ECE 534
January 18th, 2017

LECTURE 1

Overview of basic concepts in probability theory
Axiomatic approach to probability by Kolmogorov
(an axiom is a postulate assumed to be true without a proof required)
Axiomatization allows you to set a number of primitives that may be used to deductively derive more complex truths.

Example

Peano's axiomatization of natural numbers $(0,1,2,3,4,\dots)$ (1898)

S - unary function ("successor")

- There is a natural number 0
- Every natural number a has a successor Sa . ($S5=6$)
- No natural number has 0 as its successor
- Distinct numbers have distinct successors : $a \neq b \Rightarrow Sa \neq Sb$
- If a property is true of 0, and of the successor of every natural number who has that property, then the property is held by all natural numbers

Axioms should be broad enough to allow profound results and truths to be deduced, but they should not be redundant.

Axioms associated with an experiment that has random outcomes

Example

Take a die. Roll it. Observe numerical outcome.

Q1 What can I possibly observe?

Answ $\{1,2,3,4,5,6\}$ - the set of all possible outcomes
= non empty set which we refer to as sample space, Ω
$\omega \in \Omega$ is called an outcome

Q2 What am I interested in? What if I care only if the # is even or odd?

Answ Outcomes of interest are called events.
Say $E=\{1,3,5\}$, $O=\{2,4,6\}$. Events are subsets of Ω
Events of interest have to have a special structure.
If we denote the set of events by $\mathcal{F}$, we want $\mathcal{F}$ to be a σ-algebra, that is

A1 $\Omega \in \mathcal{F}$

A2 $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

A3 $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$. Also, if $A_1, A_2, \dots$ is a sequence of events $\in \mathcal{F}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

For our example, $\mathcal{F} = \{\Omega, \emptyset, \{1,3,5\}, \{2,4,6\}\}$

De Morgan's laws : $(A \cup B)^c = A^c \cap B^c$
$(A \cap B)^c = A^c \cup B^c$

Hence, have redundant axiom that if $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

Q3 How likely is the observation

Answ We need to assign a measure to every possible event that tells you how "often" you may expect to see that event

That measure is called probability, P

$P(\{1\}) = P(1) = 1/6$, $P(2) = 1/6$, ..., $P(6) = 1/6$

The measure also has to satisfy some properties, which are axiomatized

A_i, $i \in I$ are mutually exclusive if $A_i \cap A_j = \emptyset \ \forall i, j \in I$ s.t. $i \neq j$

P1 $P(A) \geq 0 \ \forall A \in \mathcal{F}$

P2 If $A, B \in \mathcal{F}$, $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$
If $A_1, A_2, \dots$ is a sequence of mutually exclusive events then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

P3 $P(\Omega) = 1$

Things we can deduce

$A \subseteq B \Rightarrow P(B) = P(A) + P(B \setminus A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ etc

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space

Later

Bertrand's paradoxes (Josept Bertrand)

Consider an equilateral triangle inscribed in a circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than the side of the triangle?

Let $\Omega = \{\omega : 0 \leq \omega \leq 1\}$

Want $\mathcal{F}$ to include closed intervals, and want "measure" to be length

i.e. $[a,b] \in \mathcal{F} \quad \forall\, a,b \in [0,1]$, $a \leq b$

$$P([a,b]) = b-a$$

$\mathcal{F}$ has to include open intervals, singletons etc. So why not take

$\mathcal{F} = 2^{\Omega}$ = power set of Ω = set of all subsets of Ω?

[Cantor's theorem states that the power set of any set has cardinality > card. of it]

Length (measure) is tricky to assign to all sets in the power set

Proof of Cantor's theorem

Take $S = \{1,2,3,\ldots\}$

$2^S = \{\emptyset, 1, 2, \ldots, \{1,2\}, \ldots\}$

same cardinality as

$\{1,2,3,\ldots\}$

Pairing

$1 \leftrightarrow \emptyset$

$2 \leftrightarrow \{1,2\}$

$3 \leftrightarrow 4$

$4 \leftrightarrow \{3,4,5\}$

etc

i is selfish if on LHS and RHS

is is unselfish if on LHS but not RHS

Since 2^S is a power set it must contain the set of all unselfish numbers. What is this set associated with?

if associated with selfish LHS, then contradiction as set is unselfish

if associated with unselfish LHS, then contradiction as LHS should then be in set, which means number should be selfish

Simply, too many sets in the power set and "similar" contradictions arise: $A \subseteq \mathbb{R}$

$0 \leq \text{length}(A) \leq \infty$

$\text{length}([a,b]) = b-a$, $a<b$

$\text{length}(A+y) = \text{length}(A) \quad \forall A \subseteq \mathbb{R}, y \in \mathbb{R}$

$A+y = \{a+y : a \in A\}$

if $A = \bigcup_{i=1}^{\infty} B_i$, $B_i \cap B_j = 0$ $\forall i \neq j$

then $\text{length}(A) = \sum_{i=1}^{\infty} \text{length}(B_i)$

$\mathbb{Q}$ – rational #s
$\mathbb{Q}$ is countably infinite

} Vitaly sets

Say $x, y \in \mathbb{R}$ are equivalent if $x - y \in \mathbb{Q}$
(equivalence : $x \sim x$, $x \sim y \Rightarrow y \sim x$, $x \sim y, y \sim z \Rightarrow x \sim z$)
Define sets $Q_x = \mathbb{Q} + x$ $\forall x \in \mathbb{R}$. For all $x \neq y$
Either $Q_x = Q_y$ or $Q_x \cap Q_y = \emptyset$

$Q_x \cap [0,1] \neq \emptyset$ $\forall x \in \mathbb{R}$

V = one element from each Q_x (Axiom of choice)

$V \subseteq [0,1]$

$Q_{[-1,1]} = \{\tilde{q}_1, \tilde{q}_2, \dots\}$ rationals in $[-1, 1]$

$V_i = V + \tilde{q}_i$

V_i's are disjoint

$[0,1] \subset \bigcup_{i=1}^{\infty} V_i \subset [-1,2]$

V_i's translations of V, all have to have the same length

$$\text{length}([0,1]) = 1 \overset{?}{\le} \text{length}\left(\bigcup_{i=1}^{\infty} V_i\right) \overset{?}{\le} \text{length}[-1,2] = 3$$

$\mathcal{F}$ = smallest σ-algebra containing all the open subsets of Ω

= intersection of all σ-algebras containing all the open subsets of Ω

= Borel σ-algebra over $[0,1]$

Not every subset of Ω is a Borel set
(Use Lebesgue measure on Borel sets as prob. measure to be "safe")

Cylinder sets: Cartesian product $X = \prod X_\alpha$ of topological spaces X_α. The canonical projection $p_\alpha : X \to X_\alpha$. Given any α open set $U \subset X_\alpha$, $p_\alpha^{-1}(U)$ is called an open cyl.
The intersection of a finite # of open cylinders is a cylinder set

$S = \{1, 2, \dots n\}$

$S^{\mathbb{Z}} = \{x = (\dots, x_{-1}, x_0, x_1, \dots) : x_k \in S \ \forall k \in \mathbb{Z}\}$

Open cylinders $C_t[a] = \{x \in S^{\mathbb{Z}} : x_t = a\}$

$C_t[a_0, \dots, a_m] = C_t[a_0] \cap \dots \cap C_{t+m}[a_m] = \{x \in S^{\mathbb{Z}} : x_t = a_0, \dots, x_{t+m} = a_m\}$

Used for repeated trials : ω = outcome = infinite sequence
events = cylinder sets

Continuity of the probability measure

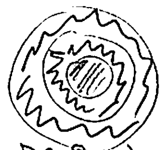
Suppose $B_1, B_2,$ is a sequence of events

a) If $B_1 \subset B_2 \subset \dots$ then $\lim_{j\to\infty} P(B_j) = P(\bigcup_{i=1}^{\infty} B_i)$

b) If $B_1 \supset B_2 \supset \dots$ then $\lim_{j\to\infty} P(B_j) = P(\bigcap_{i=1}^{\infty} B_i)$

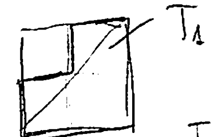
$\lim_{n\to\infty} a_n = a$: $\forall \varepsilon>0, \exists\, n_0(\varepsilon)$ s.t. $\forall n \geq n_0(\varepsilon), |a_n - a| < \varepsilon$

(a) Let $D_i = B_i \setminus B_{i-1}$, $i \geq 2$, and $D_1 = B_1$. The sets D_i are disjoint

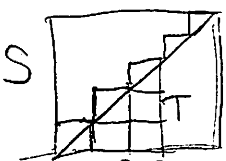


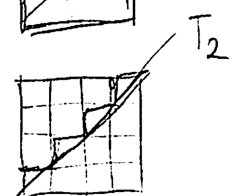
$$\lim_{j\to\infty} P(B_j) = \lim_{j\to\infty} P\left(\bigcup_{i=1}^{j} D_i\right) = \lim_{j\to\infty} \sum_{i=1}^{j} P(D_i) = \sum_{i=1}^{\infty} P(D_i)$$

$$= P\left(\bigcup_{i=1}^{\infty} D_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right)$$



Example





$\Omega = [0,1] \times [0,1]$

P - Lebesgue

events

$\mathcal{F} = \sigma(\{\square$ subsets of $S\})$
$= \sigma$(cylinder sets of S)

$T_1 \supset T_2 \supset T_4 \dots \supset T_{2^j}$, $P(T_n) = \frac{1}{n^2} + \frac{2}{n^2} + \dots + \frac{n}{n^2} = \frac{n+1}{2n}$

$\bigcap_{j=1} T_{2^j} = T$

$T \in \mathcal{F}$, and $P(T) = P(\bigcap_{j=1}^{\infty} T_{2^j}) = \lim_{j\to\infty} P(T_{2^j}) = \frac{1}{2}$

Independence and Conditional probability

Events $A_1, .., A_n$ are independent if

$$P(A_i A_j) = P(A_i)P(A_j) \qquad i \neq j$$

$$P(A_i A_j A_k) = P(A_i)P(A_j)P(A_k) \qquad i \neq j \neq k$$

$\vdots$

$$P(A_1 \dots A_n) = P(A_1) \dots P(A_n)$$

Parwise independence

Let B be an event s.t. $P(B) > 0$

Conditional prob

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Total law of probability and
Bayes formula

Let $E_1, \dots, E_n$ be events that form a partition of Ω
(i.e. $\bigcup E_i = \Omega$ and $E_i \cap E_j = \emptyset$ for $i \neq j$)

Then $P(A) = P(AE_1) + \dots + P(AE_n)$
$= P(A|E_1)P(E_1) + \dots + P(A|E_n)P(E_n)$

Since $P(E_i|A) = \dfrac{P(AE_i)}{P(A)} = \dfrac{P(A|E_i)P(E_i)}{P(A)}$

we can get to Bayes formula

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A|E_1)P(E_1) + \dots + P(A|E_n)P(E_n)}$$

Borel-Cantelli lemma

Consider a sequence of events $A_1, A_2, \dots$
and define

$$A = \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n \quad \left(\limsup_{m \to \infty} A_m\right)$$

$\forall k\ \exists n \geq k$ s.t. a sample ω belongs to some A_n

$A = \{\omega : \omega$ belongs to infinitely many A_i's$\}$

may also define

$$A = \bigcup_{k \geq 1} \bigcap_{n \geq k} A_n \qquad \liminf_{m \to \infty} A_m$$

$\exists k$ s.t. a sample ω belongs to all A_n $n \geq k$

$A = \{\omega : \omega$ belongs to all but finitely many A_i's$\}$

[Recall classical def. of limsup and liminf for sequences

$$\limsup_{n \to \infty} a_n = \lim_{n \to \infty} \sup_{k \geq n} a_k$$]

Let $p_n = P(A_n)$. If $\sum p_n < \infty$, then $P\{A_n$ infinitely often$\} = 0$
If $\sum p_n = \infty$, $A_1, \dots, A_n$ mutually independent, then $P\{A_n\ i.o.\} = 1$

Proof From continuity of probability
(as $\bigcup_{n \geq k} A_i$ is a sequence of nonincreasing sets)

$$P\{A_n \text{ infinitely often}\} \overset{\text{Cont}}{=} \lim_{k \to \infty} P\left(\bigcup_{n \geq k} A_n\right)$$

$B_n = \bigcup_{n \geq k} A_n \qquad B_1 \supset B_2 \supset \dots \quad \lim_{n \to \infty} P(B_n) = P\left(\bigcap_k \bigcup_{n \geq k} A_n\right)$

(*) $P\left(\bigcup_{n\geq k} A_n\right) \overset{\text{cont}}{=} \lim_{m\to\infty} P\left(\bigcup_{n=k}^{m} A_n\right) \leq \lim_{m\to\infty} \sum_{n=k}^{m} p_n = \sum_{n=k}^{\infty} p_n$

If $\sum_{n=1}^{\infty} p_n < \infty$, then $\lim_{k\to\infty} \sum_{n=k}^{\infty} p_n = 0$

(**) For the proof of the second part, de Morgan

$$P\left(\bigcup_{n\geq k} A_n\right) \overset{\text{cont}}{=} \lim_{m\to\infty} P\left(\bigcup_{n=k}^{m} A_n\right) = \lim_{m\to\infty} P\left(\left(\bigcap_{n=k}^{m} A_n^c\right)^c\right)$$

$$= \lim_{m\to\infty}\left[1 - P\left(\bigcap_{n=k}^{m} A_n^c\right)\right] = \lim_{m\to\infty}\left[1 - \prod_{n=k}^{m}(1-p_n)\right]$$

Now observe that $\exp(-u) = 1 - u + \frac{u^2}{2} - \dots \geq 1-u, \ \forall u$

$$\geq \lim_{m\to\infty}\left[1 - \exp\left(-\sum_{n=k}^{m} p_n\right)\right] = 1 - \exp\left(-\sum_{n=k}^{\infty} p_n\right) = 1 - \exp(-\infty) = 1$$

(*) $B_m = \bigcup_{n=k}^{m} A_n$ $\qquad B_k \subseteq B_{k+1} \subseteq \dots$

$$P\left(\bigcup_{n\geq k} A_n\right) = P(B_\infty) = \lim_{m\to\infty} P(B_m) = \lim_{m\to\infty} P\left(\bigcup_{n=k}^{m} A_n\right)$$

This result is a form of a 0-1 law of probability that often apply to tail events ($\limsup A_n$, $\liminf A_n$ are tail events) (only depend on the limiting behavior of the sequence)

Random variables

$(\Omega, \mathcal{F}, P)$, $(\mathbb{R}, \mathcal{M})$ - measurable space

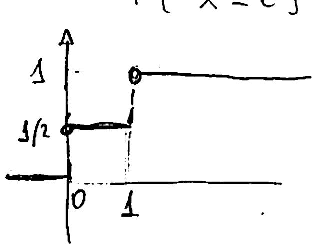
$\forall B \subseteq \mathcal{M}$, $X^{-1}(B) \in \mathcal{F}$

$(\Omega, \mathcal{F}, P)$

$X: \Omega \to \mathbb{R}$ s.t. $\{\omega: X(\omega) \leq c\} \in \mathcal{F}$ for any $c \in \mathbb{R}$

A random variable defined as above is said to be $\mathcal{F}$-measurable

$\text{CDF} = P\{\omega: X(\omega) \leq c\} = F_X(c)$
$= P\{X \leq c\} \qquad = F_X(c)$



$P\{X < c\} = ?$

Sequence $\{c_i\}_{i=1}^{\infty} \to c$ from the left $\quad c_1 < c_2 < \dots$

$\{X \leq c_j\} \supseteq \{X \leq c_i\}$ for $j > i$

$\{X \leq c_1\} \subseteq \{X \leq c_2\}$

$$P\{X < c\} = P\left\{\bigcup_{i=1}^{\infty} \{X \leq c_i\}\right\} = \lim_{i\to\infty} P\{X \leq c_i\} = \lim_{i\to\infty} F_X(c_i)$$

$$= F_X(c^-) \Rightarrow P\{X = c\} = F_X(c) - F_X(c^-)$$

A function F is a CDF of some RV iff it has the following properties

a) F is non-decreasing

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

c) F is right continuous

only if : F is a CDF of some X

a) $F(x) = P\{X \le x\} = P\{X \le z\} + P\{z < X \le x\} \Big|_{z < x} \ge P\{X \le z\}$

b) Only show $\lim_{x \to \infty} F(x) = 1$

$B_n = \{X \le n\}$, $n \in \mathbb{N}$

$B_1 \subseteq B_2 \subseteq \dots$

$\lim_{n \to \infty} F(n) = \lim_{n \to \infty} P\{B_n\} = P\{\bigcup_{n=1}^{\infty} B_n\} = P\{\Omega\} = 1$

What about the reals?

For any $\varepsilon > 0$, $\exists\, n(\varepsilon)$ s.t. $\forall x \ge n(\varepsilon)$, $F(x) \ge 1 - \varepsilon$ (*)

Hence, $\lim_{x \to \infty} F(x) = 1$

c) $A_n = \{X \le x + \frac{1}{n}\}$, $n \ge 1$

$A_1 \supset A_2 \supset A_3 \supset \dots$

$\lim_{n \to \infty} F(x + \frac{1}{n}) = \lim_{n \to \infty} P(A_n) = P\left(\bigcap_{i=1}^{\infty} A_i\right) = P\{X \le x\} = F(x)$

Same as in (*), we have that this holds for all $\mathbb{R}$.

<u>if</u> Difficult. Measure theory (see, for example, Th. 14.1 of Billingsley, Probability and Measure)

Say $\Omega = (0, 1)$

$\mathcal{F}$ = Borel σ-algebra

P = Lebesgue measure

Hinges on property of Lebesgue measure, easiest to show if $F(x)$ is monotonic, in which case there is a 1-1 mapping between $\Omega = (0,1)$ and $\mathbb{R}$, and hence a reverse mapping φ

$$P\{X \le x\} = P\{\omega \in (0,1) : \varphi(\omega) \le x\}$$
$$= P\{\omega \in (0,1) : \omega \le F(x)\} = F(x)$$

Discrete RVs : Probability mass function
$$p(x) = P\{X = x\}$$

Continuous RVs :

Absolute continuity (strong smoothness)
stronger than continuity of uniform continuity

Allows rigorous use of the fundamental theorem of calculus

$I \subseteq \mathbb{R}$, an interval

$f : I \to \mathbb{R}$ AC on I if $\forall \varepsilon > 0$, $\exists \delta > 0$

s.t. whenever for a collection of int. (x_k, y_k), $k = 1, 2, \ldots$ s.t.

$(x_k, y_k) \subset I$ $\quad (x_k, y_k) \cap (x_m, y_m) = \emptyset$, $k \neq m$

it holds
$$\sum_k (y_k - x_k) < \delta$$
then
$$\sum_k |f(y_k) - f(x_k)| < \varepsilon$$

Does not fluctuate on a set of measure zero

Maps sets of measure zero to sets of measure zero

f absolutely continuous $\Rightarrow$ f has a derivative f' almost everywhere
the derivative is Lebesgue integrable, and
$$f(x) = f(a) + \int_a^x f'(t)dt \qquad \forall x [a, b] \quad \text{(Lebesgue integral)}$$

Cantor sets - Cantor function

Lebesgue vs. Riemann integral

Later, as examples

If F is absolutely continuous, can define so called pdf $f_X(x)$ s.t.
$$F(x) = \int_{-\infty}^{x} f_X(u)du \quad (\text{why? since } F(-\infty) = 0)$$

If $f(x)$ is continuous at u, have $\quad f(u) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_u^{u+\varepsilon} f(v)dv$

We take the integrals to be Lebesgue integrals

Riemann integral

$[a, b]$ partition, say $a = x_1 < x_2 \ldots < x_n = b$ so that $[a, b] = \bigcup_{i=1}^{n-1} [x_i, x_{i+1}]$

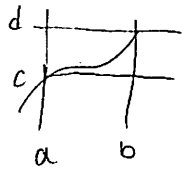
$m = \max (x_{i+1} - x_i)\Big|_{i=1}^{n-1}$ (want m to be very small)

tagged with $t_i \in [x_i, x_{i+1}]$, $i = 1, \ldots, n-1$

Riemann sum $\displaystyle\sum_{i=1}^{n-1} f(t_i)(x_{i+1} - x_i) \rightarrow$ upper sum (sup)
lower sum (inf)

take inf. over all partitions of the upper sum
take sup over all partitions of the lower sum
if they are equal, call the sum the Riemann integral

Functions that are not Riemann integrable : $\mathbb{1}_{\mathbb{Q}}$
Lebesgue integral : uses Lebesgue sums instead; we partition the range not the domain



$$[c,d] = \bigcup_{i=0}^{n-1} [y_i, y_{i+1}]$$

$$E_i = f^{-1}([y_i, y_{i+1}]) = \{x \in [a,b] \mid y_i \le f(x) \le y_{i+1}\}$$

$$y_i^* \in [y_i, y_{i+1}]$$

$$\text{sum} = \sum_{i=1}^{n-1} y_i^* \mu(E_i)$$

Expectation of a RV

* Simple RV: Discrete $E[X] = \sum_{i=1}^{m} x_i P\{X = x_i\}$

* simple : $\exists$ a finite set $\mathcal{X} = \{x_1, \dots, x_n\}$ s.t. $X(\omega) \in \mathcal{X}$ $\forall \omega$

If X has a pdf, $E[X] = \int_{-\infty}^{+\infty} x f(x) dx$ in the Lebesgue sense (mention what well defined means)

Properties

- Linearity (follows from sum-derived definition of integrals, i.e., integration being a linear operator)

  $E[X+Y] = E[X] + E[Y]$, $E[cX] = cE[X]$ for c a const.

- $E[X] = \int_0^{+\infty} (1 - F_X(x)) dx - \int_{-\infty}^{0} F_X(x) dx$ / at least one of the two integrals is finite

Function of a RV

$(\Omega, \mathcal{F}, P)$

$X : \Omega \to \mathbb{R}$ so that $\{\omega : X(\omega) \le c\} \in \mathcal{F}$ $\forall c \in \mathbb{R}$

$g : \mathbb{R} \to \mathbb{R}$ such that $\{x \in \mathbb{R} : g(x) \le c\}$ is a Borel subset of $\mathbb{R}$

$Y(\omega) = g(X(\omega))$ composition of two functions

We have $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$ in the Lebesgue sense

~~Higher order moments~~

$E[X^i]$ $i \ge 1$

* $\text{var}(X) = E[(X - E[X])^2] = E[X^2] - E^2[X]$

** Characteristic function $\Phi_X(u) = E[e^{juX}]$, $\Phi_X^{(k)}(0) = j^k E[X^k]$ if the higer order moments exist

Examples of distributions that are frequently used

Bernoulli X

$$P\{X=0\} = 1-p, \quad P\{X=1\} = p, \quad \text{where } p \in [0,1]$$

$$E[X] = p, \quad var(X) = p(1-p)$$

Binomial X

$n \geq 1$, $p \in [0,1]$

$$P\{X=i\}\Big|_{i\in[0,n]} = \binom{n}{i} p^i (1-p)^{n-i}$$

$$E[X] = np, \quad var(X) = np(1-p)$$

Poisson X

$$P\{X=i\}\Big|_{i=0}^{+\infty} = \frac{\lambda^i}{i!} e^{-\lambda}, \quad \text{where } \lambda > 0$$

$$E[X] = \lambda, \quad var(X) = \lambda$$

Geometric X

$0 \leq p \leq 1$

$$P\{X=i\}\Big|_{i=1}^{\infty} = (1-p)^{i-1} p$$

$$E[X] = \frac{1}{p}, \quad var(X) = \frac{1-p}{p^2}$$

Continuous

Gaussian $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$E[X] = \mu, \quad var(X) = \sigma^2$$

Exponential

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

$$E[X] = \frac{1}{\lambda}, \quad var(X) = \frac{1}{\lambda^2}$$

$$P\{X \geq s+t \mid X \geq s\} = P\{X \geq t\}$$
$s, t, \geq 0$

Uniform

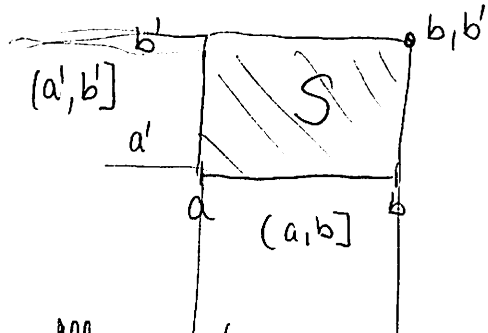
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \forall x \in [a,b] \\ 0, & \text{elsewhere} \end{cases}$$

$U(a,b)$

Conditional densities and jointly distributed RVs

$X_1, X_2, \ldots, X_m$ RVs over the same probability space $(\Omega, \mathcal{F}, P)$

$$F_{X_1 X_2 \ldots X_m}(c_1, \ldots, c_m) = P\{X_1 \le c_1, \ldots, X_m \le c_m\}$$



$$P\{(X_1, X_2) \in S\} = F_{X_1X_2}(b,b') - F_{X_1X_2}(a,b') - F_{X_1X_2}(b,a') + F_{X_1X_2}(a,a')$$

All properties/concepts carry over from 1-dim case: joint pdf $f(x_1, \ldots, x_m)$ for absolutely continuous CDFs

In particular, we are interested in joint pdf's $f_{X,Y}(x,y)$ based on which we can define conditional pdf's

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f(y)}, \text{ provided } f(y)$$

and for $-\infty < x < \infty$

Correlation and covariance

X, Y two random variables over the same probability space with finite second moments

Correlation $E[XY]$

Covariance $E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])]$

Correlation coefficient $\rho_{XY} = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

A number of useful properties of correlations/covariances follows from Schwarz's inequality

$$|E[XY]| \le \sqrt{E[X^2]E[Y^2]}$$

Note: we assumed $E[X^2], E[Y^2]$ are finite

Hölder's inequality Let $(S, \mathcal{F}, \mu)$ be a measure space, $p, q \in [1, \infty]$ s.t. $1/p + 1/q = 1$. Then, for all measurable $\mathbb{R}$- or $\mathbb{C}$-valued functions f, g over S, we have

$$\|fg\|_1 \le \|f\|_p \|g\|_q$$

where $\|f\|_p = \left(\int_S |f|^p \, d\mu \right)^{1/p}$

Minkowski's inequality $\|f+g\|_p \le \|f\|_p + \|g\|_p$

Observe that

$$(a+b)^2 = a^2+b^2+2ab = a^2+b^2+2\sqrt{a^2b^2} \overset{\downarrow \text{ geometric vs. arithmetic mean}}{\leq} a^2+b^2+\cancel{2}\frac{a^2+b^2}{\cancel{2}}$$
$$= 2a^2+2b^2$$

so that $E[X^2]$, $E[Y^2]$ being finite implies $E[(X+\lambda Y)^2]$ being finite for some constant λ

Take $\lambda = -\dfrac{E[XY]}{E[Y^2]}$

$$0 \leq E\left[\left(X - \frac{E[XY]}{E[Y^2]}Y\right)^2\right]$$
$$= E\left[X^2 - \frac{2E[XY]^2}{E[Y^2]} + \frac{E[XY]^2}{E[Y^2]^{\cancel{2}}}\cancel{E[Y^2]}\right]$$
$$= E[X^2] - \frac{E[XY]^2}{E[Y^2]}$$
$$\Rightarrow E[XY]^2 \leq E[X^2]E[Y^2]$$

Cauchy-Schwarz gives

$$|\mathrm{Cov}(X,Y)| \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$$

and consequently $-1 \leq \rho_{XY} \leq 1$

It is also useful to know that

$$\mathrm{Cov}(X+Y, U+V) = \mathrm{Cov}(X,U) + \mathrm{Cov}(X,V) + \mathrm{Cov}(Y,U) + \mathrm{Cov}(Y,V)$$

and $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$

Inequalities connecting probabilities and expectations:

Markov's inequality

If f is a monotonically increasing function on $\mathbb{R}^+$, X is a RV, $a \geq 0$, $f(a) > 0$ then

$$P\{|X| \geq a\} \leq \frac{E[f(|X|)]}{f(a)}$$

special case $P\{X \geq 0\} = 1$, $c > 0$

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

Chebyshev inequality: X has finite expectation μ, variance σ^2

$$P\{|X-\mu| \geq a\} \leq \frac{\sigma^2}{a^2}$$

Jensen's inequality

If X is a RV, f a convex function, then

$$f(E[X]) \leq E[f(X)]$$

Some examples

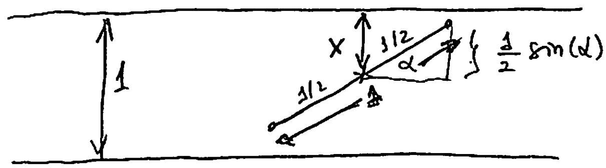
Buffon's needle problem

See MSTE (Office for Math, Science and Technology education)

Needle of length $l=1$

Regular grid of parallel lines at distance $d=1$ from each other

You throw a needle randomly



If $\frac{1}{2}\sin(\alpha) \geq X$, we will intersect the line, if $X > \frac{1}{2}\sin(\alpha)$ we will not

Joint pdf of (D, α)

$$f(x,\alpha) = \frac{2}{\pi} \cdot \frac{1}{1/2} = \frac{4}{\pi}$$

prob. of crossing a line

$$P = \int_0^{\pi/2} \int_0^{\frac{1}{2}\sin(\alpha)} \frac{4}{\pi}\, dx\, d\alpha = \frac{4}{\pi}\int_0^{\pi/2} \frac{1}{2}\sin\alpha\, d\alpha = \frac{2}{\pi}\left(-\cos\alpha\right)\Big|_0^{\pi/2} = \frac{2}{\pi}$$

Additional discussion

There is no set cardinality strictly between that of the integers and reals. Infinite sets have different cardinalities, which are described by the so-called aleph numbers ($\aleph_0, \aleph_1, \aleph_2, \ldots$)

$\aleph_0$ = cardinality of integers, rational numbers, etc

$2^{\aleph_0}$ = cardinality of reals

$\aleph_1$ = aleph-one

Continuum hypothesis (Cantor) $\aleph_1 = 2^{\aleph_0}$

We need an extension theorem from measure theory to ensure consistent def of probability

$\mathcal{F}$: Revise to accomodate the following

$$\text{if } B \subset A \in \mathcal{F} \text{ and } P(A) = 0 \Rightarrow B \in \mathcal{F} \text{ and } P(B) = 0$$

Lebesgue measure

Given a set $E \in \mathcal{F}$, and length of $I = [a, b]$ given by $\ell(I) = b - a$, the Lebesgue outer measure $\lambda^*(E)$ is defined as

$$\lambda^*(E) = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) : (I_k)_{k \in \mathbb{N}} \text{ is a sequence of open intervals with } E \subseteq \bigcup_{k=1}^{\infty} I_k \right\}$$

Note: every Borel set is Lebesgue measurable (converse is not true)

Note on axiom of choice

"Choice" is associated with distinguishable characteristic. If no such characteristic is available and we have infinitely many choices to make, we need to invoke AC.

Note on Gödel's incompleteness theorem

Any statement expressed using axioms should be provable true or false, in case that the axioms are complete. Gödel's IT asserts (vaguely) that any consistent set of axioms based on which some form of arithmetic can be carried out is incomplete

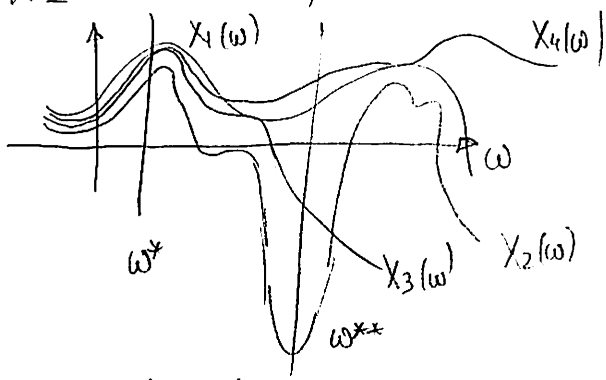
Mon/Tue
6 7 March

Mon
27 Feb

April
3/4/5
M W
April
10

Convergence of random variables

$X : \Omega \to \mathbb{R}$ over some prob space $(\Omega, \mathcal{F}, P)$
s.t. if $A \subseteq \mathcal{F}^B$ over $\mathbb{R}$, then $X^{-1}(A) \in \mathcal{F}$



"Convergence of functions" with the added prob. twist

What happens at each point?
Functions converge to another function: $X_1, \dots, X_n, \dots \to \underline{\underline{X}}$

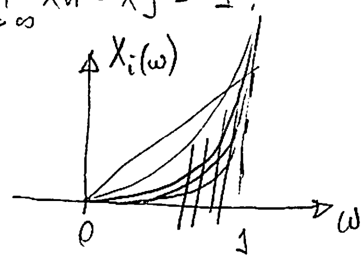
Find the set of points ω on which $X_1(\omega), X_2(\omega), \dots$ converges to some value $X(\omega)$ (e.g. ω^*). Denote this set of points as C, and let the values of the limits be summarized by a function $X(\omega)$. There may be some points on Ω (e.g., ω^{**}) where there is no convergence. There, let $X(\omega)$ be arbitrary.

We say that $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$ a.s. (almost surely) if $P(C) = 1$.

$X_n \to$ almost surely X — Alternatively, $X(\omega)$ is the almost sure limit of a sequence of random variables $X_n(\omega)$, $n = 1, 2, \dots$ defined over the same probability space if $P\{\lim_{n\to\infty} X_n = X\} = 1$.

1. Example $(X_n : n \geq 1)$

$X_n(\omega) = \omega^n$
$\omega \in [0,1] = \Omega$
$(\Omega, \mathcal{F}, P)$ — $\mathcal{F}$: "Borel", P: "length"



$$\lim_{n\to\infty} X_n(\omega) = \begin{cases} 0, & 0 \leq \omega < 1 \\ 1, & \omega = 1 \end{cases}$$

$P\{1\} = 0$

so,

$\lim_{n\to\infty} X_n(\omega) = 0$ or $X_n(\omega) \xrightarrow{a.s.} 0$

2. Example $X_1, X_2, \dots$ over $\Omega = \{H, T\}$, same prob. space

$$X_n(\omega) = \begin{cases} \frac{n}{n+1}, & \omega = H \\ (-1)^n, & \omega = T \end{cases}$$

if $\omega = H$ $X_1(\omega) = \frac{1}{2}$, $X_2(\omega) = \frac{2}{3}, \dots$ $X_n(\omega) \xrightarrow[n\to\infty]{} 1$

if $\omega = T$ $X_1(\omega) = -1$, $X_2(\omega) = 1$, $X_3(\omega) = -1, \dots$ $X_n(\omega)$ does not

So, if the coin is fair ($P\{H\} = P\{T\} = 1/2$) then $X_n(\omega)$ does not converge to any random variable.

A sequence of random variables $X_n \to X$ almost surely if $X_n|_{n=1}^{\infty}$, X are over the same probability space and

$$(**) \quad \sum_{n=1}^{\infty} P\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\} < \infty$$

but not iff

Ex Show that X_n, $n = 1, 2, \ldots$ defined according to

$$X_n = \begin{cases} -\frac{1}{n}, & \text{prob. } 1/2 \\ \frac{1}{n}, & \text{prob. } 1/2 \end{cases}$$

converges a.s. to 0

$$P\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\} = P\{\omega : |X_n(\omega)| > \varepsilon\}$$

$|X_n(\omega)| > \varepsilon$ only happens if $n < 1/\varepsilon$ (as $|X_n(\omega)| = \frac{1}{n}$)

$$\sum_{n=1}^{\infty} P\{|X_n| > \varepsilon\} = \sum_{n=1}^{\lceil 1/\varepsilon \rceil} P\{|X_n| > \varepsilon\} \leq \left\lceil \frac{1}{\varepsilon} \right\rceil < \infty$$

Convergence in probability

$X_n \xrightarrow{prob} X$, where $X_1, X_2, \ldots, X$ are over the same prob. space if for any $\varepsilon > 0$

$$\lim_{n \to \infty} P\{|X - X_n| \geq \varepsilon\} = 0$$

$$X_n \xrightarrow{P} X$$

$$\lim_{n \to \infty} P\{\omega : |X(\omega) - X_n(\omega)| \geq \varepsilon\} = 0 \quad \text{Compare to } (**)$$

Analogy : Members of a club

Attendance of meetings of the club

Almost sure : almost all members have <u>perfect</u> attendance

Probability : almost all meeting where <u>full</u>

Compare $P\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\} = 1$ vs.

$$\lim_{n \to \infty} P\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\} = 0$$

A sequence of RVs $X_1, X_2, \dots$ converges to a RV X in the MS (mean-square) sense if all variables are over the same prob. space, $E[X_n^2] < \infty$ $\forall n$ and $\lim_{n\to\infty} E[(X_n - X)^2] = 0$

$$X_n \xrightarrow{m.s.} X$$

From the definition, it follows that $X_n \xrightarrow{m.s.} X \Rightarrow E[X^2] < \infty$

$\lim_{n\to\infty} E[(X_n - X)^2] = 0 \Rightarrow \exists\, n_0$ s.t. $\forall n \geq n_0$ $E[(X_n - X)^2] \leq \varepsilon$, including $E[|X_{n_0} - X|^2] \leq \varepsilon$

Using Minkowski's inequality

$$E[X^2]^{1/2} = E[(X - X_{n_0} + X_{n_0})^2]^{1/2} \leq \underbrace{E[(X - X_{n_0})^2]^{1/2}}_{\leq \varepsilon} + \underbrace{E[(X_{n_0})^2]^{1/2}}_{< \infty} < \infty$$

Def A sequence of RVs $X_1, X_2, X_3, \dots$ converges in distribution to a RV X if $\lim_{n\to\infty} F_n(x) = F(x)$ at all points of continuity of F

$$X_n \xrightarrow{d} X$$

$X_n \xrightarrow{d} X \iff E[f(X_n)] \xrightarrow[n\to\infty]{} E[f(X)]$ for all continuous functions f bounded

$\overset{\text{Levy's cont. th.}}{\iff}$ $\phi_{X_n} \to \phi_X$ pointwise convergence $\qquad \phi =$ Characteristic function

Example

$X_1, X_2, \dots, X_n$ iid uniform $[0,1]$

$M_n = \max(X_1, \dots, X_n)$

$M_n \to ?$

We show that $M_n \xrightarrow{P} 1$

The CDF of M_n is $F_n(x) = x^n$, $x \in [0,1]$

For $\varepsilon > 0$, $P\{|M_n - 1| > \varepsilon\} = P\{M_n < 1 - \varepsilon\} = (1-\varepsilon)^n$

Hence, $\lim_{n\to\infty} P\{|M_n - 1| > \varepsilon\} = 0$

Next, we show that $n(1 - M_n)$ has an interesting limiting distribution

$$P\{n(1 - M_n) \leq x\} = P\{M_n \geq 1 - \tfrac{x}{n}\} = 1 - \left(1 - \tfrac{x}{n}\right)^n$$

$$\xrightarrow[\text{as } n\to\infty]{} 1 - \exp(-x) \;/\; \underline{\text{exponent}}$$

Relationships between different modes of convergence

$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$ $\qquad$ $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$

$X_n \xrightarrow{m.s.} X \Rightarrow X_n \xrightarrow{P} X$

PC $\quad P\{|X_n| \leq Y\} = 1$ for some RV Y with finite second moment
then $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{m.s.} X$

A sequence of RVs can have only one limit, up to differences on a set of probability zero

$X_n \xrightarrow{d} X$, $X_n \xrightarrow{d} Y \Rightarrow X, Y$ have the same distribution

Suppose that X_n has density f. f_n, $n \geq 1$, and X has density f_X

Then $f_n(x) \to f(x)$ for all but countably many x
$\Rightarrow X_n \xrightarrow{d} X$ (Scheffe's theorem)

The converse is not true

a sequence of discrete RVs can converge in distribution to a continuous RV and vice versa

Suppose that $X_1, X_2, \ldots$ are iid RVs with

$$P\{X_i = j\} = \frac{1}{10}, \; j = 0, 1, \ldots, 9$$

$$U_n = \sum_{k=1}^{n} \frac{X_k}{10^k}$$

$U_n \xrightarrow{d} U \qquad U \sim$ uniform $[0,1]$

$$P\{U_n = \frac{j}{10^n}\} = \frac{1}{10^n}, \quad j = 0, \ldots, 10^n - 1$$

For $j/10^n \leq x < (j+1)/10^n$

$$P\{U_n \leq x\} = \frac{j+1}{10^n}$$

and $|P\{U_n \leq x\} - x| \leq 10^{-n} \to 0 \quad n \to \infty$

$\Rightarrow P\{U_n \leq x\} \to x \quad \forall x \in [0,1]$

Continuous mapping theorem

Suppose that g is $\mathbb{R}$-valued and continuous

$$X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$$
$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

Slutsky's theorem

$$X_n \xrightarrow{d} X$$
$$Y_n \xrightarrow{P} \theta, \text{ a constant}$$

Then
$$X_n + Y_n \xrightarrow{d} X + \theta$$
$$X_n Y_n \xrightarrow{d} \theta X$$

Some proofs

$$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$$

For some $\varepsilon > 0$, let

$$A_n = \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}$$

Need to show that a.s $\Rightarrow P\{A_n\} \to 1$

$$B_n = \{\omega : |X_k(\omega) - X(\omega)| < \varepsilon, \ \forall k \geq n\}$$

$$B_n \subseteq A_n$$
$$B_1 \subseteq B_2 \subseteq B_3 \subseteq \ldots$$

$$\lim_{n\to\infty} P(B_n) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = P(B)$$

and $\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} \subset B$

a.s. $\Rightarrow P(B) = 1 = \lim_{n\to\infty} P(B_n) \Rightarrow \lim_{n\to\infty} P(A_n) = 1 \Leftrightarrow X_n \xrightarrow{P} X$

Suppose that $X_n \xrightarrow{m.s.} X$ and let $\varepsilon > 0$

Markov's inequality gives

$$P\{|X_n - X| \geq \varepsilon\} \leq \frac{E[|X - X_n|^2]}{\varepsilon^2} \to 0$$

and $P \geq 0$

$$\Rightarrow P\{|X_n - X| \geq \varepsilon\} \to 0$$

i.e.

$$X_n \xrightarrow{P} X$$

Notice that $\mathcal{A}$ is an MU code, since according to the **Condition 2** C is an MU code and adding runs of ones does not violate the MU property in $\mathcal{A}$. Moreover, our construction guarantees that each substring of length f of an element in $\mathcal{A}$ contains 1^{l+1} and avoids 0^{l+1} as a subword. So, there is no 0^{l+1} to bond with 1^{l+1} and form a PD product of length at least f in $\mathcal{A}$. Therefore, $\mathcal{A}$ is a binary f-APD and MU code. In addition, one can verify that $\mathcal{A}$ also inherites the Error-Correcting property from C, and both codes have the same minimum Hamming distance.

Theorem 1. *Let $A_{\text{APD\_MU}}(n, f)$ denote the maximum size of a f-APD_MU code over a binary alphabet, for positive integers $n = pf$. Then there exist constants $0 < C_6 < C_7$ such that*

$$C_6 \frac{2^n}{n^{p+1}} \leq A_{\text{APD\_MU}}(n, f) \leq C_7 \frac{2^n}{n^{p+1}}$$

Proof. To justify the lower bound we use the aforementioned construction. In this construction $|\mathcal{A}| = |C|$. When $2l \leq x - 2$ the number of binary sequences of length $x - l - 2$ containing no l consecutive zeros is at least

$$\begin{aligned}
&2^{x-l-2} - (x - 2l - 1)\, 2^{x-2l-2} \\
\geq &2^{n-l-2} - n2^{x-2l-2} \\
= &2^{n-p-2-l(p+1)} - n2^{n-p-2-l(p+2)} \\
= &2^{n-p-2} \left[2^{-l(p+1)} - n2^{-l(p+2)}\right]
\end{aligned}$$

The function $2^{-l(p+1)} - n2^{-l(p+2)}$ is maximized when $l = \log_2 \left[n\left(\frac{p+2}{p+1}\right)\right] + \delta$, where δ is chosen so that $|\delta| < 1$ and l is an integer. In this case, the value of $2^{-l(p+1)} - n2^{-l(p+2)}$ is bounded below by $\frac{C_6}{n^{p+1}}$. □