Inteligență artificială

10. Rețele bayesiene

Florin Leon

Universitatea Tehnică „Gheorghe Asachi” din Iași
Facultatea de Automatică și Calculatoare

http://florinleon.byethost24.com/curs_ia.html

Rețele bayesiene

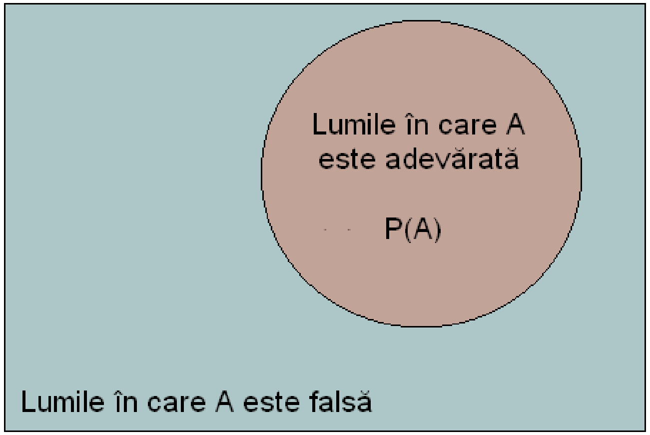
1. Probabilități
2. Rețele bayesiene
3. Inferențe exacte și aproximative
4. Rețele bayesiene dinamice
5. Filtrarea cu particule
6. Concluzii

Rețele bayesiene

1. Probabilități
    1.1. Teorema lui Bayes
    1.2. Independență și independență condiționată
2. Rețele bayesiene
3. Inferențe exacte și aproximative
4. Rețele bayesiene dinamice
5. Filtrarea cu particule
6. Concluzii

Probabilități

- Interpretarea frecventistă (număr de experimente)
  - *P*(*A*) este fracțiunea de lumi posibile în care *A* este adevărată



- Interpretarea subiectivistă (caracterizarea convingerilor, grade de încredere)

„Probabilități deterministe”

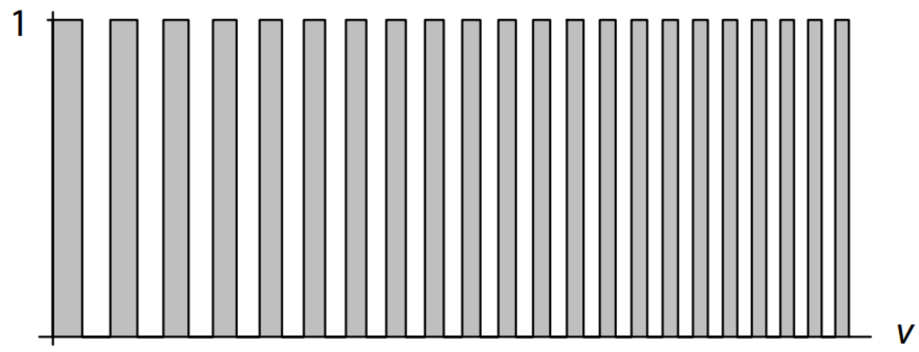


FIG. 1. Evolution function for a simple wheel of fortune, mapping initial spin speed v to either 1 (*red*) or 0 (*black*). The area under the function, corresponding to *red*-yielding values of v, is for clarity's sake shaded gray.

„Probabilități deterministe”

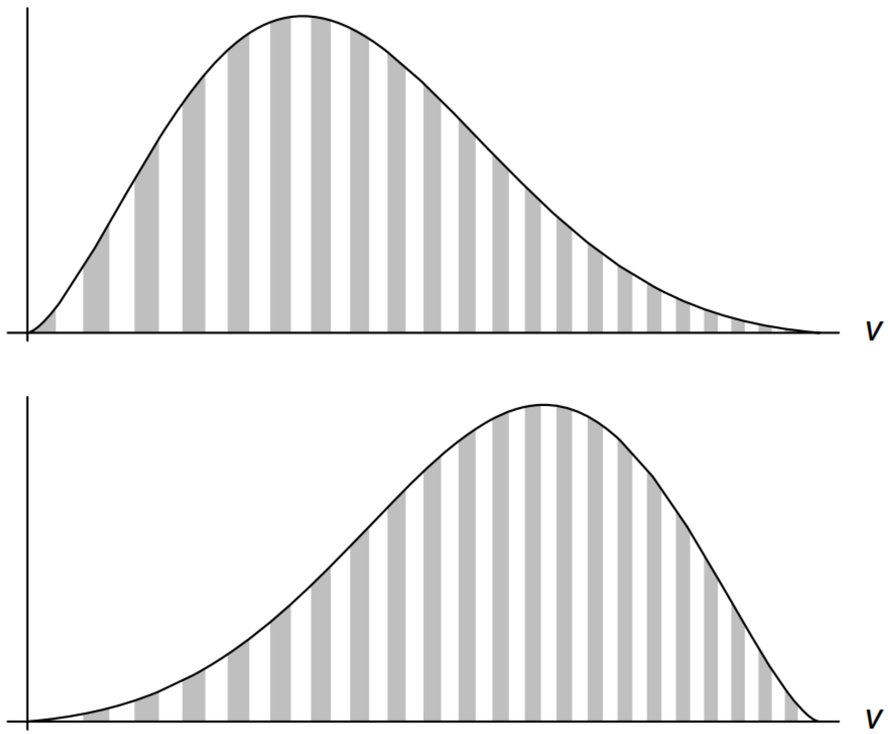


FIG. 2. Different spin speed distributions induce the same probability for *red*, equal to the strike ratio for *red* of one half.

„Probabilități deterministe”

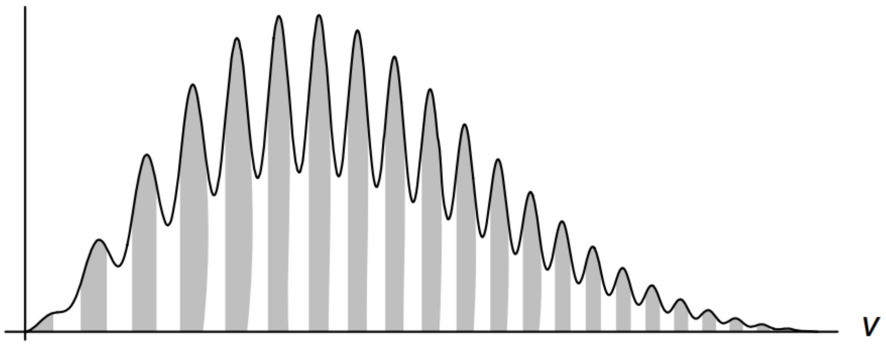
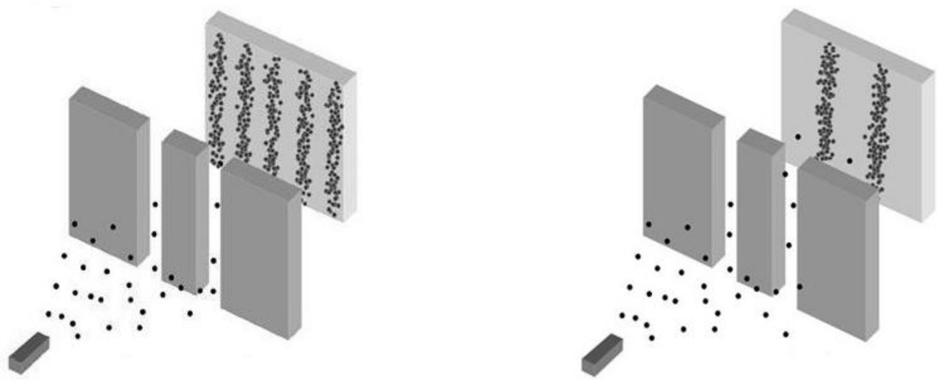


FIG. 3. Not every initial-spin-speed distribution induces a probability for *red* of one half.

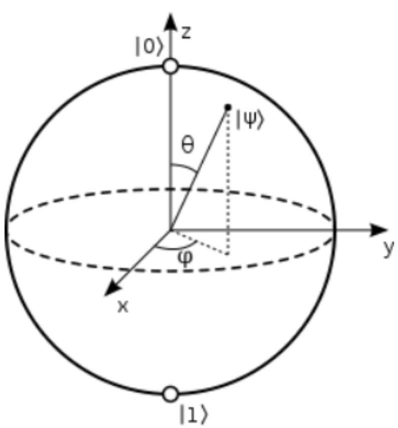
Probabilități

- Lumea macroscopică
  - Probabilitățile sunt doar o măsură a incapacității de a prezice evoluția unor procese complexe?
- Lumea cuantică
  - Procese impredictibile: descompunerea radioactivă
  - *Double-slit experiment*, *quantum pigeonholes*
  - Amplitudini de probabilitate
  - Calcul cuantic: qubiți

Lumea cuantică



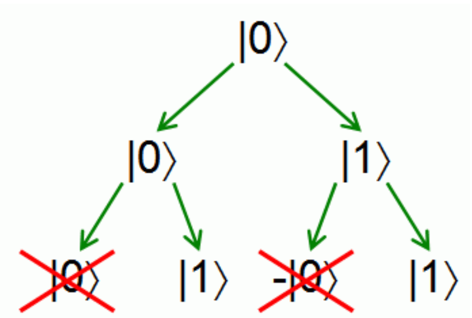
Double-slit experiment



$$|\Psi\rangle = a|0\rangle + b|1\rangle$$

Amplitudini de probabilitate

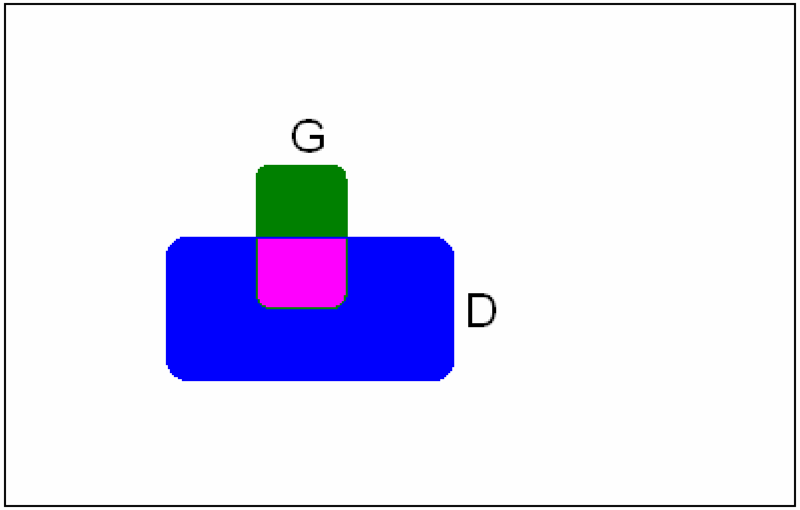
Sfera Bloch



Interferență cu poarta Hadamard

Probabilități condiționate

- $P(A|B)$ este fracțiunea de lumi posibile în care B este adevărată și atunci și A este adevărată
  - Probabilitatea lui A, dat fiind B



- D = durere de cap, $P(D) = 1/10$
- G = gripă, $P(G) = 1/40$
- $P(D|G) = 1/2$
- Dacă cineva are gripă, probabilitatea de a avea și dureri de cap este de 50%
- $P(D|G) = P(D \cap G) / P(G)$

Teorema lui Bayes

- $P(A|B) = P(A \cap B) / P(B)$
- $P(A \cap B) = P(A|B) \cdot P(B)$
- $P(A \cap B) = P(B|A) \cdot P(A)$
- $\Rightarrow \boxed{P(B|A) = P(A|B) \cdot P(B) / P(A)}$

probabilitatea a posteriori (*posterior*) — verosimilitatea (*likelihood*) — probabilitatea a priori (*prior*) — evidența (*evidence*)

Teorema lui Bayes

- $P(B|A) = P(A|B) \cdot P(B) / P(A)$

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

Teorema lui Bayes

- $P(I|E) = P(E|I) \cdot P(I) \,/\, P(E)$
  - I = ipoteza, cauza
  - E = evidența (observația), efectul

Diagnoză

- Probabilități cunoscute
  - **M**eningită: $P(M) = 0{,}002\%$
  - **G**ât înțepenit: $P(G) = 5\%$
  - Meningita cauzează gât înțepenit în jumătate din cazuri: $P(G|M) = 50\%$
- Dacă un pacient are gâtul înțepenit, care este probabilitatea să aibă meningită?
  - $P(M|G) = P(G|M) \cdot P(M) / P(G) = 0{,}02\%$

Diagnoză

- Greșeală întâlnită uneori: $P(A|B) = P(B|A)$
- Diagnostice pentru boli rare
  - Trebuie avută în vedere probabilitatea testului de a returna rezultate fals pozitive

$$P(B) = 0{,}01$$

$$P(T|B) = 0{,}99$$

$$P(T|\neg B) = 0{,}02.$$

B – boală
T – test

$$P(B|T) = \frac{P(T|B) \cdot P(B)}{P(T)} = \frac{P(T|B) \cdot P(B)}{P(T|B) \cdot P(B) + P(T|\neg B) \cdot P(\neg B)} = 0{,}33$$

Independență și independență condiționată

- Exemplul 1. Ion și Maria dau cu banul de 100 de ori. Fiecare are un ban diferit
  - Evenimente independente
  - Rezultatul unui experiment nu influențează rezultatul celuilalt experiment

Independență și independență condiționată

- Exemplul 2a. Ion și Maria dau cu același ban
  - Dacă banul nu este corect, evenimentul *A* (Ion) poate aduce informații asupra evenimentului *B* (Maria)
  - Evenimentele nu sunt independente
  - Rezultatul unui experiment poate influența cunoștințele despre rezultatul celuilalt

Independență și independență condiționată

- Exemplul 2b. Fie *C* variabila „banul este influențat în favoarea pajurei”
  - Dacă știm *C*, experimentul *A* nu mai aduce informații noi asupra lui *B*
    - $P(B|A,C) = P(B|C)$
  - *A* și *B* sunt independente condiționat dat fiind *C*
  - Situație numită „cauză comună”

Independență și independență condiționată

- Exemplul 3. Ion și Maria locuiesc în zone diferite ale orașului și vin la serviciu cu tramvaiul, respectiv mașina
  - „Ion a întârziat” și „Maria a întârziat” pot fi considerate independente
  - Dacă vatmanii sunt în grevă, atunci și traficul rutier crește. Evenimentele sunt independente *condiționat*

- Există multe situații în viața reală în care evenimente considerate independente sunt de fapt independente condiționat

Independență și independență condiționată

- Exemplul 4. Atât răceala cât și alergia îl pot determina pe Ion să strănute
  - Dacă nu știm că Ion a strănutat, răceala și alergia sunt independente
  - Dacă știm că Ion a strănutat, răceala și alergia nu mai sunt independente
  - Creșterea probabilității răcelii scade probabilitatea alergiei și viceversa
  - Situație numită revocare prin explicare (*explaining away*)

Distribuție comună de probabilitate

- 3 variabile binare: $2^3 - 1 = 7$ parametri independenți

x	y	z	$p(x,y,z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

← 1 – suma celorlalți

Reprezentarea cunoștințelor incerte

- O situație cu 5 variabile binare (exemplul următor)
  - Specifică o distribuție comună de probabilitate cu $2^5 - 1 = 31$ parametri
    - Fezabil
- Un sistem expert cu 37 de variabile binare pentru monitorizarea pacienților de la terapie intensivă
  - $2^{37} - 1 \approx 10^{11}$ parametri
    - Nefezabil

Reprezentarea distribuției comune de probabilitate

$$P(x_1, \ldots, x_n) = P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1)$$

$$\begin{aligned} P(x_1, \ldots, x_n) &= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1} | x_{n-2}, \ldots, x_1) \cdots P(x_2 | x_1) P(x_1) \\ &= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1) . \end{aligned}$$

Regula de înmulțire a probabilităților (*chain rule*)

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

Este adevărată doar dacă fiecare nod este independent condiționat de predecesorii din șirul ordonat al nodurilor, dați fiind părinții nodului

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i | Cause)$$

Dacă efectele sunt considerate independente ⇒ *Naïve Bayes*

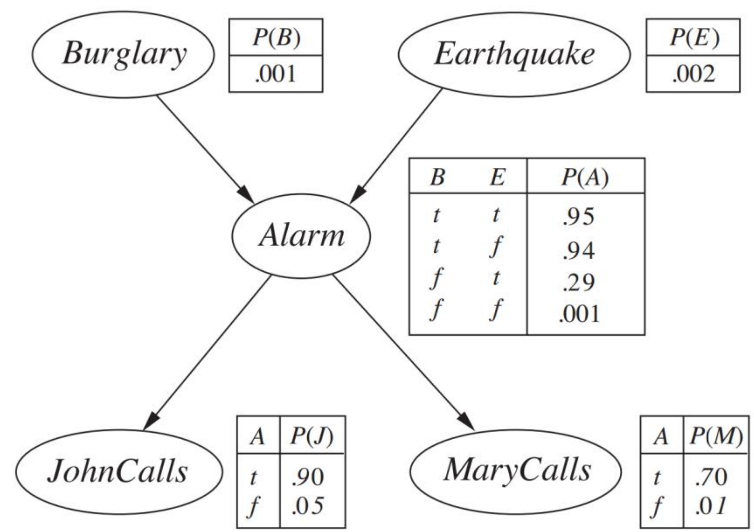
Rețele bayesiene

Rețea bayesiană

- O rețea bayesiană este un model grafic probabilistic, adică un graf cu o mulțime de noduri, care reprezintă evenimente aleatorii, conectate de arce, care reprezintă dependențe condiționate între evenimente

Exemplu

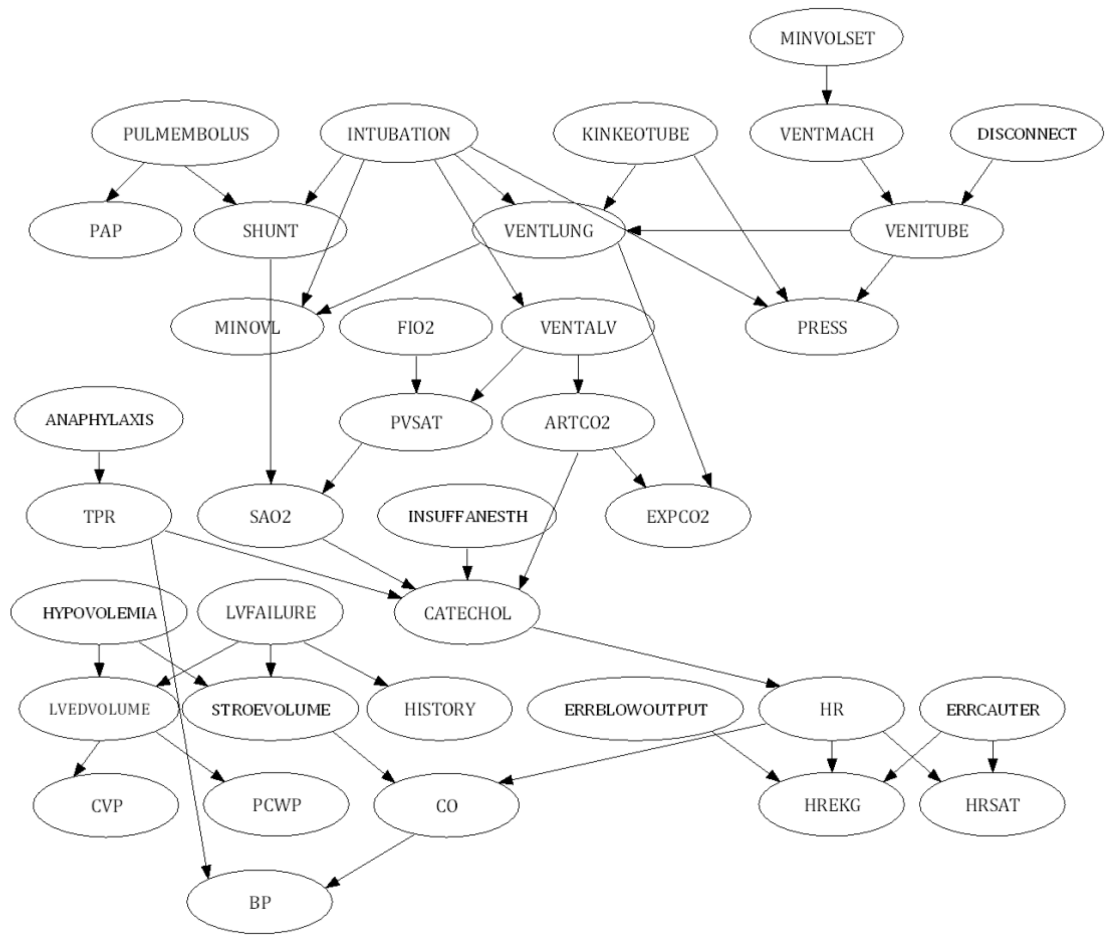
- S-a instalat un nou sistem de alarmă, care sună în cazul unei spargeri, dar și în cazul unui cutremur
- Vecinii John și Mary îl sună pe proprietar la serviciu dacă aud alarma
- 10 parametri independenți față de 31



Rețea bayesiană

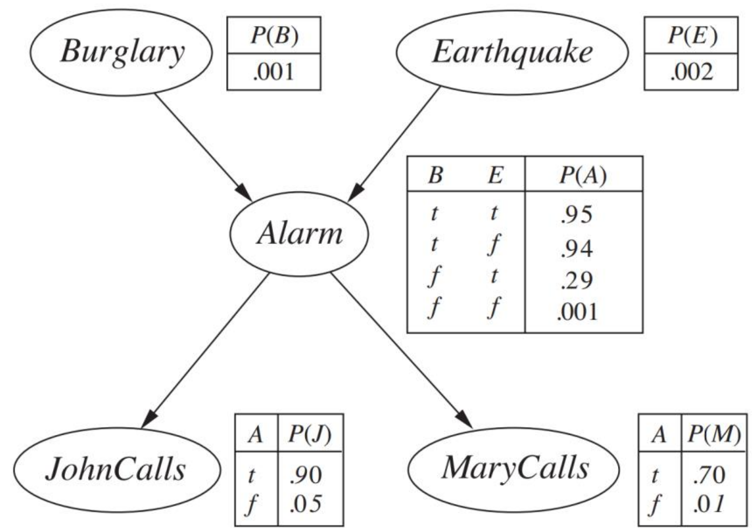
Comparație

- Sistem expert pentru monitorizarea pacienților de la terapie intensivă
- 37 variabile
- 509 parametri în loc de 10^{11} (100.000.000.000)



Interogări simple

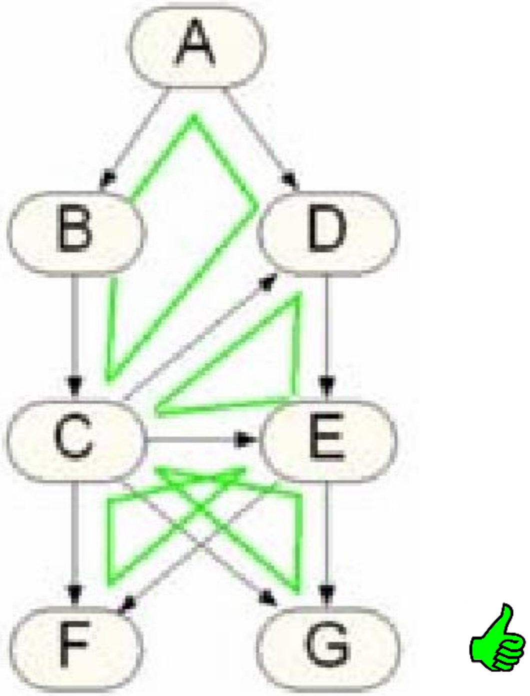
- *Care este probabilitatea ca alarma să se declanșeze fără să fi fost nicio spargere și niciun cutremur, iar John și Mary să sune?*

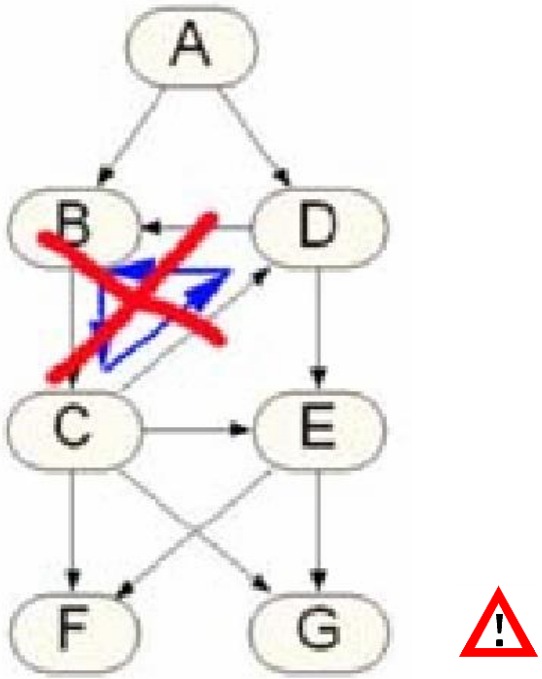


$$\begin{aligned} P(j, m, a, \neg b, \neg e) &= P(j \mid a) P(m \mid a) P(a \mid \neg b \wedge \neg e) P(\neg b) P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628 \end{aligned}$$

Validitatea unei rețele bayesiene

- O rețea bayesiană este un graf orientat aciclic
- Arcele pot forma bucle, dar nu pot forma cicluri

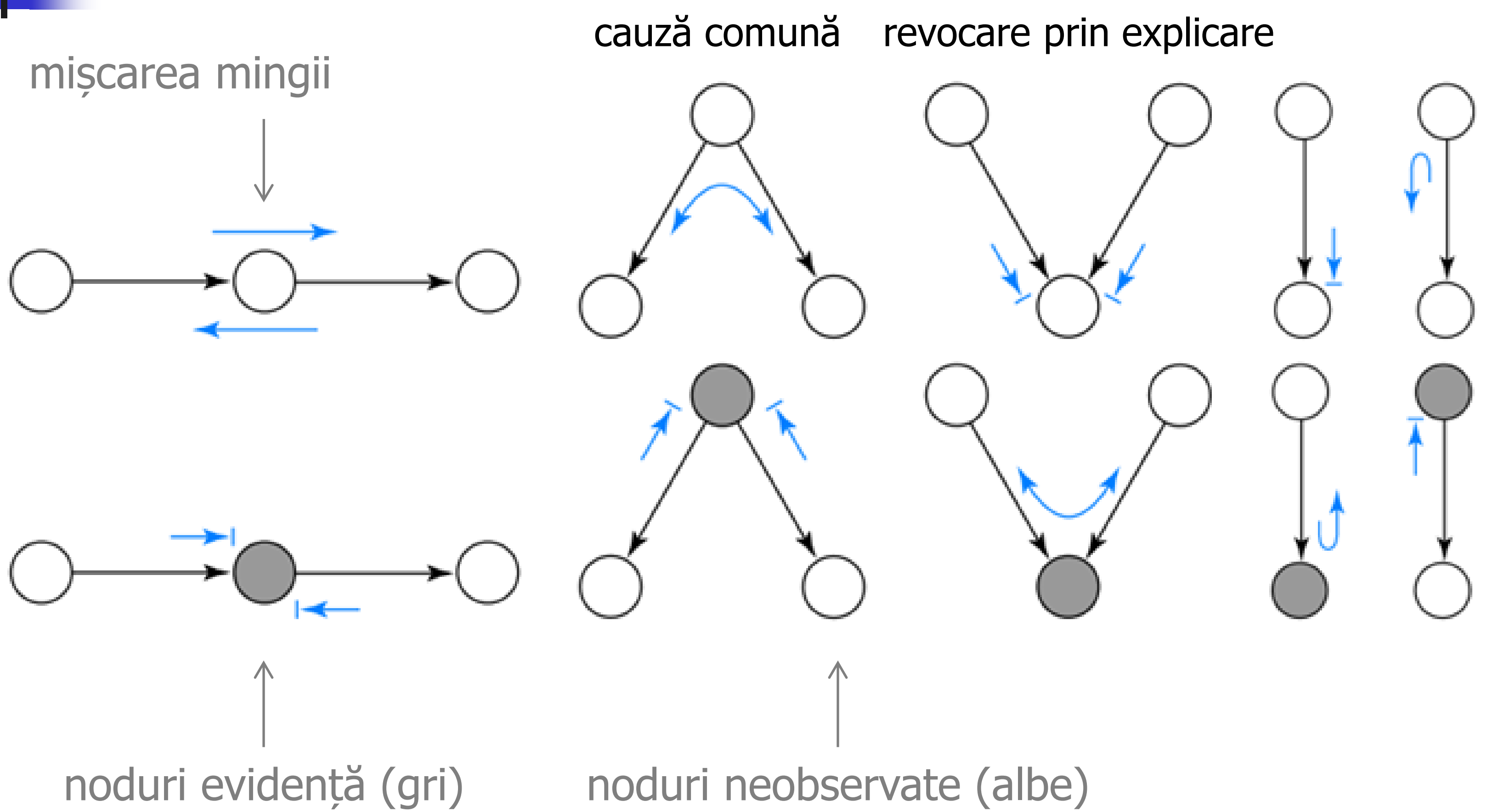




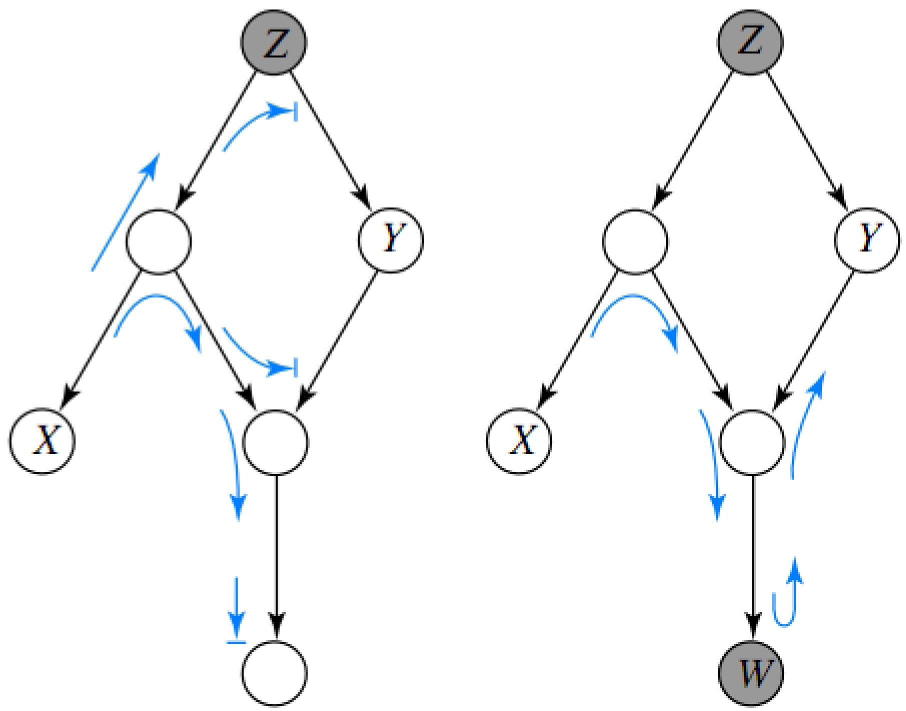
Algoritmul Bayes-Ball

- Reprezintă o modalitate simplă de a determina relațiile de independență și independență condiționată într-o rețea bayesiană
- Se presupune că o minge este trimisă dintr-un nod în rețea
- Mingea trece în moduri diferite, în funcție de cine o trimite (fiu sau părinte) și starea nodului care o primește (observat/evidență sau neobservat)
- Nodurile la care mingea nu ajunge sunt independente (condiționat) de nodul de start

Reguli de trimitere a mingii



Exemple



nicio cale activă

$X \perp\!\!\!\perp Y \mid Z$

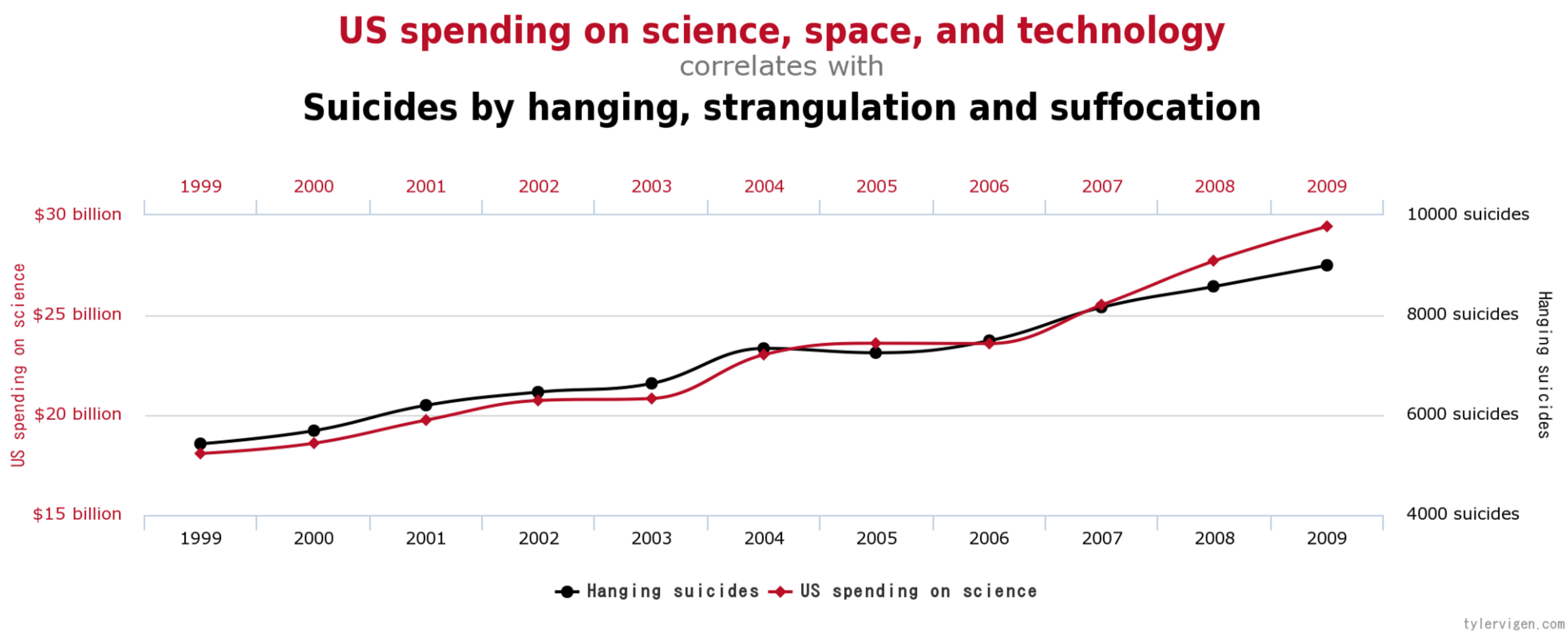
o cale activă

$X \not\perp\!\!\!\perp Y \mid \{W, Z\}$

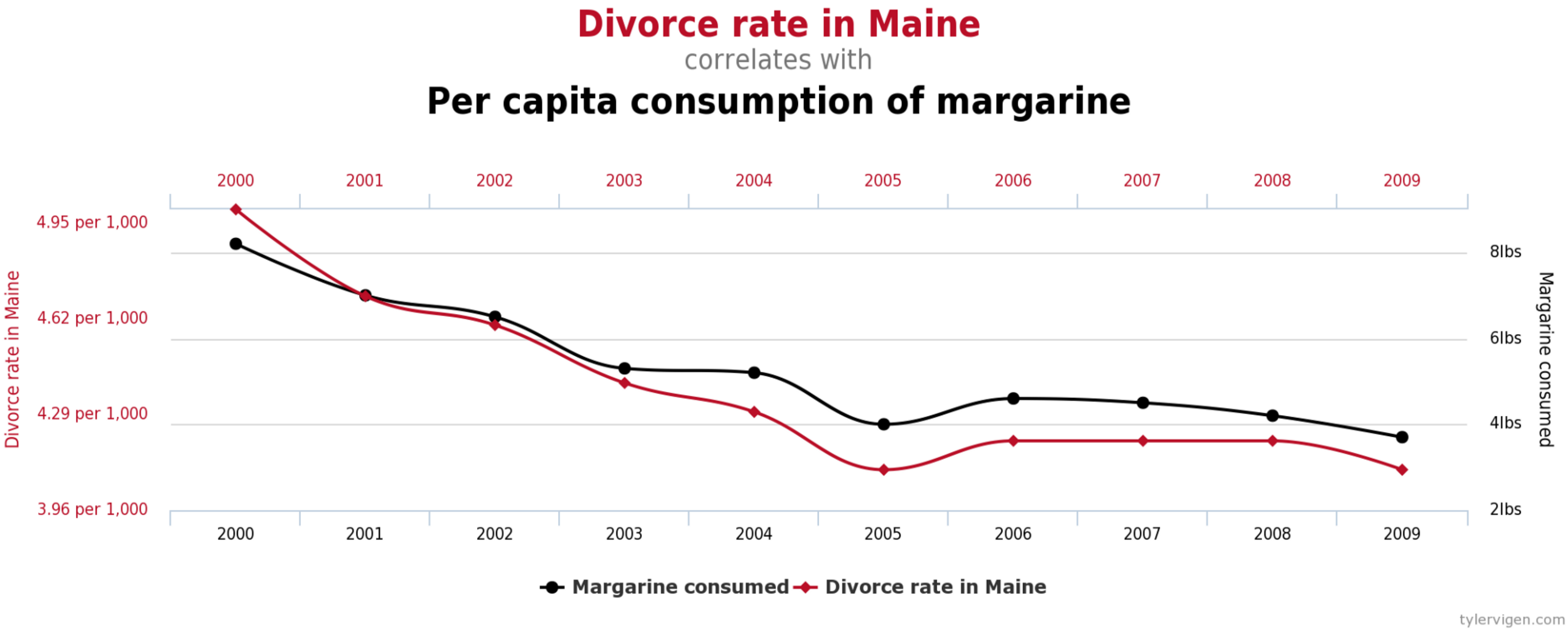
Corelație și cauzalitate

- Rețelele bayesiene codează relații de corelație
  - Probabilitatea unui eveniment *A* variază împreună cu probabilitățile altor evenimente {*B*, *C*,...}
  - *A* este corelat cu {*B*, *C*,...}
  - *A* poate fi cauzat sau nu de {*B*, *C*,...}
  - Toate evenimentele {*A*, *B*, *C*,...} ar putea fi cauzate de alte evenimente necunoscute
- Corelația $\nRightarrow$ cauzalitatea
- Cauzalitatea $\Rightarrow$ corelația

Corelații bizare



Corelații bizare



Rețele bayesiene

Sortarea topologică

- Sortarea topologică a unui graf este o ordonare liniară a nodurilor sale astfel încât, pentru fiecare arc $A \rightarrow B$, A apare înaintea lui B
- Pentru o rețea bayesiană, sortarea topologică asigură faptul că nodurile părinte vor apărea înaintea nodurilor fiu
- Algoritmii au de obicei o complexitate de timp liniară în numărul de noduri n și de arce a: $O(n + a)$
- Există mai mulți algoritmi de sortare topologică
- Dacă graful este orientat aciclic, există cel puțin o soluție
- Dacă există cicluri în graf, sortarea topologică este imposibilă

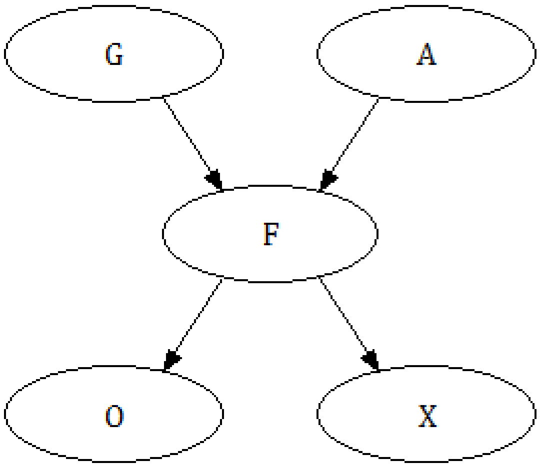
Algoritmul lui Kahn

```
L ← listă inițial vidă care va conține elementele sortate
S ← mulțimea nodurilor fără părinți

cât-timp S nu este vidă
   scoate un nod n din S
   introdu n în L
   pentru-fiecare nod m cu un arc e de la n la m
      scoate arcul e din graf
      dacă m nu are alte arce incidente atunci
         introdu m în S

dacă graful mai are arce atunci
   întoarce eroare (graful are cel puțin un ciclu)
altfel
   întoarce L (elementele sortate topologic)
```

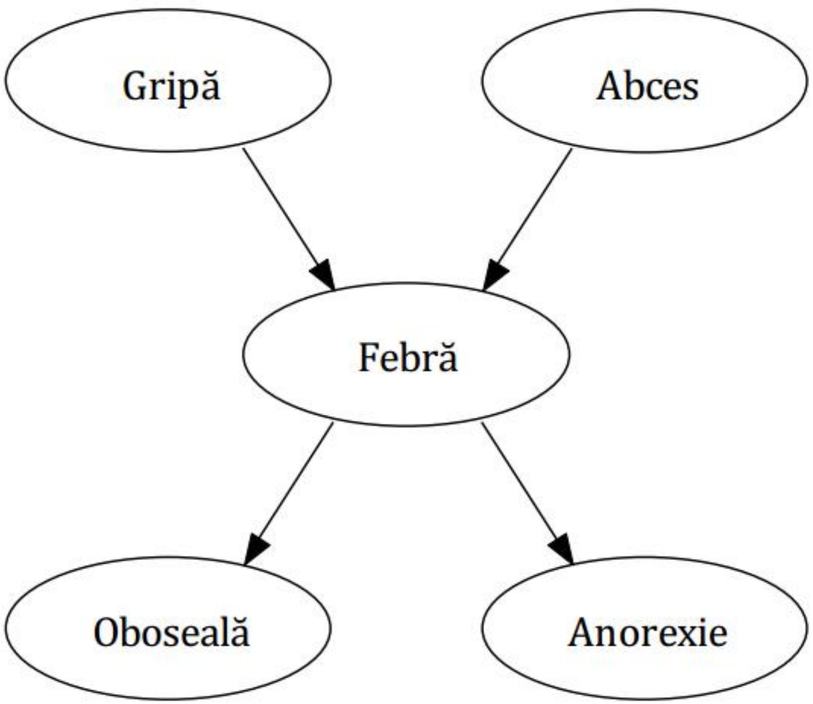
Exemplu



1. $L = \emptyset$, $S = \{G, A\}$
2. $L = \{G\}$, $S = \{A\}$
3. se elimină arcul *GF*, *F* nu poate fi adăugat în *S* pentru că mai există arcul *AF*
4. $L = \{G, A\}$, $S = \emptyset$
5. se elimină arcul *AF*, $S = \{F\}$
6. $L = \{G, A, F\}$, $S = \emptyset$
7. se elimină arcul *FO*, $S = \{O\}$
8. se elimină arcul *FX*, $S = \{O, X\}$
9. $L = \{G, A, F, O\}$, $S = \{X\}$
10. $L = \{G, A, F, O, X\}$, $S = \emptyset$

Soluția este deci: $\{G, A, F, O, X\}$.

Inferența probabilităților marginale



P(Gripă = Da)	P(Gripă = Nu)
0,1	0,9

P(Abces = Da)	P(Abces = Nu)
0,05	0,95

Gripă	Abces	P(Febră = Da)	P(Febră = Nu)
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	P(Oboseală = Da)	P(Oboseală = Nu)
Da	0,6	0,4
Nu	0,2	0,8

Febră	P(Anorexie = Da)	P(Anorexie = Nu)
Da	0,5	0,5
Nu	0,1	0,9

Nodul *Febră*

F_D = (*Febră* = *Da*)
F_N = (*Febră* = *Nu*)

Într-o rețea bayesiană, un nod poate avea oricâte valori posibile, nu doar două

$$P(F_D) =$$

$$P(F_D|G_D,A_D) \cdot P(G_D) \cdot P(A_D) +$$

$$P(F_D|G_D,A_N) \cdot P(G_D) \cdot P(A_N) +$$

$$P(F_D|G_N,A_D) \cdot P(G_N) \cdot P(A_D) +$$

$$P(F_D|G_N,A_N) \cdot P(G_N) \cdot P(A_N) =$$

$$0{,}8 \cdot 0{,}1 \cdot 0{,}05 + 0{,}7 \cdot 0{,}1 \cdot 0{,}95 + 0{,}25 \cdot 0{,}9 \cdot 0{,}05 + 0{,}05 \cdot 0{,}9 \cdot 0{,}95 = 0{,}1245 \approx 12\%.$$

$$P(F_N) = 1 - P(F_D) = 0{,}8755 \approx 88\%.$$

Nodul *Oboseală*

$$P(O_D) =$$

$$P(O_D|F_D) \cdot P(F_D) + P(O_D|F_N) \cdot P(F_N) =$$

$$0{,}6 \cdot 0{,}1245 + 0{,}2 \cdot 0{,}8755 = 0{,}2498 \approx 25\%,$$

$$P(O_N) = 1 - P(O_D) = 0{,}7502 \approx 75\%.$$

O_D = (*Oboseală* = *Da*)
O_N = (*Oboseală* = *Nu*)

Nodul *Anorexie*

$$P(X_D) =$$

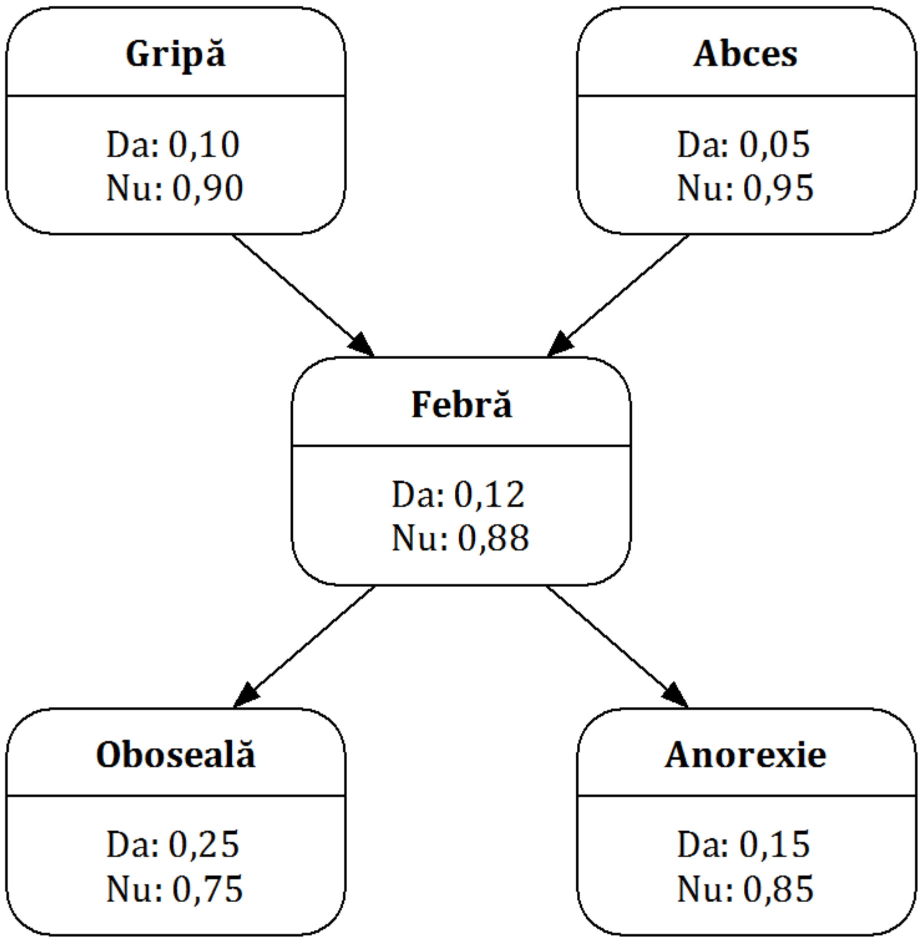
$$P(X_D|F_D) \cdot P(F_D) + P(X_D|F_N) \cdot P(F_N) =$$

$$0{,}5 \cdot 0{,}1245 + 0{,}1 \cdot 0{,}8755 = 0{,}1498 \approx 15\%,$$

$$P(X_N) = 1 - P(X_D) = 0{,}8502 \approx 85\%.$$

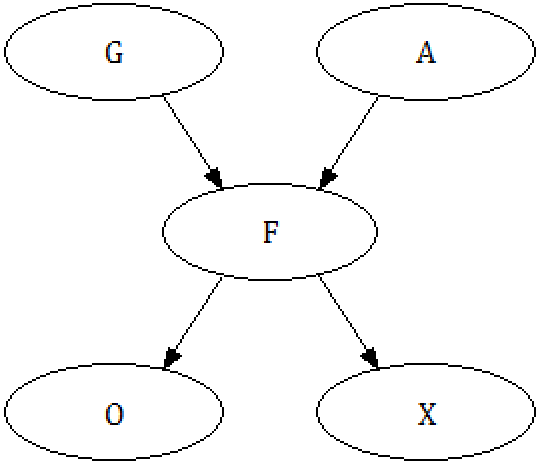
X_D = (*Anorexie* = *Da*)
X_N = (*Anorexie* = *Nu*)

Probabilitățile marginale ale nodurilor



Inferența prin enumerare

- Interogare: *Care este probabilitatea ca o persoană să aibă gripă, dacă prezintă simptome de oboseală și anorexie?*



Rezolvare

Coeficient de normalizare

$$\mathbf{P}(X|\mathbf{e}) = \alpha\,\mathbf{P}(X,\mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X,\mathbf{e},\mathbf{y})$$

Variabila interogată

Evidențele
(variabilele observate)

Variabilele neobservate

Sumă după toate valorile posibile ale variabilelor neobservate **y**, de exemplu, afirmat și negat

Rezolvare

- Vom calcula independent $P(G_D|O_D, X_D)$ și $P(G_N|O_D, X_D)$
- Pentru $P(G_D|O_D, X_D)$, variabilele rămase sunt *Abcesul* și *Febra*
- Vom suma probabilitățile corespunzătoare tuturor valorilor acestor variabile: $a \in \{A_D, A_N\}$ și $f \in \{F_D, F_N\}$
- Pentru a crește eficiența calculelor, se recomandă ca variabilele rămase să fie mai întâi sortate topologic, astfel încât părinții să apară înaintea copiilor
- În acest caz, se vor putea descompune mai ușor sumele, scoțând în față factorii care nu depind de o anumită variabilă

Rezolvare

$$P(G_D|O_D,X_D) =$$

$$\alpha \cdot \sum_{a\in\{A_D,A_N\}} \sum_{f\in\{F_D,F_N\}} P(G_D,a,f,O_D,X_D) =$$

$$\alpha \cdot \sum_a \sum_f P(G_D)\cdot P(a)\cdot P(f|G_D,a)\cdot P(O_D|f)\cdot P(X_D|f) =$$

$$\alpha \cdot P(G_D)\cdot \sum_a P(a)\cdot \sum_f P(f|G_D,a)\cdot P(O_D|f)\cdot P(X_D|f) =$$

$$\alpha \cdot P(G_D)\cdot \sum_a P(a)\cdot [P(F_D|G_D,a)\cdot P(O_D|F_D)\cdot P(X_D|F_D) +$$

$$P(F_N|G_D,a)\cdot P(O_D|F_N)\cdot P(X_D|F_N)] =$$

$$\alpha \cdot P(G_D)\cdot \{P(A_D)\cdot [P(F_D|G_D,A_D)\cdot P(O_D|F_D)\cdot P(X_D|F_D) +$$

$$P(F_N|G_D,A_D)\cdot P(O_D|F_N)\cdot P(X_D|F_N)] +$$

$$P(A_N)\cdot [P(F_D|G_D,A_N)\cdot P(O_D|F_D)\cdot P(X_D|F_D) +$$

$$P(F_N|G_D,A_N)\cdot P(O_D|F_N)\cdot P(X_D|F_N)]\} =$$

$$\alpha \cdot 0{,}1\cdot \{0{,}05\cdot [0{,}8\cdot 0{,}6\cdot 0{,}5 + 0{,}2\cdot 0{,}2\cdot 0{,}1] +$$

$$0{,}95\cdot [0{,}7\cdot 0{,}6\cdot 0{,}5 + 0{,}3\cdot 0{,}2\cdot 0{,}1]\} =$$

$$\alpha \cdot 0{,}02174.$$

P(Gripă = Da)	*P(Gripă = Nu)*
0,1	0,9

P(Abces = Da)	*P(Abces = Nu)*
0,05	0,95

Gripă	*Abces*	*P(Febră = Da)*	*P(Febră = Nu)*
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	*P(Oboseală = Da)*	*P(Oboseală = Nu)*
Da	0,6	0,4
Nu	0,2	0,8

Febră	*P(Anorexie = Da)*	*P(Anorexie = Nu)*
Da	0,5	0,5
Nu	0,1	0,9

Rezolvare

$$P(G_N|O_D,X_D) =$$

$$\alpha \cdot \sum_{a\in\{A_D,A_N\}} \sum_{f\in\{F_D,F_N\}} P(G_N,a,f,O_D,X_D) =$$

$$\alpha \cdot \sum_a \sum_f P(G_N)\cdot P(a)\cdot P(f|G_N,a)\cdot P(O_D|f)\cdot P(X_D|f) =$$

$$\alpha \cdot P(G_N)\cdot \sum_a P(a)\cdot \sum_f P(f|G_N,a)\cdot P(O_D|f)\cdot P(X_D|f) =$$

$$\alpha \cdot P(G_N)\cdot \{P(A_D)\cdot [P(F_D|G_N,A_D)\cdot P(O_D|F_D)\cdot P(X_D|F_D) +$$
$$P(F_N|G_N,A_D)\cdot P(O_D|F_N)\cdot P(X_D|F_N)] +$$
$$P(A_N)\cdot [P(F_D|G_N,A_N)\cdot P(O_D|F_D)\cdot P(X_D|F_D) +$$
$$P(F_N|G_N,A_N)\cdot P(O_D|F_N)\cdot P(X_D|F_N)]\} =$$

$$\alpha \cdot 0{,}9\cdot \{0{,}05\cdot [0{,}25\cdot 0{,}6\cdot 0{,}5 + 0{,}75\cdot 0{,}2\cdot 0{,}1] +$$
$$0{,}95\cdot [0{,}05\cdot 0{,}6\cdot 0{,}5 + 0{,}95\cdot 0{,}2\cdot 0{,}1]\} =$$

$$\alpha \cdot 0{,}03312.$$

P(Gripă = Da)	*P(Gripă = Nu)*
0,1	0,9

P(Abces = Da)	*P(Abces = Nu)*
0,05	0,95

Gripă	*Abces*	*P(Febră = Da)*	*P(Febră = Nu)*
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	*P(Oboseală = Da)*	*P(Oboseală = Nu)*
Da	0,6	0,4
Nu	0,2	0,8

Febră	*P(Anorexie = Da)*	*P(Anorexie = Nu)*
Da	0,5	0,5
Nu	0,1	0,9

Rezultat

- $P(G_D|O_D, X_D) + P(G_N|O_D, X_D) = 1$
- $P(G_D|O_D, X_D) = \alpha \cdot 0{,}02174$
- $P(G_N|O_D, X_D) = \alpha \cdot 0{,}03312$
- $\Rightarrow \alpha = 18{,}23$

- $P(G_D|O_D, X_D) = 0{,}39628 \approx 40\%$
- $P(G_N|O_D, X_D) = 0{,}60372 \approx 60\%$

Pseudocod

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for some set of variables **E**
    bn, a Bayes net
  Q ← a distribution over X, where $\mathbf{Q}(x_i)$ is $P(X=x_i)$
  for each value x_i that X can have **do**
    $\mathbf{Q}(x_i)$ ← ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$), where $\mathbf{e}_{x_i}$ is the evidence **e** plus the assignment $X=x_i$
  return NORMALIZE(**Q**)

function ENUMERATE-ALL($vars$, **e**) **returns** a probability (a real number in [0,1])
  inputs: $vars$, a list of all the variables
    e, observed values for some set of variables **E**
  if EMPTY($vars$) **then return** 1.0
  Y ← FIRST($vars$)
  if Y is assigned a value (call it y) in **e** **then**
    return $P(Y=y \mid$ values assigned to Y's parents in **e**$)$ × ENUMERATE-ALL(REST($vars$), **e**)
  else
    return $\sum_{y_i}$ [$P(Y=y_i \mid$ values assigned to Y's parents in **e**$)$ × ENUMERATE-ALL(REST($vars$), $\mathbf{e}_{y_i}$)],
      where $\mathbf{e}_{y_i}$ is the evidence **e** plus the assignment $Y=y_i$

Inferența aproximativă

- Pentru probleme din „lumea reală” au fost construite rețele bayesiene cu sute de noduri, pentru care algoritmii exacți își ating limitele, deoarece inferența este o problemă NP-dificilă
- Pentru rețele foarte complexe, inferența aproximativă este singura posibilitate de a obține un rezultat
- Pentru alte probleme în care precizia nu este un factor critic, inferența aproximativă poate crește foarte mult viteza de calcul

Inferența stohastică prin ponderarea verosimilității

- engl. “likelihood weighting”
- Se generează aleatoriu eșantionări / instanțieri ale rețelei și se calculează probabilitățile dorite ca frecvențe relative de apariție
- Nodurile fără părinți vor fi instanțiate potrivit probabilităților lor marginale
- Nodurile de evidență iau mereu valorile observate
- Valorile variabilelor neobservate au probabilități de apariție în conformitate cu probabilitățile nodurilor

Modul de lucru

- Pentru fiecare eșantionare a rețelei *s*, se calculează o pondere:

$$w(s) = \frac{\prod_{x \in U} P\left(x \middle| \pi(x)\right)}{\prod_{x \in U \setminus E} P\left(x \middle| \pi(x)\right)}$$

- unde *U* este mulțimea tuturor nodurilor, *E* este mulțimea nodurilor evidență, iar $\pi(x)$ reprezintă părinții nodului *x*
- Se repetă procesul pentru un număr prestabilit de eșantioane

Modul de lucru

- În final, are loc o fază de normalizare, în care se calculează suma ponderilor cazurilor în care o variabilă de interogare a avut o anumită valoare, împărțită la suma ponderilor tuturor cazurilor:

$$P(Var = val) = \frac{w_T}{w_S} = \frac{\sum_{s \in T} w(s)}{\sum_{s \in S} w(s)}$$

- unde *S* este mulțimea tuturor eșantioanelor, iar $T \subseteq S$ este submulțimea de eșantioane în care variabila *Var* apare cu valoarea *val*

Exemplu: aceeași problemă

- Variabila de interogare (*query*): G
- Variabile de evidență: O_D, X_D
- Variabile neobservate: A, F
- Sortarea topologică: G, A, F, O, X
- Variabilele primesc valori în mod aleatoriu, după probabilitățile din tabele
- Notație: $A \sim M(Da : 0.1, Nu : 0.9)$ înseamnă că variabila A va primi valoarea *Da* cu probabilitatea de 10% și *Nu* cu probabilitatea de 90%
  - $M(\cdot)$ = distribuția *Multinoulli* / Bernoulli generalizată / categorială ~ *roulette wheel selection*

Eșantionarea 1

- $G \sim M(Da : 0.1, Nu : 0.9) \Rightarrow G_N$
- $A \sim M(Da : 0.05, Nu : 0.95) \Rightarrow A_D$
- $F \sim M(Da : 0.25, Nu : 0.75) \Rightarrow F_N$
- O_D
- X_D
- $s_1 = (G_N, A_D, F_N, O_D, X_D)$

P(Gripă = Da)	P(Gripă = Nu)
0,1	0,9

P(Abces = Da)	P(Abces = Nu)
0,05	0,95

Gripă	Abces	P(Febră = Da)	P(Febră = Nu)
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	P(Oboseală = Da)	P(Oboseală = Nu)
Da	0,6	0,4
Nu	0,2	0,8

Febră	P(Anorexie = Da)	P(Anorexie = Nu)
Da	0,5	0,5
Nu	0,1	0,9

$$w(s_1) = \frac{P(G_N) \cdot P(A_D) \cdot P(F_N|G_N, A_D) \cdot P(O_D|F_N) \cdot P(X_D|F_N)}{P(G_N) \cdot P(A_D) \cdot P(F_N|G_N, A_D)} =$$

$$P(O_D|F_N) \cdot P(X_D|F_N) = 0.2 \cdot 0.1 = 0.02$$

Eșantionarea 2

- $G \sim M(Da : 0.1, Nu : 0.9) \Rightarrow G_D$
- $A \sim M(Da : 0.05, Nu : 0.95) \Rightarrow A_N$
- $F \sim M(Da : 0.7, Nu : 0.3) \Rightarrow F_D$
- O_D
- X_D
- $s_2 = (G_D, A_N, F_D, O_D, X_D)$

P(Gripă = Da)	P(Gripă = Nu)
0,1	0,9

P(Abces = Da)	P(Abces = Nu)
0,05	0,95

Gripă	Abces	P(Febră = Da)	P(Febră = Nu)
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	P(Oboseală = Da)	P(Oboseală = Nu)
Da	0,6	0,4
Nu	0,2	0,8

Febră	P(Anorexie = Da)	P(Anorexie = Nu)
Da	0,5	0,5
Nu	0,1	0,9

$$w(s_2) = \frac{P(G_D) \cdot P(A_N) \cdot P(F_D|G_D, A_N) \cdot P(O_D|F_D) \cdot P(X_D|F_D)}{P(G_D) \cdot P(A_N) \cdot P(F_D|G_D, A_N)} =$$

$$P(O_D|F_D) \cdot P(X_D|F_D) = 0.6 \cdot 0.5 = 0.3$$

Calcularea probabilităților finale

$$P(G_D) = \frac{w(s_2)}{w(s_1) + w(s_2)} = \frac{0.3}{0.02 + 0.3} \approx 0.94$$

$$P(G_N) = \frac{w(s_1)}{w(s_1) + w(s_2)} = \frac{0.02}{0.02 + 0.3} \approx 0.06$$

- În acest exemplu, s-au folosit doar două eșantionări, prin urmare, este normal ca rezultatele să fie diferite de valorile reale: (0.4, 0.6)

Pseudocod

function LIKELIHOOD-WEIGHTING(X, **e**, bn, N) **returns** an estimate of $\mathbf{P}(X|\mathbf{e})$
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$
    N, the total number of samples to be generated
  local variables: **W**, a vector of weighted counts for each value of X, initially zero

  for j = 1 to N **do**
    x, $w \leftarrow$ WEIGHTED-SAMPLE(bn, **e**)
    $\mathbf{W}[x] \leftarrow \mathbf{W}[x] + w$ where x is the value of X in **x**
  return NORMALIZE(**W**)

---

function WEIGHTED-SAMPLE(bn, **e**) **returns** an event and a weight

  $w \leftarrow 1$; $\mathbf{x} \leftarrow$ an event with n elements initialized from **e**
  foreach variable X_i **in** $X_1, \ldots, X_n$ **do**
    if X_i is an evidence variable with value x_i in **e**
      then $w \leftarrow w \times P(X_i = x_i \mid parents(X_i))$
      else $\mathbf{x}[i] \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
  return **x**, w

Discuție

- Precizia de estimare a probabilităților crește cu numărul de eșantionări ale rețelei
- Algoritmul nu converge bine când există multe noduri evidență

Aplicații ale rețelelor bayesiene

- Sisteme pentru diagnostic medical, de exemplu, diagnosticarea bolilor ficatului
- Predicția defectelor și a fiabilității produselor software, de exemplu, *AgenaRisk*, utilizat de Motorola, Philips și Siemens
- *TrueSkill* (Microsoft Research): sistem de ierarhizare a jucătorilor pe baza rezultatelor finale obținute de echipă

Rețele bayesiene

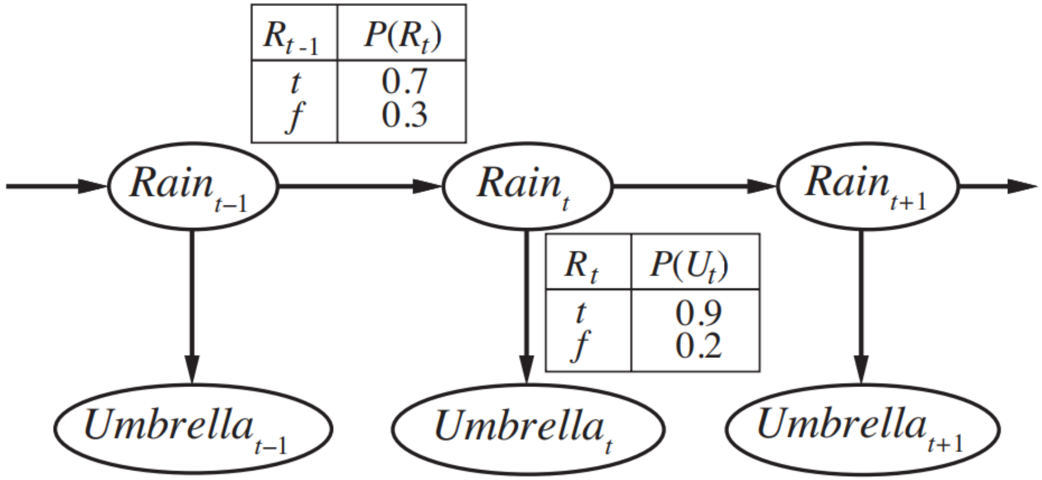
1. Probabilități
2. Rețele bayesiene
3. Inferențe exacte și aproximative
4. Rețele bayesiene dinamice
5. Filtrarea cu particule
6. Concluzii

Rețea bayesiană dinamică

- O rețea bayesiană dinamică este o rețea bayesiană care reprezintă un model de probabilitate temporal
- Este organizată în partiții (*slices*) de timp
- Fiecare partiție poate avea mai multe variabile de stare $\mathbf{X}_t$ și variabile de evidență $\mathbf{E}_t$
- Pentru a construi o RBD, trebuie specificate:
  - Distribuția a priori a variabilelor de stare $P(\mathbf{X}_0)$
  - Modelul de tranziții $P(\mathbf{X}_{t+1} \mid \mathbf{X}_t)$
  - Modelul de observații sau de senzori $P(\mathbf{E}_t \mid \mathbf{X}_t)$

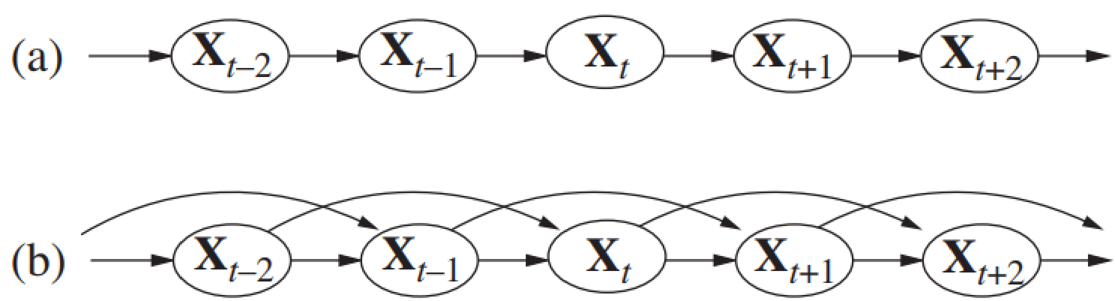
Exemplu

Consider the following example: You are the security guard stationed at a secret underground installation. You want to know whether it's raining today, but your only access to the outside world occurs each morning when you see the director coming in with, or without, an umbrella. For each day t, the set $\mathbf{E}_t$ thus contains a single evidence variable $Umbrella_t$ or U_t for short (whether the umbrella appears), and the set $\mathbf{X}_t$ contains a single state variable $Rain_t$ or R_t for short (whether it is raining).



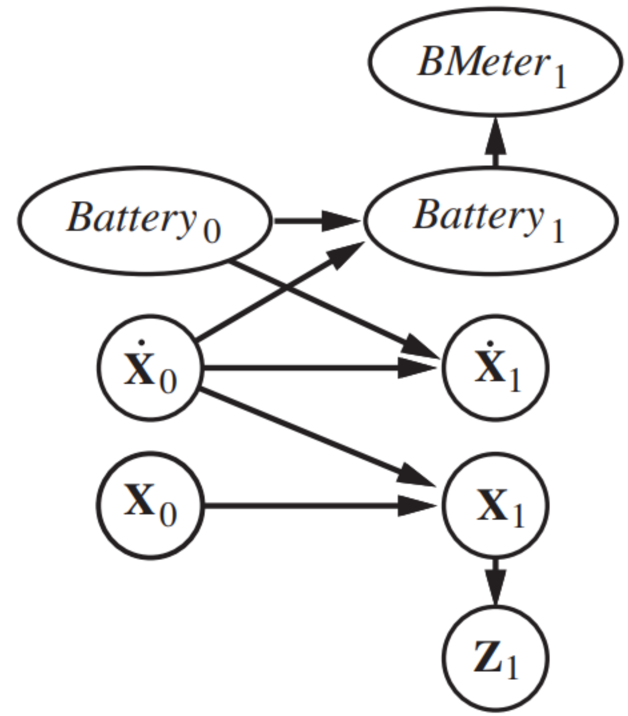
Modelarea proceselor Markov

- RBD pot modela procese Markov
- Presupunerea Markov: starea curentă depinde de un număr finit fix de stări anterioare
- În figură, se prezintă:
  - (a) un proces Markov de ordin I
  - (b) un proces Markov de ordin II



Modelarea tranzițiilor arbitrare

- Pot fi adăugate variabile care să reducă problema la un proces Markov de ordin întâi
- Exemplu: un robot cu poziția **x**, viteza $\dot{\mathbf{x}}$ și nivelul bateriei *Battery* care depinde de toate manevrele anterioare. Se adaugă o variabilă de observație (senzor) pentru baterie *BMeter*. **z** este variabila observată pentru poziție

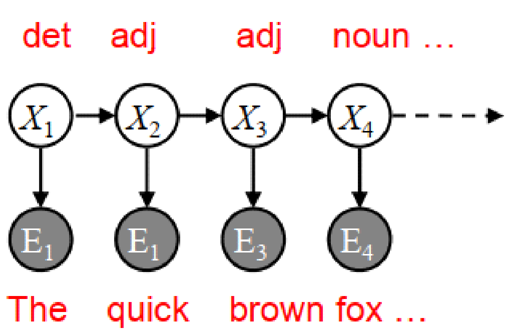


Alte presupuneri

- Schimbările din mediu sunt cauzate de un proces staționar (≠ static)
  - Modelul de tranziții $P(\mathbf{X}_{t+1} \mid \mathbf{X}_t)$ este același pentru orice t
- Senzorii urmează un proces Markov de ordin întâi
  - $P(\mathbf{E}_t \mid \mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}) = P(\mathbf{E}_t \mid \mathbf{X}_t)$
  - Observațiile curente depind doar de starea din momentul curent

Aplicații ale RBD

- Identificarea părților de vorbire
  - Observațiile: cuvintele (posibil mii)
  - Stările: etichetele părților de vorbire

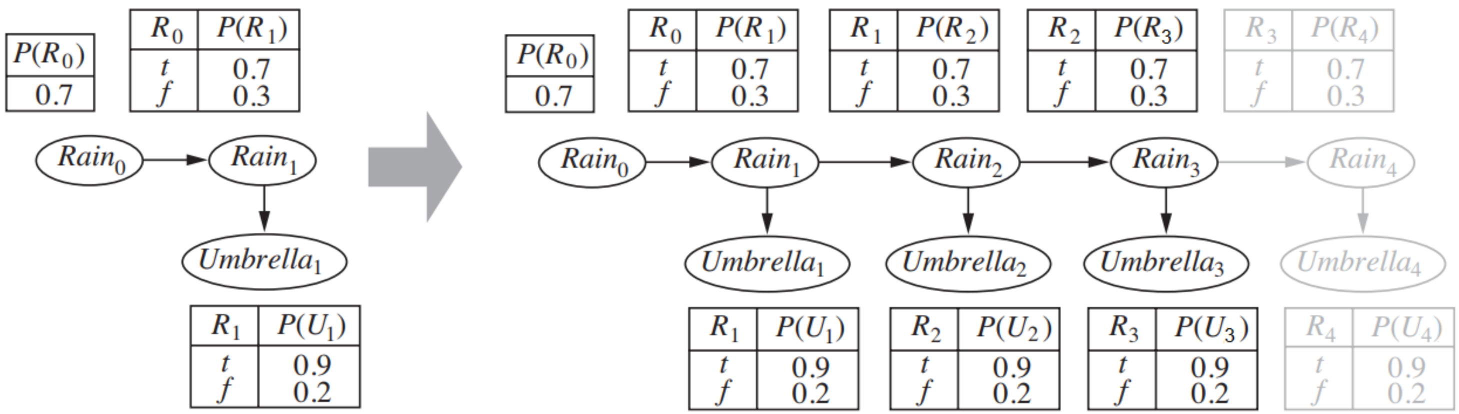


- Traducerea automată
  - Observațiile: cuvintele
  - Stările: opțiunile de traducere

- Urmărirea pozițiilor roboților sau mașinilor autonome (*tracking*)
  - Observațiile: informațiile de localizare de la senzori
  - Stările: pozițiile pe hartă

Inferența în RBD

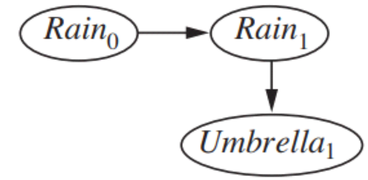
- O RBD se poate transforma într-o RB simplă prin desfășurare (*unrolling*)



- S-ar putea aplica algoritmii de inferență de la RB simple

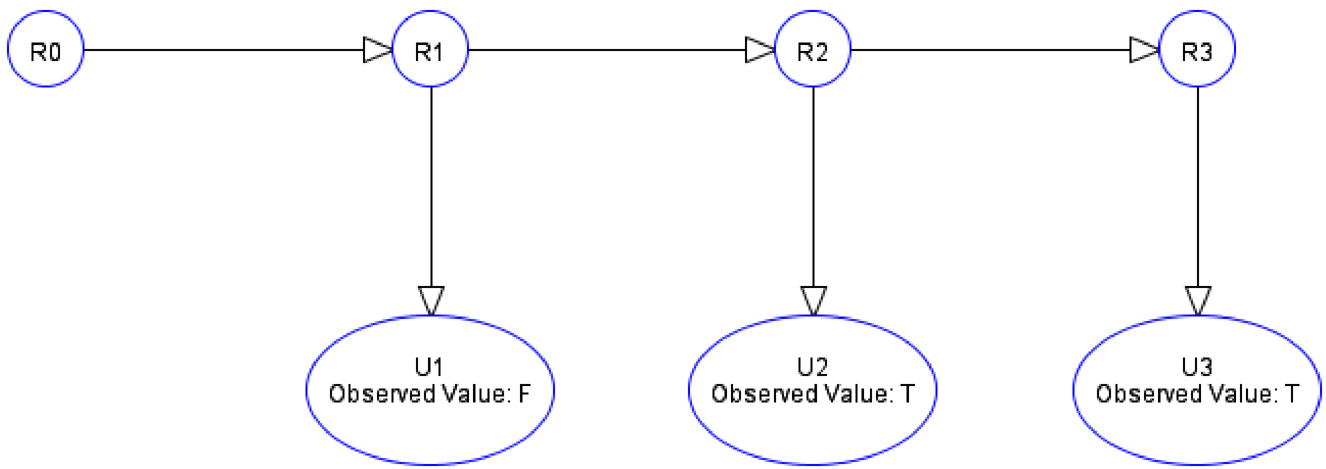
$P(R_0)$
0.7

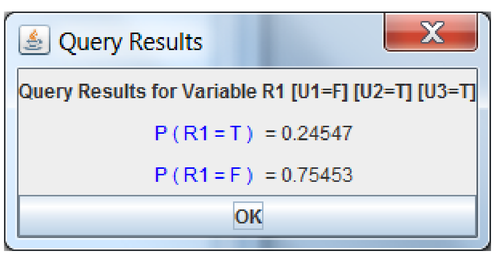
R_0	$P(R_1)$
t	0.7
f	0.3

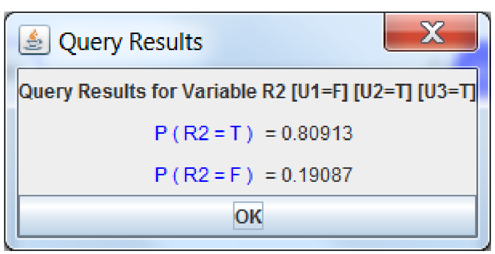


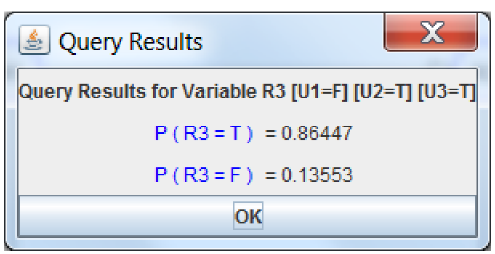
R_1	$P(U_1)$
t	0.9
f	0.2

Exemplu









Rețele bayesiene

1. Probabilități
2. Rețele bayesiene
3. Inferențe exacte și aproximative
4. Rețele bayesiene dinamice
5. Filtrarea cu particule
6. Concluzii

Filtrarea cu particule

- Complexitatea algoritmilor de inferență exactă specifici RB simple pentru RBD desfășurate este exponențială
- Există alte metode exacte și aproximative de inferență specifice RBD
- De exemplu, metoda aproximativă de filtrare cu particule (*particle filtering*)
- Filtrare înseamnă determinarea lui $P(\mathbf{X}_t \mid \mathbf{E}_{0:t})$ la momentul curent t

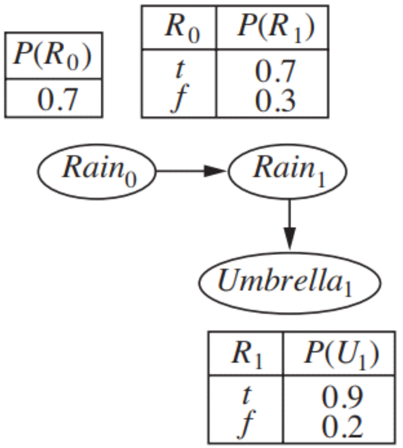
Filtrarea cu particule

- Mai întâi, se creează o populație de *n* particule cu starea eșantionată din distribuția a priori $P(\mathbf{X}_0)$
- Apoi, se repetă următoarele faze pentru fiecare moment de timp:
  - Fiecare eșantion (stare a particulei) este propagat înainte prin eșantionarea următoarei stări pe baza modelului de tranziție $P(\mathbf{X}_{t+1} \mid \mathbf{X}_t)$
  - Fiecare particulă este ponderată de probabilitatea pe care o atribuie noilor evidențe $P(\mathbf{E}_{t+1} \mid \mathbf{X}_{t+1})$
  - Populația este reeșantionată pentru a genera o nouă populație de *n* particule. Fiecare nouă particulă este selectată din populația curentă. Probabilitatea de a fi selectată o particulă este proporțională cu ponderea sa. Noile particule nu au ponderi

Exemplul cu ploaia și umbrela
Inițializarea

- $t = 0$

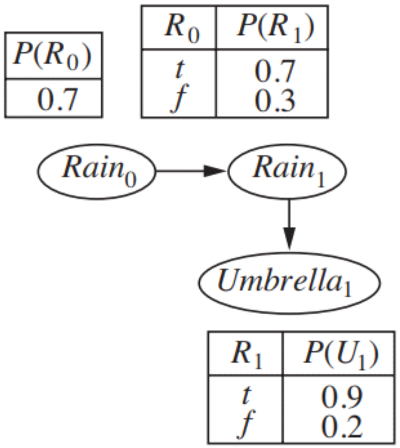
- $p_1 \sim M(T : 0.7, F : 0.3) \Rightarrow R_T$
- $p_2 \sim M(T : 0.7, F : 0.3) \Rightarrow R_F$
- $p_3 \sim M(T : 0.7, F : 0.3) \Rightarrow R_T$
- $p_4 \sim M(T : 0.7, F : 0.3) \Rightarrow R_F$



Faza 1. Propagarea

- $t = 1$

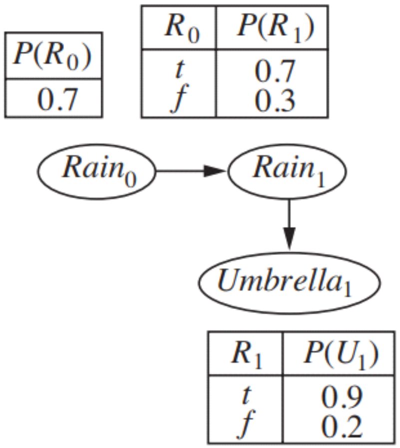
- $p_1\,(R_T) \sim M(T : 0.7,\ F : 0.3) \Rightarrow R_T$
- $p_2\,(R_F) \sim M(T : 0.3,\ F : 0.7) \Rightarrow R_F$
- $p_3\,(R_T) \sim M(T : 0.7,\ F : 0.3) \Rightarrow R_T$
- $p_4\,(R_F) \sim M(T : 0.3,\ F : 0.7) \Rightarrow R_T$



- La momentul $t = 1$, apare evidența U_F

Faza 2. Ponderarea

- $w_1 = P(U_F \mid R_T) = 0.1$
- $w_2 = P(U_F \mid R_F) = 0.8$
- $w_3 = P(U_F \mid R_T) = 0.1$
- $w_4 = P(U_F \mid R_T) = 0.1$



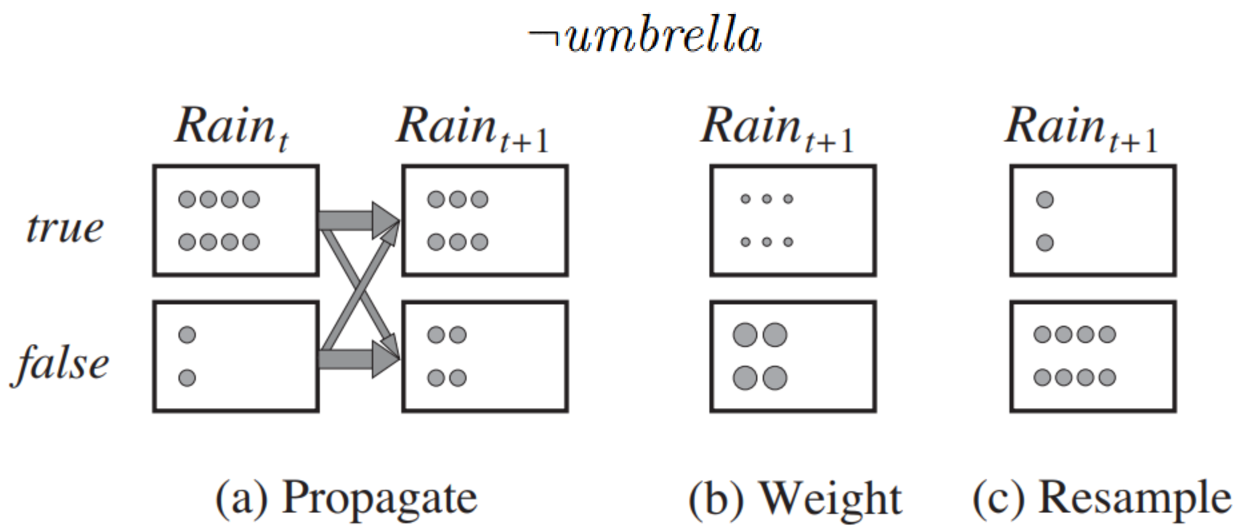
Faza 3. Reeșantionarea

- $p_i \sim M(T : 0.1, F : 0.8, T : 0.1, T : 0.1)$

- $p_1 \Rightarrow R_F$
- $p_2 \Rightarrow R_F$
- $p_3 \Rightarrow R_T$
- $p_4 \Rightarrow R_F$

- $\Rightarrow P(R_T) = 25\%, P(R_F) = 75\%$

Exemplul cu ploaia și umbrela



The particle filtering update cycle for the umbrella DBN with $N = 10$, showing the sample populations of each state. (a) At time t, 8 samples indicate $rain$ and 2 indicate $\neg rain$. Each is propagated forward by sampling the next state through the transition model. At time $t+1$, 6 samples indicate $rain$ and 4 indicate $\neg rain$. (b) $\neg umbrella$ is observed at $t+1$. Each sample is weighted by its likelihood for the observation, as indicated by the size of the circles. (c) A new set of 10 samples is generated by weighted random selection from the current set, resulting in 2 samples that indicate $rain$ and 8 that indicate $\neg rain$.

Pseudocod

function PARTICLE-FILTERING($\mathbf{e}$, N, dbn) **returns** a set of samples for the next time step
  inputs: $\mathbf{e}$, the new incoming evidence
    N, the number of samples to be maintained
    dbn, a DBN with prior $\mathbf{P}(\mathbf{X}_0)$, transition model $\mathbf{P}(\mathbf{X}_1|\mathbf{X}_0)$, sensor model $\mathbf{P}(\mathbf{E}_1|\mathbf{X}_1)$
  persistent: S, a vector of samples of size N, initially generated from $\mathbf{P}(\mathbf{X}_0)$
  local variables: W, a vector of weights of size N

  for i = 1 to N **do**
    $S[i] \leftarrow$ sample from $\mathbf{P}(\mathbf{X}_1 \mid \mathbf{X}_0 = S[i])$ /* step 1 */
    $W[i] \leftarrow \mathbf{P}(\mathbf{e} \mid \mathbf{X}_1 = S[i])$ /* step 2 */
  $S \leftarrow$ WEIGHTED-SAMPLE-WITH-REPLACEMENT(N, S, W) /* step 3 */
  return S

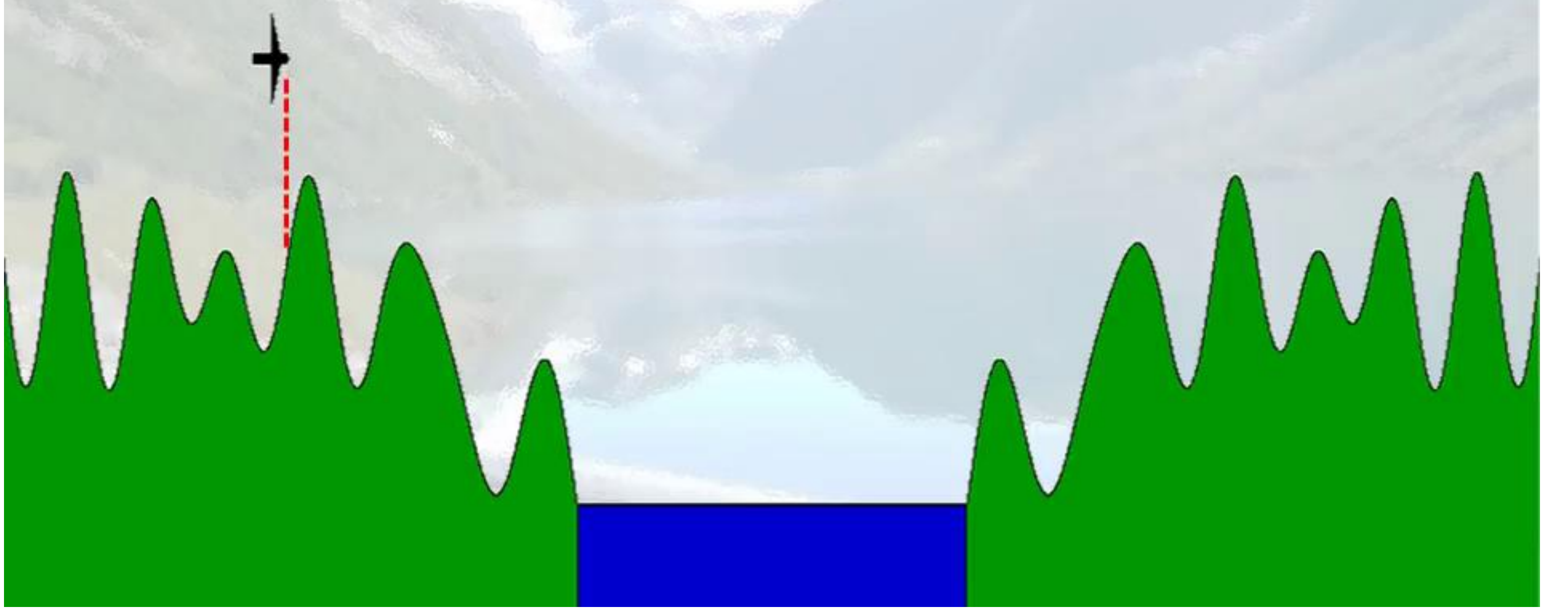
UPPSALA UNIVERSITET

So, this is our aicraft.

- We can measure the elevation (relative the sea)
- We can measure the distance to the ground
- We have the map

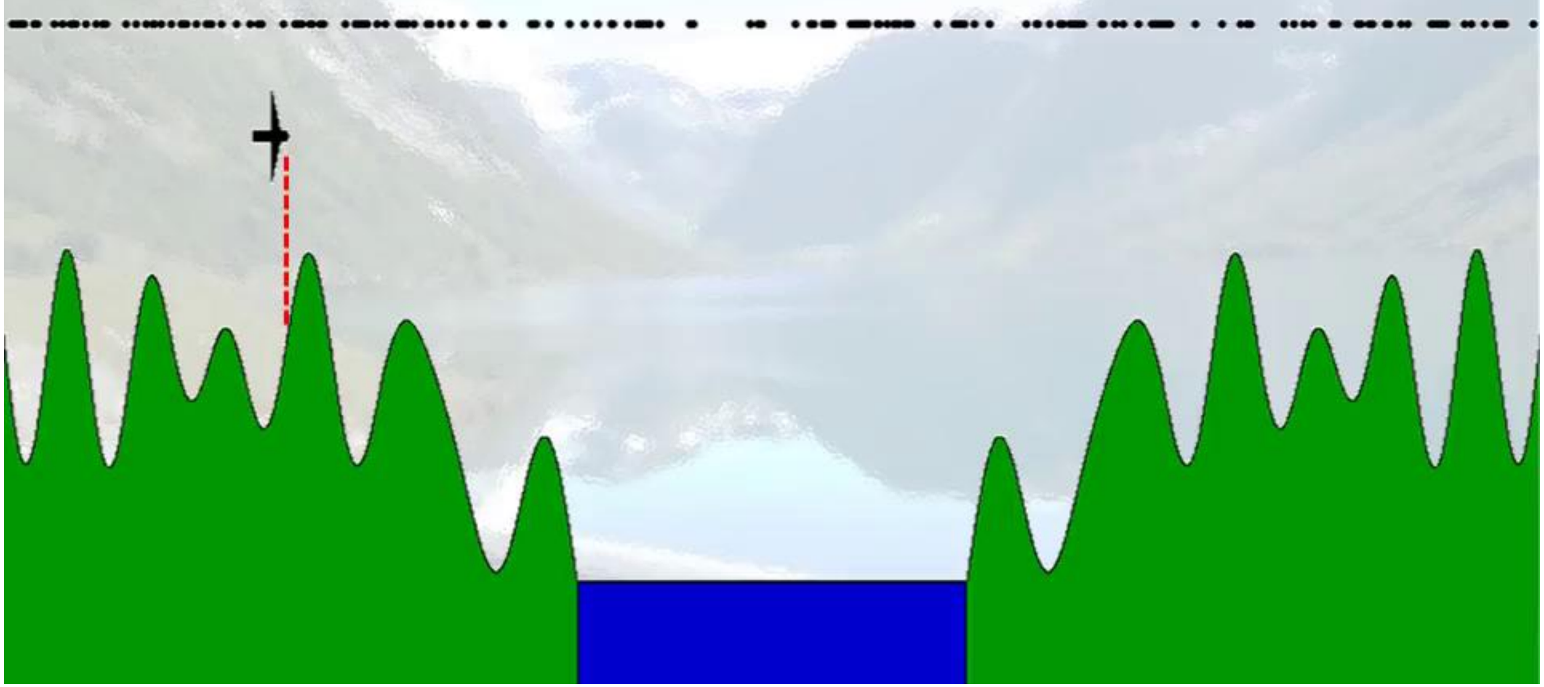
But we *don't* know where we are on the map!

(Which could be useful to know, if it's dark or foggy…)

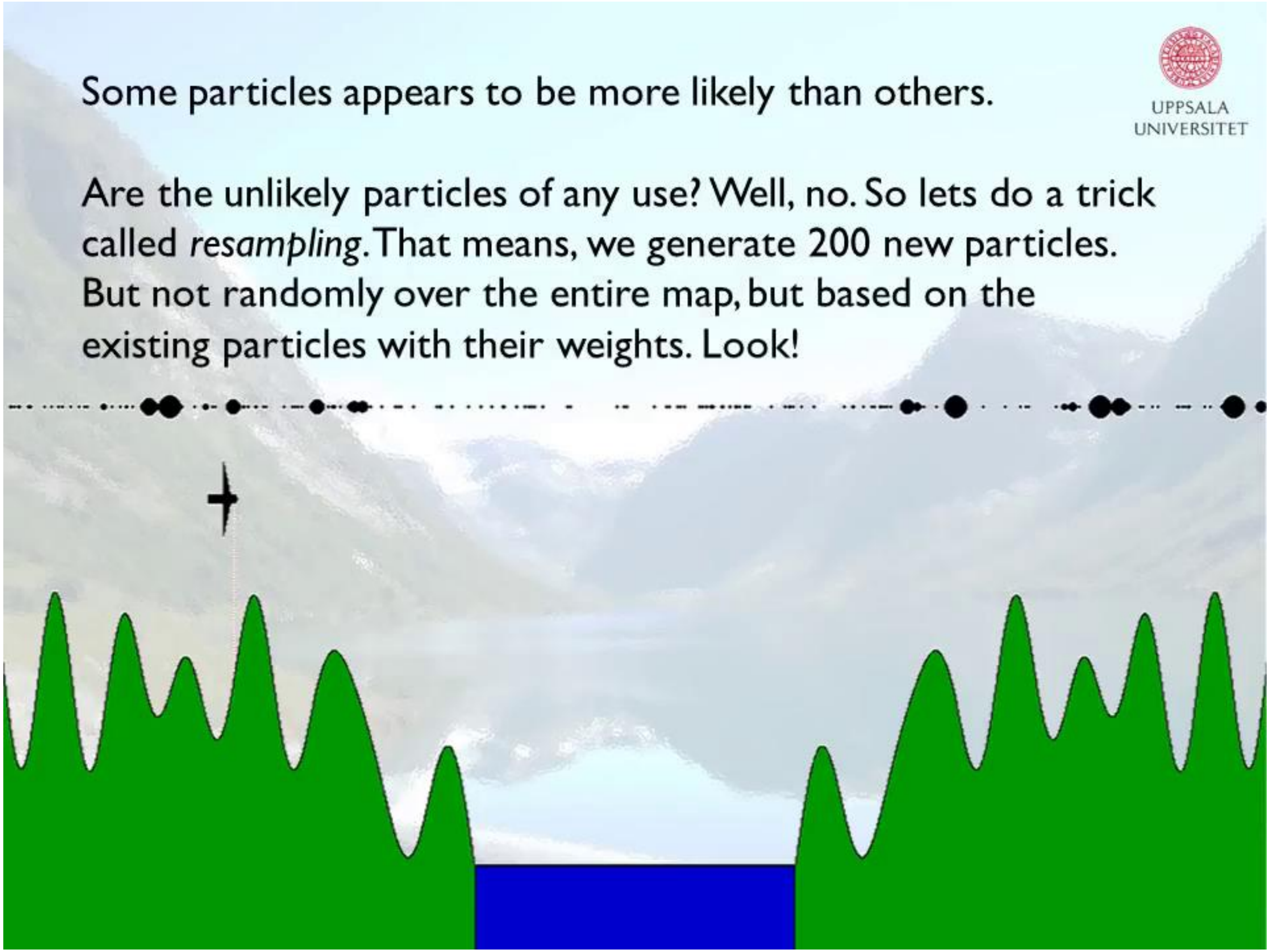


https://www.youtube.com/watch?v=aUkBa1zMKv4

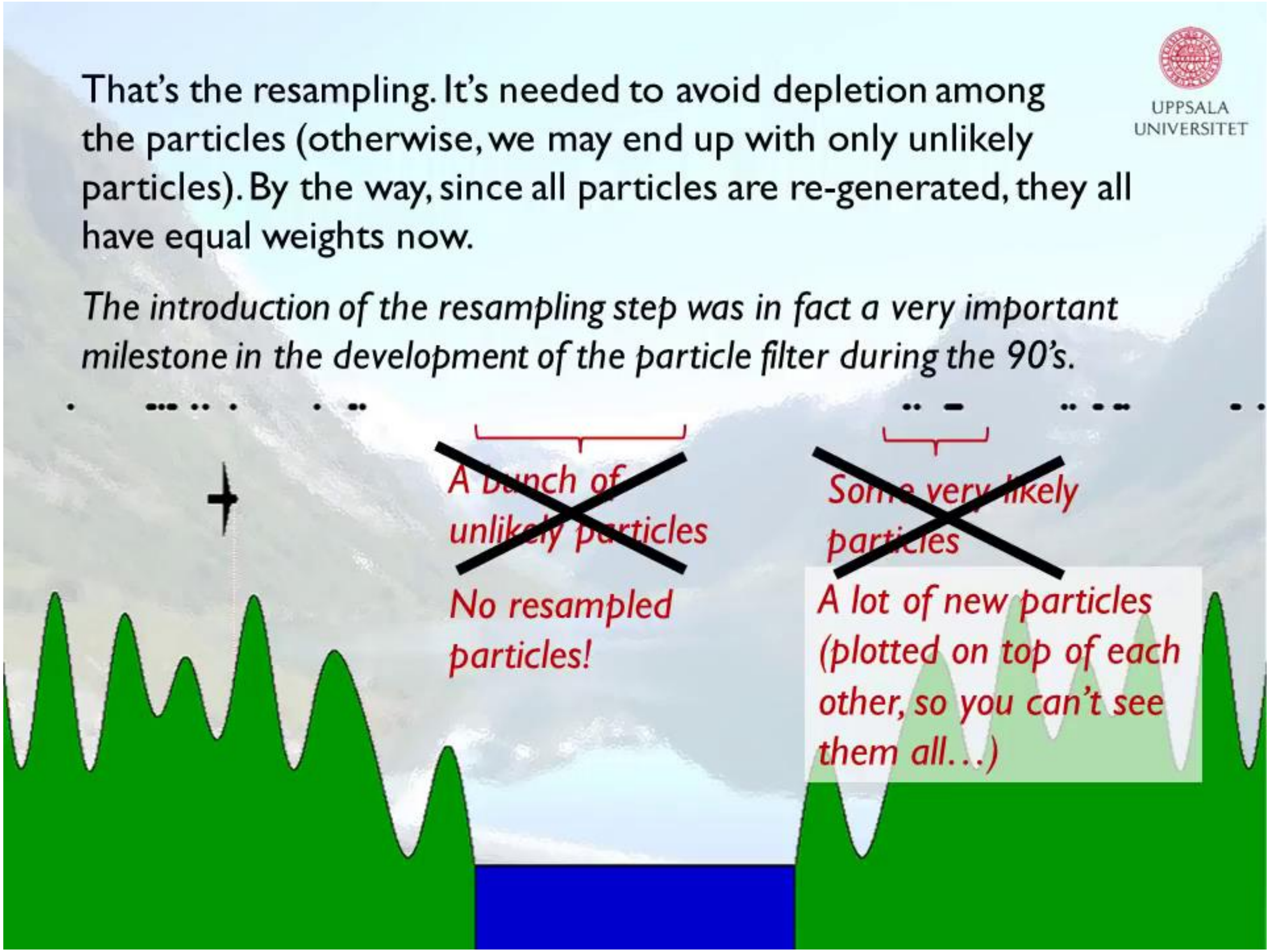
- We have measured the distance to the ground (with some uncertainty)
- We have the map
- We know how high we are flying (again with some uncertainty from the measurement device)



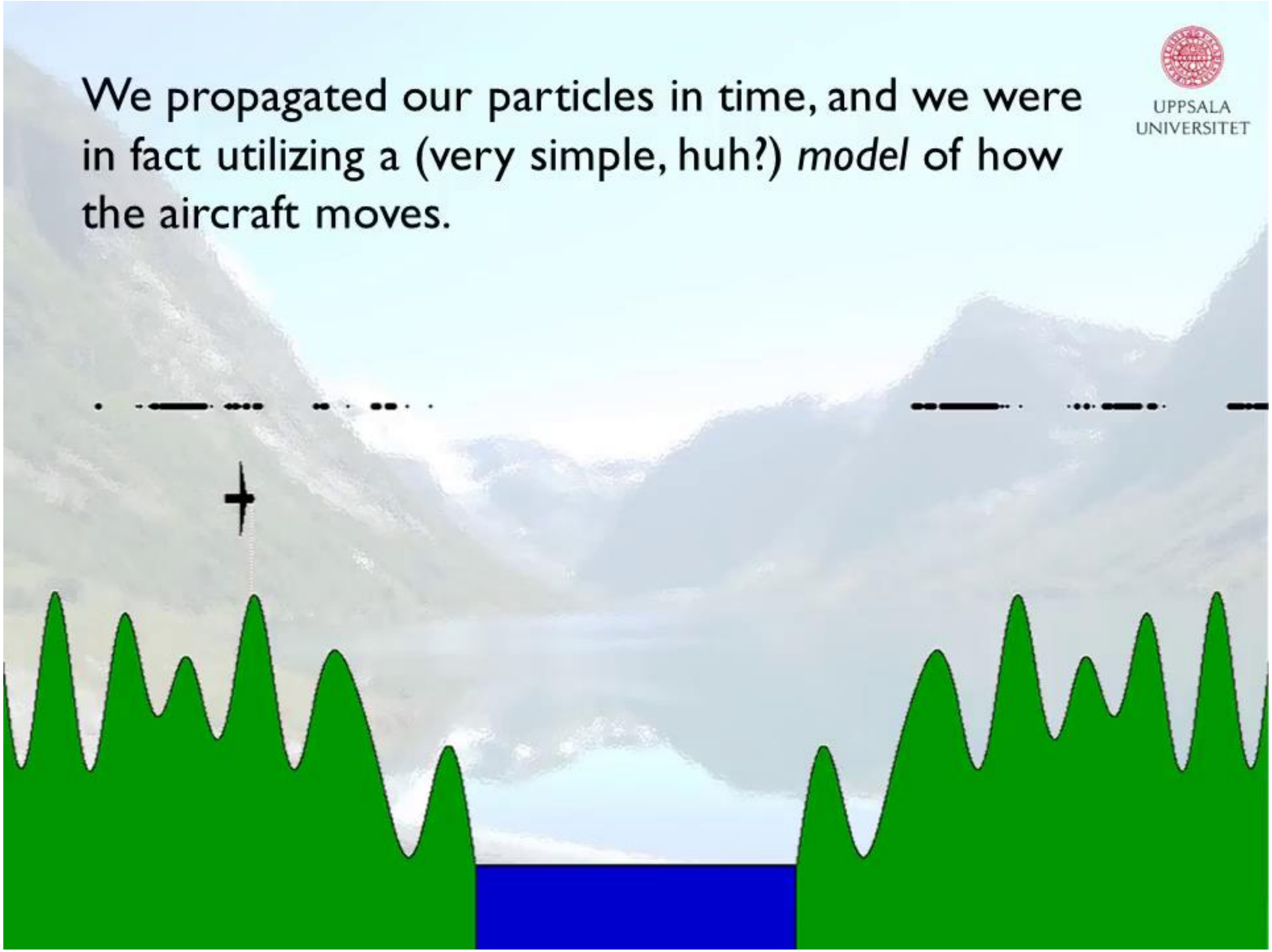
https://www.youtube.com/watch?v=aUkBa1zMKv4



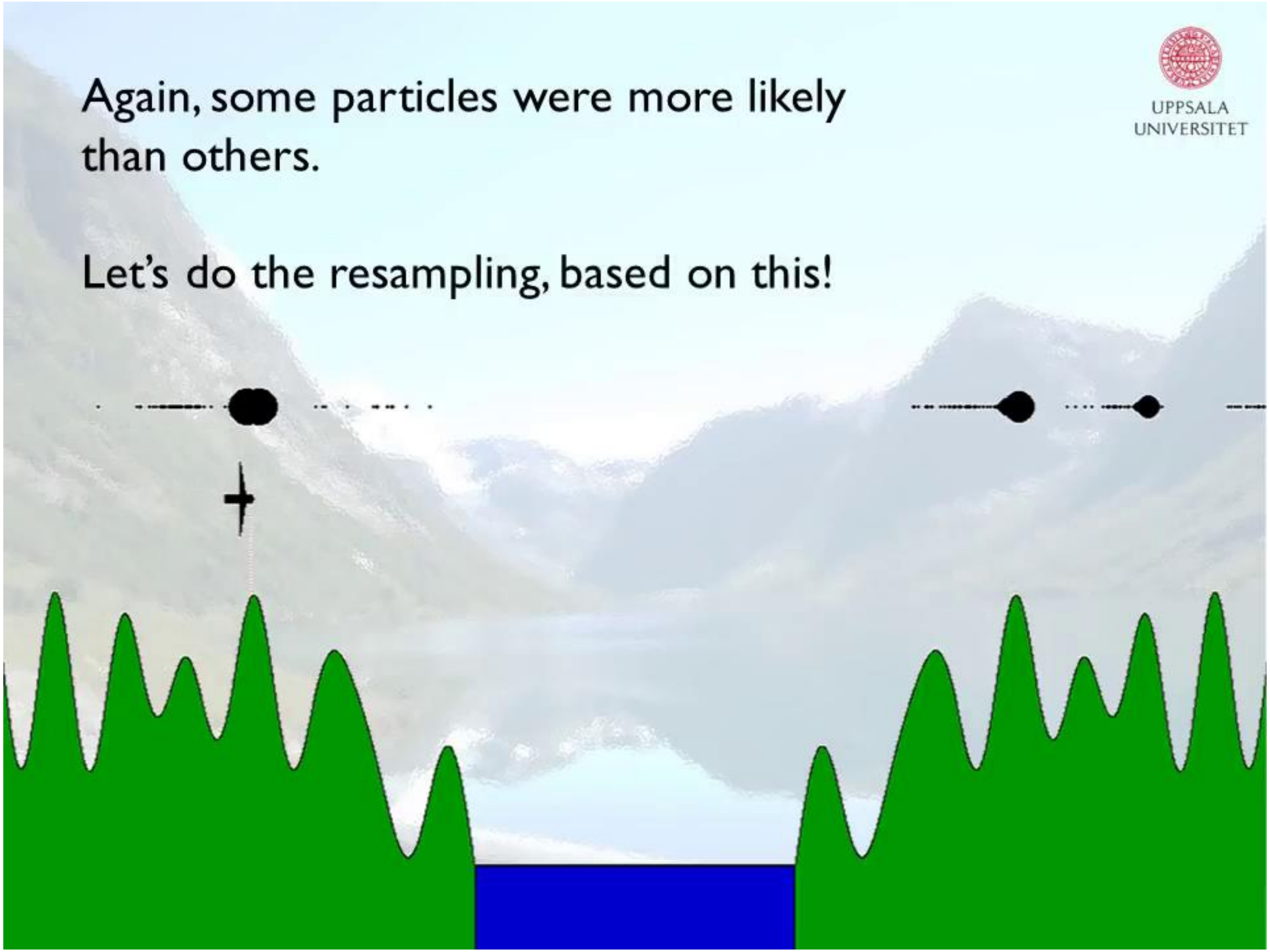
https://www.youtube.com/watch?v=aUkBa1zMKv4



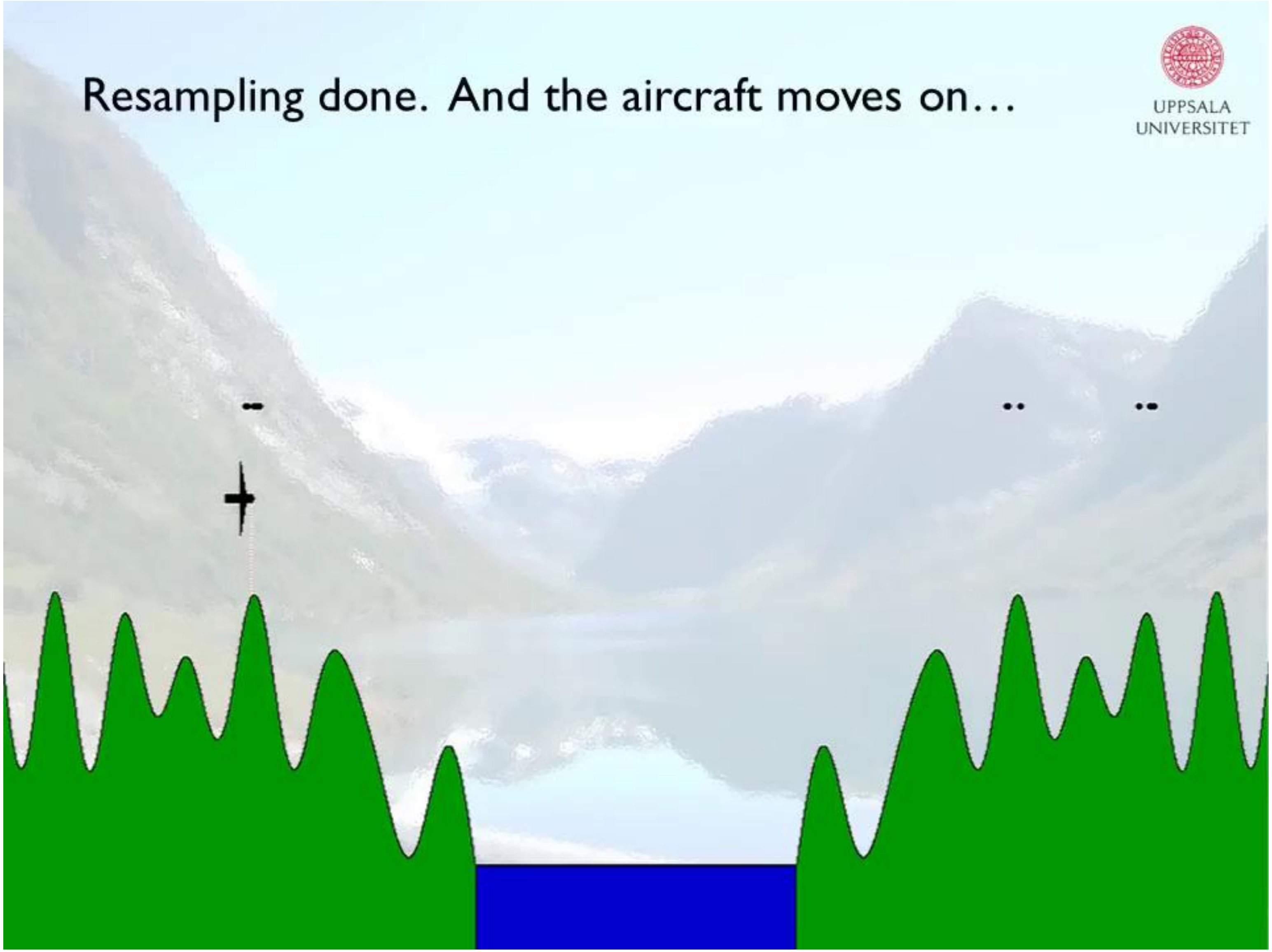
https://www.youtube.com/watch?v=aUkBa1zMKv4



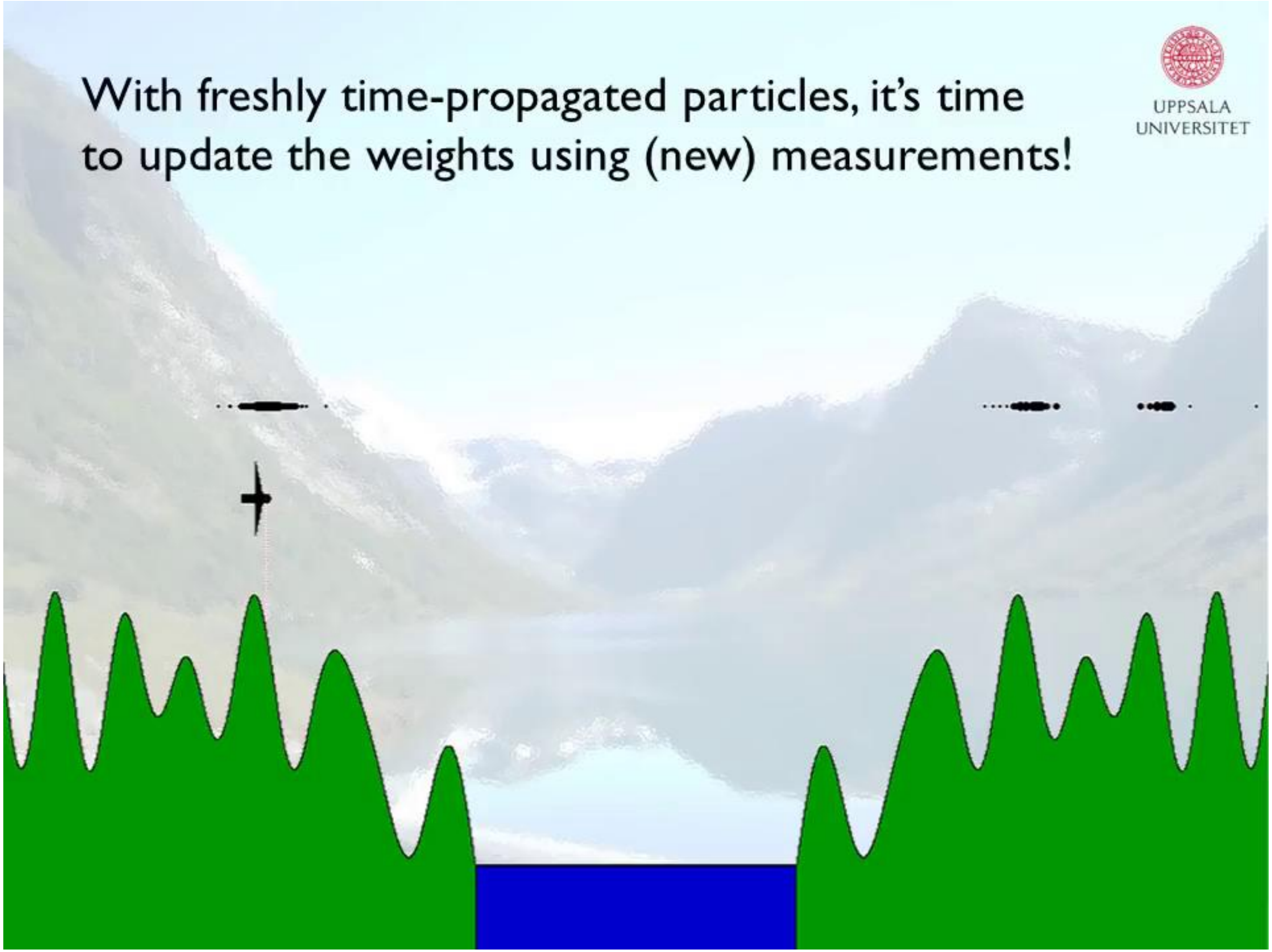
https://www.youtube.com/watch?v=aUkBa1zMKv4



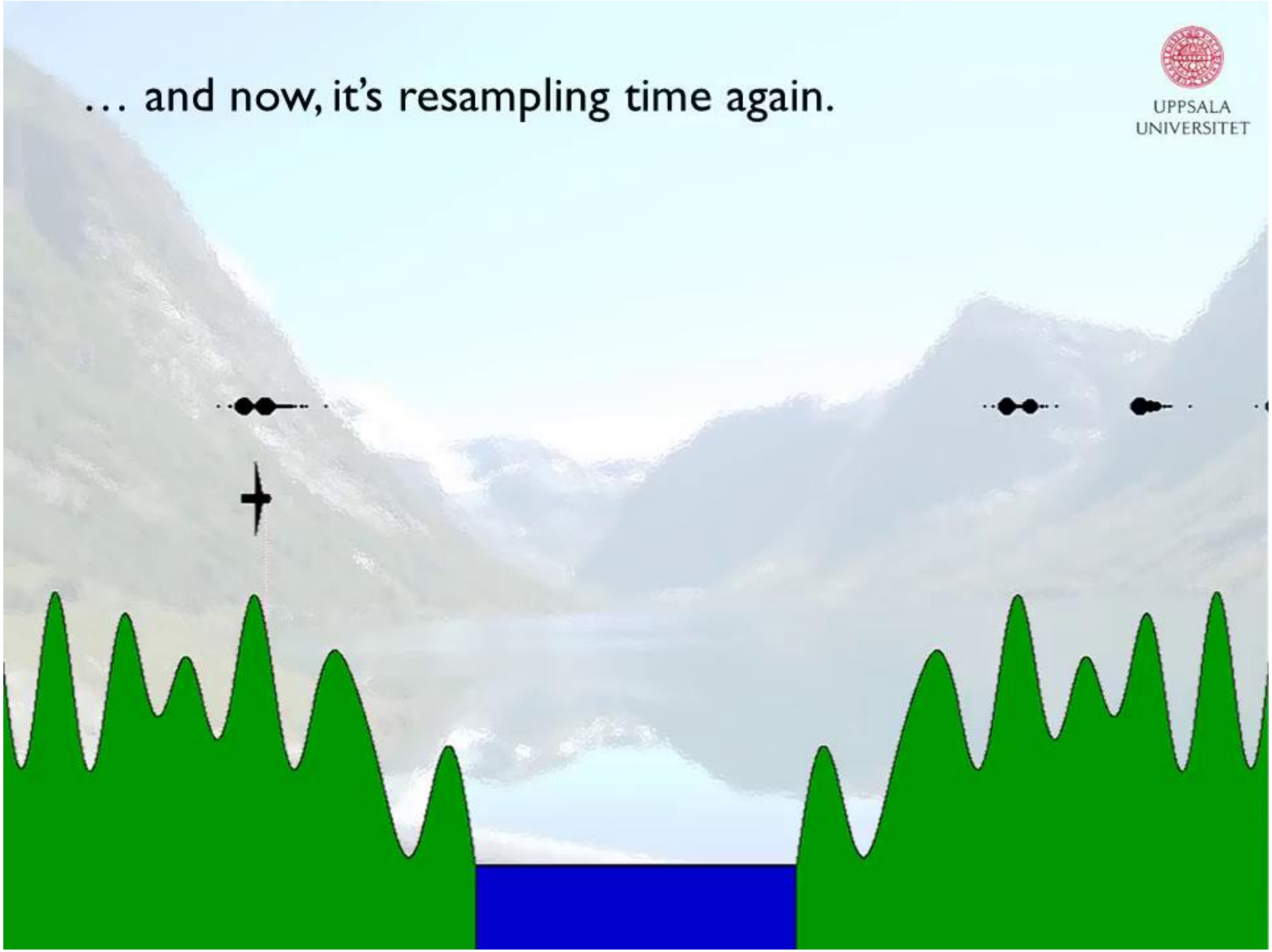
https://www.youtube.com/watch?v=aUkBa1zMKv4



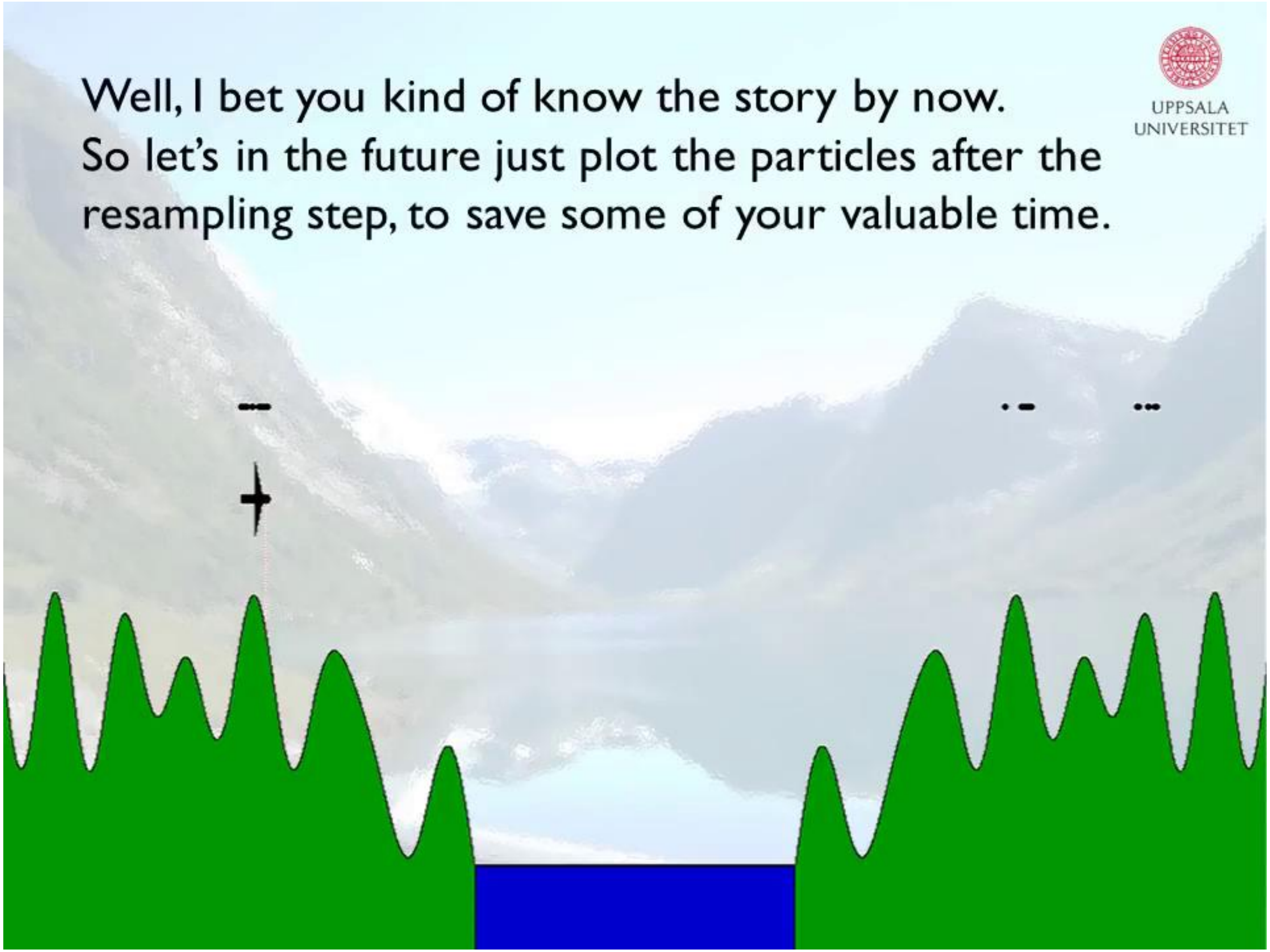
https://www.youtube.com/watch?v=aUkBa1zMKv4



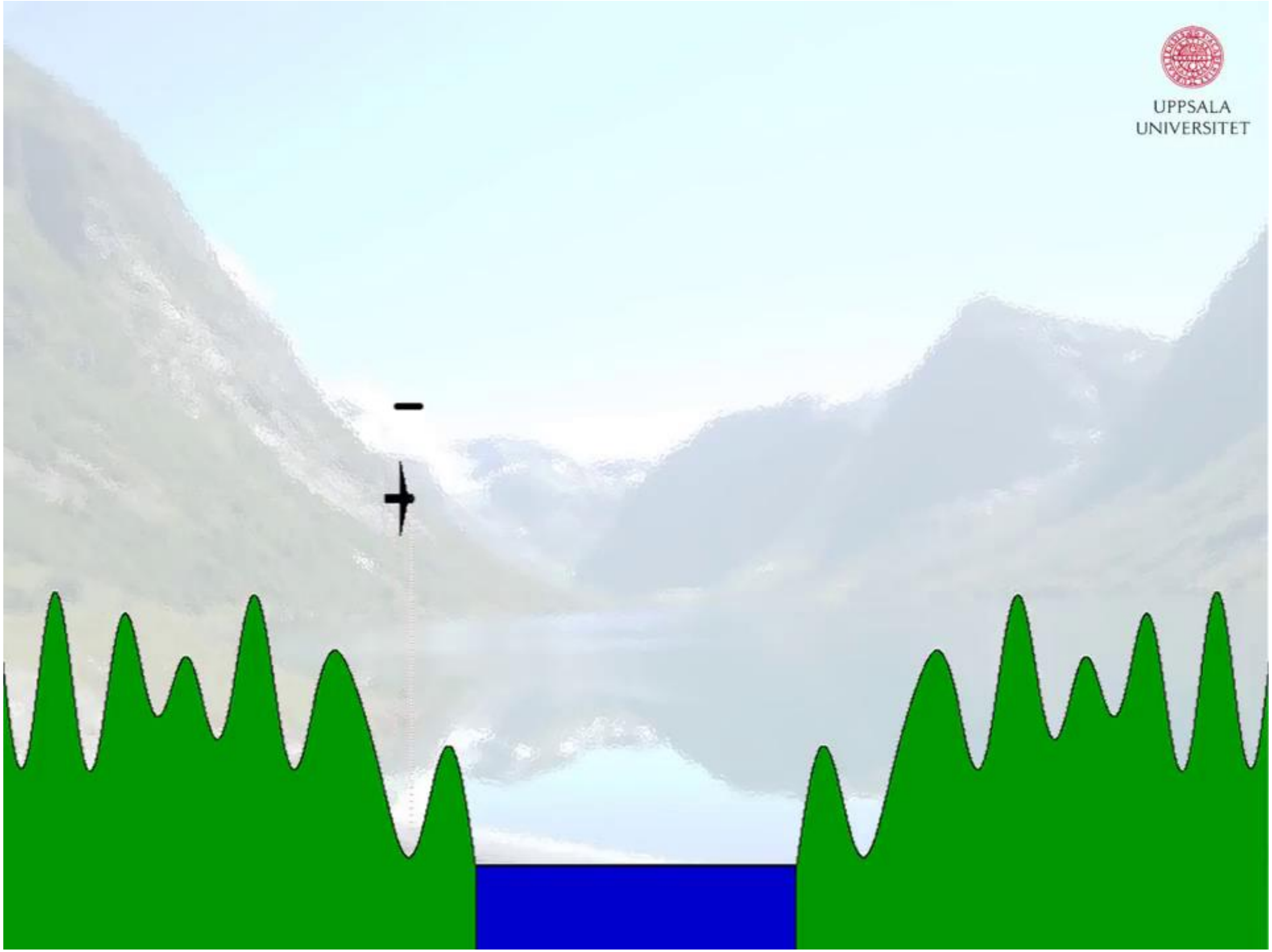
https://www.youtube.com/watch?v=aUkBa1zMKv4



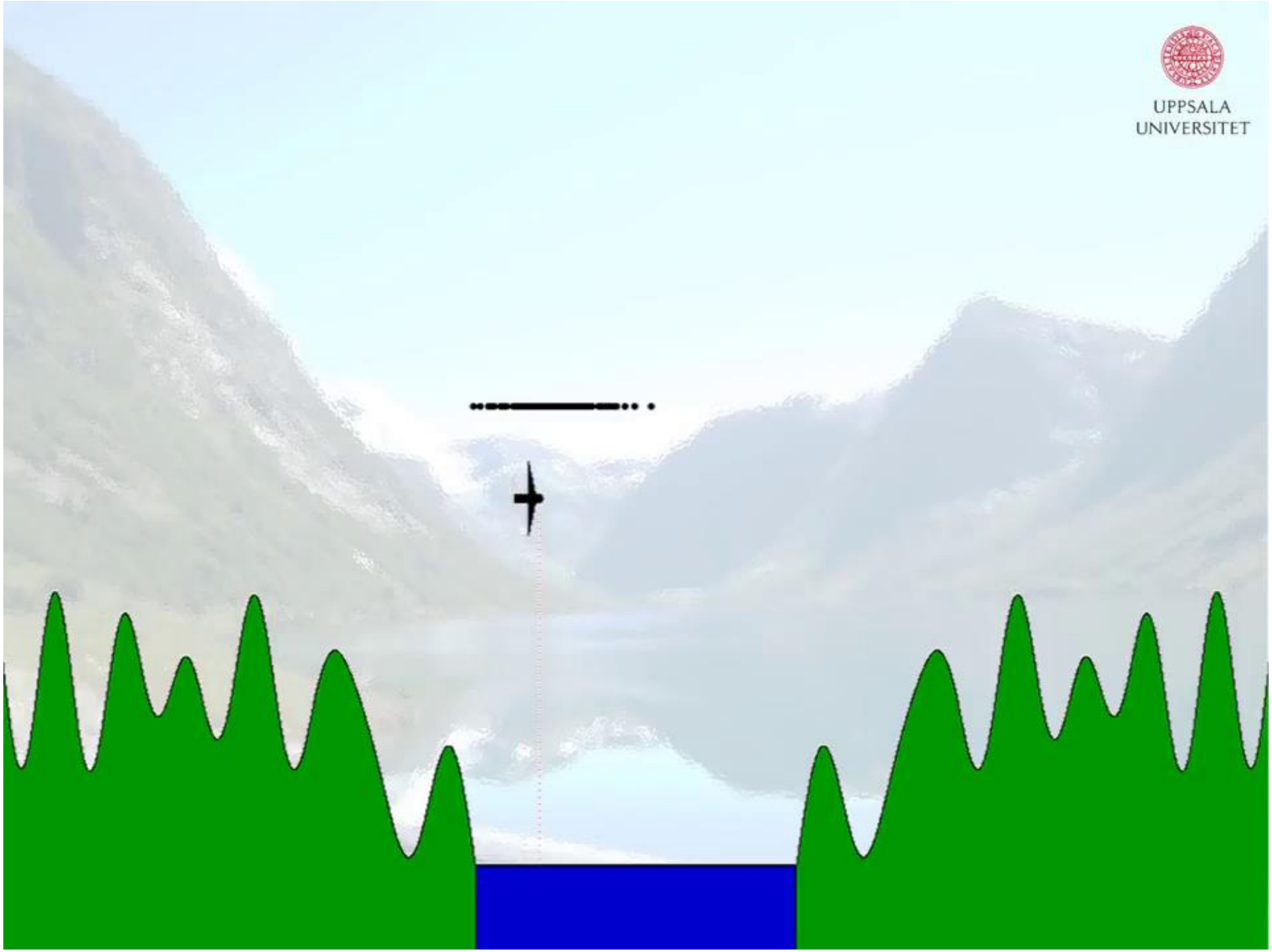
https://www.youtube.com/watch?v=aUkBa1zMKv4



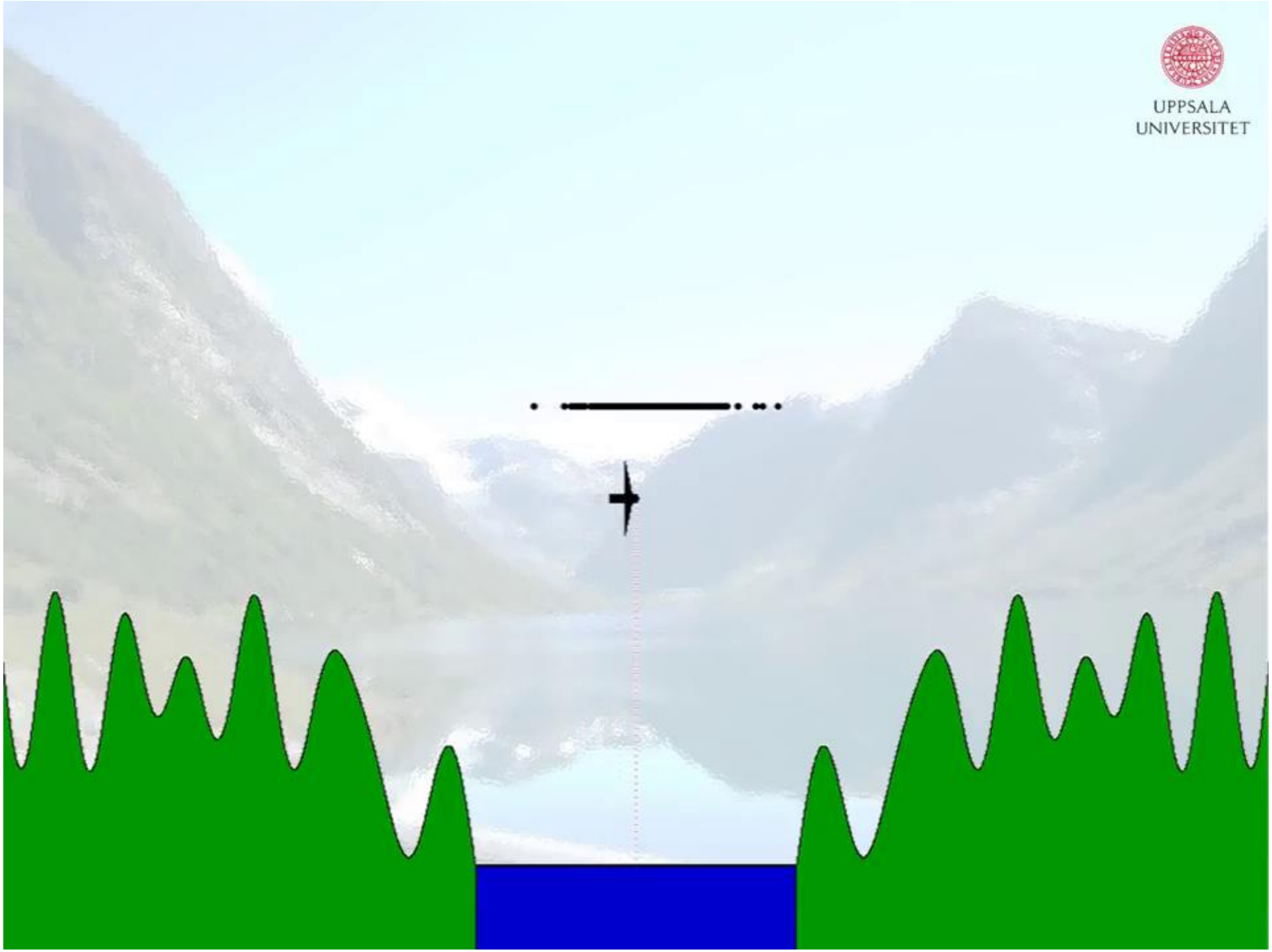
https://www.youtube.com/watch?v=aUkBa1zMKv4



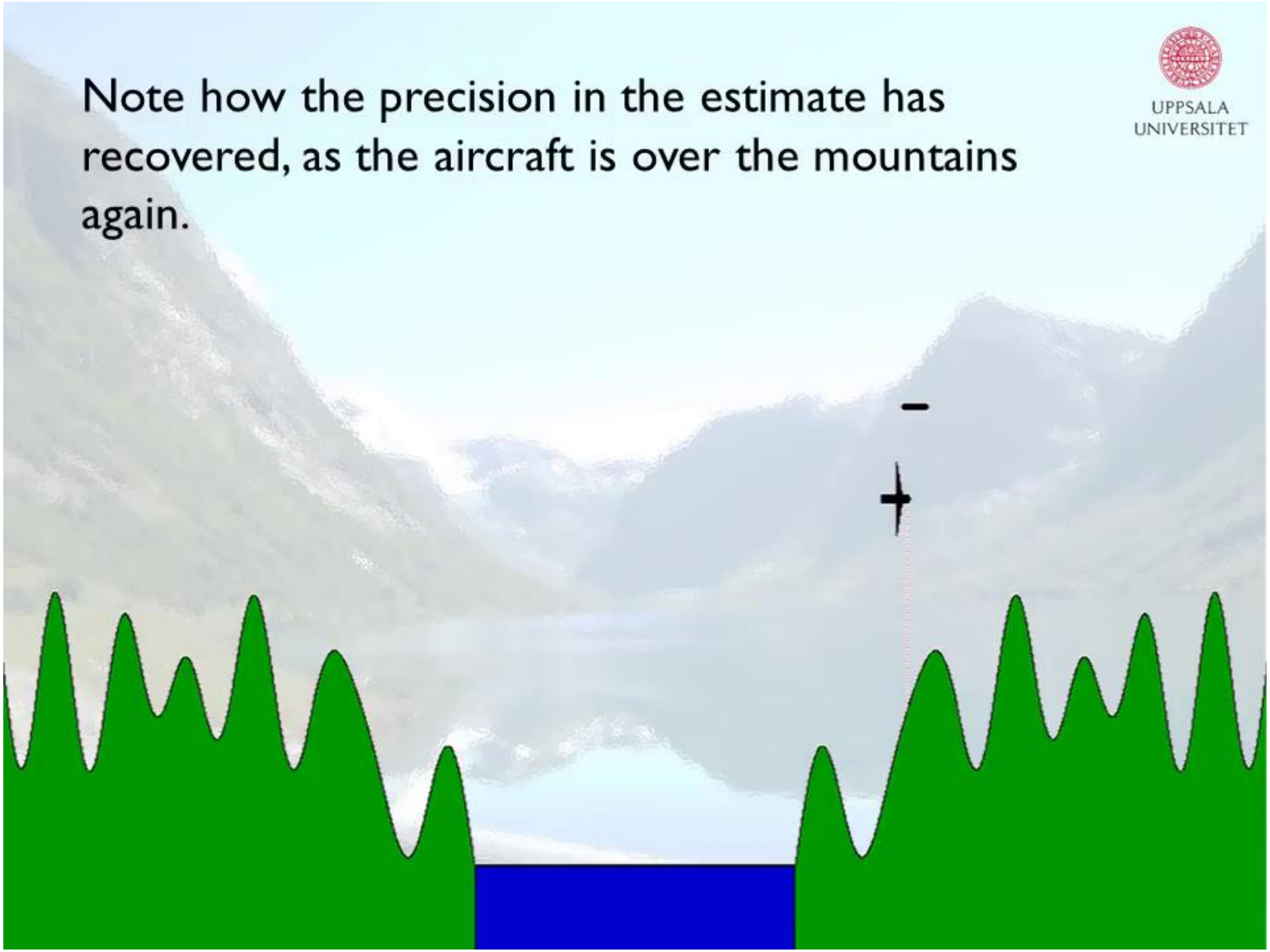
https://www.youtube.com/watch?v=aUkBa1zMKv4



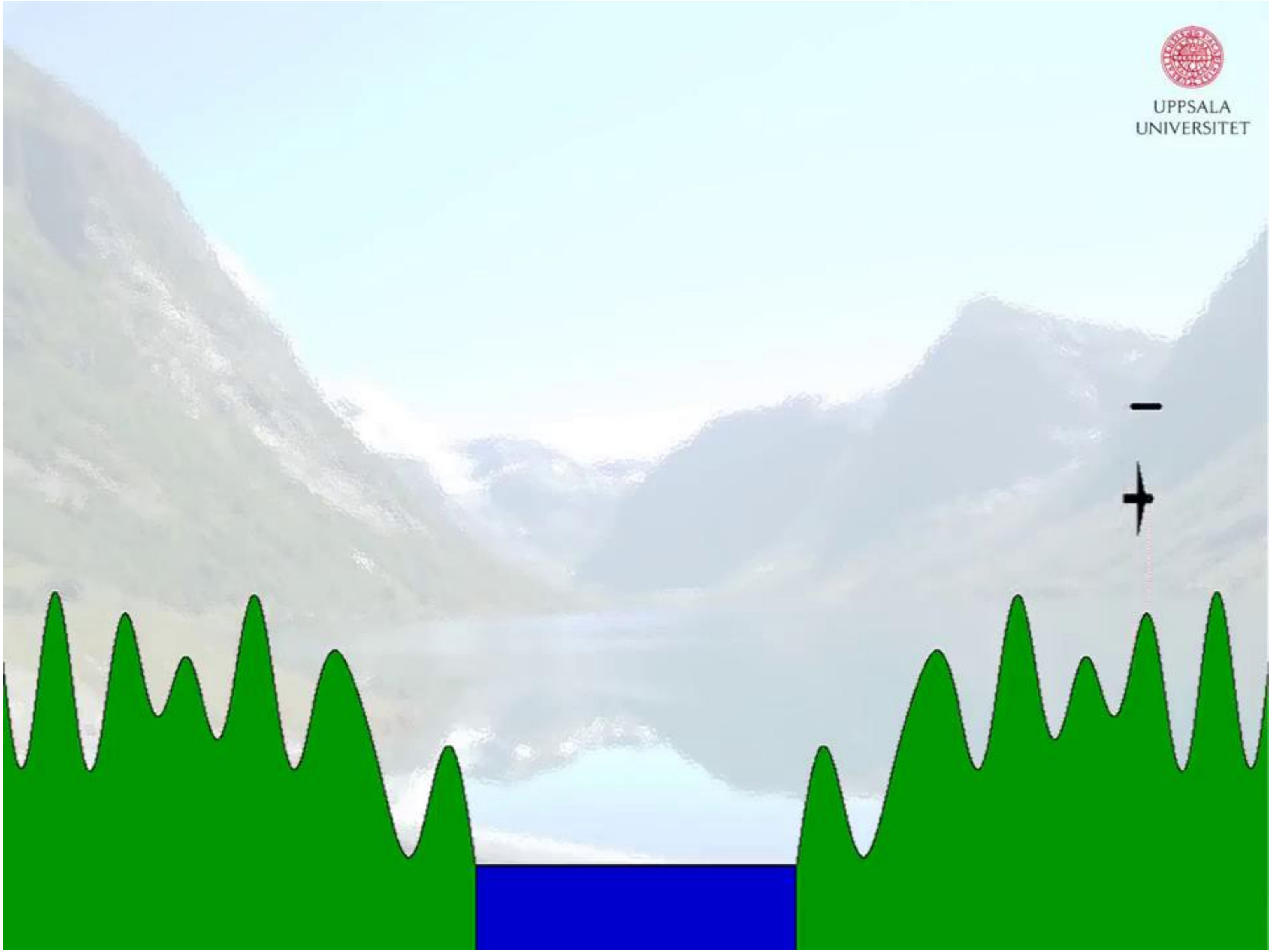
https://www.youtube.com/watch?v=aUkBa1zMKv4



https://www.youtube.com/watch?v=aUkBa1zMKv4



https://www.youtube.com/watch?v=aUkBa1zMKv4



https://www.youtube.com/watch?v=aUkBa1zMKv4

Avantaje

- Complexitate de timp constantă
- S-a observat empiric că eroarea de aproximare rămâne mărginită în timp (analiza teoretică este dificilă)
- Cu cât numărul de particule este mai mare, aproximarea probabilităților este mai bună

Concluzii

- Rețelele bayesiene asigură un mod concis de a reprezenta relațiile de independență condiționată într-un domeniu și de a face inferențe
- Rețelele bayesiane dinamice sunt rețele bayesiane care reprezintă modele de probabilitate temporale
- Există multe metode de inferență exactă sau aproximativă pentru rețele bayesiene