Probability and Random Processes

- read Chapter 1

why study probability/stochastic processes?
- interesting
- applications in statistics, mathematical finance, machine learning, physics

I Introduction

- response ω (what we "might" observe)

- sample space Ω = collection of all possible response values.

- 2^{Ω} = power set of Ω

  = collection of all possible subsets of Ω.

eg coin tossing

- $\omega = \begin{cases} 1 & \text{when a head is observed} \\ 0 & \text{when a tail is observed} \end{cases}$

$\Omega = \{0, 1\}$

$2^{\Omega} = \{\phi, \{0\}, \{1\}, \{0,1\}\}$

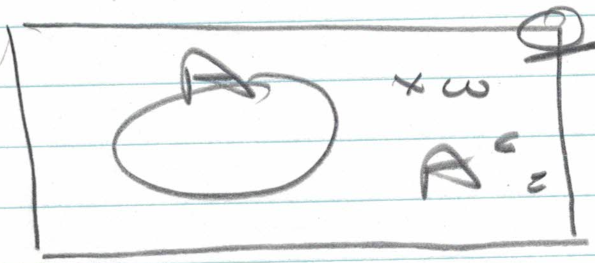
- a prob. measure (for now) any map $P : 2^{\Omega} \to [0,1]$ that satisfies $P(\Omega) = 1$ and $P(A \cup B) = P(A) + P(B)$ when $A, B \in 2^{\Omega}$ and $A \cap B = \phi$ (additive)

- what is the meaning of $P(A)$?

- $P(A)$ relates to the value of an unobserved ω, i.e. we don't know the value of ω

- then $P(A)$ measures the <u>degree of belief</u> that "$\omega \in A$" is true.

Venn Diagram



A^c = complement of A.

- so $P(A) = 1$ means we are certain $\omega \in A$
  - $= 0$ " " " " $\omega \notin A$
  - $= 1/2$ equally certain that $\omega \in A$ as $\omega \in A^c$
  - $= 3/4$ more certain that $\omega \in A$ than $\omega \in A^c$

- lots of ways P could be determined (assigned)

- in this course we suppose we have a probability measure P and consider the mathematics of this

- statistics deals with issue concerning the assignment of relevant P

eg coin tossing

$P(\{0,1\}) = 1$, $P(\phi) = P(\phi \cup \phi)$
$= P(\phi) + P(\phi)$
$\therefore P(\phi) = 0$

1 ass $P(\{0\}) = P(\{1\}) = \frac{1}{2}$ symmetric coin

another $P(\{0\}) = \frac{3}{4}$, $P(\{1\}) = \frac{1}{4}$

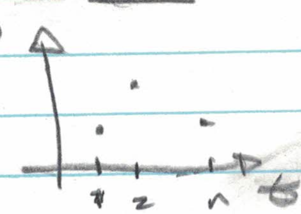
etc. $P(\{0\}) = 1-p$, $P(\{1\}) = p$ for some $p \in [0,1]$ and call this the Bernoulli(p) prob. measure (or probability distribution)

- what is a stochastic process?

- we add a "time" component via a set T (time domain) and consider a new sample space

$\Omega = \bigtimes_{t \in T} S$ = Cartesian product of a set S over the index $t \in T$ (S called the state space and T the time domain)
$\| \; S^T$

eg - $T = \{1, 2, \ldots, n\}$, S (finite time)



$\Omega = \bigtimes_{t \in \{1,\ldots,n\}} S = \{(\omega_1, \ldots, \omega_n) : \omega_i \in S\}$

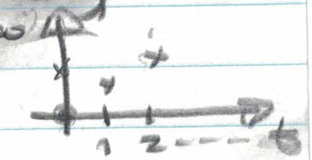
= set of all n-tuples from S.

- think of tossing a coin and recording the n outcomes in sequence to obtain $(\omega_1, \ldots, \omega_n) \in \bigtimes_{t \in \{1,\ldots,n\}} \{0,1\}$
$\{0,1\}^n$

eg $T = \{0, 1, 2, \dots\} = \mathbb{N}_0$ <u>discrete time</u>

$\underline{\Omega} = \times_{t \in T} S = \{(\omega_0, \omega_1, \dots) : \omega_i \in S\}$

- tossing a coin infinitely many times



- of course we can't do that so instead think of tossing the coin an arbitrarily large (unknown) number of times,

$\times_{t \in \{1, 2, \dots\}} \{0,1\} = \{0,1\}^{\mathbb{N}}$

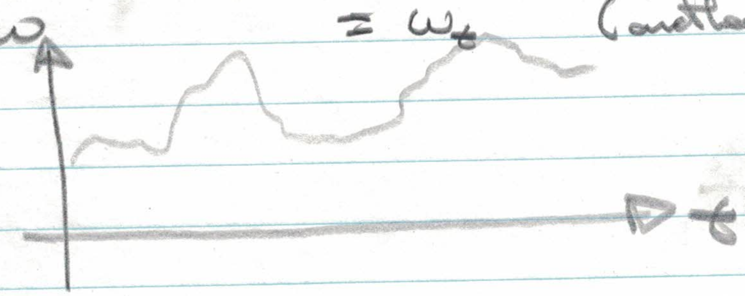
- sometimes take $T = \mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ so there is an infinite past and an infinite future.

eg $T = [0, \infty)$ - <u>continuous time</u>.

$\underline{\Omega} = \times_{t \in [0,\infty)} S = S^T = \{\omega : \omega : T \to S\}$ = set of all fns from T to S

where $\omega(t)$ = state at time t $\boxed{\text{sample function}}$

$= \omega_t$ (another notation)



- where does "stochastic" come in?

- we are uncertain about what sample function

ω that we will observe, and so we place a prob. measure on S^T

eg <u>coin tossing</u> (Bernoulli process)

- suppose a "fair" coin is tossed infinitely many times
- $T = \mathbb{N}$
- sample space $\{0,1\}^T = \{0,1\}^{\mathbb{N}}$ = set of all infinite sequences of 0's and 1's where $\omega(i) = 1$ when i-th toss is a heads and $\omega(i) = 0$ otherwise.
- what P on $\{0,1\}^{\mathbb{N}}$?
- maybe from independence
- say $A = \{(0,0,0,\omega_4,\omega_5,\ldots) : \omega_i \in \{0,1\}\ i>3\}$

  $=$ we get 3 tails on first 3 tosses.
- should have $P(A) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$
- so P is built via independence but note $P(\{(\omega_1,\omega_2,\ldots)\}) = \prod_{i=1}^{\infty} \frac{1}{2} = 0$ so every sample function has prob 0.

- think of a gambler who plays a game where the prob of winning is p on each play and 1 unit is bet to potentially win a unit.

- starts with a units.

- $T = \mathbb{N}_0$ where $\Omega = \{a\} \times \{0,1\}^{\mathbb{N}}$
  so $\omega = (a, \omega_1, \omega_2, \ldots)$

- define $X : \Omega \times T \to \mathbb{Z}$ by

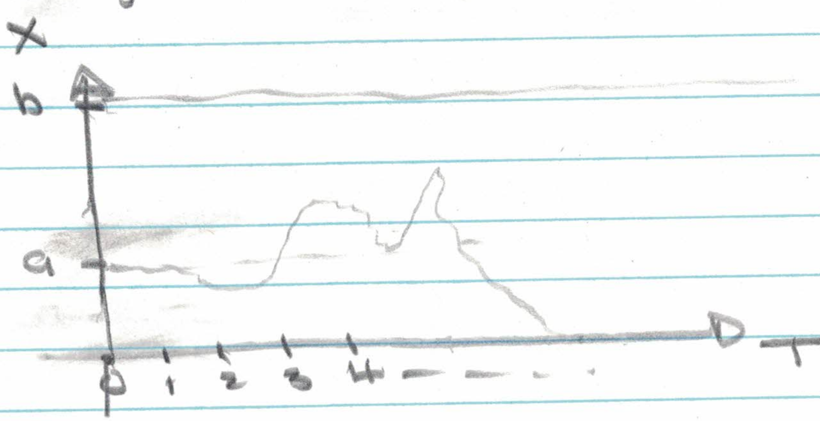
$$X(\omega, t) = a + \sum_{i=1}^{t} (2\omega_i - 1)$$

$$= \text{fortune of gambler at time } t$$

- suppose casino has b units.

- want to compute the prob. of ruin

$$P(X(\omega, t) = 0 \text{ for some } t \text{ before } X(\omega, t) = b)$$



gambler's ruin

- **note** - can think of stoch. processes from different points of view

① prob. measure on a space of functions (mathematical)

② a system evolving randomly in time (intuitive)

③ a collection of random variables indexed by a time variable, put $X_t = \omega(t)$ then replace ω by $\{X_t : t \in T\}$

(2) The Probability Model

- we define this as $(\Omega, 2^{\Omega}, P)$

- actually if Ω is a "big" set we can't define P on 2^{Ω} in the sense that it can be proved that no P exists with the right properties

- so we have to restrict the domain of P to $\mathcal{F} \subseteq 2^{\Omega}$

- so $P : \mathcal{F} \to [0,1]$

- what properties should $\mathcal{F}$ possess?

- actually we require that $\mathcal{F}$ be closed under the Boolean operations $\cup, \cap, {}^c$ and also require $\emptyset \in \mathcal{F}$ and then $\mathcal{F}$ is called a field (or algebra)

- actually we require the closure of $\mathcal{F}$ under countable infinite Boolean operations and then $\mathcal{F}$ is called a σ-field (or σ-algebra)

Def A class $\mathcal{F}$ of subsets of Ω is called a σ-field
(i) $\emptyset \in \mathcal{F}$ (ii) $A_1, A_2, \ldots \in \mathcal{F}$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
(iii) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$

Prop ① If $\mathcal{F}$ is a σ-field then
(i) $\Omega \in \mathcal{F}$
(ii) $A_1, \dots, A_n \in \mathcal{F}$ implies $\bigcup_{i=1}^{n} A_i, \bigcap_{i=1}^{n} A_i \in \mathcal{F}$
(iii) $A_1, A_2, \dots \in \mathcal{F}$ implies $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

Proof: Exercise

note – $\{\phi, \Omega\}$ is a σ-field (the smallest)

– 2^{Ω} is a σ-field (the largest)

Prop ② If $\{\mathcal{F}_i \mid i \in I\}$ is a class of σ-fields on a set Ω then $\bigcap_{i \in I} \mathcal{F}_i$ is a σ-field.

Proof: Exercise

note – if $\mathcal{A}$ is a class of subsets of Ω then

$$\mathcal{F}(\mathcal{A}) = \bigcap_{\{\mathcal{F} \mid \mathcal{A} \subseteq \mathcal{F},\ \mathcal{F} \text{ a } \sigma\text{-field on } \Omega\}} \mathcal{F}$$

is a σ-field called the σ-field generated by $\mathcal{A}$ and note that this is the smallest σ-field containing $\mathcal{A}$.

eg Borel sets $\mathcal{B}^k$

– $\Omega = \mathbb{R}^k$, $\mathcal{A} = \{ \bigtimes_{i=1}^{k} (a_i, b_i] \mid a_i \in \mathbb{R} \cup \{-\infty\}, b_i \in \mathbb{R}\}$

$\mathcal{B}^k = \mathcal{F}(\mathcal{A})$ is the Borel sets on $\mathbb{R}^k$

- suppose $\omega_1, \ldots, \omega_n \in \Omega$, define for $A \in \mathcal{F}$
$r_n(A) = \#\{\omega_i : \omega_i \in A\}/n$ (relative freq. function)

- $\forall n \in \mathbb{N}$, $r_n : \mathcal{F} \to [0,1]$ sats
(i) $r_n(\Omega) = 1$
(ii) if $A_1, A_2, \ldots \in \mathcal{F}$ and $A_i \cap A_j = \phi$, $i \neq j$, then
$r_n(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} r_n(A_i)$

and all other properties of r_n can be deduced from these.

Def A function $P : \mathcal{F} \to [0,1]$ is called a probability measure on the σ-field $\mathcal{F}$ if
(i) $P(\Omega) = 1$
(ii) $A_1, A_2, \ldots \in \mathcal{F}$ mut. disjoint implies $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

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

Prop 3 For probability model $(\Omega, \mathcal{F}, P)$
(i) $P(\phi) = 0$
(iii) if $A_1, \ldots, A_n$ are mut. disj. then
$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
(iv) $P(A^c) = 1 - P(A)$
(v) if $A \subseteq B$ then $P(A) \leq P(B)$
(vi) (Inclusion-exclusion)

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \cdots$$
$$+ (-1)^{n+1} P(A_1A_2\cdots A_n)$$

$$P(\bigcap_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cup A_j) + \cdots + (-1)^{n+1} P(A_1 \cup \cdots \cup A_n)$$

(cont'd.)

- suppose $A_1, A_2, \ldots$ are subsets of Ω

<u>Def</u> The <u>lim sup</u> of $\{A_i\}$ is

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$$

and the <u>lim inf</u> of $\{A_i\}$ is

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$$

- <u>note</u> $\overline{\lim_{n\to\infty}} A_n = \{\omega \mid \omega \in A_i \text{ for infinitely many } i\}$

$\underline{\lim_{n\to\infty}} A_n = \{\omega \mid \omega \in A_i \text{ for all but finitely many } i\}$

$$\underline{\lim_{n\to\infty}} A_n \subseteq \overline{\lim_{n\to\infty}} A_n$$

<u>Def</u> The sequence of sets $\{A_i\}$ <u>converges</u> if $\overline{\lim_{n\to\infty}} A_n = \underline{\lim_{n\to\infty}} A_n$ and the limit is this set.

<u>eg</u> - if $A_1 \subseteq A_2 \subseteq \ldots$ then

Ex

$$\overline{\lim_{n\to\infty}} A_n = \bigcup_{i=1}^{\infty} A_i = \underline{\lim_{n\to\infty}} A_n$$

Ex1
2016

- if $A_1 \supseteq A_2 \supseteq \ldots$ then

Ex

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{i=1}^{\infty} A_i = \underline{\lim_{n\to\infty}} A_n$$

- in general $\bigcup_{i=n}^{\infty} A_i \downarrow \overline{\lim_{n\to\infty}} A_n$, $\bigcap_{i=n}^{\infty} A_i \uparrow \underline{\lim_{n\to\infty}} A_n$

- **note** - if $A_1, A_2, \ldots \in \mathcal{F}$ then $\underline{\lim_{n\to\infty}} A_n \in \mathcal{F}$ and $\overline{\lim_{n\to\infty}} A_n \in \mathcal{F}$

Prop ④ (Continuity of P)

If $\mathcal{F}$ is a σ-field and $P: \mathcal{F} \to [0,1]$ sats (i) $P(\Omega)=1$ (ii) $A_1, \ldots, A_n \in \mathcal{F}$ mut disj. then $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ then P is a prob. measure iff $P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$ for every convergent seq. $\{A_n\}$.

Proof: $\Leftarrow$] $\mathcal{S}$ $A_1, A_2, \ldots \in \mathcal{F}$ mut. disj. Then

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

$$= \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$

$\Rightarrow$] First assume $A_n \uparrow A$ and put

$B_1 = A_1$, $B_i = A_i \setminus A_{i-1} = A_i A_{i-1}^c$. Then

$$P(A_n) = P\left(\sum_{i=1}^{n} B_n\right) = \sum_{i=1}^{n} P(B_i) \text{ and}$$

$$\lim_{n\to\infty} P(A_n) = \sum_{i=1}^{\infty} P(B_i) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = P(\lim_{n\to\infty} A_n).$$

A similar proof works when $A_n \downarrow A$.

Now in general observe $\bigcap_{i=n}^{\infty} A_i \subseteq A_n \subseteq \bigcup_{i=n}^{\infty} A_i$

and thus $P(\underline{\lim} A_n) = \lim P(\bigcap_{i=n}^{\infty} A_i) \leq \lim_{n\to\infty} P(A_n)$

$\leq \lim P(\bigcup^{\infty} A_i)$ $\lim P(\overline{\lim}$

3 Conditional Probability and Independence

- if $B \in \mathcal{F}$ sat. $P(B) > 0$ then define $P(\cdot \mid B) : \mathcal{F} \to [0,1]$ by $P(A \mid B) = P(AB)/P(B)$

- $P(\cdot \mid B)$ is called the <u>conditional probability measure given B</u>

Prop 5. $P(\cdot \mid B)$ is a prob. measure on $\mathcal{F}$.

Proof: Ex.

Prop 6. (<u>Theorem of Total Probability</u>)
If $B_1, B_2, \ldots \in \mathcal{F}$ are mut. disj. st. $P(B_i) > 0$ and $\bigcup_{i=1}^{\infty} B_i = \Omega$ then $P(A) = \sum_{i=1}^{\infty} P(A \mid B_i) P(B_i)$.

Proof: Ex.

- if $P(A \mid B) = P(A)$ then the information that B has occured does not affect the prob. assignment for A

Def $A, B \in \mathcal{F}$ are <u>statistically independent</u> if $P(AB) = P(A)P(B)$.

Prop 7 If $P(B) > 0$ then A and B are stat. ind. iff $P(A \mid B) = P(A)$

Proof: Ex.

$\mathcal{F}(\{A\})$ $\mathcal{F}(\{B\})$

Prop ⑧ If $A, B \in \mathcal{F}$ are stat. ind. then every element in $\{\phi, A, A^c, \Omega\}$ is stat. ind. of every element of $\{\phi, B, B^c, \Omega\}$

Proof: Ex.

- **note** $\{\phi, A, A^c, \Omega\} = \mathcal{F}(\{A\})$

Def If $\{\mathcal{F}_i \mid i \in I\}$ is a collection of σ-fields on Ω then the $\mathcal{F}_i$ are mutually stat. ind. whenever

$$P\left(\bigwedge_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$$

$\forall\, A_j \in \mathcal{F}_j$ and every finite $J \subseteq I$.

(don't do)

- we can use independence to construct probability models by taking products of probability models

- suppose we have $\{(\Omega_i, \mathcal{F}_i, P_i) \mid i \in I\}$ a collection of probability models

- define $\Omega = \underset{i \in I}{\times} \Omega_i$

$$= \{ f \mid f : I \to \bigcup_{i \in I} \Omega_i, \text{ s.t. } f(i) \in \Omega_i \}$$

- define $\mathcal{F} = \underset{i \in I}{\times} \mathcal{F}_i = \mathcal{F}(\mathcal{A})$ where

$\mathcal{A} = \{ \underset{i \in I}{\times} A_i \mid A_i \in \mathcal{F}_i$ and $A_i = \Omega_i$ except for finitely many $i \}$

- define P on $\mathcal{A}$ by $P(\underset{i \in I}{\times} A_i) = \prod_{i \in I} P(A_i)$
- <u>fact</u> - the Extension Theorem says there is a unique prob. measure on $(\mathcal{F}(\mathcal{A}))$ that agrees with P on $\mathcal{A}$.
    - this P is called the <u>product measure</u> and is denoted $\prod_{i \in I} P_i$
- $(\underset{i \in I}{\times} \Omega_i, \underset{i \in I}{\times} \mathcal{F}_i, \prod_{i \in I} P_i)$ is the product prob. model

<u>eg</u> <u>toss a fair coin infinitely many independent times</u>

- $I = \mathbb{N}$, $\Omega_i = \{0,1\}$, $\mathcal{F}_i = \{\phi, \{0\}, \{1\}, \{0,1\}\}$

  $P_i(\{0\}) = \frac{1}{2}$ $\quad \overset{\infty}{\underset{i=1}{\times}} \Omega_i = \{0,1\}^{\mathbb{N}}$

- then whenever n of the A_i equal $\{0\}$ or $\{1\}$ and the rest are $\{0,1\}$ we have $P(\overset{\infty}{\underset{i=1}{\times}} A_i) = (\frac{1}{2})^n$

- if infinitely many of the A_i equal $\{0\}$ or $\{1\}$ and the rest are $\{0,1\}$ then

$$P(\overset{\infty}{\underset{i=1}{\times}} A_i) = \lim_{n \to \infty} P(\overset{n}{\underset{i=1}{\times}} A_i \times \overset{\infty}{\underset{i=n+1}{\times}} \{0,1\}) = 0$$

- thus each element of Ω has 0 prob. mass

$- \; P(\text{"5 heads in a row"})$

$= 1 - P(\text{"5 heads in a row never observed"})$

$= 1 - P\left(\bigcap_{m=1}^{\infty} A_m\right)$ where

$$A_m = \{(\omega_{m+1}, \omega_{m+2}, \omega_{m+3}, \omega_{m+4}, \omega_{m+5}) \neq (1,1,1,1,1)\}$$

$$P(A_m) = 1 - \left(\tfrac{1}{2}\right)^5$$

$= 1$ since $\bigcap_{m=0}^{\infty} A_m \subseteq \bigcap_{m=0}^{\infty} A_{5m}$

and $\{A_5, A_{10}, A_{15}, \ldots\}$ are mut. stat. ind. and

$$P\left(\bigcap_{m=1}^{\infty} A_{5m}\right) = \lim_{n\to\infty} P\left(\bigcap_{m=1}^{n} A_{5m}\right)$$

$$= \lim_{n\to\infty} \prod_{m=1}^{n} P(A_{5m})$$

$$= \lim_{n\to\infty} \left(1 - \left(\tfrac{1}{2}\right)^5\right)^n = 0$$