Probability basics

MA 751

Probability and Measure Theory

1. Basics of probability

2 aspects of probability:

Probability = mathematical analysis

Probability = common sense

Probability basics

A set-up of the common sense view:

Define as *experiment* any event with an *outcome*.

Example 1: Toss of a die

Example 2: Record information on deaths of cancer patients.

Example 3: Measure daily high body temperature of a person on chemotherapy

Example 4: Observe the next letter in a chain of DNA:

Probability basics

$$\text{ATCTTCA} \rightarrow ?$$

Given a well-defined experiment we must know the set of all possible outcomes (this must be defined by the experimenter)

Collection of possible outcomes is denoted as

$$S = \text{ sample space.}$$

(sometimes also called a *probability space* and written Ω)

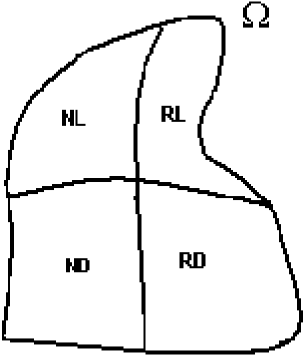
Probability basics

This is the collection of *all possible outcomes* of the experiment

Example 5: Die toss: $S = \{1, 2, 3, 4, 5, 6\}$

Example 6: Cancer outcome records:

Probability basics



4 outcomes: NL, RL, ND, RD

$R =$ received treatment; $N =$ no treatment;
$L =$ lived; $D =$ died

Probability basics

Example 7: High temperature measurement

$$S = \{t : t \quad \text{a real number}\}$$

Many other outcome (sample) spaces S are possible, for example, profiles of functions of DNA regions

So: Have practical situations and for each one we define the sample space S of possible outcomes

Def. 1: For sets A, B:
Define $A \subset B$ if A is a subcollection of B.

If $E \subset S$, E is an *event*

Probability basics

Example 8: Die tossing experiment: if E = even roll = $\{2, 4, 6\} \subset \{1, 2, 3, 4, 5, 6\}$

then E is an event.

Why an *event* ?

Intuitively, an event means something that has occurred, and above the event $A = \{2, 4, 6\}$ represents the *occurrence* of an even number

Probability basics

Again can translate between mathematical definition (subset) and intuitive notions of meanings of words (event).

Probabilist wants to assign a probability $P(E)$ (a number between 0 and 1) to every event E.

Thus, e.g., if $E = \{\text{event of an even roll}\} = \{2, 4, 6\}$ want $P(A) = \frac{1}{2}$ [Rationales can vary]

So: Ideally, want to assign numbers (probabilities) to subsets $E \subset S$.

Example 9: Die toss revisited:

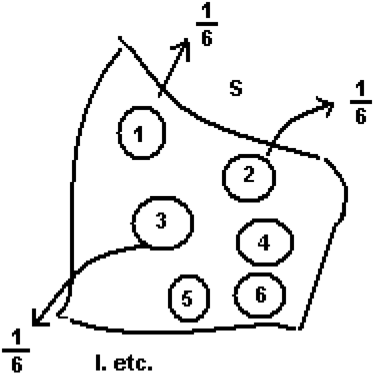
$$P(1) = \frac{1}{6}$$

$$P(2) = \frac{1}{6}$$

$$P(3) = \frac{1}{6}$$

$$P(6) = \frac{1}{6}$$

Probability basics



Each component in S has probability $\frac{1}{6}$.

Probability basics

Any subset $A \subset S$ has probability determined by adding measures of component subsets A_i.

Want $P(S) = 1$

(*why* ?)

Probability basics

Example 10: If we want to predict next base in genomic sequence, we have the sample space

$$S = \{A, T, C, G\}$$

Define

$E =$ event that next base is a purine, i.e.,

$E = \{A, G\} \subset S$.

If we expect all bases have equal probabilities, then

$$P(E) = 1/2.$$

Analytic probability

2. Sample space and probability measure: what properties do we want?

Def. 2: Recall for sets A, B, define $A \cup B$ and $A \cap B$ as the *union* and *intersection* of the sets.

Analytic probability

Desired properties of P:

(1) If A_i are disjoint (i.e., no pair of them intersects) then the probability of the union of the A_i is the sum of their probabilities:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \qquad \text{(Linearity)}$$

If $\phi =$ empty set:

(2) $$P(\phi) = 0, \quad P(S) = 1.$$

Analytic probability

An assignment of numbers $P(E)$ to subsets (events) $E \subset S$ satisfying properties (1) and (2) is called a *measure* on the set S.

However, we will sometimes need to restrict the collection of sets E to which we can assign a probability (measure) $P(E)$.

$\Rightarrow$ need a collection $\mathcal{F}$ of subsets of S whose probabilities we are allowed to measure.

Analytic probability

This collection of subsets $\mathcal{F}$ must have some specific properties:

Def. 3: The notation $E \in \mathcal{F}$ means that the object E is in the collection $\mathcal{F}$

Analytic probability

Def: 4 a collection $\mathcal{F}$ of subsets of S is a *σ-field* if

(i) $S, \phi \in \mathcal{F}$ (S and ϕ are in collection $\mathcal{F}$)

(ii) $A \Rightarrow A^c \in \mathcal{F}$ ($A^c =$ *complement of* A)

(iii) $A_i \in \mathcal{F} \ \forall i \Rightarrow \bigcup_i A_i \in \mathcal{F}$

[$\forall =$ *for all*]

where $A_1, A_2, A_3, \ldots$ form a sequence of sets.

Analytic probability

Once we have an S, $\mathcal{F}$, and a P as above, we call the triple

$$(S, \mathcal{F}, P) = \quad \text{probability space}$$

We denote the class $\mathcal{F}$ to be the *measureable sets* in S.

Analytic probability

Example 11: $S = \{1, 2, 3, 4, 5, 6\}$. In this case the collection $\mathcal{F}$ of measureable sets is usually *all* subsets of S, i.e., $\mathcal{F} = \{$all subsets of $S\}$

Example 12: $S = [0, 1]$, i.e., we are choosing a random number x in the interval $[0, 1]$; S is the set of all possible outcomes.

Here

$$\mathcal{F} = \text{collection of measureable sets}$$

is defined to be the smallest σ-field of sets including all possible open subsets of $[0, 1]$.

Analytic probability

Sometimes this collection $\mathcal{F}$ is denoted as the *Borel sets.*

We can define probabilities of sets to be, e.g.

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

This defines the probability that our random number x is in the interval (a, b).

It can be shown that we can uniquely extend this definition of probability from the collection of intervals (a, b) to the larger collection of sets $\mathcal{F}$.

Analytic probability

This full measure on the interval $[0,1]$ is called *Lebesgue measure* on $[0,1]$.

We can do the same thing on the entire real line $\mathbb{R}$ to define Lebesgue measure on $\mathbb{R}$.

On $\mathbb{R}^p$ we can do the same thing by defining Borel sets $\mathcal{B}_p$ to be the smallest σ-field of sets in $\mathbb{R}^p$ containing all open subsets. Then if we define the volume of a box

$$A = \{(x_1, \ldots, x_p) : x_1 \in (a_1, b_1), \ldots, x_p \in (a_p, b_p)\}$$

in the obvious way (i.e. as the volume

$$V(A) = (b_1 - a_1) \cdot (b_2 - a_2) \dots (b_p - a_p))$$

then we can show there is a unique measure μ defined on the sets in $\mathcal{B}_p$ such that for any box A, the measure $\mu(A)$ is the same as its volume $V(A)$.

This measure μ is called *Lebesgue measure* on $\mathbb{R}^p$.

3. Some basic notions: measurable functions

If we have a sample space S we want to define functions on it.

Definition 5: We define $\mathbb{R}$ to be the collection of all real numbers.

Definition 6: A function f on any set E is rule which assigns to each element $x \in E$ a real number denoted as $f(x)$.

In this case we write

$$f : E \to \mathbb{R}$$

Measurable functions

Assume now we have a collection of measurable sets $\mathcal{F}$ on Ω

Definition 7: A function $f : S \to \mathbb{R}$ is *measurable* if for all $x \in \mathbb{R}$, the set

$\{a : f(a) \leq x\}$ = the set of all a such that $f(a) \leq x$

is measurable (i.e. is in $\mathcal{F}$).

Example 13: Consider the function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = x^3$

Measurable functions

Example 14: Consider $f(x) = I_Q(x) =$ indicator function of rationals, defined by

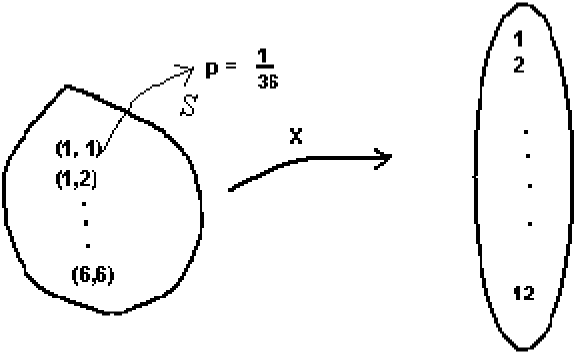
$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

4. Random Variables

Definition 8: A *random variable* X is any measurable function from the sample space S to the real numbers.

Example 15: throw 2 dice

Random variables



Suppose random variable X maps outcomes to numbers.

Random variables

Thus if the outcome is $(3,5)$ (i.e. first die is a 3 and second die is a 5) then

$$X(3,5) = 8,$$

i.e., X maps each outcome to the total.

Example 16: For a given amino acid a in a protein sequence, let $X(a)$ denote the number of other amino acids which a is bound to.

Again X maps the amino acid (in S) into a number.

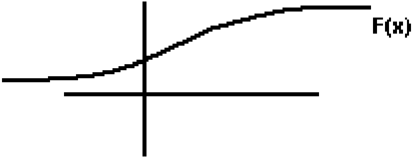
For a random variable (RV) X:

Random variables

Def. 9: If X is a random variable, we define its *distribution function* F to be

distribution function

$$F(x) : \quad P(\omega : \; X(\omega) \leq x) \; = \; P(X \leq x)$$



Random variables

Properties of F (easily derived)

(i) $F(x) \to 1$ as $x \to \infty$
$F(x) \to 0$ as $x \to -\infty$

(ii) F has at most countably many discontinuities
i.e., discontinuity points $x_1, x_2, \ldots$ can be listed

Example 17: Suppose we record high, low body temperatures for a patient on a given day; form a sample space of all possible pairs of measurements:

Random variables

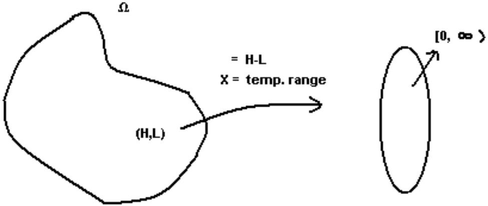
$$S = \{(H, L) : H \geq L\}$$

Define a random variable $X : S \to \mathbb{R}$ as follows:

for each element of S, let

$$X(H, L) = H - L = \text{ temperature range}$$

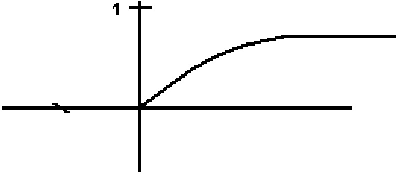
Random variables



Suppose we find that.

$$
\begin{aligned}
F(x) &= P(\,(H,L):\ (H-L) \leq x) \\
&= P(X \leq x) = \begin{cases} 1 - e^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}
\end{aligned}
$$

Random variables



can check this is a d.f.

If F has a derivative,

$$F'(x) = f(x) = \text{ \textit{density function}}$$

of X.

Example 18: here density $= f(x) = \begin{cases} e^{-x} \\ 0 \end{cases}$

Random variables

check: $$F(x) = \int_{-\infty}^{x} f(x')dx'.$$

Example 19: Normal: density is $\frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$

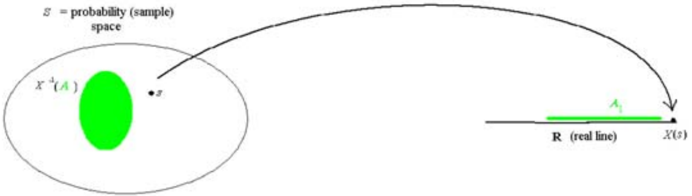
Random variables

5. Distribution measure: Given a random variable (RV) X, we define the *distribution measure* μ to be the probability measure on the basic Borel sets in $\mathbb{R}$, defined by

$$\mu(A) = P(X \in A) = P(s \in S : X(s) \in A),$$

where P is the basic probability measure on S.

Random variables



Note that if $A = (a, b) =$ interval then $\mu(A) = F(b) - F(a)$ where $F =$ distribution function of X.

6. Random vectors

Given a probability (sample) space S, if we have more than one random variable $X_1, \ldots, X_p$, we denote the function

Random variables

$$\mathbf{X}(s) = (X_1(s), \ldots, X_p(s))$$

i.e. an ordered set of functions as a *random vector*.

We define the (joint) distribution measure μ of the random vector $\mathbf{X}$ (defined on all Borel sets in $\mathbb{R}^p$) by the definition that for any (Borel) set $A \subset \mathbb{R}^p$,

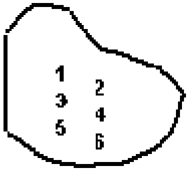
$$\mu(A) = P(\mathbf{X} \in A) = P(s \in S : \mathbf{X}(s) \in A),$$

where P is the basic probability measure on S.

More probability:

1. Conditional probability

Example 1: Die roll:



$$A = \text{ event of even roll } = \{2, 4, 6\}$$

$$B = \text{ event of roll divisible by 3 } = \{3, 6\}$$

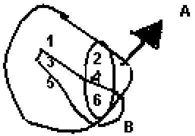
$$P(A) = \frac{1}{2}; \qquad P(B) = \frac{1}{3}$$

Define *conditional probability*

P(B|A) $=$ Prob. of B *given* A occurs

How to compute $P(B|A)$?

Create *reduced sample space* consisting only of A



and compute what portion of A is in B.

Here reduced sample space $A = \{2, 4, 6\}$.

Assuming equal probabilities:

$$P(2|A) = P(4|A) = P(6|A) = 1/3.$$

Note $P(B|A) = P(6|A) = 1/3.$

Now note $P(A \cap B) = \frac{1}{6}$; thus

$$P(B|A) = \frac{1}{3} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{P(A \cap B)}{P(A)} \, .$$

Illustrates general mathematical definition

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

2. Independence:

We say events B and A are *independent* if

$$P(B|A) = P(B) \quad \Leftrightarrow \quad \frac{P(B \cap A)}{P(A)} = P(B)$$

$$\Leftrightarrow \quad P(A \cap B) = P(A)\,P(B)$$

In general a collection of events $A_1, \ldots, A_n$ is independent if for any subcollection $A_{i_1}, \ldots, A_{i_n}$:

$$P(A_{i_1} \cap A_{i_2} \ldots \cap A_{i_n}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_n})$$

Example 2: Consider sequence $x_1, x_2, \ldots, x_k$ of bases forming gene $\boldsymbol{g}$.

Q: Are the successive bases x_i, x_{i+1} independent?

For example, given we know the base $x_i =$ A, does that change the probability $P(x_{i+1} =$ C).

More specifically, do we have

$$P(x_{i+1} = \text{C}|x_i = \text{A}) = P(x_{i+1} = \text{C})? \qquad (3)$$

If (3) also holds for all possible choices (C,G,T) replacing A and all choices replacing C, then x_{i+1} and x_i are independent.

Expect: in actual DNA exon region has more dependence from base to base than an intron does.

Why? Evolutionary structural pressures - exons are more important to survival

3. Independence of RV's

Henceforth given a subset A of sample space S, we assume A is measurable unless stated otherwise.

Definition 10: Let S be a sample (probability) space. Let $X : S \to \mathbb{R}$ be a random variable (i.e., a rule assigning a real number to each outcome in S)

If $A = [a, b] \subset \mathbb{R}$, define

$$P(X \in A) = P(a \leq X \leq b)$$

Independence of RV's

$= P($all outcomes s whose value $X(s)$ is between a and $b)$

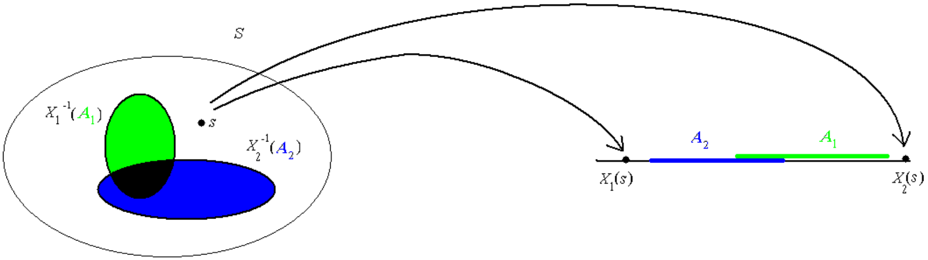
$$= P(s|\, a \le X(s) \le b).$$

Recall:

Definition 11: Let S be a sample space, and X_1, X_2 be random variables. Then $X_1,\ X_2$ are *independent* if for any sets $A_1, A_2 \subset \mathbb{R}$,

$$P(X_1 \in A_1,\ X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2)$$

Independence of RV's



Generally $X_1, \ldots, X_n$ are independent if

$$P(X_1 \in A_1, \ldots, X_n \in A_n)$$

$$= P(X_1 \in A_1)P(X_2 \in A_2)\ldots P(X_n \in A_n)$$

Independence of RV's

More to come on these topics-