

STAT 901 probability Theory

Sept 8, 2008

MC6013 A ext 35541 or better UW-ace

Sept 29 Oct 29 Nov 24, 3 assessments.

Marking Scheme $\bar{T} = \frac{T_1 + T_2 + T_3}{3}$

Textbook: A Probability Path by Sidney I. Resnick

Birkhäuser.

Tools you need set theory.

Question: How many subsets of {1, 2, 3, 4, 5} are there?

Answer $2^5 = C_5^0 + C_5^1 + \dots + C_5^5$ or examine each component
 $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5$ possibilities.Question Let $A_0 = \{\emptyset\}$ $A_1 = \{\emptyset, \{1\}\}$ $A_2 = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$ $A_n = \text{set of all subsets of } \{1, 2, \dots, n\}$ $A_\infty = \text{set of all subsets of } \{1, 2, \dots\}$

True or false?

 $A_0 \cup A_1 \cup \dots = A_\infty$ Answer: False.

For example:

 $\{1, 2, \dots\} \in A_\infty$.You will want to know some real analysis. ϵ, δ . arguments.

lub. glb: inf.

Sup.

Q1) measure-theoretic probability

— measure-theoretic prob. simplifies arguments.

(contains both discrete / continuous cases)

— essential to understanding stochastic integration

Stochastic Integration, Ito's Lemma, SDE

 $P(A|B) \stackrel{\Delta}{=} \frac{P(A \cap B)}{P(B)}$ where $P(B) > 0$. When $P(B) = 0$ undefined $f(x|y) = \frac{f(x,y)}{f_x(y)}$ $P(X \in A | Y=y) = \int_{x \in A} f(x|y) dx$ Hilary

N. Kolmogorov

Sept 10, 2008

Basic Set Theory

Ω : sample space or universal set

w : element of Ω , point in Ω (elementary outcome)

$P(\Omega)$: power set of Ω , collection of all subsets of Ω , including the empty set \emptyset and Ω itself.

$P(A)$ is the collection of all subsets of A etc.

Example:

$$P(\{1, 2\}) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$$

$A \subset \Omega$, $A \in P(\Omega)$ subsets of Ω will be written with capital letters.
(means \subseteq)

$\mathcal{A}, \mathcal{B}, \mathcal{C}$ — collections of subsets of Ω .

- written with script capitals.

$$\text{example: } \Omega = \{1, 2\} \quad P(\Omega) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$$

$$A = \{\emptyset, \{1, 2\}\} \text{ for example.}$$

A^c — complement of A

$$A^c = \{w \in \Omega : w \notin A\}$$

A_T = intersection of subsets A_t over an arbitrary index set T .

$$A_T = \{w : w \in A_t \text{ for all } t \in T\}$$

$$\text{Special cases: } \bigcap_{n=1}^{\infty} A_n \text{ or equivalently } \bigcap_{n \geq 1} A_n$$

$$A_1 \cap A_2 \cap \dots$$

$$AB = A \cap B \quad A_{1, 2} = A_1 \cap A_2 \text{ etc.}$$

A_T — union of subsets A_t over arbitrary index set T .

$$A_T = \{w : w \in A_t \text{ for some } t \in T\}$$

$A_t, t \in T$
Subsets are said to be pairwise disjoint (or mutually disjoint) if $A_t \cap A_{t'} = \emptyset$ for all $t \neq t' \in T$.

Unions of mutually disjoint sets shall be written as

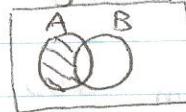
$$\sum_{T \in T} A_T$$

or $A_1 + A_2 + A_3 + \dots$ when T is positive integer.

$A \setminus B$ - set difference of A and B

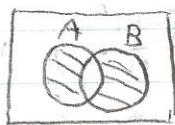
$$A \setminus B = \{w \in A : w \notin B\}$$

when $B \subseteq A$, we often write $A - B$ for this.



$A \Delta B$ - symmetric difference of A and B

$$A \Delta B = (A \setminus B) \cup (B \setminus A)$$



De-Morgan's laws.

$$(\bigcup_{t \in T} A_t)^c = \bigcap_{t \in T} A_t^c \quad (\bigcap_{t \in T} A_t)^c = \bigcup_{t \in T} A_t^c$$

Distributive Laws.

$$(\bigcup_{t \in T} A_t) \cap B = \bigcup_{t \in T} (A_t \cap B) \quad (\bigcap_{t \in T} A_t) \cup B = \bigcap_{t \in T} (A_t \cup B)$$

Properties of sets are mirrored by indicator functions!

$$A \leftrightarrow \mathbb{1}_A \text{ where } \mathbb{1}_A : \mathcal{S} \rightarrow \{0, 1\}$$

$$\mathbb{1}_A(w) = \begin{cases} 0 & w \notin A \\ 1 & w \in A \end{cases}$$

so $A \subseteq B$, if and only if $\mathbb{1}_A \leq \mathbb{1}_B$

$$A = B^c \text{ iff. } \mathbb{1}_A = 1 - \mathbb{1}_B$$

$$A = \bigcap_{t \in T} A_t \text{ iff. } \mathbb{1}_A = \prod_{t \in T} \mathbb{1}_{A_t} \text{ (product)}$$

$$= \min_{t \in T} \mathbb{1}_{A_t}$$

$$A = \bigcup_{t \in T} A_t \text{ iff. } \mathbb{1}_A = \max_{t \in T} \mathbb{1}_{A_t}$$

$$A = \sum_{t \in T} A_t \text{ iff. } \mathbb{1}_A = \sum_{t \in T} \mathbb{1}_{A_t}$$

(supremum)

Just as numbers have upper bounds, lower bounds, least upper bounds, greatest lower bound (infimum)

Given a sequence x_1, x_2, x_3, \dots of real numbers, the infimum

$\inf_{k \geq n} x_k$ is the largest real number such that

$\inf_{k \geq n} x_k \leq x_k$ for all $k \geq n$ (greatest lower bound)

PP 10.3A inf sup

Rational numbers are not complete, so sup, inf could be found in rational space.

Similarly, the supremum $\sup_{k \geq n} x_k$ is the smallest value such that $\sup_{k \geq n} x_k \geq x_k$ for all $k \geq n$. (least upper bound)

Example: $x_n = 1 + \frac{(-1)^n}{n}$, $n=1, 2, 3, \dots$

0, $3/2$, $2/3$, $5/4$, $4/5$, $7/6$, $6/7$, ...

$$\inf_{k \geq 1} x_k = 0$$

$$\inf_{k \geq 4} x_k = 4/5, \sup_{k \geq 1} x_k = 3/2 + (-1) = 2 \in \mathbb{Q}$$

$$\sup_{k \geq 4} x_k = 5/4$$

Example: $(x_k = 1 - 1/k \text{ for } k \in \mathbb{N}) \quad (\text{sets are complete})$

$$\sup_{k \geq 1} x_k = 1, \inf_{k \geq 1} x_k = 0 \quad (\text{sets are complete})$$

For sets: given a sequence of sets A_1, A_2, A_3, \dots

$$\inf_{k \geq n} A_k = \bigcap_{k \geq n} A_k \quad \sup_{k \geq n} A_k = \bigcup_{k \geq n} A_k \quad \text{trivial}$$

STAT 901 Sept 12, 2008 Small

We move next week MC 4060

Last time we defined

$$\inf_{k \geq n} A_k = \bigcap_{k=n}^{\infty} A_k \quad \sup_{k \geq n} A_k = \bigcup_{k=n}^{\infty} A_k$$

for numbers and sets. We can define a weaker form of inf and sup.

Defn. Let x_1, x_2, \dots be any sequence of reals. Define

$$\liminf_{n \rightarrow \infty} x_n := \sup_{n \geq 1} \inf_{k \geq n} x_k$$

That is, $y_n = \inf_{k \geq n} x_k$ Then $y_1 \leq y_2 \leq y_3 \leq \dots$

Def. $\liminf_{n \rightarrow \infty} x_n = \sup_{n \geq 1} y_n$

(Show) $y_n \leq y_m$ if $n < m$, $y_n = \inf_{k \geq n} x_k$

\liminf \limsup always exist!

Also

$$\limsup_{n \rightarrow \infty} x_n = \inf_{n \geq 1} \sup_{k \geq n} x_k$$

That is, $z_1 = \sup_{k \geq n} x_k$ so $z_1 > z_2 > z_3 \dots$

$$\text{Define } \limsup_{n \rightarrow \infty} x_n = \inf_{n \geq 1} z_n$$

Properties:

① $-\infty \leq \liminf_{n \rightarrow \infty} x_n \leq \limsup_{n \rightarrow \infty} x_n \leq \infty$ (prove this yourself)

② $\limsup_{n \rightarrow \infty} x_n = -(\liminf_{n \rightarrow \infty} -x_n)$

③ $\liminf_{n \rightarrow \infty} x_n = -\limsup_{n \rightarrow \infty} (-x_n)$

④ $\limsup_{n \rightarrow \infty} x_n$ is the largest cluster point of the sequence of x_n .

$\leftarrow \times \times \times (\times \times \dots \times) \rightarrow$

ininitely many times.

$\liminf_{n \rightarrow \infty} x_n$ is the smallest cluster point of the sequence of x_n .

④ When $\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n$ then $\lim_{n \rightarrow \infty} x_n$

exists, and

$$\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x_n$$

Example 1: $x_n = (-1)^n + 1/n$ $\liminf_{n \rightarrow \infty} x_n = -1$ $\limsup_{n \rightarrow \infty} x_n = 1$

Example 2: $x_n = n$ $\liminf_{n \rightarrow \infty} x_n = \infty$ $\limsup_{n \rightarrow \infty} x_n = \infty$

(3) $x_n = n(-1)^n$

$$\liminf_{n \rightarrow \infty} x_n = -\infty, \limsup_{n \rightarrow \infty} x_n = +\infty$$

Hilroy

Def. Let A_1, A_2, \dots be any sequence of subsets of \mathbb{R}

Define $\liminf_{n \rightarrow \infty} A_n = \sup_{n \geq 1} \inf_{k \geq n} A_k$

$$= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

That is let

$$B_n = \bigcap_{k=n}^{\infty} A_k \text{ Then } B_1 \subset B_2 \subset B_3 \subset \dots$$

Define $\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} B_n$

Also $\limsup_{n \rightarrow \infty} A_n = \inf_{n \geq 1} \sup_{k \geq n} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$

That is,

$$C_n = \bigcup_{k=n}^{\infty} A_k \text{ Then } C_1 \supset C_2 \supset C_3 \supset \dots$$

Define $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} C_n$

Properties:

① $\emptyset \subset \liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n \subset \mathbb{R}$

② $\limsup_{n \rightarrow \infty} A_n = (\liminf_{n \rightarrow \infty} A_n^c)^c$

$$\liminf_{n \rightarrow \infty} A_n = (\limsup_{n \rightarrow \infty} A_n^c)^c$$

③ $w \in \limsup_{n \rightarrow \infty} A_n \text{ iff. } w \in A_n \text{ for infinitely many values of } n$

For this reason, we often write

$$\limsup_{n \rightarrow \infty} A_n = \{A_n \text{ infinitely often}\}$$

or $\{A_n \text{ i.o.}\}$

e.g. coin tosses infinitely many time sequence

We $\liminf_{n \rightarrow \infty} A_n$ iff. $w \notin A_n$ for only finitely many values

If n (memberships are stronger)
having an infinite sequence with finitely many elements

Def.

when $\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$ (for now, longer insight p. 7)

then we define this common set to be $\lim A_n$

Examples $\mathbb{R} = \mathbb{R}$

$$\textcircled{1} \quad A_n = \begin{cases} [0, 2] & \text{if } n \text{ even} \\ [0, 1] & \text{if } n \text{ odd} \end{cases}$$

Then

$$\bigcup_{k=n}^{\infty} A_k = [0, 2] \quad \bigcap_{k=n}^{\infty} A_k = [0, 1]$$

$$\text{so } \bigcap_{k=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = [0, 2] \supset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = [0, 1]$$

$$\limsup_{n \rightarrow \infty} A_n$$

$$\textcircled{2} \quad A_n = [0, 1 - \frac{1}{n}]$$

$$\bigcup_{k=n}^{\infty} A_k = [0, 1]$$

$$\bigcap_{k=n}^{\infty} A_k = [0, 1 - \frac{1}{n}]$$

$$\bigcap_{k=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = [0, 1]$$

$$\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = [0, 1]$$

$$\text{so } \lim_{n \rightarrow \infty} [0, 1 - \frac{1}{n}] = [0, 1]$$

$$\textcircled{3} \quad A_n = \left\{ \frac{1}{n} \right\}$$

$$\bigcup_{k=n}^{\infty} A_k = \left\{ \frac{1}{n}, \frac{1}{n+1}, \dots \right\} \quad \bigcap_{k=n}^{\infty} A_k = \emptyset$$

$$\text{so } \bigcap_{k=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \emptyset$$

$$\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \emptyset$$

$$\text{so } \lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \right\} = \emptyset$$

Sept 15, 2008. STAT 901

Cloudy

Assigned Problems

Pages 20-22.

1, 3, 4, 6, 8, 9, 10, 14, 15.

Last time: $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$.

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

If these are equal, we call the set $\lim_{n \rightarrow \infty} A_n$

Example, when A_1, A_2, \dots is a monotone sequence of sets,

then $\lim_{n \rightarrow \infty} A_n$ always exists, for example,

$$\bigcup_{n=1}^{\infty} A_n \subset A_1 \subset A_2 \subset A_3 \subset \dots \quad (\text{monotone increasing})$$

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k$$

$$\bigcap_{k=n}^{\infty} A_k = A_n$$

Therefore,

$$\bigcap_{k=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{k=1}^{\infty} \left[\bigcup_{k=n}^{\infty} A_k \right] = \bigcup_{k=n}^{\infty} \bigcup_{k=1}^{\infty} A_k \quad (\text{im sup})$$

$$\bigcup_{k=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_n \quad (\text{im inf})$$

$$\text{So } \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} A_k$$

so this is $\lim_{n \rightarrow \infty} A_n$

Proposition, If $A_n \nearrow (A_1 < A_2 < \dots)$ then $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$

If $A_n \searrow (A_1 > A_2 > \dots)$ then $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$

The Concept of Closure

example: Let ℓ be the collection of semi-open intervals of the real line.

$$\ell = \{(a, b) : -\infty < a \leq b < +\infty\}$$

(here $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$)

Note that $(a, a] = \emptyset$

Suppose we take two elements of ℓ , and intersect them.

$$(a_1, b_1] \cap (a_2, b_2] = (\max(a_1, a_2), \min(b_1, b_2))$$

$$[(a_1, b_1)] \cap [(a_2, b_2)] = (a_1 \wedge a_2, b_1 \wedge b_2] \quad \text{another way of writing}$$

This is the interval of the same kind.

$$(a_1, b_1] \cap (a_2, b_2] \in \ell$$

We say that ℓ is closed under pairwise intersection.

Also ℓ is closed under finite intersections.
(use mathematical induction) However,

$$\bigcap_{n=1}^{\infty} (-k_n, 1] = [0, 1] \notin \ell$$

So ℓ is not closed under infinite intersections, in general.

Also, $[a_1, b_1] \cup [a_2, b_2] \notin \ell$, if $a_2 > b_1$ or $a_1 > b_2$

so ℓ is not closed under finite unions.

$$\text{Also, } [a, b]^c = (-\infty, a] \cup (b, \infty) \notin \ell$$

ℓ is not closed under complementation.

Example: $\ell = \{\emptyset, \mathbb{R}\}$ In this case ℓ is closed under complementation and arbitrary unions and intersections.

Definition: Let \mathcal{A} be a non-empty collection of subsets of \mathbb{R} , which is closed under complementation, finite intersections, and finite unions. Then \mathcal{A} is said to be a field. (Fields are also known as algebras).

A minimal set of conditions for \mathcal{A} to be a field is that -

① $\mathbb{R} \in \mathcal{A}$ ② $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$

③ If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$

$$(A \in \mathcal{A}, A^c \in \mathcal{A} \Rightarrow A \cup A^c = \mathbb{R} \in \mathcal{A})$$

①, ②, ③ are equivalent to the definition.

③ can be replaced by intersections. De Morgan's law

Definition: A σ -field or σ -algebra \mathcal{B} is a non-empty collection of subsets of \mathbb{R} that is closed under complementation, countable unions, countable intersections.

Countable = Countably infinite & finite.

Hilary

A minimal set of conditions for \mathcal{B} to be a σ -field is

- ① $\emptyset \in \mathcal{B}$
- ② $B \in \mathcal{B}$ implies $B^c \in \mathcal{B} \Rightarrow \emptyset \in \mathcal{B}$
- ③ If $B_1, B_2, \dots \in \mathcal{B}$, then $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \dots \text{ or } A \vee B \cup B \cup B \dots$$

$$A \cap B = A \cap B \cap \emptyset \cap \emptyset \dots$$

STAT 901 Prof. Small Sept 17, 2008

field / algebra \mathcal{A} : nonempty collection of subsets of Ω , closed under finite union, finite intersection, and complementations

σ -field / σ -algebra \mathcal{B} --- closed under countable unions, countable intersections & complementations.

Examples / counterexamples

① $\mathcal{B} = \mathcal{P}(\Omega)$ σ -field

② $\mathcal{B} = \{\emptyset, \Omega\}$ σ -field

③ $\mathcal{A} = \{A \subset \Omega \text{ st. } A \text{ is finite or } A^c \text{ is finite}\}$

is a field, but not generally a σ -field

proof: ① $\emptyset \in \mathcal{A}$, because $\emptyset^c = \emptyset$ which is finite. (Ω could be infinite)

of ② $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$ (obvious)

③ If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$ because

If A, B both finite then $A \cup B$ is finite.

If A^c or B^c is finite $(A \cup B)^c = A^c \cap B^c$ finite

However, it is not a σ -field in general.

for example, $\Omega = \{1, 2, 3, \dots\}$

$$A_1 = \{1\}, A_2 = \{1, 3\}, A_3 = \{1, 3, 5\} \dots$$

$A_n = \{\text{first } n \text{ odd integers}\}$ clearly, $A_n \in \mathcal{A}$ for all n .

$$\bigcup_{n=1}^{\infty} A_n = \{1, 3, 5, 7, \dots\} \notin \mathcal{A}$$

④ $\mathcal{B} = \{B \subset \Omega : B \text{ countable or } B^c \text{ countable}\}$

\mathcal{B} is a σ -field. Pages 13-14 in textbook

Idea: the σ -field generated by a collection of subsets \mathcal{C}

/ class/sefs

Let \mathcal{E} be any collection of subsets of Ω . \mathcal{E} is not necessarily a σ -field, but it is a subcollection of a σ -field. For example, $\mathcal{E} \subset P(\Omega)$ and $P(\Omega)$ is a σ -field. There may be smaller σ -fields containing \mathcal{E} . That is,

$$\mathcal{E} \subset \mathcal{B} \subset P(\Omega) \text{ where } \mathcal{B} \text{ is a } \sigma\text{-field.}$$

Definition: we define $\sigma(\mathcal{E})$, the σ -field generated by \mathcal{E} , to be the smallest σ -field such that $\mathcal{E} \subset \sigma(\mathcal{E})$

Second Definition: The σ -field generated by \mathcal{E} is a σ -field called $\sigma(\mathcal{E})$ satisfying.

- (a) $\mathcal{E} \subset \sigma(\mathcal{E})$ (b) if \mathcal{B} is any σ -field such that $\mathcal{E} \subset \mathcal{B}$, then $\sigma(\mathcal{E}) \subset \mathcal{B}$.

If $\sigma(\mathcal{E})$ satisfies the 2nd definition, it must be unique.

Suppose we had $\sigma_1(\mathcal{E})$ and $\sigma_2(\mathcal{E})$. By (b)

$$\sigma_1(\mathcal{E}) \subset \sigma_2(\mathcal{E}), \sigma_2(\mathcal{E}) \subset \sigma_1(\mathcal{E}) \Rightarrow \sigma_1(\mathcal{E}) = \sigma_2(\mathcal{E})$$

But how do we know that $\sigma(\mathcal{E})$ exists?

We will build it !!

Proposition: Let \mathcal{B}_1 and \mathcal{B}_2 be two σ -fields of subsets of Ω .

Then $\mathcal{B}_1 \cap \mathcal{B}_2$ is a σ -field.

Proof: @ $\Omega \in \mathcal{B}_1, \Omega \in \mathcal{B}_2$ because they are σ -fields, so

$$\Omega \in \mathcal{B}_1 \cap \mathcal{B}_2$$

(b) Suppose $B \in \mathcal{B}_1 \cap \mathcal{B}_2$, then $B \in \mathcal{B}_1$, so $B^c \in \mathcal{B}_1$ because \mathcal{B}_1 is a σ -field. Similarly, $B \in \mathcal{B}_2$ and so is B^c . Therefore,

$$B^c \in \mathcal{B}_1 \cap \mathcal{B}_2$$

(c) Suppose B_1, B_2, B_3, \dots are in $\mathcal{B}_1 \cap \mathcal{B}_2$, then (dntw) part (c)

B_1, B_2, \dots are all in \mathcal{B}_1 , so $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}_1$ because \mathcal{B}_1 is a σ -field. Similarly, $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}_2$, so $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}_1 \cap \mathcal{B}_2$. QED.

part c fails for Union operation.

However, $\mathcal{B}_1 \cup \mathcal{B}_2$ is not necessarily a σ -field, because (c)

Is not necessarily true for $B_1 \cup B_2$. non empty

Proposition: Let B_t , $t \in T$ be an arbitrary collection of σ -fields of subsets of \mathbb{R} . Then

$$\bigcap_{t \in T} B_t = \{B \subseteq \mathbb{R} : B \in B_t \text{ for all } t \in T\} \text{ is a } \sigma\text{-field.}$$

a σ -field.

We shall build $\sigma(\ell)$ as follows:

$$\sigma(\ell) = \bigcap_{B \in \ell} B$$

B is σ -field

$$B \subseteq \ell$$

Sept 19, 2008 STAT 901

Last time: $\sigma(\ell) = \sigma$ -field generated by ℓ

Def: ① $\ell \subseteq \sigma(\ell)$, σ -field that contains ℓ .

② if B is any σ -field such that $\ell \subseteq B$, then $\sigma(\ell) \subseteq B$

Also

$$\sigma(\ell) = \bigcap_{\substack{B \text{ is } \sigma\text{-field} \\ \ell \subseteq B}} B$$

Borel sets (Application)

Let $\mathbb{R} = \mathbb{R}$. Let $\ell = \{(a, b] : -\infty \leq a \leq b < \infty\}$

Define $B(\mathbb{R}) = \sigma(\ell)$ we call the class of subsets of \mathbb{R} which are elements of $\sigma(\ell)$, the Borel sets.

What sorts of subsets of \mathbb{R} are Borel?

① open intervals

$$(a, b) = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}] \text{ including } (-\infty, +\infty)$$

② closed intervals

$$[a, b] = \bigcap_{n=1}^{\infty} [a - \frac{1}{n}, b] \text{ including points: } \{a\} = [a, a]$$

③ any countable subset of \mathbb{R}

$$C = \{x_1, x_2, x_3, \dots\} \quad C = \bigcup_{n=1}^{\infty} \{x_n\} = \bigcup_{n=1}^{\infty} [x_n, x_n]$$

including \mathbb{Q} rationals.

④ Any open subset of \mathbb{R}

$$\bigcup_{x \in X} (x, x+1)$$

Because every open set is a countable union of open intervals.

$$\bigcup_{n=1}^{\infty} (a_n, b_n) \in \mathcal{B}(\mathbb{R})$$

⑤ Any closed subset of \mathbb{R}

If C is closed, then $C^c = \mathbb{R}$ is open. (Def)

for example, the Cantor set, is a Borel sets.

⑥ Any set of the form

$$C_f = \{x : f(x) = a\} \text{ where } f \text{ is continuous is a Borel set.}$$

because C_a is closed.

$$(a, b) = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, b) \in \mathcal{B}.$$

$$\mathcal{P}(\mathbb{R}) \supset \mathcal{B}(\mathbb{R}) \text{ (axiom of choice)}$$

on page 17, it is shown that $\mathcal{B}(\mathbb{R}) = \sigma(\ell_1) = \sigma(\ell_2) = \sigma(\ell_3) = \sigma(\ell_4)$
 $= \sigma(C_3)$

$$\ell_1 = \{(a, b) : -\infty < a < b < \infty\}$$

$$\ell_2 = \{[a, b] : -\infty < a < b < \infty\}$$

$$\ell_3 = \{[a, b] : -\infty < a \leq b < \infty\}$$

$$\ell_4 = \{(-\infty, x) : x \in \mathbb{R}\}$$

$$C_5 = \{O : O \text{ open subset of } \mathbb{R}\}$$

We can also define the Borel subsets of some interval I , let us call this $\mathcal{B}(I)$

$$\text{Basic idea: } \mathcal{B}(I) = \{I \cap B : B \in \mathcal{B}(\mathbb{R})\}$$

$$\mathcal{B}(I) = \sigma(\{(a, b) : (a, b) \in I\})$$

These are equivalent definitions.

The reason for this is formalized as Thm. 1.8.1 in textbook.

Theorem 1.8.1 has an error in it. We must also assume that

\mathbb{R}_0 is a Borel set, i.e. $\mathbb{R}_0 \in \mathcal{B}(\mathbb{R})$.

You should read:

"Let $S_0 \subset S$, where $S_0 \in \mathcal{B}$ "

Additional assigned problems from Section 1.9

17, 18, 19, 21, 22, 24, 25, 27, 28, 36.

Hilroy

C12 probability spaces:

Def: A probability space is a triple (Ω, \mathcal{B}, P) where Ω is the sample space of outcomes of a random experiment, \mathcal{B} - σ -field of subsets of Ω , these subsets are called events

P - a real-valued function on \mathcal{B} .

$P: \mathcal{B} \mapsto \mathbb{R}$ called a probability measure, satisfying

① $P(A) \geq 0$ for all $A \in \mathcal{B}$

② If A_1, A_2, A_3, \dots is a sequence of pairwise disjoint events

then

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \text{ for all } n=1, 2, 3, \dots$$

finite additivity

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (\text{or additive})$$

Book has an error again. By including finite additivity in the proof.

③ $P(\Omega) = 1$

Sept 22, 2008 STAT 901 TA: Changguo Weng

Last time: (Ω, \mathcal{B}, P) prob. space Ω - sample space, \mathcal{B} - σ -field of events. P - prob. measure satisfying

① $P(A) \geq 0$ for any $A \in \mathcal{B}$

② P finitely additive & σ -additive.

③ $P(\Omega) = 1$

Proposition: $P(A^c) = 1 - P(A)$

proof: $A \cup A^c = \Omega$

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

Proposition 2: $P(\emptyset) = 0$ $P(\emptyset) = P(\emptyset \cup \emptyset \cup \dots) = \sum P(\emptyset)$

Proposition 3: $P(A \cup B) = P(A) + P(B) - P(AB)$

proof:



$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

$$\text{So } P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) \quad \text{by finite additivity}$$

$$= [P(A \setminus B) + P(B \setminus A)] + P(A \cap B) - P(AB)$$

$+ P(AB)$

$$\therefore A = (A \setminus B) + (AB) \quad B = (B \setminus A) + A \bar{B}$$

QED.

Prop 4. Inclusion-Exclusion Formula

$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k)$$

$$\dots (-1)^k P(A_1 \cap A_2 \cap \dots \cap A_n)$$

Proof left to you use prop 3 and mathematical induction QED

prop 5. $A \subset B$ implies $P(A) \leq P(B)$

prop 6. In general,

$$P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i) \quad \text{sub additivity}$$

$$P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i) \quad \text{aka. Boole's inequalities}$$

$$\begin{aligned} \text{Proof: } P(\bigcup_{i=1}^n A_i) &= P(A_1 + (A_2 \setminus A_1) + (A_3 \setminus (A_1 \cup A_2)) + \dots \\ &\quad (A_n \setminus (\bigcup_{i=1}^{n-1} A_i))) \end{aligned}$$

$$\begin{aligned} &= P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus (A_1 \cup A_2)) + \dots + P(A_n \setminus (\bigcup_{i=1}^{n-1} A_i)) \\ &\leq \sum_{i=1}^n P(A_i) \end{aligned}$$

and proposition 5 $A_1 \setminus (\bigcup_{i=1}^{n-1} A_i) \subset A_i$

The case $n=\infty$ is similar QED.

Is P continuous? That is, is $\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n)$ true

when $\lim_{n \rightarrow \infty} A_n$ exists? Yes. We shall prove this in several steps.

prop 7. For monotone sequences, P is continuous

(i) $A_n \nearrow A$ implies $P(A_n) \rightarrow P(A)$

(ii) $A_n \searrow A$ implies $P(A_n) \rightarrow P(A)$

Equivalently, (i) if $A_1 \subset A_2 \subset A_3 \subset \dots$ then $P(\bigcup_{i=1}^{\infty} A_i) = \lim_{n \rightarrow \infty} P(A_n)$

(ii) if $A_1 \supset A_2 \supset A_3 \supset \dots$ then $P(\bigcap_{i=1}^{\infty} A_i) = \lim_{n \rightarrow \infty} P(A_n)$

Sept 24. STAT 901

$A_n \rightarrow A \quad P(A_n) \rightarrow P(A)$

Hilary

$$(a) A_n \nearrow A \text{ then } P(A_n) \nearrow P(A) \quad \liminf_{n \rightarrow \infty} P(A_n) \neq P(\liminf_{n \rightarrow \infty} A_n)$$

$$(b) A_n \searrow A \text{ then } P(A_n) \searrow P(A) \quad P(A) + (A/A) = A$$

proof: (a) Let $A_1 \subset A_2 \subset \dots$ be monotone increasing, and let

$$A = \bigcup_{n=1}^{\infty} A_n$$

$$P(A) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(A_1 + (A_2 - A_1) + (A_3 - A_2) + (A_4 - A_3) + \dots)$$

$$= P\left(\sum_{i=1}^{\infty} (A_i - A_{i-1})\right) \text{ where } A_0 = \emptyset$$

$$\sum_{i=1}^{\infty} P(A_i - A_{i-1}) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i - A_{i-1})$$

$$= \lim_{n \rightarrow \infty} P\left(\sum_{i=1}^n (A_i - A_{i-1})\right) = \lim_{n \rightarrow \infty} P(A_n)$$

part (b). If $A_n \searrow A$, then $A_n \nearrow A^c$, so

$$P(A) = 1 - P(A^c) \text{ by part (a)}$$

$$= 1 - \lim_{n \rightarrow \infty} P(A_n^c)$$

$$= \lim_{n \rightarrow \infty} [1 - P(A_n^c)] = \lim_{n \rightarrow \infty} P(A_n)$$

which is the required result, QED.

Proposition(8). (Fatou's Lemma for events)

Any sequence of events A_1, A_2, \dots
we have

$$P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n)$$

Proof You should read this on page 32.

corollary: If $\lim_{n \rightarrow \infty} A_n = A$, then

$$\lim_{n \rightarrow \infty} P(A_n) = P(A)$$

proof: Suppose $\lim_{n \rightarrow \infty} A_n = A$, then

$$\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = (\lim_{n \rightarrow \infty} A_n) = A$$

So Fatou's lemma becomes

$$P(\lim_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P(\lim_{n \rightarrow \infty} A_n) = P(A)$$

$$\text{So } P(\lim_{n \rightarrow \infty} A_n) = \liminf_{n \rightarrow \infty} P(A_n) = \limsup_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n)$$

So $\lim_{n \rightarrow \infty} P(A_n)$ exists, and

$$P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n) \quad \text{QED.}$$

Example: Let Ω be \mathbb{R} , and let P be a probability measure on the Borel sets of \mathbb{R} , so

$$(\Omega, \mathcal{B}, P) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$$

$$\text{Define } F(x) = P(-\infty, x])$$

Then ① F is nondecreasing, if $x \leq y$, then $F(x) \leq F(y)$

$$\text{② } \lim_{x \rightarrow -\infty} F(x) = 1 \quad \& \quad \lim_{x \rightarrow -\infty} F(x) = 0$$

$$\text{③ } F(x) \text{ is right continuous. } (\lim_{y \rightarrow x^+} F(y)) = F(x). \quad \lim_{y \rightarrow x^+} F(y) = F(x)$$

Proof: ① This follows from prop 5, that is

If $A \subset B$, then $P(A) \leq P(B)$ here if $x \leq y$

$$[-\infty, x] \subset (-\infty, y]$$

$$P(-\infty, x] \leq P(-\infty, y]) \text{ or } F(x) \leq F(y)$$

②, ③ follow from the continuity of P under monotone unions, and intersections. For example.

Let $x_n \nearrow \infty$. Then $(-\infty, x_n] \nearrow \bigcup_{n=1}^{\infty} (-\infty, x_n] = \mathbb{R}$.

$$\text{So } 1 = P(\mathbb{R}) = \lim_{n \rightarrow \infty} P(-\infty, x_n]) \text{ by prop 7.}$$

$$= \lim_{n \rightarrow \infty} F(x_n)$$

The rest are similar. QED.

Dynkin's π - λ Theorem.

Mathematical Induction
for entirely σ -field? to \mathbb{N} ?

Sept 26 2008.

Dynkin's π - λ Theorem.

Definitions let \mathcal{P} be any collection of subsets of Ω , then \mathcal{P} is said to be a π -system or π -class if it is closed under finite intersections. That is, if $A, B \in \mathcal{P}$ then $AB \in \mathcal{P}$.

clearly, every field is a π -system, therefore, every σ -field is a π -system.

Example $\mathcal{F} = \{\text{open intervals } (a, b) : -\infty < a < b < \infty\}$

Definition (new): Let \mathcal{L} be a collection of subsets of Ω . Then \mathcal{L} is said to be a λ -system or a λ -class if the following three axioms hold.

$$\textcircled{A_1}: \Omega \in \mathcal{L}$$

$$\textcircled{A_2}: A \in \mathcal{L} \text{ implies } A^c \in \mathcal{L}$$

$$\textcircled{A_3}: \text{If } A_1, A_2, A_3, \dots \text{ is a sequence of pairwise disjoint sets in } \mathcal{L}, \text{ then } \bigcup_{n=1}^{\infty} A_n \in \mathcal{L}. \quad (\text{Does not imply } \sigma\text{-field?})$$

clearly, every σ -field is a λ -system is both a π -system and a λ -system.

Given two prob. measures P_1 and P_2 , on what class of sets must they agree? ~~A λ -system is the class of~~

The class of sets on which P_1 and P_2 agree is a λ -system.

Definition (old):

A collection \mathcal{L} of subsets of Ω is a λ -system if and only if

$$\textcircled{X_1} = \textcircled{A_1}: \Omega \in \mathcal{L}$$

$$\textcircled{X_2} \text{ if } A, B \in \mathcal{L} \text{ and } A \subset B, \text{ then } B - A \in \mathcal{L}$$

(closed under relative complementation)

$$\textcircled{X_3} \text{ monotone unions. if } A_1 \subset A_2 \subset A_3 \subset \dots \text{ and } A_i \in \mathcal{L} \text{ for all } i, \text{ then}$$

$\bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$. Definition: Let \mathcal{L} be a class of subsets of Ω . Then a class $\mathcal{L}(\mathcal{L})$ called the λ -system generated

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \iff \begin{bmatrix} \lambda'_1 \\ \lambda'_2 \\ \lambda'_3 \end{bmatrix}$$

by \mathcal{L} , is the smallest λ -system containing \mathcal{L} .

That is,

$$① \mathcal{L}(\mathcal{L}) \text{ is a } \lambda\text{-system, and } \mathcal{L} \subset \mathcal{L}(\mathcal{L})$$

if \mathcal{L}' is any λ -system such that $\mathcal{L} \subset \mathcal{L}'$, then

$$\mathcal{L}(\mathcal{L}) \subset \mathcal{L}'$$

It can be seen that $\mathcal{L}(\mathcal{L}) = \bigcap_{\substack{\mathcal{L} \text{ a } \lambda\text{-system} \\ \mathcal{L} \supset \mathcal{L}}} \mathcal{L}$.

Dynkin's π - λ Theorem If P is a π -system, and L is a λ -system such that $P \subset L$, then $\sigma(P) \subset L$

(b) If P is a π -system, then $\sigma(P) = L(P)$

[Note: (a) and (b) are essentially the same, suppose (b) is true and L is any λ -system such that $P \subset L$, then]

$$\sigma(P) \stackrel{(b)}{=} L(P) \subset L \text{ which is statement (a).}$$

Conversely, suppose (a) is true, Then since any σ -field is a λ -system, we get $L(P) \subset \sigma(P)$

From part (a) we get $\sigma(P) \subset L(P)$, so $L(P) = \sigma(P)$, which is part (b).

Proposition: Suppose ℓ is a π -system and also a λ -system. Then ℓ is a σ -field.

Proof: First we see that ℓ is a field:

$$\textcircled{1} \quad \emptyset \in \ell \text{ by } \textcircled{1}$$

$$\textcircled{2} \quad A \in \ell \text{ implies } A^c \in \ell \text{ by } \textcircled{2}$$

$$\textcircled{3} \quad A, B \in \ell \text{ implies } AB \in \ell \text{ by def. of } \pi\text{-system}$$

so ℓ is a field. Suppose A_1, A_2, A_3, \dots is a sequence in ℓ . Then

$$\bigcup_{i=1}^{\infty} A_i = \lim_{n \rightarrow \infty} \bigcup_{i=1}^n A_i$$

Now $\bigcup_{i=1}^n A_i \in \ell$ because ℓ is a field so $\bigcup_{i=1}^{\infty} A_i \in \ell$ by

$\textcircled{1}_3$

No class on FRIDAY

Oct 1 STAT 901

π -system P closed under finite intersections

λ -system L : $\lambda_1, \lambda_2 \in L$ closed $A \in L$, implies $A^c \in L$

λ_3 : $A_1, A_2, \dots \in L$, disjoint unions $\bigcup_{i=1}^{\infty} A_i \in L$

or $\lambda'_1, \lambda'_2, \lambda'_3$

Dynkin's π - λ theorem:

(a) P , π -system, L λ -system, such that $P \subset L$

then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

(b) \mathcal{P} π -system, then $\sigma(\mathcal{P}) = \mathcal{L}(\mathcal{P})$.

proof: page 38-40. please read the proof.

Corollary: Suppose P_1 and P_2 are prob. measures on (Ω, \mathcal{B}) and suppose \mathcal{P} is a π -system such that

$$P_1(A) = P_2(A) \text{ for all } A \in \mathcal{P}.$$

Then $P_1(B) = P_2(B)$ for all $B \in \sigma(\mathcal{P})$.

Proof: Let $\mathcal{L} = \{B \in \mathcal{B}; P_1(B) = P_2(B)\}$

with $\emptyset \in \mathcal{L}$, by assumption. Also \mathcal{L} is a λ -system (axiom $\lambda_1, \lambda_2, \lambda_3$ easily checked). By Dynkin's $\pi-\lambda$ theorem part (a) we have $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Thus $P_1(B) = P_2(B)$ for all $B \in \sigma(\mathcal{P})$. QED.

Corollary: Let P_1 and P_2 be two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with distribution functions $F_1(x)$ and $F_2(x)$ respectively. suppose that $F_1(x) = F_2(x)$ for all $x \in \mathbb{R}$. Then

$$P_1(B) = P_2(B) \text{ for all } B \in \mathcal{B}(\mathbb{R}).$$

Proof: Let $\mathcal{P} = \{(-\infty, x]; x \in \mathbb{R}\}$

Then \mathcal{P} is a π -system.

$$(-\infty, x] \cap (-\infty, y] = (-\infty, \min(x, y)] \in \mathcal{P}.$$

Also, $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$. Now $A \in \sigma(\mathcal{P}) \iff A = \bigcup_{x \in A} (-\infty, x]$

$$P_1((-\infty, x]) = F_1(x) = F_2(x) = P_2((-\infty, x])$$

$$\text{so } P_1(A) = P_2(A) \text{ for all } A \in \mathcal{P}.$$

By previous corollary, $P_1(B) = P_2(B)$ for all $B \in \mathcal{B}(\mathbb{R})$. QED.

How do we construct probability models?

Start with set Ω and consider a collection \mathcal{S} of subsets of Ω . We define $P(S)$ for every $S \in \mathcal{S}$, then extend P from \mathcal{S} to $\sigma(\mathcal{S})$ and let $\mathcal{B} = \sigma(\mathcal{S})$.

problems: How do we know the extension of P from \mathcal{S} to $\sigma(\mathcal{S})$ is unique?

That is, if $P_1(A) = P_2(A)$ for every $A \in \mathcal{S}$, is it true that

$P_1(B) = P_2(B)$ for every $B \in \mathcal{G}(\mathcal{S})$?

How do we know that an extension of P from \mathcal{S} to $\sigma(\mathcal{S})$ exists?
For example, $\mathcal{J}\mathcal{R} = \{\{1, 2, 3\}\}$

$$\mathcal{S} = \{\emptyset, \{1\}, \{2\}, \{3\}, \mathcal{J}\mathcal{R}\}$$

$$\text{Suppose, } P(\{1\}) = 1/2 \quad P(\{2\}) = 1/2 \quad P(\{3\}) = 1/2$$

$$P(\emptyset) = 0 \cdot P(\mathcal{J}\mathcal{R}) = 1.$$

This function on \mathcal{S} does not extend to a prob: measure on $\sigma(\mathcal{S})$

We can solve the uniqueness prob. by requiring that \mathcal{S} be a π -system.
What about existence?

Definition: A collection \mathcal{S} of subsets of $\mathcal{J}\mathcal{R}$ is called a semi-algebra

If:

$$\textcircled{1} \emptyset, \mathcal{J}\mathcal{R} \in \mathcal{S}.$$

\textcircled{2} \mathcal{S} is a π -system.

\textcircled{3} If $A \in \mathcal{S}$, there exists a positive integer n , and disjoint sets $C_1, C_2, \dots, C_n \in \mathcal{S}$, such that

$$A^c = \sum_{i=1}^n C_i = C_1 + C_2 + \dots + C_n.$$

STAT 901 Probability Theory Oct 6 2008.

$$P: \mathcal{S} \rightarrow [0, 1] \quad \text{goal.}$$

extend $P: \sigma(\mathcal{S}) \rightarrow [0, 1]$ uniquely.

Defn. \mathcal{S} is called a semi-algebra if

$$\textcircled{1} \emptyset, \mathcal{J}\mathcal{R} \in \mathcal{S}.$$

\textcircled{2} \mathcal{S} is a π -system.

\textcircled{3} If $A \in \mathcal{S}$, there exists a positive integer n , and disjoint sets $C_1, C_2, \dots, C_n \in \mathcal{S}$, such that $A^c = C_1 + C_2 + \dots + C_n$.

Example. $\mathcal{S} = \{(a, b] : -\infty < a \leq b < \infty\}$ $\mathcal{J}\mathcal{R} = \mathbb{R}$.

— By (a, ∞) we shall mean (a, ∞)

$$\textcircled{4} \emptyset \in \mathcal{S} \quad \therefore (a, a] = \emptyset.$$

$$\mathcal{J}\mathcal{R} \in \mathcal{S} \quad \therefore (-\infty, \infty) = \mathcal{J}\mathcal{R}.$$

\textcircled{5} \mathcal{S} is a π -system.

$$\textcircled{6} (a, b]^c = (-\infty, a] + (b, \infty) \quad A^c = C_1 + C_2$$

Hilroy

proposition: Let \mathcal{S} be a semi-algebra of subsets of Ω , Define

$$\mathcal{A}(\mathcal{S}) = \left\{ \sum_{i=1}^n S_i : S_1, \dots, S_n \text{ disjoint } \in \mathcal{S} \right\}$$

Then: $\mathcal{A}(\mathcal{S})$ is the field of subsets generated by \mathcal{S} .

proof: Note $\mathcal{S} \subset \mathcal{A}(\mathcal{S})$ because $S = S_1 \in \mathcal{A}(\mathcal{S})$

We show that $\mathcal{A}(\mathcal{S})$ is a field. To show that we check the three conditions

① $\Omega \in \mathcal{A}(\mathcal{S})$ True because $\Omega \in \mathcal{S}$

② We prove closure under finite intersections next.

Suppose $\sum_{i=1}^n S_i \in \mathcal{A}(\mathcal{S})$

$$\sum_{j=1}^m S_j' \in \mathcal{A}(\mathcal{S})$$

$$\left(\sum_{i=1}^n S_i \right) \cap \left(\sum_{j=1}^m S_j' \right) = \left(\sum_{i=1}^n \sum_{j=1}^m S_i S_j' \right)$$

Now $S_i S_j' \in \mathcal{S}$, because \mathcal{S} is a π -system.

$$\text{so } \sum_{i,j} S_i S_j' \in \mathcal{A}(\mathcal{S}).$$

by induction, $\mathcal{A}(\mathcal{S})$ is closed under finite intersections.

③ Finally, we prove closure of $\mathcal{A}(\mathcal{S})$ under complementation.

suppose that $\sum_{i=1}^n S_i \in \mathcal{A}(\mathcal{S})$ Then

$$\left(\sum_{i=1}^n S_i \right)^c = \bigcap_{i=1}^n S_i^c$$

$$\text{But } S_i^c = \sum_{j=1}^{n_i} C_{ij} \text{ where } C_{ij} \in \mathcal{S}$$

$$\text{so } \left(\sum_{i=1}^n S_i \right)^c = \bigcap_{i=1}^n \left(\sum_{j=1}^{n_i} C_{ij} \right)$$

But $\sum_{j=1}^{n_i} C_{ij} \in \mathcal{A}(\mathcal{S})$ and $\mathcal{A}(\mathcal{S})$ has been shown in step (2)

to be closed under finite intersections, so

$$\bigcap_{i=1}^n \left(\sum_{j=1}^{n_i} C_{ij} \right) \in \mathcal{A}(\mathcal{S})$$

That completes the proof $\mathcal{A}(\mathcal{S})$ is a field.

Since the elements of $\mathcal{A}(\mathcal{S})$ are generated by finite unions of elements

\mathcal{S} may not be closed under unions.
now to check σ -additivity

of \mathcal{S} , this must be the field generated by \mathcal{S} . QED

first extension Theorem:

Suppose \mathcal{S} is a semi-algebra of subsets of Ω , and suppose that

$P: \mathcal{S} \rightarrow [0, 1]$ is σ -additive on \mathcal{S} and $P(\Omega) = 1$

Then there is a unique extension of P to $A(\mathcal{S})$ defined by

$$P\left(\sum_{i=1}^n S_i\right) = \sum_{i=1}^n P(S_i)$$

which is a probability measure on $A(\mathcal{S})$. That is,

P is σ -additive with $P(\Omega) = 1$.

σ -additivity of P on \mathcal{S} .

$S_1, \dots, S_n \in \mathcal{S}$ disjoint

provided $\sum_{i=1}^n S_i \in \mathcal{S}$.

$$P\left(\sum_{i=1}^n S_i\right) = \sum_{i=1}^n P(S_i) \text{ or countable case}$$

$$P\left(\sum_{i=1}^{\infty} S_i\right) = \sum_{i=1}^{\infty} P(S_i) \quad \text{if } \sum_{i=1}^{\infty} S_i \in \mathcal{S}.$$

STAT 901 Wednesday Oct 8, 2008.

First extension theorem

Suppose ① \mathcal{S} is a semi-algebra of subsets of Ω .

② $P: \mathcal{S} \rightarrow [0, 1]$ is σ -additive

③ $P(\Omega) = 1$

Then there exists a unique extension of P to $A(\mathcal{S})$ such that

$$P\left(\sum_{i=1}^n S_i\right) = \sum_{i=1}^n P(S_i) \text{ and } P \text{ is a } \sigma\text{-additive probability measure}$$

on $A(\mathcal{S})$

proof: In book

Second extension theorem:

A σ -additive probability measure P defined on a field \mathcal{A} of subsets of Ω has a unique extension to a σ -additive probability measure on $\sigma(\mathcal{A})$.

Combo extension theorem.

FET + SET.

P σ -additive probability measure extends uniquely from \mathcal{S} to a σ -additive probability measure in $\sigma(\mathcal{A})$

proof: $\mathcal{S} \subset A(\mathcal{S}) \subset \sigma(A(\mathcal{S})) = \sigma(\mathcal{S})$

first ext. thm. second ext. thm.

QED.

Examples: ① $\Omega = \mathbb{R}$, suppose F is a distribution function on \mathbb{R} .
 i.e. F is nondecreasing, $F(-\infty) = 0$, $F(\infty) = 1$, F is right continuous.
 We construct a semi-algebra which generates $\mathcal{B}(\mathbb{R})$
 $\mathcal{S} = \{(a, b] \mid -\infty < a \leq b < \infty\}$ we adopt the convention $(a, \infty] = (a, \infty)$

$$[-\infty, \infty] = \mathbb{R}$$

We see that \mathcal{S} is a semi-algebra.

We define $P((a, b]) = F(b) - F(a)$
 where $P([a, \infty)) = 1 - F(a)$ etc.

We prove that P satisfies the assumptions of the first extension theorem.
 (The hardest step is showing that P is σ -additive).

$$\begin{array}{c} \text{---} \\ a \quad a_1 \dots a_n \quad b \end{array}$$

By comb. extension theorem, P extends uniquely to a probability measure
 on all the sets of $\mathcal{B}(\mathbb{R})$.

② Lebesgue measure on $\Omega = (0, 1]$

suppose $F(x) = x$ distribution function for uniform distribution

as earlier, we can define a prob. measure on the Borel sets of
 $(0, 1]$ called $\mathcal{B}((0, 1])$ for which $F(x) = x$ is the distribution
 function. This is the unique prob. measure with distribution func.

F. we write this prob. measure as λ . In particular,

$$\lambda(\emptyset) = 0, \lambda([0, 1]) = 1$$

$$\lambda(a, b] = b - a$$

λ is called Lebesgue measure on $(0, 1]$

③ Lebesgue measure on \mathbb{R} .

Let A be any Borel set of \mathbb{R} .

Define $A_n = A \cap (n, n+1]$ where $n \in \{0, \pm 1, \pm 2, \dots\}$

Let $A'_n = \{x \in A \mid x \in A_n\}$ Then $A'_n \in \mathcal{B}((0, 1])$

Define $\lambda(A) = \sum_{n=-\infty}^{+\infty} \lambda(A'_n)$

Then $\lambda: \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty]$ indeed $\lambda(\mathbb{R}) = \infty$

properties:

① $0 \leq \lambda(B) \leq \infty$ for all $B \in \mathcal{B}(\mathbb{R})$

② If B_1, B_2, B_3, \dots are disjoint, then

$$\lambda\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \lambda(B_k)$$

25.

Henn Lebesgue.

Assigned problems: section 2.6 1, 2, 3, 6, 8, 9*, 11, 12, 14

STAT 901 Oct 10. 2008

more work

Additional problems: section 2.6 15, 17, 25

$$\mathcal{L} \subset \mathcal{A}(\mathcal{S}) \subset \sigma(\mathcal{A}(\mathcal{S})) = \sigma(\mathcal{S})$$

Proof of $\sigma(\mathcal{A}(\mathcal{S})) = \sigma(\mathcal{S})$ from definition

① $\mathcal{L} \subset \mathcal{A}(\mathcal{S})$, therefore $\sigma(\mathcal{L}) \subset \sigma(\mathcal{A}(\mathcal{S}))$. (*)

② $\mathcal{A}(\mathcal{S}) \subset \sigma(\mathcal{S})$ therefore $\sigma(\mathcal{A}(\mathcal{S})) \subset \sigma(\sigma(\mathcal{S}))$ but $\sigma(\sigma(\mathcal{S})) = \sigma(\mathcal{S})$
so $\sigma(\mathcal{A}(\mathcal{S})) \subset \sigma(\mathcal{S})$. (**)

Statements (*) (***) imply that $\sigma(\mathcal{A}(\mathcal{S})) = \sigma(\mathcal{S})$

$$\textcircled{1} \quad (a) \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

(b) - - -

$$(c) (\limsup A_n)^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \right)^c = \bigcup_{n=1}^{\infty} \left(\bigcap_{k=n}^{\infty} A_k^c \right) \text{ Dem}$$

$$= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^{cc} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n = \liminf A_n$$

2(c) "closed under monotone limits" means minimal assumptions are

④ if $A_1, A_2, A_3, \dots \in \mathcal{L}$ and either

⑤ $A_1 \subset A_2 \subset A_3 \subset \dots$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$ or

⑥ $A_1 \supset A_2 \supset A_3 \supset \dots$ then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{L}$.

$$3 \quad P(\liminf A_n) = P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right)$$

$$= \liminf_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \quad B_n = \bigcap_{k=n}^{\infty} A_k, A_n \supset \bigcap_{k=n}^{\infty} A_k$$

$\times \lim_{n \rightarrow \infty} P(A_n)$ limit may not exist. $\therefore C B_2 \subset B_3 \subset \dots$ so $P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P(A_n)$ wrong statement.

$$= \liminf_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \leq \liminf_{n \rightarrow \infty} P(A_n)$$

4. Suppose that \mathcal{F}_n are fields satisfying $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, prove that $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ is also a field.

Proof: we check the axioms.

$$C, \tau \in \bigcup_{n=1}^{\infty} \mathcal{F}_n \quad \tau \in \mathcal{F}_n \text{ which is a field}$$

Hilary

(2) suppose $A \in \bigcup_{i=1}^{\infty} \mathcal{F}_i$, then $A \in \mathcal{F}_i$ for some i

It follows that $A^c \in \mathcal{F}_i$, because \mathcal{F}_i is a field. Therefore
 $A^c \in \bigcap_{i=1}^{\infty} \mathcal{F}_i$.

(3) suppose $A, B \in \bigcup_{i=1}^{\infty} \mathcal{F}_i$, so there exists an i, j st.
 $A \in \mathcal{F}_i, B \in \mathcal{F}_j$.

wlog. $i \leq j$, then $\mathcal{F}_i \subset \mathcal{F}_j$: so $A, B \in \mathcal{F}_j$.

Since \mathcal{F}_j is a field follows that $A \cup B \in \mathcal{F}_j$.

Therefore $A \cup B \in \bigcup_{i=1}^{\infty} \mathcal{F}_i$.

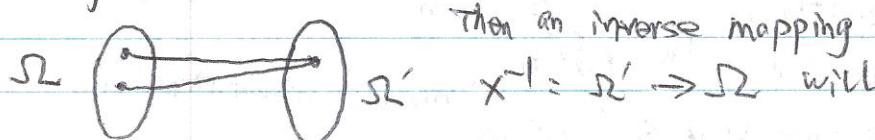
Oct 15, STAT 901 2008 MC 4060 11:30am ~ (2:20pm)

Random Variables Chapter 3

3.1 Inverse maps (mappings, functions)

Let Σ and Σ' be two sets and $X: \Sigma \rightarrow \Sigma'$ a function

In general, X is not 1-1

 Then an inverse mapping
 $\Sigma \xrightarrow{X} \Sigma' \quad X^{-1}: \Sigma' \rightarrow \Sigma$ will not in general exist. So

We define more generally.

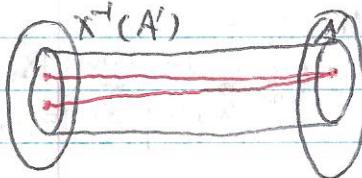
$$X^{-1}: \mathcal{P}(\Sigma') \rightarrow \mathcal{P}(\Sigma)$$

power set of Σ' power set of Σ

such that

$$X^{-1}(A') = \{w \in \Sigma: X(w) \in A'\}$$

The set $X^{-1}(A')$ is called the preimage of A'



Properties of X^{-1}

$$\textcircled{1} \quad X^{-1}(\emptyset) = \emptyset$$

$$\textcircled{2} \quad X^{-1}(\Sigma') = \Sigma$$

$$\textcircled{3} \quad X^{-1}(A'^c) = [X^{-1}(A')]^c$$

$$\textcircled{4} \quad X^{-1}\left(\bigcup_{t \in T} A'_t\right) = \bigcup_{t \in T} [X^{-1}(A'_t)]$$

$$\textcircled{5} \quad X^{-1}\left(\bigcap_{t \in T} A'_t\right) = \bigcap_{t \in T} [X^{-1}(A'_t)]$$

Definition: Let $X: \Sigma \rightarrow \Sigma'$ be a function and let \mathcal{C}' be a class of subsets of Σ' . We define

$$x^{-1}(\ell') = \{x^{-1}(c') : c' \in \ell'\}$$

Proposition: If \mathbb{B}' is a σ -field of subsets of Ω' , then $x^{-1}(\mathbb{B}')$ is a σ -field of subsets of Ω .

Proof: We check the axioms.

$$\textcircled{1} \quad \Omega = x^{-1}(\Omega') \quad \Omega' \in \mathbb{B}' \text{ because } \mathbb{B}' \text{ is a } \sigma\text{-field. So } \Omega \in$$

$$x^{-1}(\mathbb{B}')$$

\textcircled{2} Suppose that $A \in x^{-1}(\mathbb{B}')$ then there is some $A' \in \mathbb{B}'$ such that $A = x^{-1}(A')$

Now A'^c is an element of \mathbb{B}' because \mathbb{B}' is a σ -field.

$$\text{Then } A^c = [x^{-1}(A')]^c = [x^{-1}(A'^c)] \in x^{-1}(\mathbb{B}')$$

\textcircled{3} closure under countable unions.

Suppose $A_1, A_2, A_3, \dots \in x^{-1}(\mathbb{B}')$, then we can write

$$A_i = x^{-1}(A'_i) \text{ for each } i, \text{ where } A'_i \in \mathbb{B}'$$

But, $\bigcup_{i=1}^{\infty} A'_i \in \mathbb{B}'$, again, because \mathbb{B}' is a σ -field.

$$\text{so } \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} x^{-1}(A'_i) = x^{-1}\left(\bigcup_{i=1}^{\infty} A'_i\right) \in x^{-1}(\mathbb{B}') \text{ QED.}$$

$$\int A \in \Omega \quad x(A) = \{x(w) : w \in A\} \quad \text{is a } \sigma\text{-field}$$

Proposition: Let ℓ' be a class of subsets of Ω' . Then

$$x^{-1}[\sigma(\ell')] = \sigma[x^{-1}(\ell')] \quad (x^{-1}[A(\ell')] = A[x^{-1}(\ell')])$$

Proof: pages 73-74.

Definition: A pair (Ω, \mathbb{B}) where Ω is a set and \mathbb{B} is a σ -field on Ω , is called a measurable space.

Suppose (Ω, \mathbb{B}) and (Ω', \mathbb{B}') are measurable spaces, and $X: \Omega \rightarrow \Omega'$ is a mapping, then X is said to be measurable if

$$x^{-1}(\mathbb{B}') \subset \mathbb{B}.$$

Equivalently, X is measurable, if

$$x^{-1}(B') \in \mathbb{B} \quad \text{for all } B' \in \mathbb{B}'$$

Notation for this

$\therefore X$ measurable $\Leftrightarrow X: (\Omega, \mathbb{B}) \rightarrow (\Omega', \mathbb{B}')$

$$\therefore X \in \mathbb{B}/\mathbb{B}'$$

We also say that X is a random element of Ω' .

In particular, if $X: (\Omega, \mathbb{B}) \rightarrow (\mathbb{R}, \mathbb{B}(\mathbb{R}))$

then X is called a random variable.

$$P(a \leq X \leq b) = P(X^{-1}([a, b]))$$

$[a, b] \in \mathcal{B}(\mathbb{R})$, we need to calculate

$P(a \leq X \leq b)$, we want $X^{-1}(B') \in \mathcal{B}$ where $B' = [a, b]$

STAT 833 901 Oct 17

$$X: \Omega \rightarrow \Omega'$$

(Ω, \mathcal{B}) (Ω', \mathcal{B}') msf spaces

$$X \text{ msf} \Leftrightarrow X^{-1}(B') \subset B$$

we also write $X \in \mathcal{B}/\mathcal{B}'$

$$X: (\Omega, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ random variable}$$

Induced distributions

Measurable functions can be used to "transfer" probability measures from one measurable space to another.

$$\text{Suppose we have } X: (\Omega, \mathcal{B}) \rightarrow (\Omega', \mathcal{B}')$$

and suppose that (Ω, \mathcal{B}, P) is a probability space. We make (Ω', \mathcal{B}') into a probability space namely

$$(\Omega', \mathcal{B}', P \circ X^{-1})$$
 by defining

$$(P \circ X^{-1})(B') = P(X^{-1}(B')) = P\{\omega \in \Omega : X(\omega) \in B'\}$$

Then $P \circ X^{-1}$ is called induced distribution or the induced probability measure on Ω'

Example $X: (\Omega, \mathcal{B}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ i.e. X is a RV

For any $A \in \mathcal{B}(\mathbb{R})$

$$(P \circ X^{-1})(A) = P(X^{-1}(A)) = P\{\omega \in \Omega : X(\omega) \in A\}$$

$$\text{we write } (P \circ X^{-1})(A) = P(X \in A)$$

$$(P \circ X^{-1})((-\infty, x]) = P(X \leq x) = F_X(x)$$

Def. of msf $X: X^{-1}(B') \subset B$

Test for measurability of X . (prop. 3.2.1)

Suppose $X: \Omega \rightarrow \Omega'$ is a mapping, and (Ω, \mathcal{B}) (Ω', \mathcal{B}') are measurable spaces. Also, suppose

$$B' = \sigma(\ell')$$

Then X is measurable if and only if $\sigma(\ell) \circ X \in \mathcal{B}$

$$X^{-1}(l') \subset \mathbb{B}$$

Proof: check this yourself.

Corollary: $X: \Omega \rightarrow \mathbb{R}$ is a random variable if and only if

$$X^{-1}((-\infty, x]) \in \mathbb{B} \text{ for all real } x.$$

Counter example

$$\Omega = \{0, 1\} \quad \mathbb{B} = \{\emptyset, \Omega\} \quad P(\Omega) = 1 - P(\emptyset) = 0$$

$$X(w) = w$$

X not measurable.

$$X^{-1}((0, 1)) \notin \mathbb{B}.$$

So X is not a random variable.

Property of measurable functions

Proposition 3.2.2. (Important) Measurability is preserved under composition of functions. That is,

$$x_1: (\Omega_1, \mathbb{B}_1) \rightarrow (\Omega_2, \mathbb{B}_2)$$

$$x_2: (\Omega_2, \mathbb{B}_2) \rightarrow (\Omega_3, \mathbb{B}_3)$$

The composition is

$$\begin{array}{ccc} \Omega_1 & \xrightarrow{x_1} & \Omega_2 & \xrightarrow{x_2} & \Omega_3 \\ & \searrow & & \nearrow & \\ & & x_2 \circ x_1 & & \end{array}$$

$$(x_2 \circ x_1)(w_1) = x_2(x_1(w_1))$$

Then $x_2 \circ x_1$ is measurable from (Ω_1, \mathbb{B}_1) to (Ω_3, \mathbb{B}_3)

In other words,

$$x_1 \in \mathbb{B}_1 / \mathbb{B}_2 \quad x_2 \in \mathbb{B}_2 / \mathbb{B}_3 \Rightarrow x_2 \circ x_1 \in \mathbb{B}_1 / \mathbb{B}_3$$

Suppose we have a function $X: \Omega_1 \rightarrow S$ where (Ω_1, \mathbb{B}) is a measurable space and (S, d) is a metric space with metric d .

If we can make (S, d) into a measurable space, then we can say that X is a measurable mapping and therefore a random element of S .

Definition of metric space:

A metric space is a set S and a binary function $d: S \times S \rightarrow [0, +\infty)$ such that

$$(a) d(x, x) = 0 \text{ for all } x \in S$$

$$(b) d(x, y) = d(y, x) \text{ for all } x, y \in S$$

$$(c) d(x, y) + d(y, z) \geq d(x, z)$$

Examples: pseudo-metric space

$$d(x, y) = 0 \Leftrightarrow x = y$$

metric-space

$$d(x, y) = 0 \Rightarrow x = y$$

Examples: ① $S = \mathbb{R}$, $d(x, y) = |x - y|$

② $S = \mathbb{R}^k$, $d(x, y) = \|x - y\|$

③ $S = \{\text{continuous real valued functions defined on } [0, 1]\}$

$$d(f, g) = \sup_{0 \leq t \leq 1} |f(t) - g(t)|$$

To make S measurable, we first define the open sets \mathcal{O} of S .

Definition: A subset $U \subset S$ is said to be open if for all $x \in U$, there exists an $\varepsilon > 0$ such that for all $y \in S$

$$d(x, y) < \varepsilon \Rightarrow y \in U$$

Let \mathcal{O} be the class of open sets of a metric space S . We define the Borel sets of S to be

$$\mathcal{B} = \sigma(\mathcal{O})$$

Definition: Let (Ω, \mathcal{B}, P) be a probability space and (S, d) a metric space. We call a measurable function $X: \Omega \rightarrow S$ a random element of S .

$$(\mathcal{B}(\Omega), \mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}^k))$$

Special case: $X: \Omega \rightarrow \mathbb{R}^k$ where $d(x, y) = \|x - y\|$.

Then $\mathcal{B} = \mathcal{B}(\mathbb{R}^k)$ Borel sets of \mathbb{R}^k , then X is called a random vector.

Proposition: If (S_1, d_1) and (S_2, d_2) are metric spaces, and $X: S_1 \rightarrow S_2$ is continuous, then X is measurable.

proof: Let Ω_1 be the class of open sets of S_1 and Ω_2 be the class of open sets of S_2 , since $\mathcal{S}_2 = \sigma(\Omega_2)$
it suffices to show that

$$x^{-1}(\Omega_2) \subset \mathcal{S}_1$$

But this follows because Topology: preimage preserves 'open'?

$$x^{-1}(\Omega_2) \subset \Omega_1 \subset \sigma(\Omega_1) = \mathcal{S}_1$$

as required. QED.

Corollary if $g: \mathbb{R}^k \rightarrow \mathbb{R}$ is continuous, then g is measurable.

Corollary suppose $X_i: \Omega \rightarrow \mathbb{R}$ is a random variable for all $i=1, \dots, n$. Then the following are random variables.

$$\textcircled{1} \quad \sum_{i=1}^n X_i$$

$$\textcircled{2} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\textcircled{3} \quad \bigvee_{i=1}^n X_i = \max(X_1, \dots, X_n)$$

$$\textcircled{4} \quad \bigwedge_{i=1}^n X_i = \min(X_1, \dots, X_n)$$

$$\textcircled{5} \quad \sum_{i=1}^n X_i^2 \text{ etc.}$$

Proof: All of these examples can be written as

$$\Omega \xrightarrow{(X_1, \dots, X_n)} \mathbb{R}^n \xrightarrow{g} \mathbb{R}$$

So for example, in case \textcircled{1} $g(X_1, \dots, X_n) = \sum_{i=1}^n X_i$

In all cases, the function $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous.

If it can be shown that $X = (X_1, \dots, X_n)$ is a random vector, then it will follow that $g \circ X$ is a random variable since g , being continuous is measurable. This result follows from another proposition

proposition: $X = (X_1, \dots, X_n)$ is a random vector iff

X_i is a random variable for all $i=1, \dots, n$

proof see page 81.

proposition: If X_1, X_2, \dots are ^{countable} random variables, then so are

$$\textcircled{1} \quad \bigvee_{i=1}^n X_n = \sup_{n \in \mathbb{N}} X_n$$

$$\textcircled{2} \quad \bigwedge_{i=1}^n X_n = \inf_{n \in \mathbb{N}} X_n$$

③ $\limsup_{n \rightarrow \infty} X_n$ ④ $\liminf_{n \rightarrow \infty} X_n$ ⑤ $\lim_{n \rightarrow \infty} X_n$ provided this limit exists
 for all $w \in \mathbb{R}$. $\mathcal{B}(-\infty, \infty]$ extended real line

Proof of ① It suffices to show that

$\left[\bigvee_{n=1}^{\infty} X_n \leq x \right]$ is a measurable set for all x .

That is, $\{w : \bigvee_{n=1}^{\infty} X_n(w) \leq x\}$ is a measurable subset of \mathbb{R} .

Now,

$$\left[\bigvee_{n=1}^{\infty} X_n \leq x \right] = \bigcap_{n=1}^{\infty} [X_n \leq x] \text{ But}$$

$[X_n \leq x]$ is measurable because X_n is a random variable.

so $\bigcap_{n=1}^{\infty} [X_n \leq x]$ is measurable also, as required.

The others are similar. QED

σ -fields generated by X .

Suppose $X : \Omega \rightarrow \mathbb{R}$ is a random variable.

Define $\sigma(X) = " \sigma\text{-field generated by } X "$

$$= X^{-1}(\mathcal{B}(\mathbb{R}))$$

Equivalently, $\sigma(X) = \{[X \in A] : A \in \mathcal{B}(\mathbb{R})\}$

This is the smallest σ -field for which X is measurable.

Assigned problems = section 3.4 2, 3, 4, 7, 8, 11, 12, 13

STAT 901 Oct 22, 2008

$$\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$$

when X_t $t \in T$ is a collection of random variables, define

$$\sigma(X_t, t \in T) = \sigma(\bigcup_{t \in T} \sigma(X_t))$$

A comment on this. In general, $\sigma(\sigma(X_t))$ is not a σ -field. But $\sigma(\bigcup_{t \in T} \sigma(X_t))$ is the smallest σ -field containing $\sigma(X_t)$ for all $t \in T$, we

shall write $\bigvee_{t \in T} \mathcal{B}_t = \sigma(\bigcup_{t \in T} \mathcal{B}_t)$ for any indexed collection of σ -fields.

$$\text{so } \sigma(X_t, t \in T) = \bigvee_{t \in T} \sigma(X_t)$$

Example $X = 1_A$, where A is an event in \mathcal{I}_Ω .

$$1_A(w) = \begin{cases} 1 & w \in A \\ 0 & w \notin A \end{cases}$$

$$\text{Then } \sigma(1_A) = \{ [x \in B] : B \in \mathcal{B}(\mathbb{R}) \}$$

$$\text{But } [x \in B] = \begin{cases} \emptyset & 0, 1 \notin B \\ A & 0 \in B, 1 \notin B \\ A^c & 0 \in B, 1 \notin B \\ \Omega & 0 \in B, 1 \in B \end{cases}$$

$$\sigma(1_A) = \{ \emptyset, A, A^c, \Omega \}$$

$$[x \in B] = X'(B) \quad P[x \in B] = P[X'(B)]$$

4. Independence

Definition: Events A_1, A_2, \dots, A_n are independent if

$$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$$

for any subset $I \subset \{1, 2, 3, \dots, n\}$ ($2^n - n - 1$ non-trivial properties to check)

$$\text{i.e., } P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \dots P(A_{i_k})$$

for all $1 \leq i_1 < i_2 < \dots < i_k \leq n$, and all k .

Definition: Let $\ell_1, \ell_2, \dots, \ell_n$ be n collections of events on some Ω .

They are said to be independent if A_1, A_2, \dots, A_n are independent for all choices $A_i \in \ell_i, i=1, \dots, n$.

In particular, when σ -fields are independent? $\ell_1, \ell_2, \dots, \ell_n$ independent.

$$\Rightarrow \sigma(\ell_1), \sigma(\ell_2), \dots, \sigma(\ell_n) \text{ independent}$$

No, in general.

However,

Theorem: If ℓ_i is a π -system for every $i=1, \dots, n$ and if $\ell_1, \ell_2, \dots, \ell_n$ are independent, then $\sigma(\ell_1), \sigma(\ell_2), \dots, \sigma(\ell_n)$ are independent.

Proof: We use Dynkin's $\pi-\lambda$ -theorem. We start the proof by induction on the value of n .

Suppose $n=2$. Fix an event $A_2 \in \ell_2$. Define the following
 A is an event.

$$\mathcal{L} = \{A \in \mathbb{B} : P(AA_2) = P(A)P(A_2)\}$$

We show that \mathcal{L} is a λ -system.

① $\emptyset \in \mathcal{L}$, because

$$P(\emptyset A_2) = P(\emptyset) = P(A_2) \cdot P(\emptyset)$$

② If $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$ because

$$P(A^c A_2) = P(A_2) - P(AA_2) = P(A_2) - P(A_2)P(A)$$

③ If B_1, B_2, B_3, \dots are disjoint events of \mathcal{L} , then

$\sum_{i=1}^{\infty} B_i \in \mathcal{L}$, because

$$\begin{aligned} P(A_2 \sum_{i=1}^{\infty} B_i) &= P\left(\sum_{i=1}^{\infty} (A_2 B_i)\right) = \sum_{i=1}^{\infty} P(A_2 B_i) = \sum_{i=1}^{\infty} P(A_2) \cdot P(B_i) \\ &= P(A_2) \sum_{i=1}^{\infty} P(B_i) = P(A_2) P\left(\sum_{i=1}^{\infty} B_i\right) \end{aligned}$$

\mathcal{L} is a λ -system.

So \mathcal{L} is a λ -system. Also $\ell_1 \subset \mathcal{L}$, because

ℓ_1 and ℓ_2 are independent, and $A_2 \in \ell_2$.

So by Dynkin's π - λ system, $\sigma(\ell_1) \subset \mathcal{L}$

It follows that $P(A_1 A_2) = P(A_1)P(A_2)$ for all $A_1 \in \sigma(\ell_1)$ and our fixed choice for $A_2 \in \ell_2$. But the choice of $A_2 \in \ell_2$ was arbitrary, so $P(AA_2) = P(A)P(A_2)$ for all $A_1 \in \sigma(\ell_1)$ and all $A_2 \in \ell_2$.

Now repeat the argument fixed $A_1 \in \sigma(\ell_1)$ and using

$$\mathcal{L}' = \{A \in \mathbb{B} : P(A, A) = P(A)P(A)\}$$

Show \mathcal{L}' is again a λ -system and $\ell_2 \subset \mathcal{L}'$.

So by π - λ theorem, $\sigma(\ell_2) \subset \mathcal{L}'$. Therefore,

$$P(A_1 A_2) = P(A_1)P(A_2)$$

The rest of the proof by induction is left to you. QED

STAT 901

ℓ_1, \dots, ℓ_n indep. $\{\ell_i\}$'s π -system } $\Rightarrow \sigma(\ell_1), \dots, \sigma(\ell_n)$ indep.

Generalization of definition:

For infinitely many classes, say $C_t, t \in T$ we say the classes are independent, if $\ell_t, t \in I$ is an independent collection for every finite subset $I \subset T$. Note T can be uncountable. P continuous measure

Definition: Random variables $X_t, t \in T$ are independent if and only if the σ -field $\sigma(X_t)_{t \in T}$ are independent.

Factorization Criterion: Random variables $X_t, t \in T$ are independent if and only if $P(X_t \leq x_t, \text{ for all } t \in I) = \prod_{t \in I} P(X_t \leq x_t)$

for all finite subsets $I \subset T$, and all real x_t .

Proof: Suppose $X_t, t \in T$ are independent and suppose $I \subset T$ is a finite subset. Now

$$[X_t \leq x_t] \in \sigma(X_t)$$

it follows that

$$P(X_t \leq x_t \text{ for all } t \in I) = \prod_{t \in I} P(X_t \leq x_t)$$

directly from the definition of independence of random variables.

Conversely, suppose $P(X_t \leq x_t, t \in I) = \prod_{t \in I} P(X_t \leq x_t)$ for all finite $I \subset T$. Define $\ell_t = \{[X_t \leq x] : x \in \mathbb{R}\}$

Then $\ell_t, t \in T$ are independent. Furthermore ℓ_t is a π -system for all $t \in T$. So $\sigma(\ell_t), t \in T$ are independent for all finite $I \subset T$

Therefore, $\sigma(\ell_t), t \in T$ are independent.

$$\text{But } \sigma(\ell_t) = \sigma(X_t^{-1}(\mathcal{D})) \text{ where } \mathcal{D} = \{(-\infty, x] : x \in \mathbb{R}\}$$

$$= X_t^{-1}(\sigma(\mathcal{D})) = X_t^{-1}(\mathcal{B}(\mathbb{R})) = \sigma(X_t)$$

So $\sigma(X_t), t \in T$ are independent. Therefore, $X_t, t \in T$ are independent. □

Corollary: X_1, \dots, X_n independent if and only if

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(X_1 \leq x_1)P(X_2 \leq x_2) \cdots P(X_n \leq x_n)$$

Proof: We get a factorization criterion for all subsets $T \subset \{1, 2, \dots, n\}$ by letting $x_i \rightarrow \infty$ for $i \notin T$

$$\lim_{\substack{x_i \rightarrow \infty \\ i \notin T}} P(X_1 \leq x_1, \dots, X_n \leq x_n) = \lim_{\substack{x_i \rightarrow \infty \\ i \notin T}} P(X_1 \leq x_1) \cdots P(X_n \leq x_n)$$

$$\Rightarrow P(X_i \leq x_i, i \in T) = \prod_{i \in T} P(X_i \leq x_i)$$

Corollary Discrete X_1, X_2, \dots, X_n are independent if and only if their range is countable.

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) P(X_2 = x_2) \cdots P(X_n = x_n)$$

for all real x_i .

Proof: See pages 94-95: Ω is countable $\Rightarrow T \subset \Omega$ is also countable

$$\Omega = \bigcup_{C \in \mathcal{C}} C \text{ where } \mathcal{C} = \{x \in \Omega : x \in C\}$$

$$G(\omega) = \omega^{-1}(P(C))$$

$$X(\Omega) \subset \bigcup_{C \in \mathcal{C}} C \text{ where } \mathcal{C} = \{x \in \Omega : x \in C\}$$

A_1, A_2, A_3, \dots independent

$$P\left(\bigcap_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} \prod_{i=1}^n P(A_i) = \prod_{i=1}^{\infty} P(A_i)$$

skip section 4.3. Read

In section 4.4. we pick up the grouping lemma.

$\mathcal{B}_t, t \in T$ independent

Grouping Lemma. Let $\mathcal{B}_t, t \in T$ be indep. σ -fields. Let S be an index set, for each $s \in S$, suppose we have a subset $T_s \subset T$ such that

$$T_s \cap T_{s_2} = \emptyset \text{ for all } s_1 \neq s_2$$

Define $\mathcal{B}_{T_s} = \bigvee_{t \in T_s} \mathcal{B}_t = \sigma(\bigcup_{t \in T_s} \mathcal{B}_t)$, Then

$\mathcal{B}_{T_s}, s \in S$ are independent σ -fields.

(Notation $X \perp\!\!\!\perp Y$ (X, Y indep.) $\Leftrightarrow (x_1 = x_2, y_1 = y_2) \in \sigma(X, Y)$)

$\mathcal{B}_1 \perp\!\!\!\perp \mathcal{B}_2$ ($\mathcal{B}_1, \mathcal{B}_2$ indep.)

STAT 901 Oct 27 2008

Borel - Cantelli Lemma

Reminder:

$$[A_n, i_0] = \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

Let $A_n, n \geq 1$ be any sequence of events.

If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n, i_0) = 0$.

$$\text{Proof: } 0 \leq P(A_n, i_0) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \quad B_n = \bigcup_{k=n}^{\infty} A_k$$

$$= P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right)$$

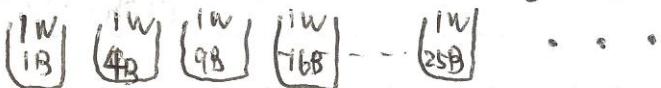
continuity

$$\leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) \quad (\text{sum over tail of a convergent series})$$

$$= 0$$

So $0 \leq P(A_n, i_0) \leq 0$ Therefore $P(A_n, i_0) = 0$ \blacksquare

Example Consider a sequence of urns filled with black and white balls.



A ball is drawn at random from each urn. Let A_n be the event that the ball drawn from urn n is white.

so $[A_n, i_0]$ is the event that infinitely many white balls are drawn

$$P(A_n, i_0) = 0 \text{ because } P(A_n) = \frac{1}{n+1} \text{ & } \sum P(A_n) < \infty$$

Borel zero-one law

Suppose $A_n, n \geq 1$ are independent, then $P(A_n, i_0) = \begin{cases} 0 & \text{if } \sum_{n=1}^{\infty} P(A_n) \\ 1 & \text{if } \sum_{n=1}^{\infty} P(A_n) = \infty \end{cases}$

proof: Since we have proved the Borel-Cantelli Lemma (without independence), it suffices to show that

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(A_n, i_0) = 1 \quad (\text{using independence}) \text{ So}$$

$$\begin{aligned} P(A_n, i_0) &= P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1 - P\left(\liminf_{n \rightarrow \infty} A_n^c\right) \quad \begin{array}{l} A_n \text{ indep.} \\ \Leftrightarrow \end{array} \\ &= 1 - P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = 1 - \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \quad \begin{array}{l} A_n^c \text{ indep.} \\ \Leftrightarrow \end{array} \\ &= 1 - \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} P(A_k^c) \quad \begin{array}{l} A_n \text{ indep.} \\ \Leftrightarrow \end{array} \\ &\quad \begin{array}{l} \text{I} \\ \text{II} \end{array} \end{aligned}$$

$\{A_n, A_n^c, \Omega\}$ indept.

$$= 1 - \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} [1 - P(A_k)]$$

$$\geq 1 - \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} e^{-P(A_k)}$$

$$= 1 - \lim_{n \rightarrow \infty} e^{-\sum_{k=n}^{\infty} P(A_k)}$$

$$= 1 - e^{-\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k)}$$

$$= 1 - e^{-\infty} = 1 - 0 = 1$$

so $P(A_n, i^c) \geq 1$ Therefore $P(A_n, i^c) = 1$

$$\infty = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{n-1} P(A_k) + \sum_{k=n}^{\infty} P(A_k) \quad \text{B.C.}$$

$$\sum_{k=n}^{\infty} P(A_k) < \infty \Rightarrow P(A_n, i^c) = 0$$

$$\text{so } \sum_{k=n}^{\infty} P(A_k) = \infty \text{ for all } n \Rightarrow$$

B-0-1 Law

$$\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = \infty \quad \text{(indep)} \quad \sum_{k=n}^{\infty} P(A_k) < \infty \Rightarrow P(A_n, i^c) = 1$$

Examples

$$\begin{array}{|c|c|c|} \hline \text{W} & \text{W} & \text{W} \\ \hline P & P & P \\ \hline 1B & 2B & 3B \\ \hline \end{array}$$

consider $p=1, 2, 3, \dots$ then $\sum_{k=1}^{\infty} \frac{1}{k+1} = \infty \Rightarrow P(A_n, i^c) = 1$

A_n = event that a white ball is drawn from urn n

Then A_n 's independent

$$\text{so } p=1 \quad \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty \Rightarrow P(A_n, i^c) = 1$$

$$p > 1: \sum_{k=1}^{\infty} \frac{1}{np+1} < \infty \Rightarrow P(A_n, i^c) = 0$$

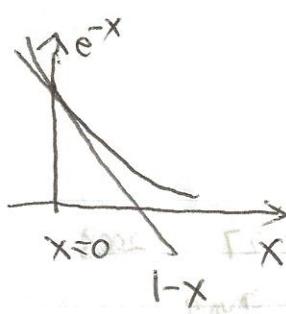
STAT 901 Oct 31, 2008

$$A_n, i^c \quad P(A_n, i^c) = \begin{cases} 0 & \sum_{k=n}^{\infty} P(A_k) < \infty \\ 1 & \sum_{k=n}^{\infty} P(A_k) = \infty \end{cases}$$

A_n independent

problems for chapter 3 Additional

3.4; 14, 16, 18, 20, 22.



$$1-x \leq e^{-x} \text{ for all real } x$$

then $\sum_{k=n}^{\infty} P(A_k) < \infty \Rightarrow \sum_{k=n}^{\infty} P(A_k) = 0$

$\sum_{k=n}^{\infty} P(A_k) = 0 \Rightarrow P(A_n, i^c) = 1$

$\sum_{k=n}^{\infty} P(A_k) > 0 \Rightarrow P(A_n, i^c) = 0$

$\sum_{k=n}^{\infty} P(A_k) = 1 \Rightarrow P(A_n, i^c) = 0$

$\sum_{k=n}^{\infty} P(A_k) < 1 \Rightarrow P(A_n, i^c) = 1$

Kolmogorov Zero-One Law

Let x_1, x_2, x_3, \dots be a sequence of random variables

Define $\mathcal{F}'_n = \sigma(x_{n+1}, x_{n+2}, \dots) = \sigma(\bigcup_{i=n+1}^{\infty} \sigma(x_i))$

The tail σ -field of $X_n, n \geq 1$ is defined as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}'_n$$

Comment The tail σ -field is the collection of events that can be described by limiting properties of the sequence $X_n, n \geq 1$. Events such as

$$[X_n \text{ converges}] = \{w \in \Omega : X_n(w) \text{ has a limit}\}$$

$$[X_n \rightarrow a] = \{w \in \Omega : \lim_{n \rightarrow \infty} X_n(w) = a\}$$

$$[\sum_{n=1}^{\infty} X_n \text{ has a limit}] \in \mathcal{T}$$

$$[\sum_{n=1}^{\infty} X_n = a] \notin \mathcal{T} \text{ generally.}$$

$$[\frac{1}{n} \sum_{i=1}^n X_i \rightarrow a \text{ as } n \rightarrow \infty] = [\lim_{n \rightarrow \infty} \bar{X}_n = a] \in \mathcal{T}$$

$$[X_n > 0 \text{ i.o.}] \in \mathcal{T}$$

Theorem: Suppose $X_n, n \geq 1$ are independent with tail σ -field \mathcal{T} , Let

$A \in \mathcal{T}$, Then

$$P(A) = 0 \text{ or } P(A) = 1$$

"Proof" Note that $\sigma(x_1, \dots, x_n)$ and $\sigma(x_{n+1}, x_{n+2}, \dots)$ are independent σ -fields, because the collections of random variables are disjoint. Since $\mathcal{T} \subset \sigma(x_{n+1}, x_{n+2}, \dots)$ it follows that

$\sigma(x_1, \dots, x_n) \perp\!\!\!\perp \mathcal{T}$ for all n .

$$\text{so } \bigcup_{n=1}^{\infty} \sigma(x_1, \dots, x_n) \perp\!\!\!\perp \mathcal{T}$$

It can be checked that $\bigcup_{n=1}^{\infty} \sigma(x_1, \dots, x_n)$ is a π -system (Do this) so

$$\sigma(X_1, X_2, \dots) = \bigvee_{n=1}^{\infty} \sigma(x_1, \dots, x_n) = \sigma\left(\bigcup_{n=1}^{\infty} \sigma(x_1, \dots, x_n)\right) \perp\!\!\!\perp \mathcal{T}$$

$$\sigma(X_1, X_2, \dots)$$

The last equality follows from the fact that

ℓ_1, ℓ_2 are π -systems, $\ell_1 \amalg \ell_2$, then $\sigma(\ell_1) \amalg \sigma(\ell_2)$

But $\sigma(X_{n+1}, X_{n+2}, \dots) \subset \sigma(X_1, X_2, \dots)$ for all n .

so $\mathcal{T} \subset \sigma(X_{n+1}, X_{n+2}, \dots) \subset \sigma(X_1, X_2, \dots)$

$$\boxed{\begin{array}{l} \sigma(X_1, X_2, \dots) \amalg \mathcal{T} \\ \mathcal{T} \subset \sigma(X_1, X_2, \dots) \end{array}} \Rightarrow \mathcal{T} \amalg \mathcal{T}$$

$$\overbrace{\bigcup_{n=1}^{\infty} \sigma(X_1, \dots, X_n)}^{\text{Pi-system}} \amalg \mathcal{T}$$

so $\Lambda \amalg \Lambda$ for all $\Lambda \in \mathcal{T}$

$$P(\Lambda \cap \Lambda) = P(\Lambda)P(\Lambda)$$

$$\text{or } P(\Lambda) = P(\Lambda)^2$$

$$\text{or } P(\Lambda) - P(\Lambda)^2 = 0$$

$$P(\Lambda) [1 - P(\Lambda)] = 0 \Rightarrow P(\Lambda) = \begin{cases} 0 \\ 1 \end{cases}$$

X_1, X_2, X_3, \dots iid $N(0, 1)$

$P(X_n > n, \omega) = 0$ Borel zero-one law.

$= \begin{cases} 0 \\ 1 \end{cases}$ Kolm 0-1 law.

$P(\bar{X}_n \rightarrow 0) = 1$ Strong law of large numbers

$= \begin{cases} 0 \\ 1 \end{cases}$ Kolm 0-1 law

Assigned problems 4-6:

1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 18, 27.

Nov. 3 2008 STAT 901

Chapter 5, integration and expectation

Simple random variables

Let $X: \Omega \rightarrow \mathbb{R}$ be a measurable function (or random variable)

We say that X is a simple function (or simple random variable) if it has a finite range:

$X(w) \in \{a_1, a_2, \dots, a_k\}$ for all $w \in \Omega$, for some k .

Define $A_i = X^{-1}(\{a_i\})$ for $i = 1, \dots, k$

Then $X = \sum_{i=1}^k a_i 1_{A_i}$

Conversely, if $X = \sum_{i=1}^k a_i 1_{A_i}$, then X is a simple random variable

(provided A_1, \dots, A_k are measurable). Note however, that in this case $\{A_1, \dots, A_k\}$ may not be the range if A_1, \dots, A_k are not disjoint.

The set of all simple random variables which we denote by \mathcal{E} forms

a vector space:

$$\textcircled{1} \quad X = \sum_{i=1}^k a_i \mathbb{1}_{A_i} \Rightarrow \alpha X = \sum_{i=1}^k (\alpha a_i) \mathbb{1}_{A_i} \in \mathcal{E}$$

$$\textcircled{2} \quad X = \sum_{i=1}^k a_i \mathbb{1}_{A_i}, Y = \sum_{j=1}^m b_j \mathbb{1}_{B_j} \Rightarrow$$

$$X + Y = \sum_{i=1}^k \sum_{j=1}^m (a_i + b_j) \mathbb{1}_{A_i B_j} \in \mathcal{E}$$

Simple random variables can be used to approximate other random variables. Suppose X is any random variable. Define

$$X_n = \begin{cases} \lfloor nX \rfloor / n & \text{when } |X| \leq n \\ n & \text{when } X > n \\ -n & \text{when } X < -n \end{cases}$$

where $\lfloor nX \rfloor$ is the greatest integer $\leq nX$. Now

$$X_n(\omega) \rightarrow X(\omega) \text{ for all } \omega \in \Omega$$

for $n > |X(\omega)|$.

$$|X_n(\omega) - X(\omega)| = \left| \frac{\lfloor nX(\omega) \rfloor}{n} - \frac{nX(\omega)}{n} \right| = \frac{1}{n} |\lfloor nX(\omega) \rfloor - nX(\omega)| \leq \frac{1}{n}.$$

$$X_n \in \mathcal{E}.$$

Measurability Theorem:

Suppose X is a random variable such that $X(\omega) \geq 0$ for all $\omega \in \Omega$.

Then, there exist simple random variables $X_n \geq 0$, such that $X_n \uparrow X$ in the sense that $X_n(\omega) \geq X(\omega)$ for all $\omega \in \Omega$.

Proof: on page 118 - 119

Expectation and Lebesgue Integrals.

Let X be a random variable, possibly into the extended real number $[-\infty, +\infty]$,

$$X: (\Omega, \mathcal{B}, P) \rightarrow (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$$

$$X: (\Omega, \mathcal{B}, P) \rightarrow (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$$

where $\tilde{\mathbb{R}} = [-\infty, +\infty] = \mathbb{R} \cup \{-\infty, +\infty\}$

We shall introduce expectation as a Lebesgue integral.

$$\mathbb{E}(X) = \int_{\Omega} X \cdot dP \quad \text{or} \quad \mathbb{E}(X) = \int_{\Omega} X(w) \cdot dP(w)$$

Let X be any simple random variable

$$X = \sum_{i=1}^k a_i \mathbf{1}_{A_i}$$

where $|a_i| < \infty$ for all i and A_1, A_2, \dots, A_k are disjoint with

$$\sum A_i = \Omega$$

Define $\mathbb{E}(X) = \sum_{i=1}^k a_i P(A_i)$

or $\int_{\Omega} X \cdot dP = \sum_{i=1}^k a_i P(A_i)$

STAT 401 Nov 5 2008 Midterm 2 solutions

1. \mathcal{P}, \mathcal{L} collections of subsets of Ω

②(a) $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$

④(b) "Old" axioms

(1) $\emptyset \in \mathcal{L}$

(2) if $A, B \in \mathcal{L}$ with $A \subset B$, then $B \setminus A \in \mathcal{L}$

(3) if $A_n \uparrow$, with $A_n \in \mathcal{L}$ for all $n \geq 1$ implies $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$

"New" axioms

(1) $\emptyset \in \mathcal{L}$

(2) $A \in \mathcal{L} \Rightarrow A^c \in \mathcal{L}$

(3) $A_n \in \mathcal{L}$, disjoint implies $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$

④(c) First part: If \mathcal{P} is a π -system and \mathcal{L} is a λ -system such that $\mathcal{P} \subseteq \mathcal{L}$ then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$

Second part: If \mathcal{P} is a π -system, then $\sigma(\mathcal{P}) = \mathcal{L}(\mathcal{P})$

③(a) A minimal set of axioms for \mathcal{S} to be a semi-algebra

(1) $\emptyset \in \mathcal{S}$

(2) \mathcal{S} is a π -system

(3) For all $A \in \mathcal{S}$, there exists a positive integer n and disjoint sets $A_1, A_2, \dots, A_n \in \mathcal{S}$ such that

$$A^c = \bigcup_{i=1}^n A_i$$

④(b) Suppose \mathcal{S} is a semi-algebra and P is a mapping

$$\mathcal{S} \rightarrow [0, 1]$$

is σ -additive on \mathcal{S} satisfying $P(\Omega) = 1$, Then there is a unique extension P' of P to $A(\mathcal{S})$ such that P' is a σ -additive measure on $A(\mathcal{S})$.

③(c) Suppose \mathcal{S} is a semi-algebra of subsets of Ω , and $p: \mathcal{S} \rightarrow [0, 1]$ is σ -additive on \mathcal{S} with $p(\Omega) = 1$. Then there is a unique probability measure P' on $A(\mathcal{S})$ that extends P .

3. $B_i \subset A_i$ for all i . So $\bigcup_i B_i \subset \bigcup_i A_i$

$$\text{Therefore } P\left(\bigcup_i A_i\right) = P\left(\bigcup_i B_i\right) = P\left(\bigcup_i A_i \setminus \bigcup_k B_k\right)$$

$$= P\left[\left(\bigcup_i A_i\right) \cap \left(\bigcup_k B_k^c\right)\right]$$

$$= P\left[\bigcup_i \left(A_i \cap \left(\bigcup_k B_k^c\right)\right)\right]$$

$$= P\left[\bigcup_i \left(A_i \cap \left(\bigcap_k B_k^c\right)\right)\right]$$

$$\leq \sum_i P(A_i \cap \bigcap_k B_k^c) \leq \sum_i P(A_i \cap B_i^c)$$

$$= \sum_i P(A_i \setminus B_i) = \sum_i [P(A_i) - P(B_i)]$$

4. Suppose H is a class of subsets of Ω , and $C \subset \Omega$ such that $C \in \sigma(H)$, show that there exists a countable class $H_C \subset H$ such that $C \in \sigma(H_C)$.

Define $G = \{C \in \sigma(H) : \text{there exists a countable } H_C \subset H \text{ st } C \in \sigma(H_C)\}$

It is sufficient to show that G is a σ -field and that $H \subset G$, it will then follow that

$\sigma(H) \subset G$ which implies $\sigma(H) = G$.

First let us show that $H \subset G$. For any $C \in H$, let

$H_C = \{C\}$, then H_C is countable, also we have

$C \in \sigma(H_C)$, therefore $H \subset G$.

Next we check that G is a σ -field.

① $\Omega \in G$, True because $\Omega \in \sigma(H_C)$ for any countable H_C .

(2) $C \in G$, $\Rightarrow C^c \in G$. True because $C \in G$ implies that there exists countable H_C such that $C \in \sigma(H_C)$ since $\sigma(H_C)$ is a σ -field, it follows that $C^c \in \sigma(H_C)$ also. so $C^c \in G$.

③ Let C_i , $i=1, 2, \dots \in G$. Then $\bigcup_{i=1}^{\infty} C_i \in G$. True. This holds because for each i , there exists H_{C_i} countable such that $C_i \in \sigma(H_{C_i})$. Then set

$$H_{\bigcup C_i} = \bigcup_{i=1}^{\infty} H_{C_i} \quad \underline{C_i \in \sigma(H_{C_i}) ?}$$

A countable union of countable sets is countable. So

$$H_{\bigcup C_i}$$
 is countable.

STAT 901 Nov 7 2008 Windy

Expectation for simple random variables

$X = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$ where A_1, \dots, A_k are disjoint, $\sum A_i = \Omega$

$$|a_1| \dots |a_k| < \infty \quad (X \in \mathcal{E}) \text{ simple RV.}$$

$$\text{Define, } E(X) = \sum_{i=1}^k a_i P(A_i)$$

properties ① $E(\mathbb{1}_A) = P(A)$ so for $A = \Omega$,

$$E(1) = 1$$

② If $x \in \mathcal{E}$ and $x \geq 0$ then $E(x) \geq 0$

③ If $x, y \in \mathcal{E}$, then $\alpha x + \beta y \in \mathcal{E}$. for all real α, β , then

$$E(\alpha x + \beta y) = \alpha E(x) + \beta E(y)$$

④ If $x, y \in \mathcal{E}$, with $x \leq y$ then $(x(w) \leq y(w) \text{ for all } w \in \Omega)$
 $E(x) \leq E(y)$

⑤ If $x_n, x \in \mathcal{E}$, with $x_n \uparrow x$ or $x_n \downarrow x$ then

$$E(x_n) \uparrow E(x) \text{ or } E(x_n) \downarrow E(x)$$

Proof: see pages 120 - 121.

Next, we extend to random variables X such that $0 \leq X < \infty$, that is, $0 \leq X(w) \leq \infty$ for all $w \in \Omega$.

From the measurability theorem, we know that for any random variable, $X \geq 0$, there exists a sequence of simple random variables $X_n \geq 0$ such that $X_n \uparrow X$, that is $X_n(w) \uparrow X(w)$ for all $w \in \Omega$.

Note that for $n \geq m$, we have $X_n \geq X_m$

so $E(X_n) \geq E(X_m)$. Thus $E(X_n)$ is an increasing sequence. Every monotonically

increasing sequence of real numbers has a limit (possibly $\pm\infty$). Define

$$E(X) = \lim_{n \rightarrow \infty} E(X_n). \quad \text{Then } 0 \leq E(X) \leq \infty$$

proposition $E(X)$ is well defined, that is, if $Y_n \uparrow X$, where $Y_n \geq 0$ is another sequence of simple random variables, then

$$\lim_{n \rightarrow \infty} E(Y_n) = \lim_{n \rightarrow \infty} E(X_n)$$

Proof: It suffices to show that

$$\lim_{n \rightarrow \infty} E(Y_n) \leq \lim_{n \rightarrow \infty} E(X_n) \quad \text{and} \quad \lim_{n \rightarrow \infty} E(Y_n) \geq \lim_{n \rightarrow \infty} E(X_n) \quad (2)$$

Now note that since $Y_n \uparrow X$, therefore $X_n \wedge Y_m \uparrow X_n \wedge X$ as $m \rightarrow \infty$ by property 5 of Expectation on E .

$$E(X_n \wedge Y_m) \uparrow E(X_n) \text{ as } m \rightarrow \infty. \text{ But}$$

$$E(X_n \wedge Y_m) \leq E(Y_m) \quad (\text{property 4})$$

$$\text{so } \lim_{m \rightarrow \infty} E(Y_m) \geq \lim_{m \rightarrow \infty} E(X_n \wedge Y_m) = E(X_n)$$

$$\text{so } \lim_{m \rightarrow \infty} E(Y_m) \geq E(X_n) \text{ for all } n.$$

$$\text{So } \lim_{m \rightarrow \infty} E(Y_m) > \lim_{n \rightarrow \infty} E(X_n) \quad \text{--- (2) is proved.}$$

To prove (1), reverse the roles of X_n, Y_m 's in the argument. QED.

Let \mathcal{E}_+ denote the class of all random variables X where $0 \leq X < \infty$

Properties of expectation on $\bar{\mathcal{E}}_+$.

- ① $0 \leq E(X) \leq \infty$
- ② For $\alpha > 0, \beta > 0; X, Y \in \bar{\mathcal{E}}_+$ then
 $\alpha X + \beta Y \in \bar{\mathcal{E}}_+$ and
 $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$

③ (Monotone convergence Theorem) allows infinity
 If $0 \leq X_n \uparrow X$, then $E(X_n) \uparrow E(X)$

Proof: pages 124-125.

$$\begin{aligned} X_n &\nearrow X \\ \Rightarrow E(X_n) &\nearrow E(X) \end{aligned}$$

$$\begin{aligned} X_n &\downarrow X \\ \underset{\geq 0}{\therefore} E(X_n) &\downarrow E(X) \end{aligned}$$

$$f_n(x) = \frac{1}{n} \quad n \geq 0$$

$$f(x) = 0$$

$$f_n(x) \nearrow f(x) \quad x \rightarrow \infty$$

$$\int_{-\infty}^{\infty} f_0 \rightarrow \int_{-\infty}^{\infty} f(x) \quad X$$

STAT 901 Nov 10 2008.

Proposition Let S be the surface of a convex body in three dimensions. Let

x_1, x_2, \dots, x_n be n points chosen from S . Then there exists a connected graph with vertices x_1, \dots, x_n whose edges are paths lying entirely in S .

Proof: We prove by induction on the value of n .

Base case $n=1$: This is immediate, a single vertex with no edges is a connected graph.

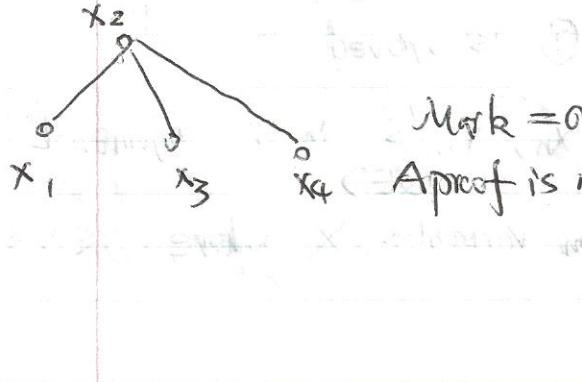
The induction step: We assume the result is true for n and prove it for $n+1$. Given x_1, \dots, x_{n+1} divide the points into two groups:

first $\{x_1, \dots, x_n\}$ and secondly $\{x_{n+1}\}$ connected

By the induction hypothesis, there exists a graph with vertices $\{x_1, \dots, x_n\}$ and edges entirely in S . Call this G_1 . Similarly, there exists a graph G_2 with vertices $\{x_{n+1}\}$ and edges in S .

Consider the graph $G_1 \cup G_2$, which has vertices $\{x_1, \dots, x_n\} \cup \{x_{n+1}\} = \{x_1, \dots, x_{n+1}\}$, edges which are the unions of the two sets of edges.

This is a connected graph, which is required graph on $\{x_1, \dots, x_{n+1}\}$.

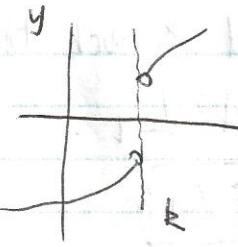


Mark = 0

A proof is not a seq. of true statements

$$\textcircled{5} \quad k = \sup \{x : f(x) \leq y\}$$

$$f^{-1}(-\infty, y] = \{x : f(x) \leq y\}$$



What is k where $\{x : f(x) \leq y\} = \emptyset$?

\textcircled{6} Every continuous function is measurable wrt. Borel sigma field (\mathcal{B})

\textcircled{6} if \mathcal{B} is a σ -field which makes every continuous function msl, then $(\mathcal{B}(R) \subset \mathcal{B}$. (Identity function)

STAT 901 Nov 12, 2008

$E(X)$ for $X \in \bar{\Sigma}_+$ Most interesting nontrivial property

MCT If $0 \leq X_n \uparrow X$ then $0 \leq E(X_n) \uparrow E(X)$

Extending expectation to other random variables.

Let X be any random variable, with $\mathbb{X} = [-\infty, +\infty]$

Define a new random variable

$$X^+(w) = \begin{cases} X(w) & \text{if } X(w) \geq 0 \\ 0 & \text{if } X(w) < 0 \end{cases}$$

Equivalently, $X^+ = X \vee 0$

Define

$$X^-(w) = \begin{cases} -X(w) & \text{if } X(w) \leq 0 \\ 0 & \text{if } X(w) > 0 \end{cases}$$

Equivalently, $X^- = (-X)^+$

The random variables X^+, X^- are called the positive part and the negative part of X , respectively.

Also, $X^+, X^- \in \bar{\Sigma}_+$

Note that $X = X^+ - X^-$. Also

$$|X| = X^+ + X^-$$

Definition: We say that X is integrable if $E(X^+) < \infty$ and $E(X^-) < \infty$.

Now, X integrable $\Rightarrow E(X^+) < \infty, E(X^-) < \infty \Rightarrow E(X^+) + E(X^-) < \infty$

$$\Rightarrow E(X^+ + X^-) < \infty \Rightarrow E(|X|) < \infty$$

reverse also applies: X integrable iff $E(|X|) < \infty$

Traditionally, the set of all X such that

$E(|X|) < \infty$ is called L_1 .

X integrable $\Leftrightarrow X \in L_1$

Def'n. If X is integrable, we define $E(X) = E(X^+) - E(X^-)$

Properties of $E(X)$, for $X \in L_1$

① X integrable $\Rightarrow P(X=\infty) = P(X=-\infty) = 0$

② if X is integrable, then cX is integrable, and
 $E(cX) = cE(X)$

③ If X, Y are integrable, then $X+Y$ is integrable, and

$$E(X+Y) = E(X) + E(Y)$$

④ Suppose that $X_n, n \geq 1$ is a sequence of random variables

such that X_m is integrable for some positive integer m

if $X_n \uparrow X$, then $E(X_n) \uparrow E(X)$

if $X_n \downarrow X$, then $E(X_n) \searrow E(X)$

⑤ If X is integrable, then

$$|E(X)| \leq E(|X|)$$

$$(cX)^+ = \begin{cases} cX^+ & c \geq 0 \\ -cX^- & c < 0 \end{cases}$$

$$(cX)^- = \begin{cases} cX^- & c \geq 0 \\ -cX^+ & c < 0 \end{cases}$$

Proof of ③ Note that

$$|X+Y| \leq |X| + |Y|$$

Suppose that X and Y are integrable, so

$$E(|X+Y|) \leq E(|X| + |Y|) = E(|X|) + E(|Y|) < \infty$$

so $X+Y$ is integrable.

$$\text{Next, } X+Y = (X^+ - X^-) + (Y^+ - Y^-)$$

$$\text{and } X+Y = (X+Y)^+ - (X+Y)^-$$

$$\text{so } (X+Y)^+ - (X+Y)^- = (X^+ - X^-) + (Y^+ - Y^-)$$

$$\text{Therefore } (X+Y)^+ + X^- + Y^- = (X+Y)^- + X^+ + Y^+$$

$$\text{Then } E(X+Y)^+ + E(X^-) + E(Y^-) = E(X+Y)^- + E(X^+) + E(Y^+)$$

∴ Thus

$$E(X+Y)^- - E(X+Y)^+ = E(X^+) - E(X^-) + E(Y^+) - E(Y^-)$$

It follows that

$$E(X+Y) = E(X) + E(Y) \quad \text{QED.}$$

Proof of 5.

$$|E(X)| = |E(X^+) - E(X^-)| \leq |E(X^+)| + |E(X^-)|$$

$$= E(X^+) + E(X^-) = E(X^+ + X^-) = |E(X)| \quad \text{QED.}$$

Series Version of MCT

If $X_n, n \geq 1$ are non-negative random variables, then

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n)$$

Proof: Let $Y_n = \sum_{i=1}^n X_i$. Then $0 \leq Y_n \uparrow \sum_{i=1}^{\infty} X_i$.

By MCT, $E(Y_n) \uparrow E\left(\sum_{i=1}^{\infty} X_i\right)$

Thus $E\left(\sum_{n=1}^{\infty} X_n\right) \uparrow E\left(\sum_{i=1}^{\infty} X_i\right)$

So $\sum_{i=1}^n E(X_i) \uparrow E\left(\sum_{i=1}^{\infty} X_i\right)$

or

taking limit as $n \rightarrow \infty$, $\sum_{n=1}^{\infty} E(X_n) = E\left(\sum_{i=1}^{\infty} X_i\right)$ QED.

5.3 Limit Theorems

STAT 901

MCT $0 \leq X_n \uparrow X \Rightarrow 0 \leq E(X_n) \uparrow E(X)$

Series Version of MCT $X_n \geq 0$, for all n

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n)$$

Fatou's Lemma

If $X_n \geq 0$ for all n , then $E(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} E(X_n)$

Proof: Note that

$\bigwedge_{k=n}^{\infty} X_k \uparrow \liminf_{n \rightarrow \infty} X_n$ so by the MCT.

$$E(\liminf_{n \rightarrow \infty} X_n) = \lim_{n \rightarrow \infty} E\left(\bigwedge_{k=n}^{\infty} X_k\right)$$

$$= \liminf_{n \rightarrow \infty} E\left(\bigwedge_{k=n}^{\infty} X_k\right) \leq \liminf_{n \rightarrow \infty} E(X_n) \quad \text{QED.}$$

VERY POWERFUL

Extension of Fatou's Lemma.

We can replace the condition $X_n \geq 0$ for all n , by

$X_n \geq Y$ for all n where Y is integrable.

"proof". Apply Fatou to $X_n - Y \geq 0$

Also if $X_n \leq Z$, for all n , where Z is integrable,

then $E(\limsup_{n \rightarrow \infty} X_n) \geq \limsup_{n \rightarrow \infty} E(X_n)$

Pointed Convergence Theorem (PCT)

Suppose $X_n \rightarrow X$ and suppose there exists an integrable random variable Z , such that

$|X_n| \leq Z$ for all n . Then

$E(X_n) \rightarrow E(X)$ and $E|X_n - X| \rightarrow 0$

Proof: We have $-Z \leq X_n \leq Z$ for all n

where Z and $-Z$ are integrable. Also

$$X = \liminf_{n \rightarrow \infty} X_n = \limsup_{n \rightarrow \infty} X_n$$

$$\text{So } E(X) = E(\liminf_{n \rightarrow \infty} X_n)$$

$$\leq \liminf_{n \rightarrow \infty} E(X_n) \quad \text{Fatou's, using } -Z \leq X_n$$

$$\leq \limsup_{n \rightarrow \infty} E(X_n)$$

$$\leq E(\limsup_{n \rightarrow \infty} X_n) \quad \text{Fatou's using } X_n \leq Z$$

$$= E(X)$$

So all \leq signs become $=$

$$\text{so } E(X_n) \rightarrow E(X)$$

Next, $|X_n - X| \leq |X_n| + |X| \leq Z + Z = 2Z$, which is integrable

furthermore, $|X_n - X| \rightarrow 0$ so by the first part of PCT, which

we proved

$$E|X_n - X| \rightarrow E(0) = 0 \quad \text{QED.}$$

$$X_n \rightarrow X \quad |X_n| \leq Z \Rightarrow X \leq Z$$

$$\frac{\partial E(X_t)}{\partial t} = E\left(\frac{\partial X_t}{\partial t}\right)$$

regularity condition. Cramer-Yao?

$$\frac{\partial E(X_t)}{\partial t}$$

$$E\left(\frac{\partial X_t}{\partial t}\right)$$

$$\lim_{h \rightarrow 0} \frac{E(X_{t+h}) - E(X_t)}{h} = E\left(\lim_{h \rightarrow 0} \frac{X_{t+h} - X_t}{h}\right)$$

$$\lim_{h \rightarrow 0} E\left[\underbrace{\frac{X_{t+h} - X_t}{h}}_{y_h}\right] = E\left[\lim_{h \rightarrow 0} \frac{X_{t+h} - X_t}{h}\right]$$

$$\lim_{h \rightarrow 0} E[y_h] = E[\lim_{h \rightarrow 0} y_h]$$

$$E(X_n) \rightarrow E(X) \quad \lim_{n \rightarrow \infty} E(X_n) \Rightarrow E(\lim_{n \rightarrow \infty} X_n)$$

In Lebesgue integral notation,

$$E(X) = \int x dP \quad \text{or} \quad E(X) = \int_{\Omega} x dP$$

We shall also define :

$$\int_A x dP = E(X \cdot \mathbb{1}_A) \quad \text{for any event } A \subset \Omega$$

Properties. Assume $X \geq 0$ (important)

$$\textcircled{1} \quad 0 \leq \int_A x dP \leq E(X)$$

$$\textcircled{2} \quad \int_A x dP = 0 \quad \text{if and only if} \quad P(X > 0 \text{ and event } A \text{ occurs}) = 0$$

That is. $P([X > 0] \cap A) = 0$

\textcircled{3} If A_1, A_2, \dots are disjoint, then

$$\int_{\bigcup_{n=1}^{\infty} A_n} x dP = \sum_{n=1}^{\infty} \int_{A_n} x dP$$

$$\textcircled{4} \quad A_1 \subset A_2 \Rightarrow \int_{A_1} x dP \leq \int_{A_2} x dP$$

\textcircled{5} If $A_n \nearrow A$, then $\int_{A_n} x dP \nearrow \int_A x dP$
Also, if X is integrable, then

$$A_n \nearrow A \text{ implies } \int_{A_n} x dP \nearrow \int_A x dP$$

$$(x \mathbb{1}_{A_n}) \leq x$$

dominating . special case .

see pages 134-135

STAT 901 Nov 17, 2008 Monday

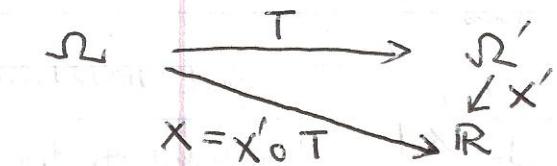
$$\int_{\Omega} x dP = E(X) \quad \int_A x \cdot dP = E(X \mathbf{1}_A)$$

5.5 Transformation Theorem

Suppose $T: (\Omega, \mathcal{B}, P) \rightarrow (\Omega', \mathcal{B}')$ is measurable. Earlier, we saw that T induces a prob. measure on (Ω', \mathcal{B}') such that

$$P'(A') = (P \circ T^{-1})(A) = P[T^{-1}(A')]$$

Let X' be an integrable random variable on $(\Omega', \mathcal{B}', P')$



Transformation Theorem:

Then X is integrable, and

$$\int_{\Omega} x dP = \int_{\Omega'} x' dP'$$

$$\text{or } E(X) = E(X')$$

This is just a change of variable formula because we can write (formally)

$$\begin{aligned} \int_{\Omega} x dP &= \int_{\Omega} (X' \circ T)(\omega) \frac{dP}{dP'} dP' \\ &= \int_{\Omega'} X(\omega') \frac{dP}{dP'}(\omega') dP'(\omega') \quad \text{Tacobian} \end{aligned}$$

A special case is important. Take $T = X$, and

$X' = I$ which is the identity function.

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathbb{R} \\ & \searrow & \downarrow I \\ & & \mathbb{R} \end{array} \quad I: x \mapsto x$$

The transformation theorem tells us

$$\int_{\Omega} x dP = \int_{\mathbb{R}} I(x) d(P \circ X^{-1})$$

The integral ~~on~~ the RHS. of X on \mathbb{R}

$$= \int_{-\infty}^{+\infty} x dF_X(x), \text{ where}$$

F_x is the distribution func. of X

Riemann Stieltjes integral.

$$\text{or } E(X) = \int_{-\infty}^{+\infty} x dF_x(x)$$

Special cases:

1. X discrete. $dF_x(x) = f(x)$, prob. function of X .

$$\text{So } E(X) = \sum x f(x)$$

2. X continuous $dF_x(x) = f(x)dx$ where f is the density of X .

$$\text{So } E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Assigned problems

5.10 : 1, 2, 3, 4, 6, 7, 8, 11, 21, 26, 29

Ch.6. Convergence of Random Variables

6.1 Almost sure convergence.

A property is said to hold almost surely (abbreviated a.s.) or almost everywhere (a.e.) or almost certainly (a.c.) if there is some event $N \subset \Omega$, such that $P(N)=0$ and the property holds on N^c .

E.g. $X=X'$ a.s. if $P(X=X')=1$ where X, X' are random variables.

We say that $X_n \rightarrow X$ a.s. if the sets

$$N = \{w : X_n(w) \rightarrow X(w)\}$$

$N^c = \{w : X_n(w) \not\rightarrow X(w)\}$ have probability

$$P(N) = 0 \text{ and } P(N^c) = 1$$

We also write $X_n \xrightarrow{\text{a.s.}} X$. Thus $P(\lim_{n \rightarrow \infty} X_n = X) = 1$

6.2 Convergence in Probability

We say that X_n converges in probability to X and write

$X_n \xrightarrow{P} X$ if, for all $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1$$

Equivalently,

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

Theorem: If $x_n \xrightarrow{a.s.} X$, then $x_n \xrightarrow{P} X$

proof: Suppose $x_n \xrightarrow{a.s.} X$. Then for all $\epsilon > 0$

$$0 = P(|x_n - X| > \epsilon, i.o.)$$

STAT 901 Theory of probability Nov 19, 2008

$x_n \xrightarrow{a.s.} X \quad P(\lim_{n \rightarrow \infty} x_n = X)$

$x_n \xrightarrow{P} X \quad \text{for all } \epsilon > 0 \quad \lim_{n \rightarrow \infty} P(|x_n - X| \leq \epsilon) = 1$

Theorem: If $x_n \xrightarrow{a.s.} X$, then $x_n \xrightarrow{P} X$

Proof: Suppose $x_n \xrightarrow{a.s.} X$, then for all $\epsilon > 0$

$$P(|x_n - X| > \epsilon, i.o.) = 0$$

$$= P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [x_k - X > \epsilon]\right)$$

$$= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} [x_k - X > \epsilon]\right)$$

$$= \limsup_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} [x_k - X > \epsilon]\right)$$

$$\geq \limsup_{n \rightarrow \infty} P(|x_n - X| > \epsilon) \geq 0$$

$$\text{so } \limsup_{n \rightarrow \infty} P(|x_n - X| > \epsilon) = 0$$

$$\text{but } 0 \leq \liminf_{n \rightarrow \infty} P(|x_n - X| > \epsilon) \leq \limsup_{n \rightarrow \infty} P(|x_n - X| > \epsilon) = 0$$

Therefore, $\lim_{n \rightarrow \infty} P(|x_n - X| > \epsilon) = 0$ since this is true for all $\epsilon > 0$

it follows that $x_n \xrightarrow{P} X$. QED.

counter example: $x_n \xrightarrow{P} X$ does not imply $x_n \xrightarrow{a.s.} X$ in general.

To see this:

Let $X \equiv 0$. Let $x_n, n \geq 1$ be independent, with

$$P(x_n = 0) = 1 - \frac{1}{n}, \quad P(x_n = 1) = \frac{1}{n}$$

Then $P(x_n = 1, i.o.) = P(x_n = 1 \text{ for infinitely many } n)$

$$= 1 \quad (\text{because } \sum_{n=1}^{\infty} P(x_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty, \text{ Borel-Cantelli law})$$

$$\text{so } P(\lim_{n \rightarrow \infty} x_n \neq 0) \geq P(x_n = 1, i.o.) = 1$$

Therefore $P(\lim_{n \rightarrow \infty} X_n \neq 0) = 1$ or $X_n \xrightarrow{a.s.} 0$

However, $\lim_{n \rightarrow \infty} P(|X_n - 0| > \varepsilon) = \lim_{n \rightarrow \infty} P(X_n = 1)$
for $\varepsilon > 0$,

$$= \lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

Therefore, $X_n \xrightarrow{P} 0$.

Statistical estimation, strong consistency of $\hat{\theta}_n \Rightarrow$ weak consistency
but not the converse.

Definition: A sequence X_n is said to be Cauchy in probability if
for all $\varepsilon > 0$ and all $\delta > 0$, there exists a positive integer
 n_0 such that

$$P(|X_r - X_s| > \varepsilon) < \delta \text{ for all } r, s \geq n_0$$

Equivalently, $X_n - X_m \xrightarrow{P} 0$ as $n, m \rightarrow \infty$ (double limit)

Proposition: X_n converges in probability to some random variable if
and only if X_n is Cauchy in probability.

Proof: See pages 172.

Proposition: $X_n \xrightarrow{P} X$ if and only if every subsequence X_{n_1}, X_{n_2}, \dots
 $\sim X_{n_3}, \dots$ (with $n_1 < n_2 < \dots$) has a sub-subsequence

$X_{n_{k_1}}, \dots, X_{n_{k_2}}, \dots, X_{n_{k_3}}, \dots$ such that
 $X_{n_{k_j}} \xrightarrow{a.s.} X$ as $j \rightarrow \infty$

Proof: P173.

Corollary: (Exercise 6.7.1 part (a))

Let X_n be monotone, then $X_n \xrightarrow{P} X$ if and only if
 $X_n \xrightarrow{a.s.} X$.

Proof: We already know $X_n \xrightarrow{a.s.} X$ implies $X_n \xrightarrow{P} X$

We prove the converse.

Suppose $X_n \nearrow$ wlog, and $X_n \xrightarrow{P} X$.

So there exists a subsequence i.e. X_{n_j} ($j \geq 1$) such that

$X_{n_j} \xrightarrow{a.s.} X$ as $j \rightarrow \infty$. So with probability one, for all
 $\varepsilon > 0$, there exists a random variable

J. such that $|x_{n_j} - x| < \varepsilon$ for all $j \geq J$. But $x_{n_1} \leq x_{n_2} \leq x_{n_3} \leq \dots \leq x$

so for all $n \geq n_J$

$$x_{n_J} \leq x_n \leq x_{n_J+k} \leq x \quad \text{Also, } x - x_{n_J} < \varepsilon$$

Therefore for all $n \geq n_J$, we have

$$|x_n - x| = x - x_n < \varepsilon$$

Therefore $x_n \xrightarrow{\text{a.s.}} x$ QED

STAT901 Nov 21, 2008

Conditional Expectation (Section 10.2 onwards)

$$P(A|B) = \frac{P(AB)}{P(B)}, P(B) > 0$$

- If $P(B)=0$, this definition does not work

- when x, y are jointly continuous - then

$$P(x \in A | Y=y) = \int_A \frac{f(x,y)}{f_y(y)} dx$$

- this is conditional on an event $Y=y$ of probability zero which violates our original restriction.

- We have two definitions

Steps to a general theory of conditioning

① Replace probability by expectation, because we can write

$$P(A|B) = E(\mathbb{1}_A | \mathbb{1}_B = 1)$$

So it is sufficient to define

$E[X | Y=y]$ for any integrable X .

② Conditional expectation is a random variable.

Let $h(y) = E(X | Y=y)$

for each value of y in the range of the random variable Y

Now consider the random variable $Z = h(Y)$. For any $w \in \Omega$, let $Y(w) = y$, then

$$Z(w) = h(Y(w)) = h(y) = E(X | Y=y)$$

Thus Z takes the value $E(X|Y=y)$ whenever Y takes the value y . Henceforth, we write $E(X|Y)$ for this random variable.

- ③ Consider a transformation of Y which is 1-1. Say

$$W = Y^3$$

Set $E(X|Y) = Z_1$, and $E(X|W) = Z_2$.

Say Consider an $\omega \in \Omega$, Let $Y(\omega) = y$ and $W(\omega) = w = y^3$, then

$$Z_1(\omega) = E(X|Y=y) = E(X|Y^3=y^3)$$

$$= E(X|W=w) = Z_2(\omega)$$

omega

That is $E(X|Y) = E(X|W)$

To get this, we needed Y, W in 1-1 measurable correspondence.

More generally, when $\sigma(Y) = \sigma(W)$, then we want

$$E(X|Y) = E(X|W), \text{ so we can write}$$

$$E[X|\sigma(Y)] \text{ for the two when } \sigma(Y) = \sigma(W)$$

- ④ We do NOT need the random variable Y . We can write

$$E(X|\mathcal{G}) \quad \mathcal{G} \subset \mathcal{B} \text{ on } (\Omega, \mathcal{B}, P)$$

\mathcal{G} -field

\mathcal{G} -measurable, $\sigma(E(X|\mathcal{G})) \subset \mathcal{G}$.

DEFINITION. Let (Ω, \mathcal{B}, P) be a probability space, and let

$\mathcal{G} \subset \mathcal{B}$, where \mathcal{G} is a σ -field.

Let X be an integrable random variable. We define the conditional expectation of X with respect to \mathcal{G} to be a random variable,

call $E(X|\mathcal{G})$, such that

① $E(X|\mathcal{G})$ is integrable and \mathcal{G} -measurable.

② For all $G \in \mathcal{G}$,

$$\int_G X dP = \int_G E(X|\mathcal{G}) dP$$

(That is, $E(X \mathbb{1}_G) = E[E(X|\mathcal{G}) \mathbb{1}_G]$)

$E(X|\mathcal{G})$ exists and is a.s. unique.

To say a.s. unique, we mean if U, V are integrable, \mathcal{G} -measurable and

$$\int_G x \cdot dP = \int_G U \cdot dP = \int_G V \cdot dP$$

(They both satisfy definition. Then $U=V$ a.s.)

$G=\Omega$ case.

$$E(X\mathbf{1}_\Omega) = E(X) = E[E(X|G)]$$

STAT 901 Nov 26, 2008

Conditional Expectation

X integrable on (Ω, \mathcal{B}, P)

G σ -field, $\mathcal{G} \subset \mathcal{B}$

Then $E(X|G)$ is a random variable such that

- ① $E(X|G)$ is \mathcal{G} -measurable, integrable
- ② $\int_G x \cdot dP = \int_G E(X|G) \cdot dP$ for all $G \in \mathcal{G}$.

X integrable

$x \in L^1(\Omega, \mathcal{B}, P)$

Examples

① $\mathcal{G} = \{\emptyset, \Omega\}$, In this case, $E(X|G)$ is a constant random variable because it is \mathcal{G} -measurable and \mathcal{G} is trivial. Let

$$E(X|G) = c, \text{ say}$$

consider $G = \Omega \in \mathcal{G}$, Then

$$\int_\Omega x \cdot dP = \int_\Omega c \cdot dP \Rightarrow E(X) = c$$

$$\text{so } E(X| \{\emptyset, \Omega\}) = E(X)$$

② $\mathcal{G} = \mathcal{B}$. Then

$$\int_B x \cdot dP = \int_B E(X|\mathcal{B}) \cdot dP \text{ for all } B \in \mathcal{B}.$$

Do we know a random variable which satisfies this for

$E(X|\mathcal{B})$ which is \mathcal{B} -measurable.

Yes, set $E(X|\mathcal{B}) = X$, This works.

Any other X' , such that $X' = X$ almost surely a.s. will also work

$$[\int_B x \cdot dP = \int_B x \cdot dP \text{ for all } B \in \mathcal{B}] \Leftrightarrow$$

$$P(X = X') = 1.$$

If X is \mathcal{G} -measurable, then $E(X|\mathcal{G}) = X$ a.s. In particular $E(X|\sigma(X)) = X$ a.s. Also $E(X|\sigma(X, Y)) = X$ a.s. etc

④ $\mathcal{G} = \{\emptyset, A, A^c, \Omega\}$ where $A \in \mathcal{B}$

The \mathcal{G} -measurable random variables have the form

$$C_1 \mathbb{1}_A + C_2 \mathbb{1}_{A^c} \text{ why?}$$

If say Y is \mathcal{G} -measurable, suppose Y is constant, then

we have $C_1 = C_2$. If Y is not constant, let C_1, C_2 be in the range of Y with $C_1 \neq C_2$. Consider the events

$$\{Y=C_1\} \in \{\emptyset, A, A^c, \Omega\}$$

$$\{Y=C_2\} \in \{\emptyset, A, A^c, \Omega\}$$

So $\{Y=C_1\} = A$ and $\{Y=C_2\} = A^c$
or vice versa.

Let us set $E(X|\mathcal{G}) = C_1 \mathbb{1}_A + C_2 \mathbb{1}_{A^c}$

$$\int_A X dP = \int_A E(X|\mathcal{G}) dP = \int_A (C_1 \mathbb{1}_A + C_2 \mathbb{1}_{A^c}) dP$$

$$\begin{aligned} &= C_1 \int_A \mathbb{1}_A dP + C_2 \int_{A^c} \mathbb{1}_{A^c} dP \\ &= C_1 E(\mathbb{1}_A \mathbb{1}_A) + C_2 E(\mathbb{1}_A \mathbb{1}_{A^c}) \\ &= C_1 P(A) + 0 \end{aligned}$$

$$\text{so } C_1 = \int_A X dP / P(A) \quad (= E(X|A))$$

$$\text{Similarly; } C_2 = \int_{A^c} X dP / P(A^c) \quad (= E(X|A^c))$$

$$\text{so } E(X|\{\emptyset, A, A^c, \Omega\}) = \mathbb{1}_A E(X|A) + \mathbb{1}_{A^c} E(X|A^c)$$

properties:

① If X, Y integrable, then

$$E(\alpha X + \beta Y | \mathcal{G}) = \alpha E(X|\mathcal{G}) + \beta E(Y|\mathcal{G}), \text{ a.s.}$$

proof: Since $E(X|\mathcal{G})$ and $E(Y|\mathcal{G})$ are integrable, and \mathcal{G} -measurable it follows that $\alpha E(X|\mathcal{G}) + \beta E(Y|\mathcal{G})$ is integrable and \mathcal{G} -measurable.

so it suffices to check the integral identity

$$\int_G (\alpha X + \beta Y) dP = \alpha \int_G X dP + \beta \int_G Y dP$$

$$= \alpha \int_G \mathbb{E}(x|G) dP + \beta \int_G \mathbb{E}(y|G) dP$$

$$= \int_G [\alpha \mathbb{E}(x|G) + \beta \mathbb{E}(y|G)] dP$$

as required.

STAT 901 Nov 28 2008

$x \in L_1$, $G \in \mathcal{B}$ σ -field.

$\mathbb{E}(x|G) \in L_1$, also G -measurable

$$\int_G \mathbb{E}(x|G) dP = \int_G x dP \text{ for all } G \in \mathcal{G}$$

$x \in L_1 \Leftrightarrow X$ integrable

X G -measurable $\Leftrightarrow \mathcal{G}(X) \subseteq G$.

book uses notation $X \in \mathcal{G}$, but I don't
properties:

① Linearity If X, Y integrable, then

$$\mathbb{E}(\alpha X + \beta Y | G) = \alpha \mathbb{E}(X|G) + \beta \mathbb{E}(Y|G) \text{ a.s.}$$

② If $X \geq 0$, X integrable, then

$$\mathbb{E}(X|G) \geq 0 \text{ a.s.}$$

Proof: Page 345

③ If X integrable, then

$$|\mathbb{E}(X|G)| \leq \mathbb{E}(|X| | G) \text{ a.s.}$$

④ Smoothing Suppose $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{B}$ and suppose

$X \in L_1$. Then

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G}_2) | \mathcal{G}_1] = \mathbb{E}(X|\mathcal{G}_1) \text{ a.s.}$$

Proof: Both sides are \mathcal{G}_1 -measurable. So it suffices to show that

$$\int_{\mathcal{G}_1} \mathbb{E}(X|\mathcal{G}_2) dP = \int_{\mathcal{G}_1} X dP$$

if we have shown this then

why? We need to show that

$$\int_{\mathcal{G}_1} \mathbb{E}[\mathbb{E}(X|\mathcal{G}_2) | \mathcal{G}_1] dP = \int_{\mathcal{G}_1} \mathbb{E}(X|\mathcal{G}_2) dP$$

$$\stackrel{?}{=} \int_{\mathcal{G}_1} X dP = \int_{\mathcal{G}_1} \mathbb{E}(X|\mathcal{G}_1) dP$$

$y, z \in \mathcal{G}_1$ - measurable

$$\int_{\mathcal{G}_1} y \, dP = \int_{\mathcal{G}_1} z \, dP \quad \text{for } \mathcal{G}_1 \Rightarrow y = z \text{ a.s.}$$

So this would establish that

$$\mathbb{E}[\mathbb{E}(x|\mathcal{G}_2)|\mathcal{G}_1] = \mathbb{E}(x|\mathcal{G}_1) \text{ a.s.}$$

To see that

$$\int_{\mathcal{G}_1} \mathbb{E}(x|\mathcal{G}_2) \, dP = \int_{\mathcal{G}_1} x \, dP$$

Note that

$$\int_{\mathcal{G}_2} \mathbb{E}(x|\mathcal{G}_2) \, dP = \int_{\mathcal{G}_2} x \, dP \quad \text{for all } \mathcal{G}_2 \in \mathcal{G}_2$$

But $\mathcal{G}_1 \in \mathcal{G}_2$, because $\mathcal{G}_1 \subset \mathcal{G}_2$, so the result follows.

Also, if $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{B}$, then

$$\mathbb{E}[\mathbb{E}(x|\mathcal{G}_1)|\mathcal{G}_2] = \mathbb{E}(x|\mathcal{G}_1) \text{ a.s.}$$

which is straightforward.

Note a consequence of smoothing

$$\mathbb{E}[\mathbb{E}(Y|x_1, x_2, \dots)|x_1] = \mathbb{E}(Y|x_1) \text{ a.s.}$$

[Let $\mathcal{G}_2 = \sigma(x_1, x_2, \dots)$ and $\mathcal{G}_1 = \sigma(x_1)$]

⑤ MCT, Fatou, DCT all hold.

MCT: $0 \leq x_n \uparrow X$ a.s. $\Rightarrow 0 \leq \mathbb{E}(x_n|\mathcal{G}) \uparrow \mathbb{E}(x|\mathcal{G})$ a.s.

Fatou: $0 \leq x_n \in \mathcal{L}_1$. Then

$$\mathbb{E}(\liminf x_n|\mathcal{G}) \leq \liminf \mathbb{E}(x_n|\mathcal{G}) \text{ a.s.}$$

DCT: $x_n \rightarrow X$, $|x_n| \leq Y \in \mathcal{L}_1$. Then

$$\mathbb{E}(x_n|\mathcal{G}) \rightarrow \mathbb{E}(x|\mathcal{G}) \text{ a.s.}$$

Proof of MCT, Fatou, DCT. see pages 346-347

⑥ Suppose that X, Y random variables, such that $X \in \mathcal{L}_1$, and $X, Y \in \mathcal{L}_1$. Suppose also that Y is \mathcal{G} -measurable.

Then $\mathbb{E}(XY|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$ a.s.

Proof:

since $E(X|G)$ is G -measurable and L_1
and Y is G -measurable and L_1 , it follows that

$YE(X|G)$ is G -measurable and L_1 . (can be shown that
since $XY \in L_1$, then $YE(X|G) \in L_1$)

It suffices to show that

$$\int_G YE(X|G) dP = \int_G XY dP \quad \text{for all } G \in \mathcal{G}$$

We prove this first for $Y = \mathbb{1}_H$ where $H \in \mathcal{G}$.

Then

$$\begin{aligned} \int_G YE(X|G) dP &= \int_G \mathbb{1}_H E(X|G) dP \\ &= \int_{\mathcal{G} \cap H} E(X|G) dP = \int_{\mathcal{G} \cap H} X dP = \int_G \mathbb{1}_H X dP = \int_G XY dP \text{ as required} \end{aligned}$$

suppose that Y is a simple random variable.

write $Y = \sum_{i=1}^n c_i \mathbb{1}_{H_i}$ where $H_i \in \mathcal{G}$.

$$\begin{aligned} \int_G YE(X|G) dP &= \int_G \sum_{i=1}^n c_i \mathbb{1}_{H_i} E(X|G) dP \\ &= \sum_{i=1}^n c_i \int_G \mathbb{1}_{H_i} E(X|G) dP \\ &= \sum_{i=1}^n c_i \int_G \mathbb{1}_{H_i} X dP \\ &= \int_G Y X dP \text{ as required.} \end{aligned}$$

Next. prove this for all $Y \geq 0$, using

$\sum_{i=1}^n c_i \mathbb{1}_{H_i} \nearrow Y$. Then prove the result for general Y using

$Y = y^+ - y^-$. This is left for you. QED.

STAT 901 Dec. 1 2008

⑥ Y integrable, $A_m, m \geq 1$ such that $P(A_m) \rightarrow 0$

prove that $\int_{A_m} Y dP \rightarrow 0$

Define $Y_m = Y \mathbb{1}_{A_m}$ since $P(A_m) \rightarrow 0$, therefore $Y_m \rightarrow 0$ a.s.
 Also $|Y_m| \leq |Y|$, so by PCT $\int Y_m dP \rightarrow \int 0 dP = 0$ still wr

$Y_m \rightarrow 0$ $Y_m = Y \mathbb{1}_{A_m} \Rightarrow P(Y_m \neq 0) \leq P(A_m) \rightarrow 0$
 $\Rightarrow P(Y_m \neq 0) \rightarrow 0$ $(Y_m \text{ outside converge in probability})$

So $Y_m \xrightarrow{P} 0$ subsequence a.s. converges.

$$\int Y_{m_k} \rightarrow 0 \quad \int_{A_{m_k}} Y dP \xrightarrow{as j \rightarrow 0} 0$$

problem 2(c) (iii) The event $\left[\sum_{n=1}^{\infty} X_n = c \right]$ is not tail event
 $\uparrow \quad \uparrow$

What it converges to is not tail event
 it converges that is tail event.