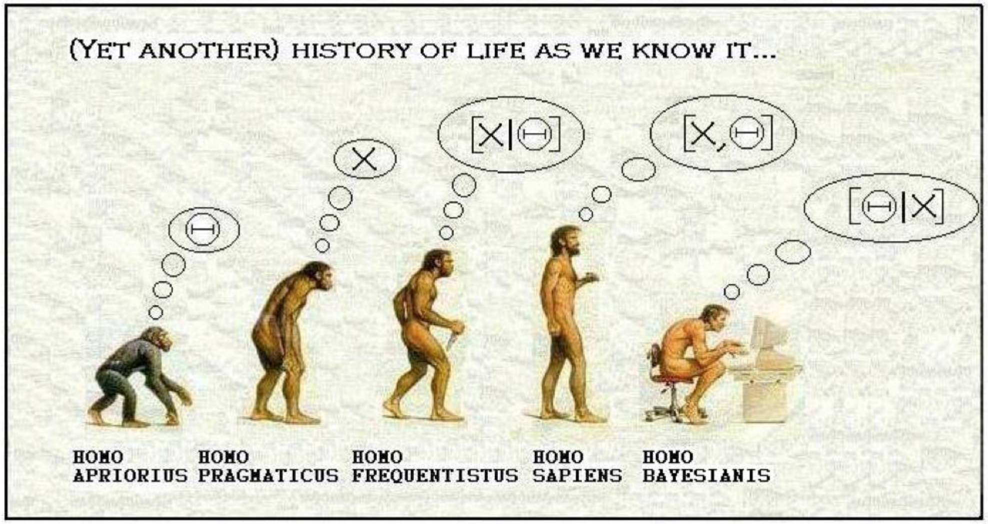


Aprendre amb les lleis de la probabilitat

Coneixement, Raonament i Incertesa.

El contingut d'aquest document s'ha derivat de material provinent de
Tom Mitchell, William Cohen, Andrew Moore, Aarti Singh, Eric Xing, Carlos Guestrin.

Que tenim fins ara?

Hem viscut en un mon determinista

$$f: X \longrightarrow Y$$

I si no n'estem segurs?

Recordatori de probabilitat

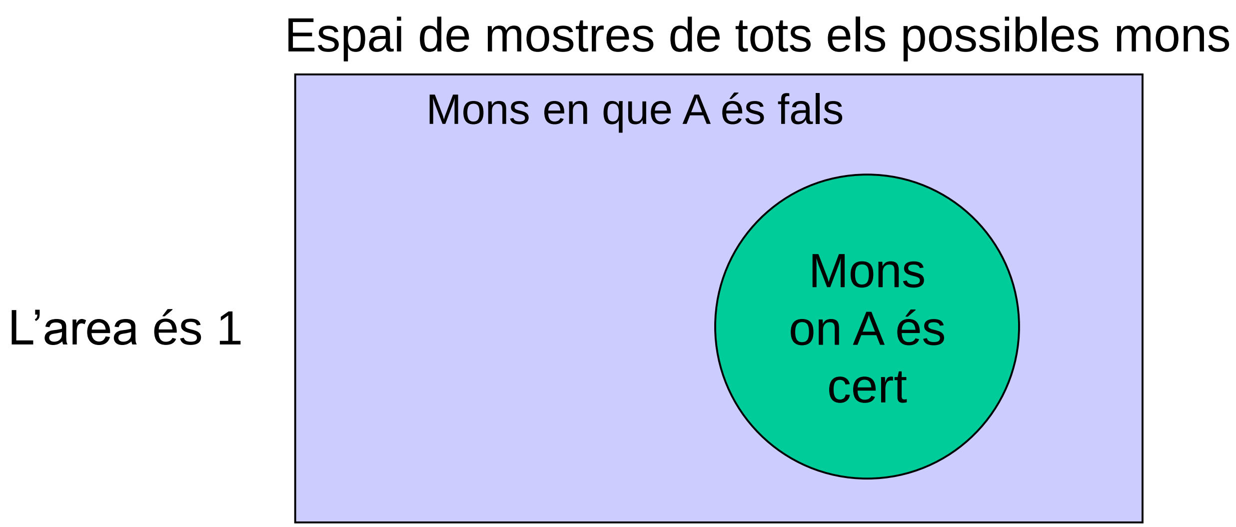
- Variables aleatòries (Random variables)
- Axiomes de la probabilitat
- Events Independents
- Probabilitat condicionada
- Regla de Bayes i creença (beliefs)
- Joint probability distribution
- Expectatives
- Independència, Independència Condicional

Conceptes bàsics

- Un espai de mostres (sample space) *S* és el conjunt de tots els possibles resultats d'un experiment conceptual o físic repetible. (*S* pot ser finit o infinit.)
  - P.ex., *S* podria ser els possibles resultats d'una tirada de daus
- Un event A és un subconjunt de *S*.
  - P.ex., A= l'event que al tirar el dau surti un número < 3.

Probabilitat

- Una *probabilitat* *P(A)* és una funció que mapeja un event *A* en un interval *[0, 1]*. *P(A)* també s'anomena mesura de probabilitat o 'probability mass' de *A*.



P(A) és l'àrea del cercle

Variables Aleatòries

- Informalment, A és una variable aleatòria si
  - A denota alguna cosa sobre el que no estem segurs

    Potser la sortida d'un experiment aleatori

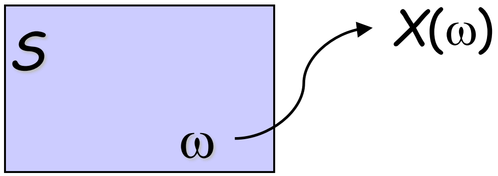
- Exemples
  - A = si una persona aleatòria de la classe es dona/home
  - A = la ciutat natal de una persona aleatòria de la classe
  - A = si dues persones aleatòries de la classe fan anys el mateix dia

- Definim P(A) com "la fracció de possibles mons en els que A és veritat" o la seva aproximació "la fracció de vegades que A és cert, en una seqüència aleatòria d'execucions d'un experiment"
  - una variable aleatòria A és una funció definida sobre S

$$A:S \rightarrow \{0,1\}$$

Variables Aleatòries

- una *variable aleatòria (r.v.)* és una funció que associa un únic número amb el resultat d'un experiment.



- r.v. discreta:
    - El resultat de tirar un dau: D={1,2,3,4,5,6}
  - Event binari i variable indicadora:
    - Veure un "6" en una tirada ⇒ X=1, si no X=0.
    - Això descriu el resultat verdader o fals d'un event aleatori.
- r.v continua:
    - El resultat d'observar la posició mesurada d'un avió.

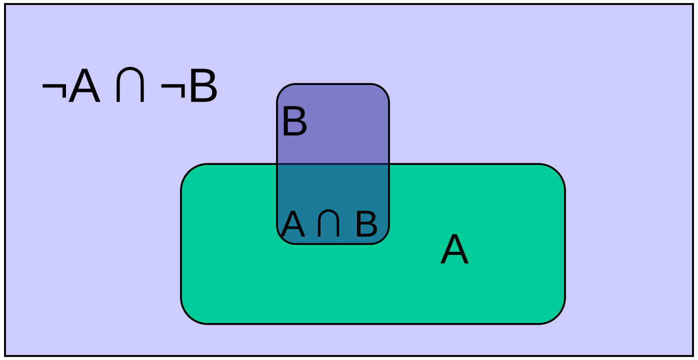
Axiomes de Kolmogorov

- Totes les probabilitats estan entre 0 i 1

$$0 \leq P(A) \leq 1$$

- $P(S) = 1$
- $P(\Phi)=0$
- La probabilitat de la disjunció ve donada per

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



A ∩ B ?

Teoremes útils (derivats dels axiomes)

- Totes les probabilitats estan entre 0 i 1

  $0 \leq P(A) \leq 1$
- $P(S) = 1$
- $P(\Phi)=0$
- La probabilitat de la disjunció ve donada per

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

$\rightarrow$ P(not A) = P(¬A) = 1-P(A)

Teoremes útils (derivats dels axiomes)

Axioma: $0 <= P(A) <= 1$, $P(True) = 1$, $P(False) = 0$,

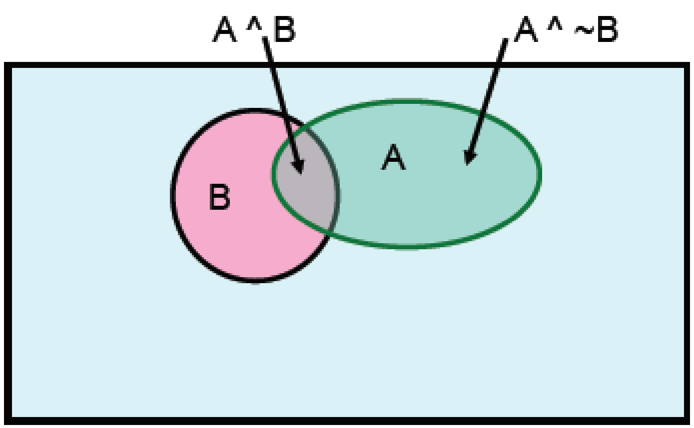
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\boxed{\rightarrow P(A) = P(A \cap B) + P(A \cap \neg B)} \; *$$

$$A = [A \cap (B \cup \neg B)] = [(A \cap B) \cup (A \cap \neg B)]$$

$$P(A) = P(A \cap B) + P(A \cap \neg B) - P((A \cap B) \cap (A \cap \neg B))$$

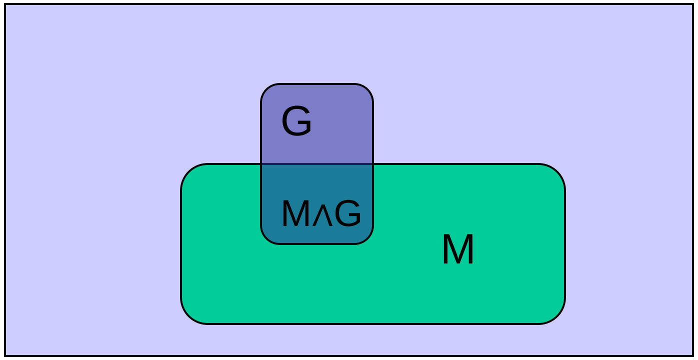
$$P(A) = P(A \cap B) + P(A \cap \neg B) - \cancel{P(A \cap B \cap A \ \neg B)}$$



* llei de probabilitat total
(versió simplificada)

Probabilitat condicionada

- P(X|Y) = Fracció dels mons en els que X és cert donat que Y és cert
  - M = “tenir maldecap"
  - G = “tenir la grip"
    - P(M)=1/10
    - P(G)=1/40
    - P(M|G)=1/2
  - P(M|G) = fracció de maldecap donat que tenim la grip
    = P(M∧G)/P(G)



- Definició: $P(X\,|\,Y)=\dfrac{P(X\wedge Y)}{P(Y)}$

Probabilitat condicionada

Definició:

$$P(X \mid Y) = \frac{P(X \wedge Y)}{P(Y)}$$

Corolari: La regla de la cadena

$$P(X \wedge Y) = P(X \mid Y)P(Y)$$

$$P(X \wedge Y \wedge Z) = P(Z \mid X \wedge Y)P(X \mid Y)P(Y)$$

Exemples Probabilitat condicionada

1. Escollim dues cartes a l'atzar d'un joc de 52 cartes. Quina és la probabilitat de treure un as en la segona carta si en la primera hem tret un as?

2. En un curs les dades sobre un cert període són: 23 alumnes passen el primer examen, 25 l'han passat en el segon intent, 7 en el tercer i 8 han suspès definitivament. Quina és la probabilitat (risc) de suspendre el curs si l'estudiant ha suspès els dos primers exàmens?

3. (2005) 84% de les famílies de Suïssa tenen ordinador a casa, el 81% de les famílies de Suïssa tenen ordinador i internet a casa. Quina és la probabilitat que una família suïssa tingui internet a casa si sabem que tenen ordinador?

Exemples Probabilitat condicionada

1. Escollim dues cartes a l'atzar d'un joc de 52 cartes. Quina és la probabilitat de treure un as en la segona carta si en la primera hem tret un as?

$$\boxed{P(2=\text{as}\,|\,1=\text{as})=\frac{3}{51}}$$

quina seria la probabilitat que els dos siguin asos?

3. En un curs les dades sobre un cert periode són: 23 alumnes passen el primer exàmen, 25 l'han passat en el segon intent, 7 en el tercert i 8 han suspes definitivament. Quina és la probabilitat (risc) de suspendre el curs si l'estudiant ha suspès els dos primers exàmens?

4. (2005) 84% de les families de Suissa tenen ordinador a casa, el 81% de les families de Suissa tenen ordinador i internet a casa. Quina és la probabilitat que una familia suissa tingui internet a casa si sabem que tenen ordinador?

Exemples Probabilitat condicionada

1. Escollim dues cartes a l'atzar d'un joc de 52 cartes. Quina és la probabilitat de treure un as en la segona carta si en la primera hem tret un as?

$$P(2=\text{as}\,|\,1=\text{as})=\frac{3}{51}=5.8\%$$

$$P(2=\text{as},1=\text{as})=P(2=\text{as}\,|\,1=\text{as})P(1=\text{as})=\frac{3}{51}\cdot\frac{4}{52}=0.5\%$$

3. En un curs les dades sobre un cert periode són: 23 alumnes passen el primer exàmen, 25 l'han passat en el segon intent, 7 en el tercert i 8 han suspes definitivament. Quina és la probabilitat (risc) de suspendre el curs si l'estudiant ha suspès els dos primers exàmens?

4. (2005) 84% de les families de Suissa tenen ordinador a casa, el 81% de les families de Suissa tenen ordinador i internet a casa. Quina és la probabilitat que una familia suissa tingui internet a casa si sabem que tenen ordinador?

Exemples Probabilitat condicionada

1. Escollim dues cartes a l'atzar d'un joc de 52 cartes. Quina és la probabilitat de treure un as en la segona carta si en la primera hem tret un as?

$$\boxed{P(2=\text{as}\,|\,1=\text{as})=\frac{3}{51}}$$

$$\boxed{P(2=\text{ace},1=\text{as})=P(2=\text{as}\,|\,1=\text{as})P(1=\text{as})=\frac{3}{51}\cdot\frac{4}{52}=0.5\%}$$

3. En un curs les dades sobre un cert període són: 23 alumnes passen el primer examen, 25 l'han passat en el segon intent, 7 en el tercer i 8 han suspès definitivament. Quina és la probabilitat (risc) de suspendre el curs si l'estudiant ha suspès els dos primers exàmens?

$$\boxed{P(\text{suspendre curs}\,|\,\text{suspes E}1\,\&\,2)=\frac{P(\text{suspendre curs, suspendre E}1\,\&\,2)}{P(\text{suspes E}1\,\&\,2)}=\frac{8/63}{15/63}=53\%}$$

4. (2005) 84% de les families de Suissa tenen ordinador a casa, el 81% de les families de Suissa tenen ordinador i internet a casa. Quina és la probabilitat que una familia suissa tingui internet a casa si sabem que tenen ordinador?

Exemples Probabilitat condicionada

1. Escollim dues cartes a l'atzar d'un joc de 52 cartes. Quina és la probabilitat de treure un as en la segona carta si en la primera hem tret un as?

$$\boxed{P(2=\text{as}\,|\,1=\text{as})=\frac{3}{51}}$$

$$\boxed{P(2=\text{ace},1=\text{as})=P(2=\text{as}\,|\,1=\text{as})P(1=\text{as})=\frac{3}{51}\cdot\frac{4}{52}=0.5\%}$$

3. En un curs les dades sobre un cert periode són: 23 alumnes passen el primer exàmen, 25 l'han passat en el segon intent, 7 en el tercert i 8 han suspes definitivament. Quina és la probabilitat (risc) de suspendre el curs si l'estudiant ha suspès els dos primers exàmens?

$$\boxed{P(\text{suspendre curs}\,|\,\text{suspendre OE }1\,\&\,2)=\frac{P(\text{suspendre curs, suspendre E}1\,\&\,2)}{P(\text{suspendre OE}1\,\&\,2)}=\frac{8/63}{15/63}=53\%}$$

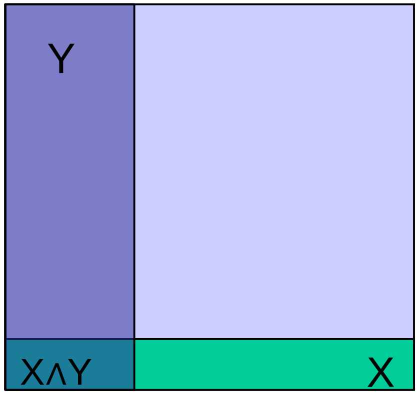
4. (2005) 84% de les families de Suissa tenen ordinador a casa, el 81% de les families de Suissa tenen ordinador i internet a casa. Quina és la probabilitat que una familia suissa tingui internet a casa si sabem que tenen ordinador?

$$\boxed{P(\text{Internet}\,|\,\text{Ordinador})=\frac{P(\text{Internet, Ordinador})}{P(\text{Ordinador})}=\frac{0.81}{0.84}=96\%}$$

Events Independents

- Dues variables aleatories X i Y es diu que son independents si:

  $P(X \cap Y) = P(X)*P(Y)$



- Alternativament,

  $P(X \mid Y) = P(X)$ i $P(Y \mid X) = P(Y)$

- Intuitivament, això vol dir que Y passi no ens dona cap informació del valor de B (i a l'inreves).

- Nota: Això no vol dir que X i Y siguin dijuntes!!!

Regles d'independencia - exemple

- P(Virus | veure_cervesa) = **P(Virus)**
  iff Virus és independent de veure_cervesa

- P(Grip | Virus; veure_cervesa) = **P(Grip|Virus)**
  iff Grip és independent de veure_cervesa, donat Virus

- P(mal_de_cap | grip;Virus;cervesa) = **P(Mal_de_cap|Grip;Cervesa)**
  iff Mal_de_cap és independent de Virus, donat Grip i Cervesa

La Regla de Bayes

$$P(X \mid Y) = \frac{P(X \wedge Y)}{P(Y)}$$

$$P(X \wedge Y) = P(X \mid Y)P(Y)$$

$$P(Y \mid X)P(X) = P(X \wedge Y) = P(X \mid Y)P(Y)$$

La Regla de Bayes

$$P(Y \mid X) = \frac{P(X \mid Y)\,p(Y)}{P(X)}$$

Anomenem P(Y) «prior»
Anomenem P(X|Y) «posterior»

Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53:370-418**

De forma general la Regla de Bayes és:

$$P(Y \mid X) = \frac{P(X \mid Y)\,p(Y)}{P(X \mid Y)\,p(Y) + P(X \mid \neg Y)\,p(\neg Y)}$$

$$P(Y = y_i \mid X) = \frac{P(X \mid Y = y_i)\,p(Y = y_i)}{\sum_{j \in S} P(X \mid Y = y_j)\,p(Y = y_j)}$$

…by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter…. necessary to be considered by any that would give a clear account of the strength of *analogical* or *inductive reasoning*…

Aplicant la regla de Bayes

$$P(Y \mid X) = \frac{P(X \mid Y)p(Y)}{P(X)} \qquad P(Y \mid X) = \frac{P(X \mid Y)p(Y)}{P(X \mid Y)p(Y) + P(X \mid \neg Y)p(\neg Y)}$$

A = tens la grip, B = acabes de tossir

Assumim:

P(grip) = 0.05

P(tossir | grip) = 0.80

P(tossir | ~grip) = 0.1

Quina és la probabilitat que tinguis grip si acabes de tossir?

Que tenim de nou ara?

Hem viscut en un mon determinista

$$f: X \longrightarrow Y$$

I ara afegim la incertesa en les nostres decisions

$$P(Y|X)$$

Com?

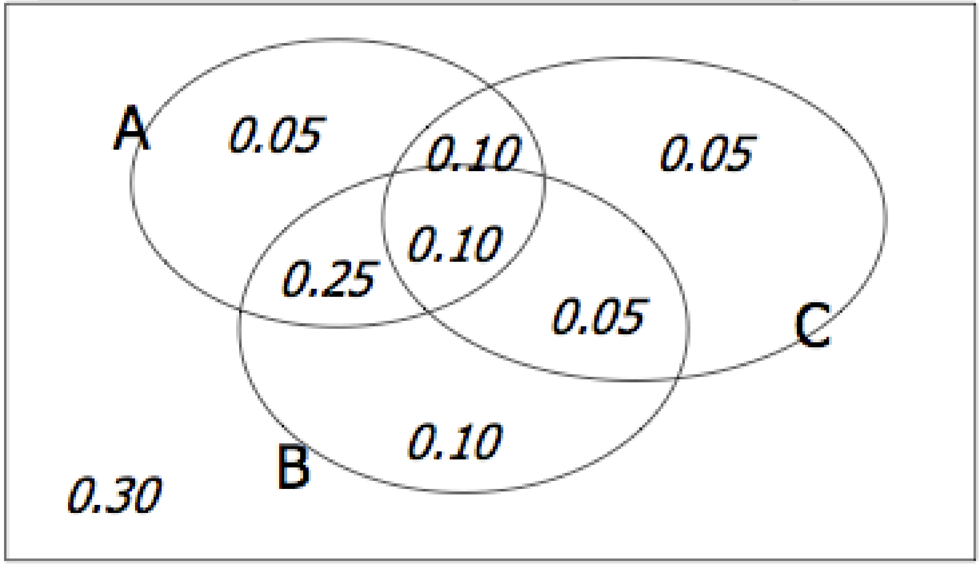
Joint Distribution

Recepta per a fer la joint distribution de M variables:

1. Fer una taula de veritat amb totes les combinacions dels valors de les variables (si hi ha M variables booleanes llavors tindrem 2^M files).
2. Per a cada combinació calcular com és de probable.
3. Aquestes probabilitats han de sumar 1 segons els axiomes de la probabilitat.

Variables booleanes A, B and C

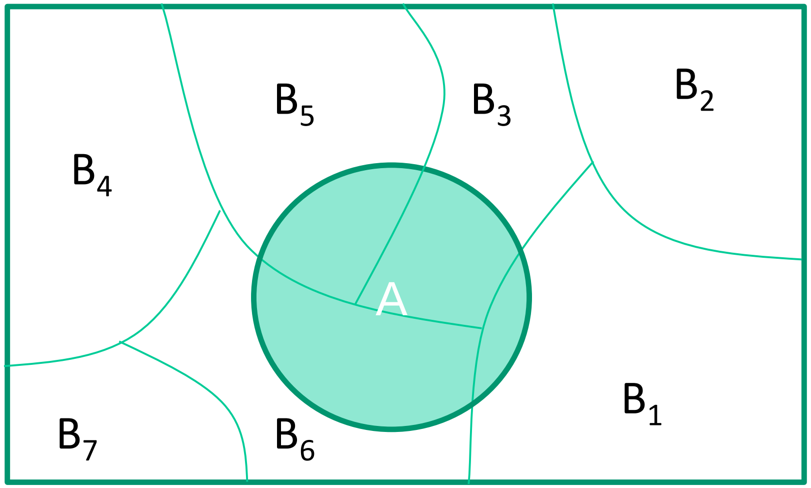
A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10



Marginalització

- Si coneixem p(X, Y), que és P(X=x)?
- Podem usar la llei de probabilitat total

$$p(x) = \sum_{y} P(x, y)$$
$$= \sum_{y} P(y) P(x \mid y)$$



$$p(x) = \sum_{y,z} P(x, y, z)$$
$$= \sum_{z,y} P(y, z) P(x \mid y, z)$$

Us de la Joint Distribution

Un cop tenim la JD podem preguntar per la probabilitat de qualsevol expressió lògica sobre les nostres variables

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(E) = \sum_{files\ correspoents\ a\ E} P(fila)$$

Dades de la UCI "Adult" Cens

Us de la Joint Distribution

P(Poor) = 0.7604

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(E) = \sum_{files\ correspoents\ a\ E} P(fila)$$

Us de la Joint Distribution

P(Poor ; Male) = 0.4654

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(E) = \sum_{files\ correspoents\ a\ E} P(fila)$$

Inferencia amb la Joint Distribution

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(E_1|E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{files\ correspoents\ a\ E_1\ i\ a\ E_2} P(fila)}{\sum_{files\ correspoents\ a\ E_2} P(fila)}$$

P(Male | Poor) = 0.4654 / 0.7604 = 0.612

Aprenentatge amb la Joint Distribution

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

Suposem que volem aprendre la funció f: <G, H> → W

De forma equivalent, P(W | G, H)

Solució: aprendre la ‘joint distribution’ de les dades, calcular P(W | G, H)

P.ex:., P(W=rich | G = female, H = 40.5-) =

D’on surten les ‘Join Distribution’

- Idea 1: Humans Experts
- Idea 2: fets probabilistics simples + algebra

Exemple: Suposem que coneixem P(A) = 0.7

P(B|A) = 0.2 P(B|~A) = 0.1

P(C|A^B) = 0.1 P(C|A^~B) = 0.8

P(C|A^~B) = 0.8 P(C|~A^B) = 0.3

P(C|~A^~B) = 0.1

Llavors podem calcular la JD usant la regla de la cadena

P(A=x ^ B=y ^ C=z) =

P(C=z|A=x^ B=y) P(B=y|A=x) P(A=x)

Les xarxes bayesianes intenten automatitzar-ho

D’on surten les ‘Join Distribution’

- Idea 3: Extreure-la de les dades

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

On som?

- Hem après a fer inferencies
  - Em fa mal el coll: quina possibilitat hi ha de tenir meningits?
  - La PN ha trobat un arsenal d'armes. Quina possibilitat hi ha que el TS ho declari com a delicte de terrorisme?
  - Aquesta persona esta llegint un llibre sobre pianistes: quina possibilitat hi ha que vulgui comprar un piano?

És un gran repte, ja que la inferència és al nucli de molta part de la 'industria'. Predir enquestes, la borsa, optimitzar la localització de anuncis, etc poden potencialment guanyar diners. Predir una passa de meningitis, pot ajudar el mon ☺.

Ja hem acabat?

On som?

1. Necessitem 2^m files en la joint distribution per poder fer inferència

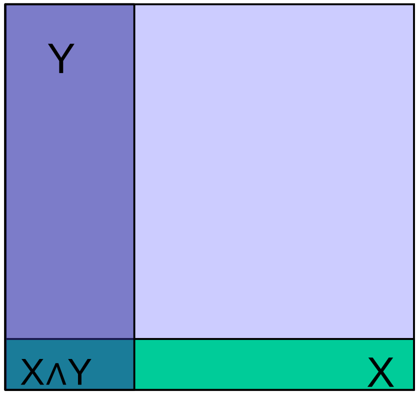
p.ex: m=40 ➔ 1.099.511.627.776 files

Solució?

Events Independents

- Dues variables aleatories X i Y es diu que son independents si:

  $P(X \cap Y) = P(X)*P(Y)$



- Alternativament,

  $P(X \mid Y) = P(X)$ i $P(Y \mid X) = P(Y)$

- Formalment escriurem $X \perp Y$

- Intuitivament, això vol dir que Y passi no ens dona cap informació del valor de B (i a l'inreves).

- Nota: Això no vol dir que X i Y siguin dijuntes!!!

Independencia Condicional

P(Engega*, Benzina, Tard*) té 2^3 entrades

Si el cotxe no engega, la probabilitat de no tenir benzina no depèn de si fem tard a la feina:

(1) **P**(Benzina=no | *Tard=si, Engega=no*) = **P**(Benzina=no | *Engega=no*)

La mateixa independència és certa si engega:

(2) **P**(Benzina=no | *Tard=si,* Engega) = **P**(Benzina=no | Engega)

La variable ‘Benzina’ és condicionalment independent de ‘Tard’ donat ‘Engega’ :

P(Benzina | Tard, Engega) = **P**(Benzina | Engega)

De forma equivalent:
P(Tard | Benzina, Engega) = **P**(Tard| Engega)

P(Tard, Benzina| Engega) = **P**(Tard| Engega) **P**(Benzina | Engega)

Independència Condicional

De forma general dos events X,Y són condicionalment independents d'un event Z sii:

P(X,Y|Z) = P(X|Z)P(Y|Z) o bé

P(X|Y,Z) = P(X|Z) i de forma equivalent **P(Y|X,Z) = P(Y|Z)**

Formalment escriurem $X \perp Y|Z$

Independència Condicional

Independència absoluta (marginal) implica independència condicional?

Tirem dues monedes a l'aire. A és l'event de que la primera moneda sigui cara i B l'event de que la segona moneda sigui cara. Sigui C l'event representant el fet de que les dues monedes obtenen el mateix resultat (les dues cara o les dues creu). A i B són clarament independents però no són independents donat C

$P(A|B) = P(A)$
$P(A|B,C) <> P(A|C)$

Independència condicional implica independència absoluta (marginal)?

Truco a dos amics (A i B) i els hi dic el mateix número del 1 al 10 (N). Donat el soroll de la linia els dos de forma independent fan una deducció sobre el número que han sentit Na i Nb

Na i Nb són marginalment independents?

és lògic pensar que $P(Na=x|Nb=x) > P(Na=x)$

Na i Nb són condicionalment independents?

$P(Na=x|Nb=x,N=2) = P(Na=x|N=2)$

un cop sabem el que he dit, tant li fa el que ha escoltat l'altre !!!

Independència Condicional

Escrivim la joint distribution completa usant la regla de la cadena:

P(Tard*, Benzina, Engega*)
= **P**(Tard | *Benzina, Engega*) **P**(*Benzina, Engega*)

= **P**(Tard | *Benzina, Engega*) **P**(*Benzina* | *Engega*) **P**(*Engega*)

= **P**(Tard | *Benzina*) **P**(*Benzina* | *Engega*) **P**(*Engega*)

I.e., 2 + 2 + 1 = 5 números independent

En molts casos, l'ús de la independència condicional redueix el tamany de la representació de la 'joint distribution' d'exponencial a lineal.

On som?

1. Necessitem 2^m files en la joint distribution per poder fer inferencia (m és el número de variables)

p.ex: m=40 ➔ 1.099.511.627.776 files

Solució?

Si les variables són independents necessitem 40 files

On som?

1. Necessitem 2^m files en la joint distribution per poder fer inferència (m és el número de variables)

   Solució? No sempre podem assegurar independència

2. No sempre tenim informació de tots els casos

   Solució? Buscar maneres alternatives a la 'joint distribution'