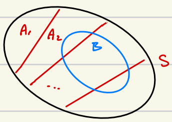
Week 3

Recall: Bayes Formula ; $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$ for $P(B) > 0$, $A, B \in \mathcal{B}$.

> **Prop.** Let $A_1, A_2, \ldots$ be a partition of S, i.e. $S = \bigcup_{i=1}^{\infty} A_i$
>
> Then $P(B) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i)P(A_i)$
>
> $$\Rightarrow \boxed{P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)} \ (*)}$$



corollary: $\forall i \quad \underbrace{P(A_i|B)}_{\text{Posterior dist.}} \propto \underbrace{P(B|A_i)}_{\text{likelihood}} \underbrace{P(A_i)}_{\text{Prior dist.}}$

proportional to (up to multiplying constant)

Bayesian stats

observed data

Remark: (*) is the law of total prob.

Independence

We want to be able to convey that an event B has no effect on A by insisting that

$$\underline{P(A|B) = P(A)} \quad (\text{provided } P(B) > 0)$$

$$\Rightarrow P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{\cancel{P(A)}P(B)}{\cancel{P(A)}} = P(B), \text{ provided } P(A) > 0.$$

Moreover,

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

> **Def:** two events are <u>statistically independent</u> if $\boxed{P(A \cap B) = P(A)P(B)}$

Remarks: · does not require $P(A)$ or $P(B) > 0$
· A & B independent does <u>not</u> mean $A \cap B = \emptyset$ (disjoint def.)

Ex: Dice rolling ; $S = \{1, \ldots, 6\}$. Here, $\tilde{S} = \{(r_1, \ldots, r_4)\} \in \{1, \ldots, 6\}^4$

Q: $P(\{\text{of throwing at least one 6 in 4 rolls}\})$

Assuming outcome of roles are independent

$$= 1 - P(\text{no 6 in 4 rolls})$$
$$= 1 - \prod_{i=1}^{4} (\text{no 6 in roll } i)$$
$$= 1 - (1 - \tfrac{1}{6})^4 = 1 - (5/6)^4 = 0.518$$

> **Thm.** If $A \overset{\text{ind.}}{\perp\!\!\!\perp} B$, then
> i) $A \perp\!\!\!\perp B^c$
> ii) $A^c \perp\!\!\!\perp B$
> iii) $A^c \perp\!\!\!\perp B^c$

Proof (i): $P(A \cap B^c) = P(A) - P(A \cap B)$, AoP(iii)
$= P(A) - P(A)P(B)$, ind.
$= P(A)(1 - P(B))$
$= P(A)P(B^c)$. ∎

Mutual Independence of Multiple Events

> **Def:** A collection of events $A_1, \ldots, A_n$ are <u>mutually independent</u>, if from any subcollection $A_{i_1}, \ldots, A_{i_k}$, we have
> $$\boxed{P(A_{i_1} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})}$$

Remark: · mutual independence much stronger than pairwise ind.

> corollary: Pairwise ind. does <u>not</u> imply mutual ind., i.e.
> given 3 events A, B, C s.t.
> $P(A)P(B) = P(A \cap B)$
> $P(A)P(C) = P(A \cap C)$ $\quad \not\Rightarrow P(A \cap B \cap C) = P(A)P(B)P(C)$
> $P(B)P(C) = P(B \cap C)$
>
> conversely, in general,
> $P(A \cap B \cap C) = P(A)P(B)P(C) \not\Rightarrow P(A \cap B) = P(A)P(B)$

EX(1): Let $S = \begin{cases} aaa & bbb & ccc \\ abc & bca & cba \\ acb & bac & cab \end{cases}$ s.t. each element in S is equally (1/9) probable.
Let $A_i = \{i^{th} \text{ place in the triple is occupied by } a\}$.

Then, $P(A_1) = P(A_2) = P(A_3) = 1/3$,

$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3)$
$= P(A_1 \cap A_2 \cap A_3) = P(\{aaa\}) = 1/9.$

Thus, A_1, A_2, A_3 are pairwise independent, but
$P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$ and, therefore, not mutually independent.

EX(2): Assume three coin tosses, each of which produces H or T.

$S = \begin{Bmatrix} HHH, HHT, HTH, HTT, \\ THH, THT, TTH, TTT \end{Bmatrix}$ Suppose equal prob. (1/8) on each element of S.

Let $A_i :=$ i^{th} toss is H, $i = 1, 2, 3$.

Then, $P(A_1) = P(\{HHH, HHT, HTH, HTT\}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$.

$P(A_1 \cap A_2) = (\{HHH, HHT\}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

$P(A_1 \cap A_2 \cap A_3) = P(\{HHH\}) = \frac{1}{8}$.

Likewise $P(A_i) = 1/2 \quad \forall i = 1, 2, 3$

$P(A_i \cap A_j) = 1/4 \quad \forall i, j \in \{1, 2, 3\}$

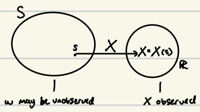
So $\begin{cases} P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \\ P(A_i \cap A_j) = P(A_i)P(A_j) \ \forall i \neq j \end{cases}$

Hence, A_1, A_2, A_3 are pairwise and mutually independent

1.4: Random variables

Def: A random variable is a function from a sample space to the real numbers (*)

Remark: may change to other spaces from $\mathbb{R}$, e.g. $\mathbb{Z}$, set of images, etc.



EX: Experiments — Random Variables

1) Toss 2 dice — X = sum of 2 numbers

$S = \{1, \dots, 6\} \times \{1, \dots, 6\}$

- Suppose $s = (s_1, s_2)$. Then $X := X(s) = s_1 + s_2$

2) Toss a coin 25 times — X = # heads

$S = \{H, T\}^{25}$

- Suppose $s = \{(c_1, \dots, c_{25})\}$. Then, $X := X(s) = \sum_{i=1}^{25} \mathbb{1}(c_i = H)$

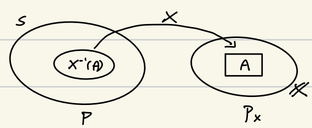
note: indicator function $\mathbb{1}(A) = \begin{cases} 1, & A \text{ true} \\ 0, & \text{otherwise} \end{cases}$

3) Apply different amt of fertilizer to flower plants — X = yield/acre

Probability Distribution

Def: Suppose the range of random variable X is $\mathcal{X}$. For any subset $A \subset \mathcal{X}$, the probability distribution of X is a probability function, denoted P_X, on (a sigma algebra of) subsets of $\mathcal{X}$ s.t.

$\forall$ such subset $A \subset \mathcal{X}$ $\boxed{P_X(X \in A) = P(X^{-1}(A)) = P(\{s \in S : X(s) \in A\})}$



Remark: Equivalently, $P_X(X = x_i) = P(\{s_j \in S : X(s_j) = x_i\})$

EX. Let X: number of heads out of tossing coin 3 times

$s \in S$ =	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(s)$	3	2	2	1	2	1	1	0

So, the range of (RV) X is:

$\mathcal{X}$	0	1	2	3
P_X	1/8	3/8	3/8	1/8

EX: Let X = # heads after tossing coin 50 times

$\mathcal{X} = \{0, \dots, 50\}$. So $P_X(X = i) = \dfrac{\binom{50}{i}}{2^{50}} \quad \forall i \in \mathcal{X}$.

Remark: For complicated space (e.g. uncountable), we need to specify prob.'s on "simple" sets which generate the associated sigma algebra

Distribution Functions

Def: the cumulative distribution function (cdf) of a RV X, denoted $\boxed{F_X(x) := P_X(X \leq x), \ \forall x \in \mathcal{X}}$, $\mathcal{X} = \mathbb{R}, \mathbb{Z}, \mathbb{N}$, for e.g.

Thm: The function $F: \mathcal{X} \to \mathbb{R}$ is a cdf i.f.f.

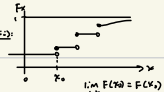
i) $\underline{\lim_{x\to-\infty} F(x) = 0}$ and $\underline{\lim_{x\to\infty} F(x) = 1}$

ii) F is a <u>non-decreasing function</u> in x

iii) F is right-continuous, i.e.

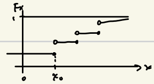
$$\lim_{x \downarrow x_0} F(x) = F(x_0) \quad \forall x_0 \in \mathbb{R}$$

<u>e.g. (satisfying i-iii)</u>:



$\lim_{x\downarrow x_0} F(x_0) = F(x_0)$ and thus right-continous

whereas



$\lim_{x\uparrow x_0} F(x_0) = F(x_0)$ and thus left-continuous, but $\lim_{x\downarrow x_0} F(x_0) \neq F(x_0) \Rightarrow$ <u>not</u> right-continuous

<u>Ex(i)</u>: Tossing coins for a head

Let $p :=$ prob. that a coin turns head

$X :=$ # independent coin tosses needed to get a head

$\Leftrightarrow$ # (ind.) trials until first head

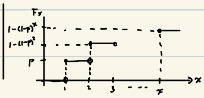
$\Leftrightarrow$ # (ind.) trials on which first head occurs

Then, $X \in \{1, 2, \dots\}$. So, $P(X=x) = \begin{cases} (1-p)^{x-1} p, & \forall x \in \{1,2,\dots\} \\ 0, & \text{otherwise} \end{cases}$

<u>note</u>: <u>$X \sim Geo(p)$.</u> ; <u>remark</u>: some authors define $X \sim Geo(p)$ $\begin{cases} X^* := \text{\# failures before first success} \\ X := \text{\# trials on which first success occurs} \end{cases}$

$\Rightarrow X^* = X - 1 \Rightarrow P(X = x_0) = P(X^* = x_0 - 1)$.

$$\begin{aligned} \Rightarrow F_X(x) &= P(X \le x), \ \forall x \ge 0 \\ &= \sum_{i=1}^{x} (1-p)^{i-1} p \\ &= p \sum_{k=0}^{x-1} (1-p)^k, \text{ letting } k = i-1 \\ &= \cancel{p} \frac{1-(1-p)^x}{1-\cancel{(1-p)}} \\ &= 1 - (1-p)^x, \ x \ge 1 \end{aligned}$$



<u>recall: (Geometric series)</u>

<u>Prop.</u> $\forall r \neq 1 \in \mathbb{R}$, $\sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}$

<u>Prop.</u> $\forall |r| < 1$, i.e. $r \in (-1, 1)$, $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$; otherwise, series diverges.

<u>Ex (continuous cdf)</u>

Let $F_X(x) = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$. Then $\frac{d}{dx} F_X(x) = \frac{e^{-x}}{(1+e^{-x})^2} > 0 \Rightarrow F_X$ increasing (and therefore) nondecreasing

and $\begin{cases} F_X \downarrow 0 \text{ as } x \to -\infty \\ F_X \uparrow 1 \text{ as } x \to \infty \end{cases}$

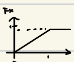
<u>Remarks</u>: To show F_X is nondecreasing $\begin{cases} \text{If } F_X \text{ is continuous and differentiable,} \\ \quad \text{then } F'(x) = f(x) \ge 0 \Rightarrow F_X \text{ is nondecreasing.} \\ \text{Generally, and only approach for a discrete distribution,} \\ \quad \text{If } x_1 < x_2, \text{ then } \{X \le x_1\} \subseteq \{X \le x_2\} \Rightarrow F(x_1) \le F(x_2) \end{cases}$

where F_X is always nondecreasing, not necessarily strictly increasing.

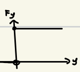
Def: A (RV X) is <u>continuous</u> if its cdf is a continuous function.
X is <u>discrete</u> if its cdf is a step function.

<u>Remark</u>: cdf can also be a "mixture" of continuous segments and jumps

e.g.s



where $X \sim$ Uniform s.t. $F_X(x) = \begin{cases} x, & x \in [0,1] \\ 0, & x \le 0 \\ 1, & x \ge 1 \end{cases}$



s.t. $F_Y(y) = \begin{cases} 0, & y < 0 \\ 1, & y \ge 0 \end{cases}$

<u>Notice</u>: $F_Z(z) = \frac{1}{2} F_X(z) + \frac{1}{2} F_Y(z)$

still satisfies (i)-(iii) $\Rightarrow$ valid cdf.

and in context,

flip a fair coin $C = \{H, T\}$ s.t. $\begin{cases} \text{if } C = H, \ Z = X \\ \text{if } C = T, \ Z = Y \end{cases}$

<u>Remark</u>: motivates RV model construction from constituent RV mixtures

Identical Distributions

<u>Def</u>: Two RVs X and Y are <u>identically distributed</u> if

$\boxed{P(X \in A) = P(Y \in A), \ \forall A \in \mathcal{B}}$, i.e.

X and Y have the same distribution

<u>Remark</u>: sometimes we write $\underline{X \overset{d}{=} Y}$ but NOT $X = Y$.

<u>Ex(i)</u>: Toss a fair coin 3 times.

Then $X =$ # heads and $Y =$ # tails

$\Rightarrow X \overset{d}{=} Y$ since $P(X = i) = P(Y = i) \ \forall i \in S$.

<u>Thm</u>. Let X, Y be RV's associated w/ sample space S.
$\boxed{X \overset{d}{=} Y \text{ i.f.f. } F_X(a) = F_Y(a) \ \forall a \in \mathbb{R}.}$

<u>Proof</u>.

"only if": if $X \overset{d}{=} Y$, then $\forall A \in \mathcal{B}$, $P_X(A) = P_Y(A)$

In particular, $\underbrace{P_X((-\infty, a])}_{F_X(a)} = \underbrace{P_Y((-\infty, a])}_{F_Y(a)}$ $\forall a \in \mathbb{R}$.

"if": If $F_X(a) = F_Y(a)$ $\forall a \in \mathbb{R}$

$\Rightarrow P_X((-\infty, a]) = P_Y((-\infty, a])$ $\forall a \in \mathbb{R}$

But we know that $\{(-\infty, a]\}$ generates the Borel-sigma algebra of subsets of $\mathbb{R}$.

Combining w/ AOP, it follows that

$$P_X(A) = P_Y(A) \quad \forall \ A \subset \mathcal{B}. \quad \square$$