Stochastic Processes

Andrea Tancredi

Sapienza University of Rome

Week 2

Distributive property of the intersection: Set intersection is distributive over set union

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

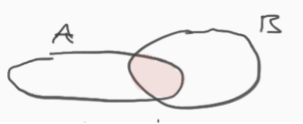
More generally

$$A \cap \bigcup_i B_i = \bigcup_i (A \cap B_i)$$

This property holds for whatever sets A and $B_1,\ B_2, \dots$

Conditional Probability

If we know that B occurs, A occurs if and only if $A \cap B$ occurs



Given B the experiment outcomes are those of B, and those of A can be only $A \cap B$

Definition If $A, B \in \mathcal{F}$ and $P(B) > 0$ the **conditional probability** of B given A is denoted with $P(A|B)$ and defined by

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

Theorem: the conditional probability is a probability measure
If $B \in \mathcal{F}$ and $P(B) > 0$, then $(\Omega, \mathcal{F}, Q)$ is a probability space, where $Q : \mathcal{F} \to R$ is defined by

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

Three things to prove...proof in class

- $Q(\Omega) = 1$
- $Q(A) \geq 0$
- $Q\left(\bigcup A_i\right) = \sum_i Q(A_i)$

The proof

- $Q(\Omega) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1$
- $Q(A) = P(A \cap B)/P(B) \geq 0$
- Let $A_1, A_2 ..$ disjoint events

$$
\begin{aligned}
Q\left(\bigcup_i A_i\right) &= \frac{1}{P(B)} P\left(\left(\bigcup_i A_i\right) \cap B\right) = \frac{1}{P(B)} P\left(\bigcup_i (A_i \cap B)\right) \\
&= \frac{1}{P(B)} \sum_i P(A_i \cap B) = \sum_i P(A_i|B) = \sum_i Q(A_i)
\end{aligned}
$$

Since Q is a probability measure, if $A \in \mathcal{F}$ and $P(B) > 0$

$$Q(A^c) = P(A^c|B) = 1 - Q(A) = 1 - P(A|B)$$

Without exploiting the previous theorem we could observe that, since $B = (B \cap A) \cup (B \cap A^c)$

$$P(A^C|B) = \frac{P(A^c \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)}$$

Suppose that $P(A) > 0$ and $P(B) > 0$. Then

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

and we have that

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

and (ex 1.35)

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

Ex. 1.34 Show that, if $P(B \cap C) > 0$

$$P(A \cap B \cap C) = P(A|B, C)P(B|C)P(C)$$

Note that since $P(B \cap C) > 0$ we can write

$$P(A \cap B \cap C) = P(A \cap (B \cap C)) = P(A|B, C)P(B \cap C)$$

Moreover $(B \cap C) \subseteq C$. Then $P(B \cap C) > 0 \Rightarrow P(C) > 0$ and we can write that

$$P(B \cap C) = P(B|C)P(C)$$

Then

$$P(A \cap B \cap C) = P(A|B, C)P(B|C)P(C)$$

Independent events

Two event are independent if the occurrence of one of them does not affect the probability that the other occurs. That is if $P(A) > 0$ and $P(B) > 0$ then

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B) \qquad *$$

Independence can be defined directly in the following way

Definition ** Events A and B of a probability space are called **independent** if

$$P(A \cap B) = P(A)P(B)$$

and **dependent** otherwise

- Since $P(A|B) = P(A \cap B)/P(B)$, condition * $\iff$ definition **
- Definition ** is slightly more general since it allows A and B to have non zero probabilities

Example: Toss n times a coin, equiprobable events.
$|\Omega| = 2^n$, $A =$ "the i-th coin is head", $B =$ "the j-th coin is head", $i \neq j$
$|A| = |B| = 2^{n-1}$ $|A \cap B| = 2^{n-2}$. Then

$$P(A) = \sum_{\omega \in A} \frac{1}{2^n} = \frac{1}{2^n} 2^{n-1} = \frac{1}{2} \quad P(B) = \sum_{\omega \in B} \frac{1}{2^n} = \frac{1}{2^n} 2^{n-1} = \frac{1}{2}$$

$$P(A \cap B) = \sum_{\omega \in B} \frac{1}{2^n} = \frac{1}{2^n} 2^{n-2} = \frac{1}{4}$$

Then A and B are independent since

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(A)P(B)$$

If $C =$"k heads", $|C| = \binom{n}{k}$, $|A \cap C| = \binom{n-1}{k-1}$ and

$$P(A \cap C) = \frac{1}{2^n}\binom{n-1}{k-1} \neq \frac{1}{2}\frac{1}{2^n}\binom{n}{k} = P(A)P(C)$$

Hence A and C are dependent

Note also that

$$\begin{aligned} P(C|A) &= \frac{P(C \cap A)}{P(A)} = \frac{\binom{n-1}{k-1}/2^n}{1/2} = \binom{n-1}{k-1}\frac{1}{2^{n-1}} \\ &= \binom{n-1}{k-1}\left(\frac{1}{2}\right)^{k-1}\left(\frac{1}{2}\right)^{n-1-(k-1)} \end{aligned}$$

Exercise 1.43 If A and B are independent and disjoint what can be said about $P(A)$ and $P(B)$?
We have

$$0 = P(A \cap B) = P(A)P(B)$$

Then at least one has 0 probability

Exercise 1.44 Prove that A and B are independent iff A and B^c are independent
Suppose A and B independent. Note that $(A \cap B) \cup (A \cap B^c) = A$

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) = P(A)P(B^c) \end{aligned}$$

Then A and B^c are independent. Moreover if A and B^c are independent, also A and $(B^c)^c = B$ are independent

- A family $\mathcal{A} = \{A_i \; i \in I\}$ of events is called independent if for all finite subset J of I

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{j \in J} P(A_i) \quad *$$

- $\mathcal{A}$ is said pairwise independent if * holds whenever $|J| = 2$
- $A_1, A_2, \ldots, A_m$ are independent iff $A_1^c, A_2^c, \ldots, A_m^c$ are independent

Exercise 1.46 (a) If $A_1, A_2, \ldots, A_m$ are independent and $P(A_1) = p$ for $i = 1, \ldots, m$ find the probability that none of A_i occur

$P(\text{none of the } A_i \text{ occur}) \quad = P(A_1^c \cap A_2^c \cap \cdots A_m^c) = \prod_{i=1}^m P(A_i^c) = (1-p)^m$

Note that

$P(\text{at least one of the } A_i \text{ occurs }) = 1 - P(\text{none of the } A_i \text{ occur}) = 1 - (1-p)^m$

Definition A partition of Ω is a collection $\{B_i : i \in I\}$ of disjoint events such that

$$\Omega = \bigcup_{i \in I} B_i$$

If $\{B_i : i \in I\}$ is a partition of Ω

$$A = A \cap \Omega = A \cap \bigcup_{i \in I} B_i = \bigcup_{i \in I} (A \cap B_i)$$

Partition Theorem If $\{B_1, B_2, \ldots\}$ is a partition of Ω and $P(B_i) > 0 \, \forall i$

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

Proof in class

$$P(A) \quad = \quad P(A \cap \Omega)...$$

- Example 1.51 (Modified).

  A="Tomorrow I'll be late,B="Tommorow it will rain".

  Suppose that $P(B) = 3/5$, $P(A|B) = 3/5$ and $P(A|B^c) = 1/5$

  Then

  $$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c) = \frac{3}{5}\frac{3}{5} + \frac{1}{5}\frac{2}{5} = \frac{11}{25}$$

Suppose that the occurring of A represents some evidence and $B_1, B_2, \ldots$ possible states of nature. If we know the conditional probabilities $P(A|B_i)$ we can obtain easily the conditional probabilities for the states B_i given A

Bayes Theorem If $\{B_1, B_2, \ldots\}$ is a partition of Ω and $P(B_i) > 0 \, \forall i$, then $\forall \, A \in \mathcal{F}$ with $P(A) > 0$

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}$$

- Example 1.51.

  A disease has incidence $1/10^5$, that is $P(D) = 1/10^5$. If you have the disease and do a diagnostic test, the test is positive P with probability $9/10, P(P|D) = 9/10$, if you do not have the disease $P(P|D^c) = 1/20$ (false positive).

  You are positive...

$$P(D|P) = \frac{P(P|D)P(D)}{P(P|D)P(D) + P(P|D^c)P(D^c)} = \frac{\frac{9}{10}\frac{1}{10^5}}{\frac{9}{10}\frac{1}{10^5} + \frac{1}{20}\frac{10^5-1}{10^5}} \approx 0.0002$$

- Exercise 1.52 At home

- A sequence A_1, A_2...of events of $\mathcal{F}$ is called increasing if $A_n \subseteq A_{n+1}$ for $n = 1, 2, ...,$
- The union

$$A = \bigcup_{i=1}^{\infty} A_i$$

  of an increasing sequence of events is called the limit of the sequence
- $A \in \mathcal{F}$ (in fact is a countable union)
- Theorem: Continuity of probability measures. Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $A_1, A_2...$ is an increasing sequence of events in $\mathcal{F}$ with limit A

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

Example 1.55 Consider an infinite set of tosses of a coin.

- Here a single ω is a infinite sequence like

$$\omega = THTHHTHHHHTHT......$$

  and is it possible to prove that Ω is uncountable (since Ω is one to one with $[0,1]$...)

- Let A_n be the event that the first n tosses of the coin yield at least one head. We have $A_n \subseteq A_{n+1}$ for $n = 1, 2, \ldots$.

- For a finite set like A_n we can safely take $P(A_n) = 1 - (1/2)^n$

- Note that $A = \cup_{i=1}^{\infty} A_i$ is the event that a head will occur sooner or later.

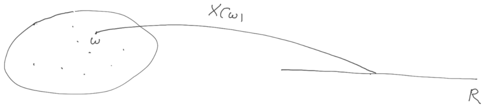
$$P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} 1 - (1/2)^n = 1$$

  and $P(A^c) = 0$ where A^c is the event "no head ever appears"

Discrete random variables

- Given a probability spaces we are interested on the values X of a real valued function acting on Ω
- A discrete random variable X on $(\Omega, \mathcal{F}, P)$ is a mapping

$$X : \Omega \rightarrow R$$

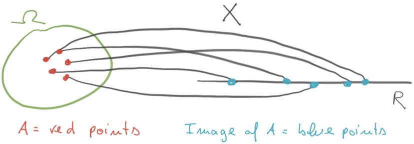


such that

(a) the image $X(\Omega)$ is a countable subset of R
(b) the set $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F} \; \forall x \in R$

If $X : \Omega \to R$, the image of A is the set

$$X(A) = \{X(\omega) : \omega \in A\}$$



- We may call X discrete since $X(\Omega)$ takes only countably many values (condition (a))
- The condition (b) says that the results of X is random since $X = x$ is a event, that is

$$X = x \text{ occurs} \iff X^{-1}(x) = \{\omega \in \Omega : X(\omega) = x\} \text{ occurs}$$

and the subsets $X^{-1}(x)$ belong to $\mathcal{F}$ $\forall x$ and are therefore assigned a probability

- The event $\{\omega \in \Omega : X(\omega) = x\}$ will be denoted with $(X = x)$
- $\operatorname{Im} X$ will denote the image of Ω under X

Definition The probability mass function (p.m.f.) of a discrete r.v. is the function $P_X : R \to [0, 1]$ defined by

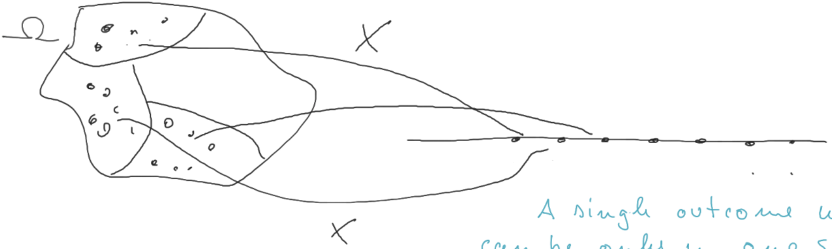
$$p_X(x) = P(X = x)$$

Properties of the p.m.f.

- $p_X(x) = 0$ if $x \notin \operatorname{Im} X$
- $$\sum_{x \in R} p_X(x) = \sum_{x \in Im\, X} p_X(x) = 1$$

  In fact

$$\begin{aligned} 1 &= P(\Omega) = P\left(\bigcup_{x \in Im\, X} \{\omega \in \Omega : X(\omega) = x\}\right) \\ &= \sum_{x \in Im\, X} p_X(x) = \sum_{x \in R} p_X(x) \end{aligned}$$



$$\Omega = \bigcup_{x \in \mathrm{Im}(X)} \{\omega : X(\omega) = x\}$$

A single outcome ω can be only in one set $\{\omega : X(\omega) = x\}$ otherwise this point ω would be such that $X(\omega) = x$ and $X(\omega) = x'$

Theorem Let $S = \{s_i : i \in I\}$ be a countable set of distinct real numbers and let $\{\pi_i : i \in I\}$ be a collection of real numbers such that

$$\pi_i \geq 0 \;\; \forall\, i \in I \quad \text{and} \quad \sum_{i \in I} \pi_i = 1$$

There exists a probability space and a discrete random variable X such that $p_X(s_i) = \pi_i$ for $i \in I$ and $p_X(s) = 0$ for $s \notin S$

Proof. Take $\Omega = S$, $\mathcal{F}$ the power set of Ω and

$$P(A) = \sum_{i : s_i \in A} \pi_i$$

P is a probability measure, see Exercise 1.17 (and example 1.16 that we did in class)

Take X as the identity function i.e. $X : \Omega \to R$ with $X(\omega) = \omega$

Examples of discrete random variables

- **Bernoulli**

  We say that X has the Bernoulli distribution (X is a Bernoulli r.v.) with parameter p if the image of X is $\{0,1\}$ and

  $$P(X=0)=q=1-p \quad P(X=1)=p$$

  The p.m.f. of X is then

  $$p_X(0)=q, \quad p_X(1)=p, \quad p_X(x)=0 \text{ if } x\neq 0,1$$

  Note that

  $$p+q=p+(1-p)=1$$

- **Binomial**

  We say that X has the Binomial distribution (X is a Binomial r.v.) with parameters n and p if the image of X is $\{0, 1, \dots, n\}$ and

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \quad \text{for } k = 0, 1, \dots, n \quad \mathrm{q} = 1 - p$$

  Remember (or learn..) the Binomial theorem that is for $x \in \mathbb{R}$

$$(1 + x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$$

  then

$$(a + b)^n = b^n(1 + a/b)^n = b^n \sum_{k=0}^{n} \binom{n}{k} (a/b)^k = \sum_{k=0} \binom{n}{k} a^k b^{n-k}$$

  Hence

$$1 = (p + q) = (p + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

Let X be a Binomial(n, p).

$$P(X = \text{even}) \quad = \quad \frac{1}{2}(1 + (q-p)^n)$$

In fact

$$\begin{aligned} 1 = (p+q)^n &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \\ (q-p)^n &= \sum_{k=0}^{n} \binom{n}{k} q^k (-1)^{n-k} p^{n-k} = \sum_{k=0}^{n} \binom{n}{n-k} q^k (-1)^{n-k} p^{n-k} \end{aligned}$$

Summing the two equations on the left we have $1 + (q-p)^n$ on the right we have two times the sum of the binomial probabilities for k even (since the odd terms cancel out)

- **Poisson**

  We say that X has the Poisson distribution (X is a Poisson r.v.) with parameter $\lambda > 0$ if X takes value $\{0, 1, \dots\}$ and

$$P(X = k) = \frac{1}{k!}\lambda^k e^{-\lambda} \quad \text{for } k = 0, 1, \dots$$

  Remember (or learn..) that for $x \in \mathbb{R}$

$$e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = \sum_{k=0}^{\infty}\frac{x^k}{k!}$$

  Hence

$$\sum_{k=0}^{\infty}\frac{1}{k!}\lambda^k e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{1}{k!}\lambda^k = e^{-\lambda}e^{\lambda} = 1$$