{aligned}$$

and

$$\frac{1}{\sqrt{n}} M_n = \left(\frac{1}{n} \sum_{k=1}^{n} H'(X_{k-1},\beta_0) \right) \sqrt{n}(\hat{\beta}_n-\beta_0) + \frac{1}{n^{3/2}} (\cdot) \left(\sqrt{n}(\hat{\beta}_n-\beta_0)\right)^2$$

Therefore, assuming $\int x^2 \pi(dx) < \infty$ and $\int H'(x,\beta_0)\pi(dx) \neq 0$,

$$\sqrt{n}(\hat{\beta}_n-\beta_0) \Rightarrow \frac{Y}{\int H'(x,\beta_0)\pi(dx)}$$

Y normal with $E[Y]=0$ and $Var(Y) = \int x^2\pi(dx) - \int H(x,\beta_0)^2\pi(dx)$.

18. Brownian motion

- Random variables in $C[0,1]$
- Convergence in distribution in $C[0,1]$
- Construction of Brownian motion by Donsker invariance
- Markov property
- Transition density and heat semigroup
- Strong Markov property
- Sample path properties
- Lévy characterization

The space $C[0,1]$

Define

$$d(x,y) = \sup_{s\leq 1} |x(s) - y(s)| \wedge 1$$

Lemma 18.1 $(C[0,1], d)$ *is a complete, separable metric space.*

Proof. If $\{x_n\}$ is Cauchy, there exists a subsequence such that $d(x_{n_k}, x_{n_{k+1}}) \leq 2^{-k}$. Defining $x(t) = \lim_{n\to\infty} x_n(t)$,

$$|x(t) - x_{n_k}(t)| \leq 2^{-k+1}$$

and hence $x \in C[0,1]$ and $\lim_{n\to\infty} d(x_n, x) = 0$.

To check separability, for $x \in C[0,1]$, let x_n be the linear interpolation of the points $(\frac{k}{n}, \frac{\lfloor x(\frac{k}{n})n \rfloor}{n})$, so

$$x_n(t) = \frac{\lfloor x(\frac{k}{n})n \rfloor}{n} + n(t - \frac{k}{n})\frac{\lfloor x(\frac{k+1}{n})n \rfloor - \lfloor x(\frac{k}{n})n \rfloor}{n}, \quad \frac{k}{n} \le t \le \frac{k+1}{n}.$$

Then

$$|x_n(t) - x(t)| \le |x(\frac{\lfloor nt \rfloor}{n}) - x(t)| + |x(\frac{\lfloor nt \rfloor}{n}) - \frac{\lfloor x(\frac{\lfloor nt \rfloor}{n})n \rfloor}{n}| + \frac{|\lfloor x(\frac{k+1}{n})n \rfloor - \lfloor x(\frac{k}{n})n \rfloor}{n}$$

and $\lim_{n\to\infty} \sup_{0 \le t \le 1} |x_n(t) - x(t)| = 0.$ □

Borel subsets of $C[0,1]$

Let $\pi_t x = x(t)$, $0 \leq t \leq 1$. and define $\mathcal{S} = \sigma(\pi_t, 0 \leq t \leq 1)$, that is, the smallest σ-algebra such that all the mappings $\pi_t : C[0,1] \to \mathbb{R}$ are measurable.

Lemma 18.2 $\mathcal{S} = \mathcal{B}(C[0,1])$.

Proof. Since π_t is continuous, $\mathcal{S} \subset \mathcal{B}(C[0,1])$. Since for $0 < \epsilon < 1$,

$$\bar{B}_\epsilon(y) = \{x | d(x,y) \leq \epsilon\} = \cap_{t \in \mathbb{Q} \cap [0,1]} \{x : |x(t) - y(t)| \leq \epsilon\} \in \mathcal{S},$$

and since each open set is a countable union of balls, $\mathcal{B}(C[0,1]) \subset \mathcal{S}$.
□

A convergence determining set

Lemma 18.3 *Let (S,d) be a complete, separable metric space, and let $C_u(S)$ denote the space of bounded, uniformly continuous functions on S. Then $C_u(S)$ is convergence determining.*

Proof. For $g \in \bar{C}(S)$, define

$$g_l(x) = \inf_y (g(y) + ld(x,y)), \quad g^l(x) = \sup_y (g(y) - ld(x,y))$$

and note that $g_l(x) \leq g(x) \leq g^l(x)$ and

$$\lim_{l\to\infty} g_l(x) = \lim_{l\to\infty} g^l(x) = g(x).$$

Then

$$g_l(x_1) - g_l(x_2) \geq \inf_y l(d(x_1,y) - d(x_2 y) \geq -ld(x_1,x_2),$$

and it follows that $|g_l(x_1) - g_l(x_2)| \leq ld(x_1,x_2)$, so $g_l \in C_u(S)$. Similarly, $g^l \in C_u(S)$.

Suppose $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ for each $f \in C_u(S)$. Then for each l,

$$\begin{aligned} E[g_l(X)] = \lim_{n\to\infty} E[g_l(X_n)] &\leq \liminf_{n\to\infty} E[g(X_n)] \leq \limsup_{n\to\infty} E[g(X_n)] \\ &\leq \lim_{n\to\infty} E[g^l(X_n)] = E[g^l(X)]. \end{aligned}$$

But $\lim_{l\to\infty} E[g_l(X)] = \lim_{l\to\infty} E[g^l(X)] = E[g(X)]$, so

$$\lim_{n\to\infty} E[g(X_n)] = E[g(X)].$$

□

Tightness of probability measures

Lemma 18.4 *Let (S,d) be a complete, separable metric space. If $\mu \in \mathcal{P}(S)$, then for each $\epsilon > 0$ there exists a compact $K_\epsilon \subset S$ such that $\mu(K_\epsilon) \geq 1-\epsilon$.*

Proof. Let $\{x_i\}$ be dense in S, and let $\epsilon > 0$. Then for each k, there exists N_k such that

$$\mu(\cup_{i=1}^{N_k} B_{2^{-k}}(x_i)) \geq 1 - \epsilon 2^{-k}.$$

Setting $G_{k,\epsilon} = \cup_{i=1}^{N_k} B_{2^{-k}}(x_i)$, define K_ϵ to be the closure of $\cap_{k\geq 1} G_{k,\epsilon}$. Then

$$\mu(K_\epsilon) \geq 1 - \mu(\cup_k G_{k,\epsilon}^c) \geq 1 - \epsilon.$$

□

Prohorov's theorem

Theorem 18.5 $\{\mu_{X_\alpha}, \alpha \in \mathcal{A}\} \subset \mathcal{P}(S)$ *is relatively compact in the weak topology if and only if for each* $\epsilon > 0$*, there exists a compact* $K_\epsilon \subset S$ *such that*

$$\inf_{\alpha \in \mathcal{A}} P\{X_\alpha \in K_\epsilon\} \geq 1 - \epsilon. \text{ tightness}$$

Corollary 18.6 *Suppose that for each* k, $\{X_\alpha^k\}$ *is relatively compact in convergence in distribution in* (S_k, d_k)*. Then* $\{(X_\alpha^1, X_\alpha^2, \cdots)\}$ *is relatively compact in* $(\prod S_k, d)$*,*

$$d(x,y) = \sum_k 2^{-k} d_k(x_k, y_k) \wedge 1.$$

Convergence based on approximation

Lemma 18.7 *Let* $\{X_n\}$ *be a sequence of* S*-valued random variables. Suppose that for each* $\epsilon > 0$*, there exists* $\{X_n^\epsilon\}$ *such that* $E[d(X_n, X_n^\epsilon) \wedge 1] \leq \epsilon$ *and* $X_n^\epsilon \Rightarrow X^\epsilon$*. Then* $\{X^\epsilon\}$ *converges in distribution to a random varible* X *as* $\epsilon \to 0$ *and* $X_n \Rightarrow X$*.*

Proof. Let $X_n^k = X_n^{2^{-k}}$. Then $\{(X_n^1, X_n^2, \ldots)\}$ is relatively compact in S^∞ and any limit point $(X^1, X^2, \ldots)$ will satisfy $E[d(X^l, X^{l+1}) \wedge 1] \leq 2^{-l} + 2^{l+1}$. Consequently,

$$X = X^1 + \sum_{l=1}^{\infty} (X^{l+1} - X^l)$$

exists.

Let $g \in C_u(S)$, and let $w(\delta) = \sup_{d(x,y)\leq\delta} |g(x) - g(y)|$. Then for $0 < \epsilon < 1$,

$$\begin{aligned} |E[g(X_n^\epsilon)] - E[g(X_n)]| &\leq E[w(d(X_n, X_n^\epsilon))] \\ &\leq w(\sqrt{\epsilon}) + 2\|g\|_\infty \sqrt{\epsilon}. \end{aligned}$$

It follows that

$$\lim_{n\to\infty} E[g(X_n)] = \lim_{\epsilon\to 0} \lim_{n\to\infty} E[g(X_n^\epsilon)] = E[g(X)].$$

□

Convergence in distribution in $C[0,1]$

Let $P_k x$ be the linear interpolation of $(x(0), x(2^{-k}), \ldots, x((2^k-1)2^{-k}), x(1))$, that is,

$$P_k x(t) = x(l2^{-k}) + 2^k(t - l2^{-k})(x((l+1)2^{-k}) - x(l2^{-k})), \quad l2^{-k} \leq t \leq (l+1)2^{-k}.$$

Theorem 18.8 *Let $\{X_n\}$ be $C[0,1]$-valued random variables. Then $X_n \Rightarrow X$ if and only if $(X_n(t_1), \ldots, X_n(t_m)) \Rightarrow (X(t_1), \ldots, X(t_m))$, for all $t_1, \ldots, t_m \in [0,1]$ (the finite dimensional distributions converge), and*

$$\lim_{k\to\infty} \sup_n E[d(X_n, P_k X_n)] = 0.$$

Proof. The theorem is an immediate consequence of Lemma 18.7. $\square$

Kolmogorov criterion

Lemma 18.9 *Suppose that X takes values in $C[0,1]$, and there exist $C,\beta>0$ and $\theta>1$ such that*

$$E[|X(t)-X(s)|^\beta\wedge 1]\leq C|t-s|^\theta,\quad 0\leq t,s\leq 1.$$

Then

$$E[d(X,P_kX)]\leq 2C^{1/\beta}\frac{2^{-k\frac{\theta-1}{\beta}}}{1-2^{-\frac{\theta-1}{\beta}}}$$

Proof. If $l2^{-k}\leq t\leq (l+1)2^{-k}$, then

$$|X(t)-X(l2^{-k})|\leq\sum_{m=k}^\infty |X(2^{-(m+1)}\lfloor t2^{m+1}\rfloor)-X(2^{-m}\lfloor t2^m\rfloor)|$$

and

$$|X(t)-X((l+1)2^{-k})|\leq\sum_{m=k}^\infty |X(2^{-(m+1)}\lceil t2^{m+1}\rceil)-X(2^{-m}\lceil t2^m\rceil)|,$$

and

$$|X(t) - P_kX(t)| \leq |X(t) - X(l2^{-k})| + |X(t) - X((l+1)2^{-k})|.$$

Let

$$\eta_m = \sum_{l<2^m} |X((l+1)2^{-m}) - X(l2^{-m})|^\beta \wedge 1.$$

Then

$$|X(t) - P_kX(t)| \wedge 1 \leq 2\sum_{m=k}^{\infty} \eta_m^{1/\beta},$$

and hence

$$E[d(X, P_kX)] \leq 2\sum_{m=k}^{\infty} E[\eta_m]^{1/\beta} \leq 2\sum_{m=k}^{\infty} (2^m C 2^{-m\theta})^{1/\beta} = 2C^{1/\beta}\sum_{m=k}^{\infty} 2^{-m\frac{\theta-1}{\beta}}.$$

□

Construction of Brownian motion by Donsker invariance

$\xi_1, \xi_2, \ldots$ iid $E[\xi] = 0$, $Var(\xi) = 1$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \xi_i + \sqrt{n}(t - \frac{\lfloor nt \rfloor}{n})\xi_{\lfloor nt \rfloor + 1}.$$

Then $X_n \Rightarrow W$, standard Browian motion.

W is continuous

W has independent increments

$E[W(t)] = 0$, $\quad Var(W(t)) = t$, $\quad Cov(W(t), W(s)) = t \wedge s$

W is a martingale.

Proof. For simplicity, assume that $E[\xi_k^4] < \infty$. Then, assuming $t - s > n^{-1}$,

$$
\begin{aligned}
&E[(X_n(t) - X_n(s))^4] \\
&\quad = E[(\sqrt{n}(\frac{\lceil ns \rceil}{n} - s)\xi_{\lfloor ns \rfloor + 1} + \frac{1}{\sqrt{n}} \sum_{k=\lfloor ns \rfloor + 2}^{\lfloor nt \rfloor} \xi_k + \sqrt{n}(t - \frac{\lfloor nt \rfloor}{n})\xi_{\lfloor nt \rfloor + 1})^4] \\
&\quad \le C_1((\frac{\lfloor nt \rfloor - \lfloor ns \rfloor + 1}{n})^2 + \frac{1}{n}\frac{\lfloor nt \rfloor - \lfloor ns \rfloor + 1}{n}) \\
&\quad \le C_2|t - s|^2.
\end{aligned}
$$

For $0 < t - s \le n^{-1}$,

$$E[(X_n(t) - X_n(s))^4] \le C(t - s)^4 n^2 \le C(t - s)^2.$$

□

Markov property

$X(t) = X(0) + W(t)$, $X(0)$ independent of W.

$$T(t)f(x) \equiv E[f(x + W(t))] = \int_{-\infty}^{\infty} f(y)\frac{1}{\sqrt{2\pi t}}e^{-\frac{(y-x)^2}{2t}}dy$$

$$E[f(X(t+s))|\mathcal{F}_t^X] = E[f(X(t)+W(t+s)-W(t))|\mathcal{F}_t^X] = T(s)f(X(t))$$

and for $0 < s_1 < s_2$

$$\begin{aligned} &E[f_1(X(t+s_1))f_2(X(t+s_2))|\mathcal{F}_t^X] \\ &\quad = E[f_1(X(t+s_1))T(s_2-s_1)f_2(X(t+s_1))|\mathcal{F}_t^X] \\ &\quad = T(s_1)[f_1T(s_2-s_1)f_2](X(t)) \end{aligned}$$

Theorem 18.10 *If* $P_x(B) = P\{x + W(\cdot) \in B\}$, $B \in \mathcal{B}(C[0,\infty))$, *then*

$$E[\mathbf{1}_B(X(t+\cdot))|\mathcal{F}_t] = P_{X(t)}(B)$$

Transition density

The transition density is

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}$$

which satisfies the Chapman-Kolmogorov equation

$$p(t+s,x,y) = \int_{\mathbb{R}} p(t,x,z)p(s,z,y)dz$$

Note that

$$\frac{\partial}{\partial t} T(t)f(x) = \frac{1}{2}\frac{d^2}{dx^2} T(t)f(x)$$

Right continuous filtration

$$\begin{aligned} E[f(X(t+s))|\mathcal{F}_{t+}^X] &= \lim_{h\to 0} E[f(X(t+s))|\mathcal{F}_{t+h}^X] \\ &= \lim_{h\to 0} T(s-h)f(X(t+h)) = T(s)f(X(t)) \end{aligned}$$

Lemma 18.11 *If Z is bounded and measurable with respect to $\sigma(X(0), W(s), s \geq 0)$, then*

$$E[Z|\mathcal{F}_t^X] = E[Z|\mathcal{F}_{t+}^X] \quad a.s.$$

Proof. Consider

$$E[\prod_i f_i(X(t_i))|\mathcal{F}_{t+}^X] = E[\prod_i f_i(X(t_i))|\mathcal{F}_t^X]$$

and apply the Dynkin-class theorem. □

Corollary 18.12 *Let $\bar{\mathcal{F}}_t^X$ be the completion of $\mathcal{F}_t^X$. Then $\bar{\mathcal{F}}_t^X = \bar{\mathcal{F}}_{t+}^X$.*

Proof. If $C \in \mathcal{F}_{t+}^X$, then $E[\mathbf{1}_C|\mathcal{F}_t^X] = \mathbf{1}_C$ a.s. Consequently, setting

$$C^o = \{E[\mathbf{1}_C|\mathcal{F}_t^X] = 1\} \qquad P(C^o \triangle C) = 0$$

□

Approximation of stopping times by discrete stopping times

Lemma 18.13 *Every stopping time is the limit of a decreasing sequence of discrete stopping times.*

Proof. If τ is a $\{\mathcal{F}_t^X\}$-stopping time, define

$$\tau_n = \begin{cases} \sum_{k=1}^{\infty} \frac{k}{2^n} \mathbf{1}_{[(k-1)2^{-n},k2^{-n})}(\tau), & \tau < \infty \\ \infty & \tau = \infty. \end{cases}$$

Then

$$\{\tau_n \le t\} = \{\tau_n \le \frac{[2^n t]}{2^n}\} = \{\tau < \frac{[2^n t]}{2^n}\} \in \mathcal{F}_t^X.$$

□

Strong Markov Property

Theorem 18.14 *Let* τ *be a* $\{\mathcal{F}_t^X\}$*-stopping time with* $P\{\tau < \infty\} = 1$*. Then*

$$E[f(X(\tau + t))|\mathcal{F}_\tau] = T(t)f(X(\tau), \tag{18.1}$$

and more generally, if $P_x(B) = P\{x + W(\cdot) \in B\}$*,* $B \in \mathcal{B}(C[0,\infty))$*, then*

$$E[\mathbf{1}_B(X(\tau + \cdot))|\mathcal{F}_\tau] = P_{X(\tau)}(B)$$

Proof. Prove first for discrete stopping times and take limits. Let τ_n be as above. Then

$$\begin{aligned} E[f(X(\tau_n+t))|\mathcal{F}_{\tau_n}] &= \sum_k E[f(X(\tau_n+t))|\mathcal{F}_{k2^{-n}}]\mathbf{1}_{\{\tau_n=k2^{-n}\}} \\ &= \sum_k E[f(X(k2^{-n}+t))|\mathcal{F}_{k2^{-n}}]\mathbf{1}_{\{\tau_n=k2^{-n}\}} \\ &= \sum_k T(t)f(k2^{-n})\mathbf{1}_{\{\tau_n=k2^{-n}\}} \\ &= T(t)f(X(\tau_n). \end{aligned}$$

Assume that f is continuous so that $T(t)f$ is continuous. Then

$$E[f(X(\tau_n+t))|\mathcal{F}_\tau] = E[T(t)f(X(\tau_n))|\mathcal{F}_\tau]$$

and passing to the limit gives (18.1). The extension to all bounded, measurable f follows by Corollary 21.4. □

Lemma 18.15 *If* $\gamma \geq 0$ *is* $\mathcal{F}_\tau$*-measurable, then*

$$E[f(X(\tau+\gamma))|\mathcal{F}_\tau] = T(\gamma)f(X(\tau)).$$

Proof. First, assume that γ is discrete. Then

$$\begin{aligned} E[f(X(\tau+\gamma))|\mathcal{F}_\tau] &= \sum_{r\in\mathcal{R}(\gamma)} E[f(X(\tau+\gamma))|\mathcal{F}_\tau]\mathbf{1}_{\{\gamma=r\}} \\ &= \sum_{r\in\mathcal{R}(\gamma)} E[f(X(\tau+r))|\mathcal{F}_\tau]\mathbf{1}_{\{\gamma=r\}} \\ &= \sum_{r\in\mathcal{R}(\gamma)} T(r)f(X(\tau))\mathbf{1}_{\{\gamma=r\}} \\ &= T(\gamma)f(X(\tau)). \end{aligned}$$

Assuming that f is continuous, general γ can be approximated by discrete γ. □

Reflection principle

Lemma 18.16

$$P\{\sup_{s\leq t} W(s) > c\} = 2P\{W(t) > c\}$$

Proof. Let $\tau = t \wedge \inf\{s : W(s) \geq c\}$, and $\gamma = (t - \tau)$. Then setting $f = \mathbf{1}_{(c,\infty)}$,

$$E[f(W(\tau + \gamma))|\mathcal{F}_\tau] = T(\gamma)f(W(\tau)) = \frac{1}{2}\mathbf{1}_{\{\tau<t\}}$$

and hence, $P\{\tau < t\} = 2P\{W(t) > c\}$. □

Extension of martingale results to continuous time

If X is a $\{\mathcal{F}_t\}$-submartingale (supermartingale, martingale), then $Y_k^n = X(k2^{-n})$ is $\{\mathcal{F}_k^n\}$-submartingale (supermartingale, martingale), where $\mathcal{F}_k^n = \mathcal{F}_{k2^{-n}}$. Consequently, each discrete-time result should have a continuous-time analog, at least if we assume X is right continuous.

Optional sampling theorem

Theorem 18.17 *Let X be a right-continuous $\{\mathcal{F}_t\}$-submartingale and τ_1 and τ_2 be $\{\mathcal{F}_t\}$-stopping times. Then*

$$E[X(\tau_2 \wedge c)|\mathcal{F}_{\tau_1}] \geq X(\tau_1 \wedge \tau_2 \wedge c)$$

Proof. For $i = 1, 2$, let τ_i^n be a decreasing sequence of discrete stopping times converging to τ_i. Then, since $X \vee d$ is a submartingale, by the discrete-time optional sampling theorem

$$E[X(\tau_2^n \wedge c) \vee d|\mathcal{F}_{\tau_1^n}] \geq X(\tau_1^n \wedge \tau_2^n \wedge c) \vee d.$$

Noting that $\{X(\tau_2^n \wedge c) \vee d\}$ is uniformly integrable, conditioning on $\mathcal{F}_{\tau_1}$ and passing to the limit, we have

$$E[X(\tau_2 \wedge c) \vee d|\mathcal{F}_{\tau_1}] \geq X(\tau_1 \wedge \tau_2 \wedge c) \vee d.$$

Letting $d \to -\infty$, the theorem follows. □

Exit distribution for W

Let $a, b > 0$ and $\tau = \inf\{t : W(t) \notin (-a, b)\}$. Since

$$\{\tau \leq t\} = \cap_n \cup_{s\in[0,t]\cap\mathbb{Q}} \{W(s) \notin (-a + n^{-1}, b - n^{-1})\},$$

τ is a $\{\mathcal{F}_t^W\}$-stopping time. Since $\lim_{t\to\infty} P\{W(t) \in (-a, b)\} = 0$, $\tau < \infty$ a.s. For each $c > 0$,

$$E[W(\tau \wedge c)] = 0$$

Letting $c \to \infty$, by the bounded convergence theorem

$$E[W(\tau)] = -aP\{W(\tau) = -a\} + bP\{W(\tau) = b\} = 0,$$

and

$$P\{W(\tau) = b\} = \frac{a}{a+b}.$$

Doob's inequalities

Theorem 18.18 *Let X be a nonnegative, right-continuous submartingale. Then for $p > 1$,*

$$E[\sup_{s \le t} X(s)^p] \le \left(\frac{p}{p-1}\right)^p E[X(t)^p]$$

Corollary 18.19 *If M is a right-continuous square integrable martingale, then*

$$E[\sup_{s \le t} M(s)^2] \le 4E[M(t)^2].$$

Proof. By Theorem 11.25,

$$E[\max_{k \leq 2^n} X(k2^{-n}t)^p] \leq \left(\frac{p}{p-1}\right)^p E[X(t)^p],$$

and the result follows by the monotone convergence theorem. □

Samplepath properties

Finite, nonzero quadratic variation

$$\lim \sum (W(t_{i+1}) - W(t_i))^2 = t.$$

Brownian paths are nowhere differentiable.

Law of the Iterated Logarithm

$$\limsup_{t\to\infty} \frac{W(t)}{\sqrt{2t \log\log t}} = 1$$

$\hat{W}(t) = tW(1/t)$ is Brownian motion. $Var(\hat{W}(t)) = t^2\frac{1}{t} = t$ Therefore

$$\limsup_{t\to 0} \frac{W(1/t)}{\sqrt{2t^{-1}\log\log 1/t}} = \limsup_{t\to 0} \frac{\hat{W}(t)}{\sqrt{2t\log\log 1/t}} = 1$$

Consequently,

$$\limsup_{h\to 0} \frac{W(t+h) - W(t)}{\sqrt{2h\log\log 1/h}} = 1$$

See [2], Theorem 13.18.

The tail of the normal distribution

Lemma 18.20

$$\int_a^\infty e^{-\frac{x^2}{2}}dx < a^{-1}e^{-\frac{a^2}{2}} = \int_a^\infty (1+x^{-2})e^{-\frac{x^2}{2}}dx$$
$$< (1+a^{-2})\int_a^\infty e^{-\frac{x^2}{2}}dx$$

Proof. Differentiate

$$\frac{d}{da}a^{-1}e^{-\frac{a^2}{2}} = -(a^{-2}+1)e^{-\frac{a^2}{2}}.$$

□

Modulus of continuity

Theorem 18.21 *Let* $h(t) = \sqrt{2t \log 1/t}$. *Then*

$$P\{\lim_{\epsilon \to 0} \sup_{t_1, t_2 \in [0,1], |t_1 - t_2| \le \epsilon} \frac{|W(t_1) - W(t_2)|}{h(|t_1 - t_2|)} = 1\} = 1$$

Proof.

$$P\{\max_{k \le 2^n}(W(k2^{-n}) - W((k-1)2^{-n})) \le (1-\delta)h(2^{-n})\} = (1-I)^{2^n} < e^{-2^n I}$$

for

$$I = \int_{(1-\delta)\sqrt{2\log 2^n}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx > C \frac{1}{\sqrt{n}} e^{-(1-\delta)^2 \log 2^n} > \frac{C}{\sqrt{n}} 2^{-(1-\delta)^2 n}$$

so $2^n I > 2^{n\delta}$ for n sufficiently large and Borel-Cantelli implies

$$P\{\limsup_{n \to \infty} \max_{k \le 2^n}(W(k2^{-n}) - W((k-1)2^{-n}))/h(2^{-n}) \ge 1\} = 1.$$

For $\delta > 0$ and $\epsilon > \frac{1+\delta}{1-\delta} - 1$

$$\begin{aligned}
&P\{\max_{0<k\leq 2^{n\delta}, 0\leq i\leq 2^n-2^{n\delta}} \frac{|W((i+k)2^{-n}) - W(i2^{-n})|}{h(k2^{-n})} \geq (1+\epsilon)\} \\
&\quad\leq \sum 2(1-\Phi((1+\epsilon)\sqrt{2\log(2^n/k)})) \\
&\quad\leq C\sum \frac{1}{(1+\epsilon)\sqrt{2\log(2^n/k)}} e^{-2(1+\epsilon)^2\log(2^n/k))} \\
&\quad\leq C\frac{1}{\sqrt{n}} 2^{n(1+\delta)} 2^{-2n(1-\delta)(1+\epsilon)^2}
\end{aligned}$$

and the right side is a term in a convergent series. Consequently, for almost every ω, there exists $N(\omega)$ such that $n \geq N(\omega)$ and $0 < k \leq 2^{n\delta}, 0 \leq i \leq 2^n - 2^{n\delta}$ implies

$$|W((i+k)2^{-n}) - W(i2^{-n})| \leq (1+\epsilon)h(k2^{-n})$$

If $|t_1 - t_2| \leq 2^{-(N(\omega)+1)(1-\delta)}$,

$$\begin{aligned} |W(t_1) - W(t_2)| \;\leq\; & |W([2^{N(\omega)}t_1]2^{-N(\omega)}) - W([2^{N(\omega)}t_2]2^{-N(\omega)})| \\ & + \sum_{n \geq N(\omega)} |W([2^n t_1]2^{-n}) - W([2^{n+1}t_1]2^{-(n+1)})| \\ & + \sum_{n \geq N(\omega)} |W([2^n t_2]2^{-n}) - W([2^{n+1}t_2]2^{-(n+1)})| \end{aligned}$$

so

$$\begin{aligned} |W(t_1) - W(t_2)| \;\leq\; & 2(1+\epsilon) \sum_{n=N(\omega)+1}^{\infty} h(2^{-n}) \\ & +(1+\epsilon)h(|[2^{N(\omega)}t_1] - [2^{N(\omega)}t_2]|2^{-N(\omega)}) \end{aligned}$$

□

Lévy characterization

Theorem 18.22 *Let M be a continuous martingale such that $M^2(t) - t$ is also a martingale. Then M is a standard Brownian motion*

Proof.

$$\begin{aligned}
&E[e^{i\theta(M(t+r)-M(t))}|\mathcal{F}_t] \\
&\quad = 1 + \sum_{k=0}^{n-1} E[(e^{i\theta(M(s_{k+1})-M(s_k)} - 1 - i\theta(M(s_{k+1}) - M(s_k)) \\
&\qquad\qquad + \frac{1}{2}\theta^2(M(s_{k+1}) - M(s_k))^2)e^{i\theta(M(s_k)-M(t))}|\mathcal{F}_t] \\
&\qquad - \frac{1}{2}\theta^2 \sum_{k=0}^{n-1}(s_{k+1} - s_k)E[e^{i\theta(M(s_k)-M(t))}|\mathcal{F}_t]
\end{aligned}$$

The first term converges to zero by the dominated convergence theorem, so we have

$$E[e^{i\theta(M(t+r)-M(t))}|\mathcal{F}_t] = 1 - \frac{1}{2}\theta^2 \int_0^r E[e^{i\theta(M(t+s)-M(t))}|\mathcal{F}_t]ds$$

and $E[e^{i\theta(M(t+r)-M(t))}|\mathcal{F}_t] = e^{-\frac{\theta^2 r}{2}}$. □

19. Problems

1. Let M be a set and $\{\mathcal{M}_\alpha, \alpha \in \mathcal{A}\}$ be a collection of σ-algebras of subsets of M. Show that $\cap_{\alpha\in\mathcal{A}}\mathcal{M}_\alpha$ is a σ-algebra.

2. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X be a real-valued function defined on Ω. Show that $\{B \subset \mathbb{R} : \{X \in B\} \in \mathcal{F}\}$ is a σ-algebra.

3. Note that if $\theta_1, \theta_2 \in \{0,1\}$, then $\max\{\theta_1, \theta_2\} = \theta_1 + \theta_2 - \theta_1\theta_2$. Find a similar formula for $\max\{\theta_1, \ldots, \theta_m\}$ and prove that it holds for all choices of $\theta_1, \ldots, \theta_m \in \{0,1\}$. Noting that

$$\max\{\mathbf{1}_{A_1}, \ldots, 1_{A_m}\} = \mathbf{1}_{\cup_{i=1}^m A_i},$$

use the identity to prove the inclusion-exclusion principle (that is, express $P(\cup_{i=1}^m A_i)$ in terms of $P(A_{i_1} \cap \cdots \cap A_{i_l})$.

4. Six couples are seated randomly at a round table. (All 12! placements are equally likely.) What is the probability that at least one couple is seated next to each other?

5. Let $\{a_k^n\}$ be nonnegative numbers satisfying

$$\lim_{n\to\infty} a_k^n = a_k$$

for each k. Suppose that for each $K \subset \{1, 2, \ldots\}$,

$$\nu_K = \lim_{n\to\infty} \sum_{k\in K} a_k^n$$

exists and is finite. Show that

$$\nu_K = \sum_{k\in K} a_k.$$

6. Let $(M, \mathcal{M})$ be a measurable space, and let $\mu_1, \mu_2, \ldots$ be probability measures on $\mathcal{M}$. Suppose that

$$\mu(A) = \lim_{n\to\infty} \mu_n(A)$$

for each $A \in \mathcal{M}$. Show that μ is a measure on $\mathcal{M}$.

7. Find σ-algebras $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ such that $\mathcal{D}_1$ is independent of $\mathcal{D}_3$, $\mathcal{D}_2$ is independent of $\mathcal{D}_3$, and $\mathcal{D}_1$ is independent of $\mathcal{D}_2$, but $\mathcal{D}_1 \vee \mathcal{D}_2$ is not independent of $\mathcal{D}_3$.

8. Let (S, d) be a metric space. Show that $d \wedge 1$ is a metric on S giving the same topology as d and that (S, d) is complete if and only if $(S, d \wedge 1)$ is complete.

9. Give an example of a sequence of events $\{A_n\}$ such that $\sum_n P(A_n) = \infty$ but $P(B) = 0$ for $B = \cap_n \cup_{m\geq n} A_m$.

10. Give and example of a nonnegative random variable X and a σ-algebra $\mathcal{D} \subset \mathcal{F}$ such that $E[X] = \infty$ but $E[X|\mathcal{D}] < \infty$ a.s.

11. Let $\mathcal{D}$, $\mathcal{G}$, and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. Suppose $\mathcal{G}$ and $\mathcal{H}$ are independent, $\mathcal{D} \subset \mathcal{G}$, X is an integrable, $\mathcal{G}$-measurable random variable, and Y is an integrable, $\mathcal{H}$-measurable random variable.

    (a) Show that
    $$E[X|\mathcal{D} \vee \mathcal{H}] = E[X|\mathcal{D}] \,,$$
    where $\mathcal{D} \vee \mathcal{H}$ is the smallest σ-algebra containing both $\mathcal{D}$ and $\mathcal{H}$.

    (b) Show that
    $$E[XY|\mathcal{D}] = E[Y]E[X|\mathcal{D}] \,.$$

    (c) Show by example, that if we only assume $\mathcal{H}$ is independent of $\mathcal{D}$ (not $\mathcal{G}$), then the indentity in Part 11b need not hold.

12. Let $Z \in L^1$, and let τ be a finite $\{\mathcal{F}_n\}$-stopping time. Show that
$$E[Z|\mathcal{F}_\tau] = \sum_{n=0}^{\infty} E[Z|\mathcal{F}_n]\mathbf{1}_{\{\tau=n\}}.$$

13. Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be independent uniform $[0,1]$ random variables, and define $\mathcal{F}_n = \sigma(X_i, Y_i : i \leq n)$. Let

$$\tau = \min\{n : Y_n \leq X_n\}.$$

Show that τ is a $\{\mathcal{F}_n\}$-stopping time, and compute the distribution function

$$P\{X_\tau \leq x\}.$$

14. Let $\{X_n\}$ be adapted to $\{\mathcal{F}_n\}$. Show that $\{X_n\}$ is an $\{\mathcal{F}_n\}$-martingale if and only if

$$E[X_{\tau \wedge n}] = E[X_0]$$

for every $\{\mathcal{F}_n\}$-stopping time τ and each $n = 0, 1, \ldots$.

15. Let X_1 and X_2 be independent and Poisson distributed with parameters λ_1 and λ_2 respectively. ($P\{X_i = k\} = e^{-\lambda_i} \frac{\lambda_i^k}{k!}$, $k = 0, 1, \ldots$) Let $Y = X_1 + X_2$. Compute the conditional distribution of X_1 given Y, that is, compute $P\{X_1 = i|Y\} \equiv E[\mathbf{1}_{\{X_1 = i\}}|Y]$.

16. A family of functions $H \subset B(\mathbb{R})$ is *separating* if for finite measures μ and ν, $\int f d\mu = \int f d\nu$, for all $f \in H$, implies $\mu = \nu$. For example, $C_c^\infty(\mathbb{R})$ and $\{f : f(x) = e^{i\theta x}, \theta \in \mathbb{R}\}$ are separating families. Let X and Y be random variables and $\mathcal{D} \subset \mathcal{F}$. Let H be a separating family.

(a) Show that $E[f(X)|\mathcal{D}] = E[f(X)]$ for all $f \in H$ implies X is independent of $\mathcal{D}$.

(b) Show that $E[f(X)|Y] = f(Y)$ for all $f \in H$ implies $X = Y$ a.s.

17. Let $\{X_n\}$ be $\{\mathcal{F}_n\}$-adapted, and let $B, C \in \mathcal{B}(\mathbb{R})$. Define

$$A_n = \{X_m \in B, \text{ some } m > n\},$$

and suppose that there exists $\delta > 0$ such that

$$P(A_n|\mathcal{F}_n) \geq \delta \mathbf{1}_C(X_n) \quad a.s.$$

Show that

$$\{X_n \in C \text{ i.o.}\} \equiv \cap_n \cup_{m>n} \{X_m \in C\} \subset \{X_n \in B \text{ i.o.}\}.$$

18. Let $X_1, X_2, \ldots$ be random variables. Show that there exist positive constants $c_k > 0$ such that $\sum_{k=1}^{\infty} c_k X_k$ converges a.s.

19. Let $\{a_n\} \subset \mathbb{R}$. Suppose that for each $\epsilon > 0$, there exists a sequence $\{a_n^\epsilon\}$ such that $|a_n - a_n^\epsilon| \leq \epsilon$ and $a^\epsilon = \lim_{n\to\infty} a_n^\epsilon$ exists. Show that $a = \lim_{\epsilon\to 0} a^\epsilon$ exists and that $a = \lim_{n\to\infty} a_n$.

20. Let $X_1, X_2, \ldots$ be iid on $(\Omega, \mathcal{F}, Q)$ and suppose that $\mu_{X_k}(dx) = \gamma(x)dx$ for some strictly positive Lebesgue density γ. Define $X_0 = 0$ and for $\rho \in \mathbb{R}$, let

$$L_n^\rho = \prod_{k=1}^{n} \frac{\gamma(X_k - \rho X_{k-1})}{\gamma(X_k)}.$$

(a) Show that $\{L_n^\rho\}$ is a martingale on $(\Omega, \mathcal{F}, Q)$.

(b) Let $\mathcal{F}_N = \sigma(X_1, \ldots, X_N)$ and define P_ρ on $\mathcal{F}_N$ by $dP_\rho = L_N^\rho dQ$. Define $Y_k^\rho = X_k - \rho X_{k-1}$. What is the joint distribution of $\{Y_k^\rho, 1 \leq k \leq N\}$ on $(\Omega, \mathcal{F}_N, P_\rho)$?

21. (a) Let $\{M_n\}$ be a $\{\mathcal{F}_n\}$-martingale. Assume that $\{M_n\}$ is $\{\mathcal{G}_n\}$-adapted and that $\mathcal{G}_n \subset \mathcal{F}_n$, $n = 0, 1, \ldots$. Show that $\{M_n\}$ is a $\{\mathcal{G}_n\}$-martingale.

(b) Let $\{U_n\}$ and $\{V_n\}$ be $\{\mathcal{F}_n\}$-adapted and suppose that

$$U_n - \sum_{k=0}^{n-1} V_k$$

is a $\{\mathcal{F}_n\}$-martingale. Let $\{\mathcal{G}_n\}$ be a filtration with $\mathcal{G}_n \subset \mathcal{F}_n$, $n = 0, 1, \ldots$. Show that

$$E[U_n|\mathcal{G}_n] - \sum_{k=0}^{n-1} E[V_k|\mathcal{G}_k] \tag{19.1}$$

is a $\{\mathcal{G}_n\}$-martingale. (Note that we are not assuming that $\{U_n\}$ and $\{V_n\}$ are $\{\mathcal{G}_n\}$-adapted.)

22. Let $r > 0$, and let $\{\xi_1, \ldots, \xi_m\}$ be independent, uniform $[0, r]$ random variables. Let $\rho > 1$, and define

$$X_n^{(k)} = \rho^n \xi_k$$

and $N_n = \#\{k : X_n^{(k)} \leq r\}$. Let $\mathcal{F}_n = \mathcal{F}_0 \equiv \sigma(\xi_1, \ldots, \xi_m)$. For $g \in C_b[0, \infty)$ satisfying $g(x) = 1$ for $x \geq r$, define

$$U_n = \prod_{k=1}^{m} g(X_n^{(k)}), \quad V_n = \prod_{k=1}^{m} g(X_{n+1}^{(k)}) - \prod_{k=1}^{m} g(X_n^{(k)}).$$

Then (trivially) $\{U_n\}$ and $\{V_n\}$ are $\{\mathcal{F}_n\}$-adapted and

$$U_n - \sum_{k=0}^{n-1} V_k = U_0$$

is a $\{\mathcal{F}_n\}$-martingale. Let $\mathcal{G}_n = \sigma(N_k, k \leq n)$. Compute the martingale given by (19.1).

23. Let $X_1 \geq X_2 \geq \cdots \geq 0$ and $E[X_1] < \infty$. Let $\{\mathcal{F}_n\}$ be a filtration. ($\{X_n\}$ is not

necessarilty adapted to $\{\mathcal{F}_n\}$.) Show that with probability one

$$\lim_{n\to\infty} E[X_n|\mathcal{F}_n] = \lim_{n\to\infty} E[\lim_{k\to\infty} X_k|\mathcal{F}_n] = \lim_{k\to\infty}\lim_{n\to\infty} E[X_k|\mathcal{F}_n].$$

(If you have not already completed Problem 17, you may want to apply the result of this problem.)

24. Suppose that $Y_n^\epsilon \leq X_n \leq Z_n^\epsilon$, $Y_n^\epsilon \Rightarrow Y^\epsilon$ and $Z_n^\epsilon \Rightarrow Z^\epsilon$ as $n \to \infty$, and $Y^\epsilon \Rightarrow X$ and $Z^\epsilon \Rightarrow X$ as $\epsilon \to \infty$. Show that $X_n \Rightarrow X$.

25. For each n, let $\{X_k^n\}$ be a sequence of random variables with $\mathcal{R}(X_k^n) = \{0, 1\}$. Assume $\{X_k^n\}$ is adapted to $\{\mathcal{F}_k^n\}$ and define $Z_k^n = E[X_{k+1}^n | \mathcal{F}_k^n]$. Suppose that $\lambda > 0$, $\sum_k Z_k^n \to \lambda$ in probability, and $E[\max_k Z_k^n] \to 0$. Show that $\sum_k X_k^n \Rightarrow Y$ where Y is Poisson distributed with parameter λ.

    Hint: There is, no doubt, more than one way to solve this problem; however, you may wish to consider the fact that $X_k^n \in \{0, 1\}$ implies

$$e^{i\theta X_k^n} = 1 + X_k^n(e^{i\theta} - 1) = \frac{1 + X_k^n(e^{i\theta} - 1)}{1 + Z_{k-1}^n(e^{i\theta} - 1)}(1 + Z_{k-1}^n(e^{i\theta} - 1))$$

20. Exercises

1. Let X be a $\mathbb{R}$-valued function defined on Ω. Show that $\{\{X \in B\} : B \in \mathcal{B}(\mathbb{R})\}$ is a σ-algebra.
2. Let $\mathcal{D}_1 \subset \mathcal{D}_2$, and $X \in L^2$. Suppose that $E[E[X|\mathcal{D}_1]^2] = E[E[X|\mathcal{D}_2]^2]$. Show that $E[X|\mathcal{D}_1] = E[X|\mathcal{D}_2]$ a.s.

Glossary

Borel sets. For a metric space (E, r), the collection of *Borel sets* is the smallest σ-algebra containing the open sets.

Complete metric space. We say that a metric space (E, r) is *complete* if every Cauchy sequence in it converges.

Complete σ-probability space. A probability space $(\Omega, \mathcal{F}, P)$ is *complete*, if $\mathcal{F}$ contains all subsets of sets of probability zero.

Conditional expectation. Let $\mathcal{D} \subset \mathcal{F}$ and $E[|X|] < \infty$. Then $E[X|\mathcal{D}]$ is the, essentially unique, $\mathcal{D}$-measurable random variable satisfying

$$\int_D X dP = \int_D E[X|\mathcal{D}] dP, \quad \forall D \in \mathcal{D}.$$

Consistent. Assume we have an arbitrary state space $(E, \mathcal{B})$ and an index set I. For each nonempty subset $J \subset I$ we denote by E^J the product set $\prod_{t \in J} E$, and we define $\mathcal{B}^J$ to be the product-σ-algebra $\otimes_{t \in J} \mathcal{B}$. Obviously, if $J \subset H \subset I$ then there is a projection map

$$p_J^H : \; E^H \to E^J.$$

If for every two such subsets J and H we have

$$P_J = p_J^H(P_H)$$

then the family $(P_J)_{\emptyset \neq J \subset H}$ is called *consistent*.

Metric space (E, r) is a *metric space* if E is a set and $r : E \times E \to [0, \infty)$ satisfies

a) $r(x, y) = 0$ if and only if $x = y$.

b) $r(x, y) = r(y, x)$, $x, y \in E$

c) $r(x, y) \leq r(x, z) + r(z, y)$ (*triangle inequality*)

Separable. A metric space (E, r) is called *separable* if it contains a countable dense subset; that is, a set with a countable number of elements whose closure is the entire space. Standard example: **R**, whose countable dense subset is **Q**.

Separating set A collection of function $M \subset \bar{C}(S)$ is *separating* is $\mu, \nu \in \mathcal{M}_f(S)$ and $\int g d\nu = \int g d\mu$, $g \in M$, implies that $\mu = \nu$.

σ-finite A measure μ on $(M, \mathcal{M})$ is *σ-finite* if there exist $A_i \in \mathcal{M}$ such that $\cup_i A_i = M$ and $\mu(A_i) < \infty$ for each i.

Uniform equicontinuity A collection of functions $\{h_\alpha, \alpha \in \mathcal{A}\}$ is *uniformly equicontinuous* if for each $\epsilon > 0$, there exists a $\delta > 0$ such that $|x - y| \leq \delta$ implies

$\sup_{\alpha \in \mathcal{A}} |h_\alpha(x) - h_\alpha(y)| \leq \epsilon.$

21. Technical lemmas

- Product limits
- Open sets in separable metric spaces
- Closure of collections of functions

Product limits

Lemma 21.1 *For* $|x| \leq \frac{1}{5}$, $e^{-x-x^2} \leq 1 - x$. *Consequently, if* $\lim_{n\to\infty} \sum_k a_{kn} = c$ *and* $\lim_{n\to\infty} \sum_k (a_{kn})^2 = 0$, *then*

$$\lim_{n\to\infty} \prod_k (1 - a_{kn}) = e^{-c}.$$

Proof. Let $h(x) = 1 - x - e^{-x-x^2}$, and note that $h(0) = h'(0) = 0$ and for $|x| \leq \frac{1}{5}$, $h''(0) \geq 0$. Since $\lim_{n\to\infty} \sum_k (a_{kn})^2 = 0$, for n sufficiently large, $\max_k a_{kn} \leq \frac{1}{5}$ and hence

$$\begin{aligned} e^{-c} = \lim_{n\to\infty} \exp\{-\sum_k a_{kn} - \sum_k (a_{kn})^2\} &\leq \lim_{n\to\infty} \prod_k (1 - a_{kn}) \\ &\leq \lim_{n\to\infty} \exp\{-\sum_k a_{kn}\} = e^{-c} \end{aligned}$$

□

Open sets in separable metric spaces

Lemma 21.2 *If (S, d) is a separable metric space, then each open set is a countable union of open balls.*

Proof. Let $\{x_i\}$ be a countable dense subset of S and let G be open. If $x_i \in G$, define $\epsilon_i = \inf\{d(x_i, y) : y \in G^c\}$. Then

$$G = \cup_{i:x_i \in G} B_{\epsilon_i}(x_i).$$

$\square$

Closure of collections of functions

Theorem 21.3 *Let H be a linear space of bounded functions on S that contains constants, and let $\mathcal{S}$ be a collection of subsets of S that is closed under intersections. Suppose $\mathbf{1}_A \in H$ for each $A \in \mathcal{S}$ and that H is closed under convergence of uniformly bounded increasing sequences. Then H contains all bounded, $\sigma(\mathcal{S})$-measurable functions.*

Proof. $\{C \subset S : \mathbf{1}_C \in H\}$ is a Dynkin class containing $\mathcal{S}$ and hence $\sigma(\mathcal{S})$. Consequently, H contains all $\sigma(\mathcal{S})$-measurable simple functions and, by approximation by increasing sequences of simple functions, all $\sigma(\mathcal{S})$-measurable functions. □

Corollary 21.4 *Let H be a linear space of bounded functions on S that contains constants. Suppose that H is closed under uniform convergence, and under convergence of uniformly bounded increasing sequences. Suppose $H_0 \subset H$ is closed under multiplication. Then H contains all bounded, $\sigma(H_0)$-measurable functions.*

Proof. If $p(z_1, \ldots, z_m)$ is a polynomial and $f_1, \ldots, f_m \in H_0$, then $p(f_1, \ldots, f_m) \in H$. Since any continuous function h on a product of intervals can be approximated uniformly by polynomials, it follows that $h(f_1, \ldots, f_m) \in H$. In particular, $g_n = \prod_{i=1}^m [(1 \wedge f_i - a_i) \vee 0]^{1/n} \in H$. Since g_n increases to $\mathbf{1}_{\{f_1 > a_1, \ldots, f_m > a_m\}}$, the indicator is in H and hence, by Theorem 21.3, H contains all bounded, $\sigma(H_0)$-measurable functions. □

References

[1] A. D. Barbour, Lars Holst, and Svante Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1992. ISBN 0-19-852235-5. , Oxford Science Publications.

[2] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[3] A. N. Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Co., New York, 1956. URL http://www.clrc.rhul.ac.uk/resources/fop/index.htm. Translation edited by Nathan Morrison, with an added bibliography by A. T. Bharucha-Reid.

[4] Sundar Sethuraman. A martingale central limit theorem. URL http://orion.math.iastate.edu/sethuram/papers/wi_mart.pdf. 2002.