Arbres de décision

Bertrand MARESCHAL
bmaresc@ulb.ac.be
http://homepages.ulb.ac.be/~bmaresc/

1

Plan

- « Decision analysis »
  - Décision en avenir incertain
  - Arbres de décision
  - Espérance mathématique
- Probabilités
  - Cas discret
  - Cas continu

2

« Decision Analysis »

- Prise de décisions en avenir incertain.
- Modèle : arbre de décision.
- Méthode d'analyse générale.
- Besoin du calcul des probabilités.

3

Un 1er exemple simple

Exercice 1.1

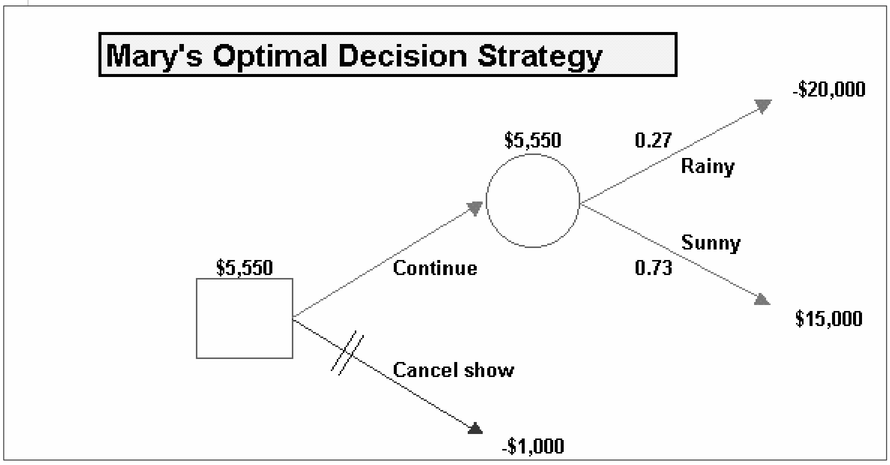
EXERCISE 1.1 Mary is organizing a special outdoors show which will take place on August 15. The earnings from the show will depend heavily on the weather. If it rains on August 15, the show will lose \$20,000; if it is sunny on August 15, the show will earn \$15,000. Historically, the likelihood of it raining on any given day in mid-August is 27%. Suppose that today is July 31. Mary has the option of canceling the show by the end of the day on July 31, but if she does so, she will then lose her \$1,000 deposit on the facilities.

(a) What is Mary's optimal decision strategy?

(b) Suppose that Mary can also cancel the show on August 14, but if she waits until then to do so, she must pay a fee of \$10,000. The advantage of waiting until August 14 is that she can listen to the weather forecast for the next day on the local news station. According to station records, the weather was forecast to be sunny 90% of the days in mid-August in previous years. Also, when the weather was forecast to be sunny, it turned out to be sunny 80% of the time. When the weather was forecast to be rainy, it turned out to be rainy 90% of the time. What is Mary's optimal decision strategy in this case?

4

Un 1er exemple simple
Exercice 1.1



5

Noeuds

- Nœuds de décision :
  - Représentés par des carrés
  - chaque branche correspond à une décision possible
- Nœuds de la nature (événements) :
  - représentés par des cercles
  - chaque branche correspond à un état possible de la nature (résultat, « outcome »), avec sa probabilité de réalisation

6

Un critère numérique : l'EMV

- « Expected Monetary Value » :
  - Espérance de « gain ».
  - Espérance mathématique.
  - L'EMV associée à un événement incertain est la somme de tous les résultats numériques possibles pondérés par leurs probabilités de réalisation respectives.
- Autres critères possibles :
  - Utilité, attitude face au risque.

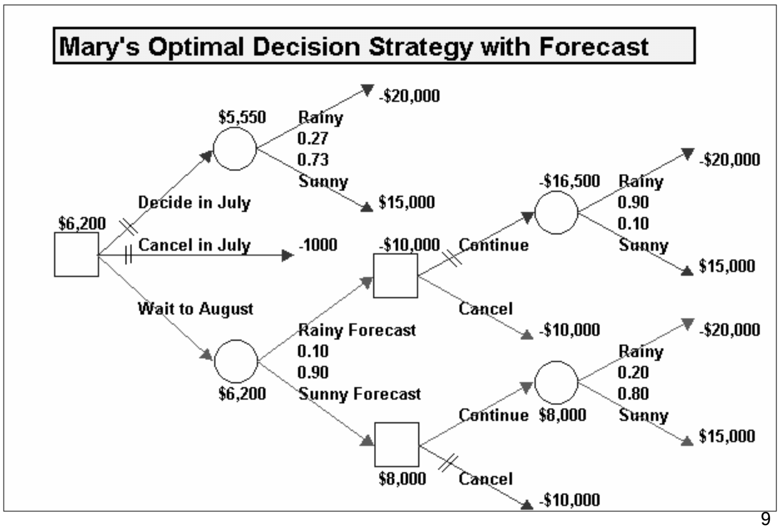
7

Analyse de l'arbre de décision (« Backwards induction »)

- Commencer par les noeuds terminaux de l'arbre
  - Pour un nœud événement, l'EMV est la moyenne pondérée des EMV de chaque branche issue de ce nœud, pondérées par leurs probabilités.
  - Pour un nœud de décision, l'EMV correspond à la branche issue de ce noeud qui permet d'obtenir la meilleure EMV.
- L'EMV du nœud initial de l'arbre correspond à la stratégie optimale de décision.

8

Un exemple un peu moins simple
Exercice 1.1 (suite)



9

Probabilités
Cas discret

- Résultats, probabilités et événements.
- Lois de probabilité.
- Tableaux de contingence.
- Variables aléatoires.
- Distributions de probabilité.
- Paramètres.

10

Résultats, probabilités et événements

◈ Expérience aléatoire :

◈ Résultats:
- pile ou face,
- 1, 2, 3, 4, 5 ou 6,
- une parmi 52 cartes.

11

Résultats, probabilités et événements

◈ Résultats :
- mutuellement exclusifs
- exhaustifs.

◈ Probabilité : mesure de la vraisemblance qu'un résultat soit observé.

◈ Evénement : sous-ensemble de résultats.

◈ Evénements mutuellement exclusifs : s'ils n'ont aucun résultat en commun.

12

Lois du calcul des probabilités

- $1^{ère}$ Loi :

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

- $2^{ème}$ Loi : si *A* et *B* sont des événements mutuellement exclusifs, alors :

$$P(A \text{ ou } B) = P(A) + P(B)$$

13

Lois du calcul des probabilités

- $3^{ème}$ Loi : si *A* et *B* sont deux événements, on a :

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

- Evénements indépendants : *A* et *B* tels que :

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

$$\rightarrow P(A \text{ et } B) = P(A) \times P(B)$$

14

Probabilité conditionnelle

- Exemple:
  - classe de 100 étudiants : 40 – 60
  - 70 Français : 25 – 45
  - 30 d'autres nationalités : 15 – 15
  - un étudiant est choisi au hasard :

  (F: Français, I: international, M: garçon, W: fille)

$$P(F\,|\,M) = \frac{P(F \text{ et } M)}{P(M)}$$

15

Tableau de contingence

- Distribution des étudiants de la classe :

	Français	Internat.	Total
Garçons	25	15	40
Filles	45	15	60
Total	70	30	100

- Table de probabilités :

	Français	Internat.	Total
Garçons	0.25	0.15	0.40
Filles	0.45	0.15	0.60
Total	0.70	0.30	1.00

16

Variables aléatoires

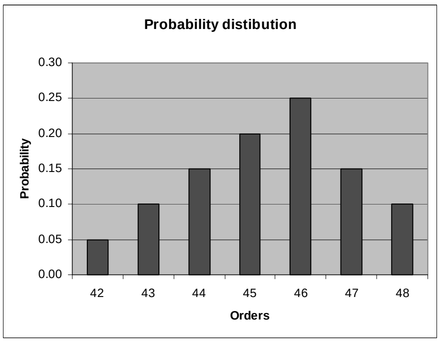
- Lorsque les résultats d'une expérience sont numériques.
- Variable aléatoire discrète :
  - Valeurs possibles dénombrables (nombre fini, ou infini dénombrable).
- Variable aléatoire continue :
  - Peut prendre n'importe quelle valeur dans un intervalle réel.

17

Distributions de probabilité discrètes

- Liste des valeurs possibles (Xi), et des probabilités correspondantes (Pi).

Orders	Probability
42	0.05
43	0.10
44	0.15
45	0.20
46	0.25
47	0.15
48	0.10



18

La distribution binomiale

- Expérience : *n* essais indépendants.
- Chaque essai mène à l'un de 2 résultats possibles :
  - Succès : P (succès) = p
  - Echec : P (échec) = $1 - p$
- Variable aléatoire :
  X = nombre de succès
  ➔ distribution binomiale

19

Paramètres des distributions de probabilité

- Variable aléatoire discrète X
- Valeurs possibles :

$$x_1, x_2, \ldots, x_i, \ldots, x_n$$

- Probabilités associées :

$$p_1, p_2, \ldots, p_i, \ldots, p_n$$

$$p_i = P(X = x_i)$$

$$\sum_{i=1}^{n} p_i = 1$$

20

Paramètres : Moyenne ou espérance

- Moyenne, valeur espérée, espérance mathématique... de *X*
- Définition:

$$\mu_X = E(X) = \sum_{i=1}^{n} p_i \times x_i$$

- Exemple:

$$E(X) = 0.05 \times 42 + 0.10 \times 43$$
$$+ 0.15 \times 44 + 0.20 \times 45 + 0.25 \times 46$$
$$+ 0.15 \times 47 + 0.10 \times 48 = 45.35$$

21

Paramètres : Variance – écart-type

- Ecart « moyen » de *X* autour de sa moyenne.
- Définition:

$$\sigma_X^2 = \sum_{i=1}^{n} p_i \times (x_i - \mu_X)^2$$

- Ecart-type :

$$\sigma_X = \sqrt{\sigma_X^2}$$

22

Cas 2
San Carlos mud slides

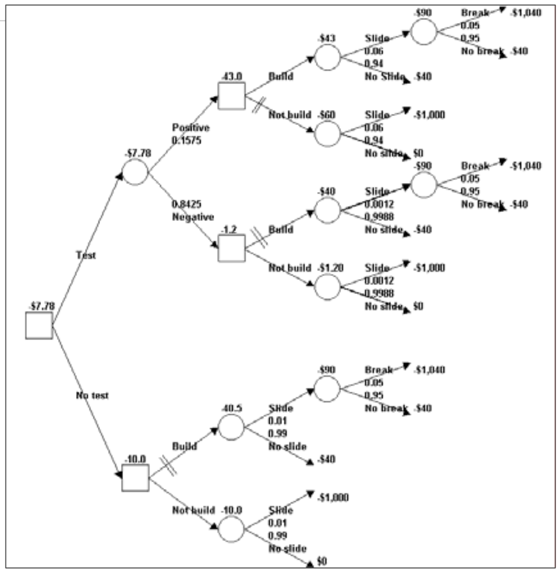
Extensive logging has exposed a hillside in San Carlos to the possibility of a mudslide. Reforestation is underway, but it will be a year before the new vegetation will be mature enough to remove the danger. If a slide occurs in the interim, human injuries will be avoided because mud moves slowly. The damage from such a slide would be limited to the road that passes beneath the hill. Construction of a retaining wall on the uphill side of the road has been suggested as a possible step to prevent this damage.

The Mayor of San Carlos is puzzled by the uncertainty concerning the issue. He has consulted with an expert who states that there is only one chance in 100 that a slide will occur within the next year. The expert adds that roughly 5% of all such slides break through retaining walls like the one proposed. The retaining wall would cost $40,000 to build. The road would cost about $1,000,000 to repair if damaged by a mudslide.

The expert points out that she can better assess the likelihood of a slide occurring in the next year if she conducts a geological test of the igneous rock layer below the hillside. Like any test, this one is imperfect. A positive test outcome would indicate a higher chance of a slide than a negative test outcome. The test has been conducted at sites at which slides eventually occurred and at sites at which slides did not subsequently occur. The information from these previous tests can be summarized as follows. Positive test outcomes had been reported on 90% of the sites at which slides subsequently occurred. Negative test results had been reported at 85% of the sites at which slides did not subsequently occur.

23

Cas 2
San Carlos mud slides



24

Probabilités
Cas continu

- Variables aléatoires continues
- Fonction de densité de probabilité (pdf)
- Fonction de répartition (cdf)
- Distribution normale (Gaussian distr.)

25

Variable aléatoire continue

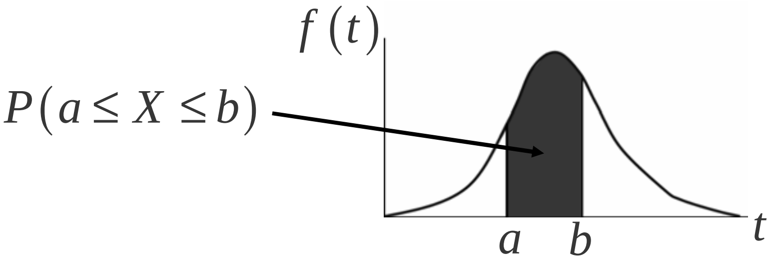
- Prend n'importe quelle valeur dans un intervalle donné.
- Exemples :
  - largeur d'une plaque d'acier,
  - poids d'un étudiant,
  - temps d'attente au bureau de poste, ...
- Probabilités associées à des sous-intervalles et non à des valeurs individuelles :

$$P(\text{temps d'attente} = 32.142 \text{ sec}) = 0$$

26

Fonction de densité de probabilité (pdf)

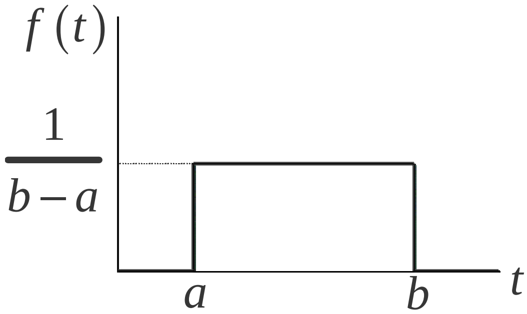
- Fonction non-négative $f(t)$
- Surface sous la courbe pdf égale à 1
- Probabilités = surfaces sous la courbe :



27

Distribution uniforme continue

- Si X is equally likely to take on any value in a given interval.



28

Fonction de répartition (cdf)

$$F(t) = P(X \leq t)$$

$$P(X > t) = 1 - F(t)$$

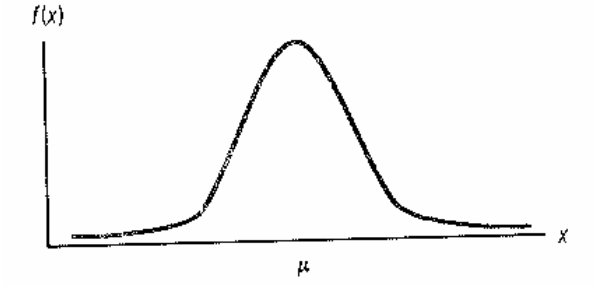
$$P(a \leq X \leq b) = F(b) - F(a)$$

29

Distribution normale

- Forme en cloche (gaussienne).
- Caractérisée par 2 paramètres :

μ et σ $\qquad X \sim N(\mu, \sigma)$



30

Standard normal distribution

$$Z = N(0,1)$$

$$X = N(\mu,\sigma) \longrightarrow Z = \frac{X-\mu}{\sigma}$$

$$P(a \le X \le b) = P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right)$$

31

Standard normal table

Percentiles of the Normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0.90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1.00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60
2.70	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974	2.70
2.80	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981	2.80

32