

Théorie des probabilités

Alexis Nasr - supports adaptés par Carlos Ramisch

Plan

- Ensemble fondamental et événements
- Probabilité
- Probabilités conditionnelles
- Formules des probabilités totales
- Loi de Bayes

Sources :

- **Initiation aux probabilités**, Sheldon Ross, édité aux Presses Polytechniques et universitaires romanes, 1994
- **All of Statistics**, Larry Wasserman, Springer, 2004

Ensemble fondamental

On considère une expérience dont l'issue n'est pas prévisible mais dont l'ensemble des résultats possibles (on dit aussi **issues**) est connu et appelé **ensemble fondamental**, noté Ω .

Exemples :

- si le résultat de l'expérience équivaut à la détermination du sexe d'un nouveau né, alors :

$$\Omega = \{g, f\}$$

- si le résultat est l'ordre d'arrivée d'une course entre 7 chevaux numérotés de 1 à 7, alors :

$$\Omega = \{\text{toutes les permutations de } 1 \dots 7\}$$

Événement

Tout sous-ensemble E de Ω est appelé un **événement**.

Si le résultat de l'expérience est compris dans E , alors on dit que E est **réalisé**.

Exemples :

- Dans le premier exemple, si $E = \{g\}$ alors E est l'événement *l'enfant est un garçon*.
- Dans le second exemple, si $E = \{\text{tous les résultats de } \Omega \text{ commençant par } 3\}$ alors E correspond à l'événement *le cheval 3 remporte la course*.

Variables aléatoires

- Une variable aléatoire est une fonction qui associe un nombre réel à tous les éléments de Ω
- Les variables aléatoires permettent ainsi de passer d'événements, qui peuvent être de natures très diverse, à des nombres, et d'effectuer des calculs sur ces derniers.

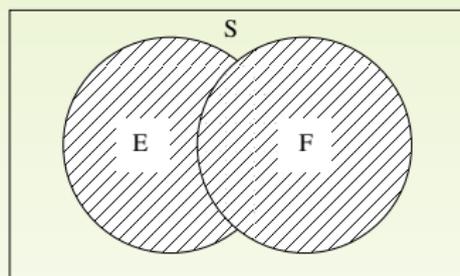
Opérations sur les événements

Un événement étant un ensemble, on peut combiner des événements grâce aux opérateurs ensemblistes :

- Union
- intersection
- complémentation

Union

L'événement $E \cup F$ est composé des résultats appartenant à E ou à F . L'événement $E \cup F$ est réalisé si soit E soit F l'est.



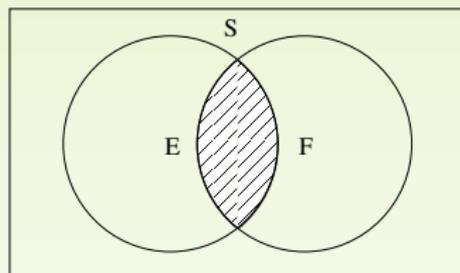
Exemple :

- Dans le cas du premier exemple, si $E = \{g\}$ et $F = \{f\}$, alors $E \cup F = \{f, g\} = \Omega$

intersection

L'événement $E \cap F$ (noté EF) est composé des résultats appartenant à E et à F .

EF est réalisé si E est réalisé et F l'est aussi.

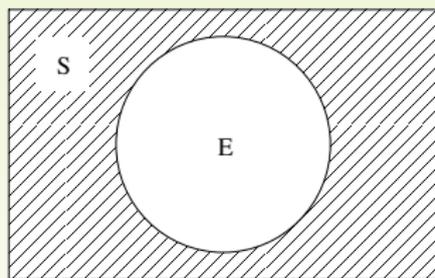


Exemple :

- Si $E = \{PP, PF, FP\}$ (au moins une pièce est pile) et $F = \{PF, FP, FF\}$ (au moins une pièce est face), alors, l'événement $EF = \{PF, FP\}$ est l'événement *une pièce est pile et l'autre face*.

complémentation

Pour chaque événement E , l'événement \bar{E} est composé des résultats qui sont dans Ω et qui ne sont pas dans E .
 \bar{E} est appelé le complémentaire de E dans Ω .



Exemple :

- Si $E = \{g\}$ alors $\bar{E} = \{f\}$.

Événement vide

Si $E = \{g\}$ et $F = \{f\}$, alors l'événement EF ne pourra jamais être réalisé, on appellera un tel événement l'**événement vide**, noté \emptyset .

Si $EF = \emptyset$, on dit que E et F sont **mutuellement exclusifs**.

On considère que pour chaque événement E de l'ensemble fondamental Ω , il existe une valeur $P(E)$ appelée **probabilité de E** qui vérifie les trois axiomes suivants :

①

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

②

$$P(\Omega) = 1$$

③ Pour chaque séquence d'événements mutuellement exclusif $E_1, E_2 \dots E_n$ ($E_i E_j = \emptyset$ si $i \neq j$),

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

Exemple

En supposant qu'à l'issue du lancer d'un dé les six faces ont les mêmes chances d'apparaître, on aura :

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

Du troisième axiome, il résulte que la probabilité de tirer un nombre pair est :

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6}$$

Quelques théorèmes élémentaires

- $P(\bar{E}) = 1 - P(E)$
- Si $E \subset F$, alors $P(E) \leq P(F)$
- $P(E \cup F) = P(E) + P(F) - P(EF)$

$$P(\bar{E}) = 1 - P(E)$$

Démonstration :

$$\begin{aligned} 1 &= P(\Omega) \\ &= P(E \cup \bar{E}) \\ &= P(E) + P(\bar{E}) \quad (\text{Axiome 2}) \end{aligned}$$

Si $E \subset F$, alors $P(E) \leq P(F)$

Démonstration :

Du fait que $E \subset F$, on peut écrire :

$$F = E \cup (\bar{E} \cap F)$$

E et $\bar{E} \cap F$ étant mutuellement exclusifs, par l'axiome 3, on tire :

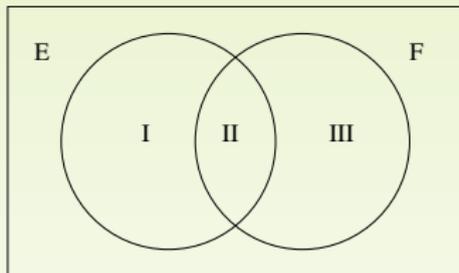
$$P(F) = P(E) + P(\bar{E} \cap F)$$

or, d'après l'axiome 1 $P(\bar{E} \cap F) \geq 0$, d'où :

$$P(E) \leq P(F)$$

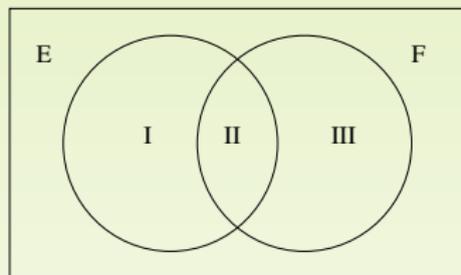
$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Démonstration :



$$\begin{array}{ll} I = E\bar{F} & E \cup F = I \cup II \cup III \\ II = EF & E = I \cup II \\ III = \bar{E}F & F = II \cup III \end{array}$$

Démonstration (suite)



I, II et III étant disjoints, on a d'après l'axiome 2 :

$$P(E \cup F) = P(I) + P(II) + P(III)$$

$$P(E) = P(I) + P(II)$$

$$P(F) = P(II) + P(III)$$

On a donc :

$$P(E \cup F) = P(E) + P(F) - P(II)$$

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Ens. fond. à événements élémentaires équiprobables

Pour de nombreuses expériences, il est naturel d'admettre que chaque événement élémentaire a la même probabilité d'apparaître.

Si $\Omega = \{1, 2 \dots N\}$, on a :

$$P(\{1\}) = P(\{2\}) = \dots = P(\{N\})$$

Ce qui implique du fait des axiomes 2 et 3 :

$$\forall i \ 1 \leq i \leq N, P(\{i\}) = \frac{1}{N}$$

De ceci et de l'axiome 3, il résulte que pour tout événement E :

$$P(E) = \frac{\text{nombre d'éléments dans } E}{\text{nombre d'éléments dans } \Omega}$$

Exemple 1

Si deux dés sont jetés, quelle est la probabilité que la somme des faces soit 7 ?

Solution : On fait l'hypothèse que les 36 issues possibles sont équiprobables.

Puisqu'il y a 6 issues qui donnent une somme de 7 :
(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)

la probabilité est $6/36 = 1/6$.

Exemple 2

Si deux boules sont tirées au hasard d'un bol en contenant 6 blanches et 5 noires, quelle est la probabilité qu'une des boules tirées soit blanche et l'autre noire ?

Solution : On considère que l'ordre dans lequel les boules sont choisies est significatif

l'ensemble fondamental comprend $11 \times 10 = 110$ points.

Il y a $6 \times 5 = 30$ manières de tirer pour lesquelles la première boule est blanche et l'autre noire.

et $5 \times 6 = 30$ manières de tirer pour lesquelles la première boule est noire et la seconde blanche.

Si les 110 points de l'ensemble fondamental ont la même probabilité, la probabilité cherchée est :

$$\frac{30 + 30}{110} = \frac{6}{11}$$

Probabilités conditionnelles

Un des concepts les plus importants de la théorie des probabilités. L'importance de ce concept est de deux ordres :

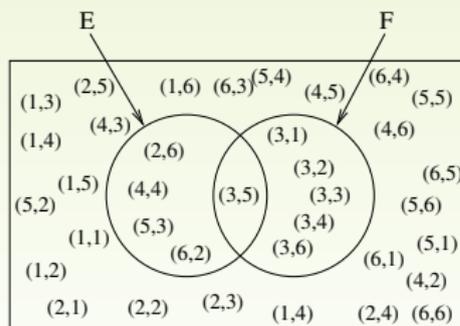
- On s'intéresse souvent à calculer des probabilités lorsqu'une partie de l'information concernant le résultat de l'expérience est disponible dans une telle situation, les probabilités cherchées sont justement des probabilités conditionnelles.
- Même lorsqu'aucune information partielle n'est disponible, il est quelquefois avantageux d'utiliser un détour par certaines probabilités conditionnelles pour réussir le calcul des probabilités cherchées.

Présentation intuitive

- On jette deux dés, chacun des 36 événements élémentaires a la même probabilité de survenir, soit $\frac{1}{36}$.
- Si l'on sait que le premier dé donne un 3, quelle est la probabilité que la somme des deux dés donne 8 ?
- le dé initial étant un 3, il ne peut plus y avoir que 6 événements dans notre expérience, à savoir :
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6).
- Puisque chacun de ces événements a originellement la même probabilité d'apparaître, ils auront encore des probabilités égales : $\frac{1}{6}$

Présentation intuitive (suite)

- Si nous désignons respectivement par E et F les événements *la somme des dés est 8* et *le premier dé donne 3*, la probabilité précédente est appelée probabilité conditionnelle que E apparaisse sachant que F est réalisée, elle est notée $P(E|F)$ (*probabilité de E sachant F*).



- $P(E|F) = \frac{1}{6}$

Généralisation

On s'inspire de la même démarche pour dériver une formule générale donnant $P(E|F)$ pour tout événement E et F :

- Si F est réalisé, alors E apparaîtra chaque fois que l'on aura affaire à un événement de E et de F à la fois, en d'autres termes, ce sera un élément de EF .
- Par ailleurs, comme nous savons que F est réalisé, cet ensemble devient le nouvel ensemble fondamental, appelé *ensemble fondamental réduit*.
- La probabilité conditionnelle de l'événement E sera donnée par comparaison de la probabilité non conditionnelle de EF avec la probabilité non conditionnelle de F .

Généralisation - 2

- On débouche ainsi sur la définition suivante : Si $P(F) > 0$, la probabilité conditionnelle de E sera :

$$P(E|F) = \frac{P(EF)}{P(F)}$$

- En multipliant par $P(F)$ les deux membres de l'équation, on obtient :

$$P(EF) = P(F)P(E|F)$$

- Cette équation signifie que la probabilité que E et F apparaissent à la fois est égale à la probabilité que F apparaisse multipliée par la probabilité conditionnelle de E si l'on sait que F est survenu.

Exemple

Une urne contient 8 boules rouges et 4 blanches. On tire sans remise 2 boules de l'urne et on admet qu'à chaque étape tous les tirages possibles sont équiprobables.

Quelle est la probabilité que les 2 boules tirées soient rouges ?

Solution :

- $R_1 =$ la première boule est rouge
- $R_2 =$ la seconde est rouge

$$P(R_2R_1) = P(R_2|R_1)P(R_1)$$

- $P(R_1) = 8/12$
- $P(R_2|R_1) = 7/11$;

donc

$$P(R_2R_1) = 2/3 \times 7/11 = 14/33$$

Événements indépendants

- En général $P(E|F) \neq P(E)$
- Le fait de savoir que F est survenu influence la probabilité de E .
- Dans les cas où $P(E|F)$ est bien égal à $P(E)$, E est dit **indépendant** de F .
- Du fait que $P(E|F) = \frac{P(EF)}{P(F)}$, l'indépendance de E et F équivaut à :

$$P(EF) = P(E)P(F)$$

- Cette équation est symétrique en E et F , il en résulte que lorsque E est indépendant de F , F l'est aussi de E .

Exemple - 1

On tire au hasard une carte d'un paquet de 52 cartes à jouer ordinaires.

- $E =$ la carte tirée est un as
- $F =$ c'est un pique

E et F sont indépendants. En effet :

- $P(EF) = 1/52$
- $P(E) = 4/52$
- $P(F) = 13/52.$

On a bien :

$$P(EF) = P(E)P(F)$$

Exemple - 2

On jette deux pièces et on suppose que les 4 résultats possibles sont équiprobables.

- $A =$ la première pièce est pile
- $B =$ la seconde est face.

A et B sont indépendants. En effet :

- $P(AB) = P(\{(P, F)\}) = 1/4$
- $P(A) = P(\{(P, P), (P, F)\}) = 1/2$
- $P(B) = P(\{(P, F), (F, F)\}) = 1/2$

On a bien :

$$P(AB) = P(A)P(B)$$

Événements indépendants

- Il est important de ne pas confondre les deux notions d'événements mutuellement exclusifs et d'événements indépendants
- Des événements mutuellement exclusifs ne sont pas indépendants.
- En effet, si l'on sait que E et F sont mutuellement exclusifs alors on sait que si E est réalisé, F ne peut pas l'être, et vice versa.
- Par conséquent, la connaissance de E va modifier la connaissance de F , les deux événements ne sont donc pas indépendants.

La règle de multiplication

On peut généraliser la règle des probabilités conditionnelles à l'intersection d'un nombre arbitraire d'événements. Cette règle est appelée règle de multiplication.

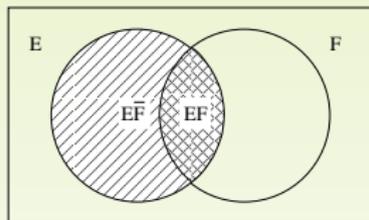
$$P(E_1 E_2 \dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \dots P(E_n|E_1 \dots E_{n-1})$$

On démontre cette loi en appliquant la définition des probabilités conditionnelles au membre de droite :

$$P(E_1) \frac{P(E_2 E_1)}{P(E_1)} \frac{P(E_3 E_1 E_2)}{P(E_2 E_1)} \dots \frac{P(E_1 \dots E_n)}{P(E_1 \dots E_{n-1})} = P(E_1 E_2 \dots E_n)$$

Formule des probabilités totales - 1

Soient E et F deux événements quelconques. Nous pouvons écrire E sous la forme $E = EF \cup E\bar{F}$.



EF et $E\bar{F}$ étant mutuellement exclusifs, on peut écrire :

$$\begin{aligned}P(E) &= P(EF) + P(E\bar{F}) \\ &= P(E|F)P(F) + P(E|\bar{F})P(\bar{F}) \\ &= P(E|F)P(F) + P(E|\bar{F})[1 - P(F)]\end{aligned}$$

Formule des probabilités totales - 2

$$P(E) = P(E|F)P(F) + P(E|\bar{F})[1 - P(F)]$$

La probabilité de l'événement E est une moyenne pondérée

- de la probabilité conditionnelle de E lorsque F est apparu
- et de la probabilité de E lorsque F n'est pas apparu
- les poids étant les probabilités des événements conditionnants.

L'intérêt de cette formule est qu'elle permet de déterminer la probabilité d'un événement en commençant par le conditionner selon l'apparition ou non d'un autre événement.

Exemple

Une compagnie d'assurance estime que les gens peuvent être répartis en deux classes :

- ceux qui sont enclins aux accidents
- ceux qui ne le sont pas.

Les statistiques montrent qu'un individu enclin aux accidents a une probabilité de 0,4 d'en avoir un dans l'espace d'un an ; cette probabilité vaut 0,2 pour les gens à risque modéré.

On suppose que 30% de la population appartient à la classe à haut risque.

Quelle est la probabilité qu'un nouvel assuré soit victime d'un accident durant l'année qui suit la signature de son contrat ?

Solution

Nous obtiendrons la probabilité de l'événement cité en le conditionnant selon que le signataire de la police est ou n'est pas enclin aux accidents.

- X = le signataire aura un accident dans l'année qui suit l'établissement du contrat
- A = le signataire est enclin aux accidents.

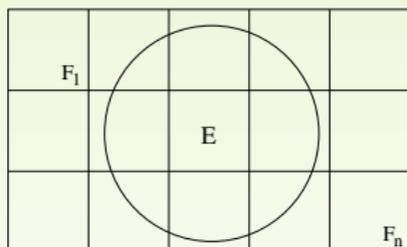
La formule des probabilités totales nous donne :

$$\begin{aligned}P(X) &= P(X|A)P(A) + P(X|\bar{A})P(\bar{A}) \\ &= 0.4 \times 0.3 + 0.2 \times 0.7 = 0.26\end{aligned}$$

Formule des probabilités totales - Généralisation

Si $F_1 \dots F_n$ sont des événements s'excluant mutuellement, tels que :

$$\bigcup_{i=1}^n F_i = \Omega$$



On peut écrire :

$$E = \bigcup_{i=1}^n EF_i$$

Formule des probabilités totales - Généralisation

En utilisant le fait que les événements EF_i s'excluent mutuellement, on peut écrire :

$$\begin{aligned} P(E) &= \sum_{i=1}^n P(EF_i) \\ &= \sum_{i=1}^n P(E|F_i)P(F_i) \end{aligned}$$

Cette équation montre qu'étant donné un jeu d'événements $F_1 \dots F_n$ desquels un et un seul surviendra, on peut calculer $P(E)$ en commençant par conditionner selon les F_i .

Théorème de Bayes

- Supposons que E se soit réalisé et que nous cherchions à déterminer la probabilité que l'un des F_j se soit aussi réalisé.
- On déduit de la formule des probabilités totales le théorème suivant, appelé théorème de Bayes :

$$\begin{aligned}P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}\end{aligned}$$

Exemple 1

On considère deux urnes, l'une contient une bille noire et une blanche, et l'autre deux noires et une blanche.

On désigne une urne au hasard, de laquelle on tire une bille.

- Quelle est la probabilité qu'elle soit noire ?
- Si l'on sait que la bille est blanche, quelle est la probabilité que ce soit la première urne qui ait été désignée ?

Loi des probabilités totales

- Quelle est la probabilité qu'elle soit noire ?
- $N =$ la bille choisie est noire
- $B =$ la bille choisie est blanche
- $U_1 =$ l'urne choisie est l'urne 1
- $U_2 =$ l'urne choisie est l'urne 2

$$P(N) = P(N|U_1)P(U_1) + P(N|U_2)P(U_2) = \frac{1}{2} \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{2} = \frac{7}{12}$$

Loi de Bayes

- Si l'on sait que la bille est blanche, quelle est la probabilité que ce soit la première urne qui ait été désignée ?
- $N =$ la bille choisie est noire
- $B =$ la bille choisie est blanche
- $U_1 =$ l'urne choisie est l'urne 1
- $U_2 =$ l'urne choisie est l'urne 2

$$P(U_1|B) = \frac{P(U_1B)}{P(B)} = \frac{P(B|U_1)P(U_1)}{P(\bar{N})} = \frac{12}{20}$$

Exemple 2

- Un laboratoire d'analyse du sang assure avec une fiabilité de 95% la détection d'une certaine maladie lorsqu'elle est effectivement présente.
- Cependant, le test indique aussi un résultat faussement *positif* pour 1% des personnes réellement saines à qui on le fait subir (une personne saine testée sera déclarée malade une fois sur cent).
- Si 0,5% de la population porte effectivement la maladie, quelle est la probabilité qu'une personne soit vraiment malade lorsqu'on la déclare telle sur la base du test ?

Solution

- D = la personne soumise au test est porteuse de la maladie
- E = le résultat du test est positif

La formule des probabilités totales nous donne :

$$\begin{aligned}P(D|E) &= \frac{P(DE)}{P(E)} \\ &= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|\bar{D})P(\bar{D})} \\ &= \frac{.95 \times .005}{.95 \times .005 + .01 \times .995} \simeq .323\end{aligned}$$

Ainsi 32% seulement des personnes dont le résultat au test est positif ont vraiment la maladie !

Surprise

- Soit E , l'un des événements pouvant survenir à la suite d'une expérience.
- A quel point serions-nous surpris d'apprendre que E a effectivement eu lieu ?
- La surprise liée à E est liée à la probabilité de E .
- Elle est d'autant plus élevée que la probabilité est faible.

Suprise

La surprise est formalisée sous la forme d'une fonction $S(p)$, où p est une probabilité différente de 0.

On désire que cette fonction réponde aux conditions suivantes :

- 1 $S(1) = 0$ il n'y a pas de surprise à apprendre qu'un événement certain est effectivement arrivé.
- 2 S est une fonction strictement décroissante de p . (si $p < q$ alors $S(p) > S(q)$). Plus un événement est improbable plus grande sera notre surprise.
- 3 $S(pq) = S(p) + S(q)$

Suprise

- $S(pq) = S(p) + S(q)$
 - Soient deux événement indépendants E et F de probabilités respectives p et q .
 - Comme $P(EF) = pq$, la surprise correspondant à l'apparition simultanée de E et F est $S(pq)$.
 - On apprend que E est survenu, puis, plus tard, que F est survenu lui aussi. La surprise additionnelle créée par le fait que F survienne vaut :

$$S(pq) - S(p)$$

- F étant indépendant de E , l'information sur E ne change pas la probabilité de F , la surprise conditionnelle doit donc être $S(q)$, d'où la condition :

$$S(pq) = S(p) + S(q)$$

- Une fonction vérifiant les conditions précédentes est la fonction suivante :

$$S(p) = -\log_2(p)$$

La surprise est mesurée en **bits** *binary digits*. Un bit est la quantité d'information correspondant à une probabilité de $1/2$.

- Exemple : on tire une carte d'un jeu de 52 cartes. Soit E l'événement *la carte tirée est l'as de pique*.
 - $P(E) = \frac{1}{52}$
 - $S\left(\frac{1}{52}\right) = -\log_2\left(\frac{1}{52}\right) = \log_2(52) = 5.7\text{bits}$
 - 6 bits : 2 pour la couleur et 4 pour la valeur.

Entropie d'une variable aléatoire

- Soit X une variable aléatoire qui prend ses valeurs dans l'ensemble $\{x_1, x_2, \dots, x_n\}$ avec les probabilités correspondantes p_1, p_2, \dots, p_n ($P(X = x_i) = p_i$).
- La grandeur $-\log_2(p_i)$ représentant la mesure de la surprise associée à l'événement $X = x_i$, la surprise moyenne créée lorsqu'on apprend quelle valeur X a prise est :

$$H(X) = - \sum_{i=1}^n p_i \log_2(p_i)$$

- Cette quantité est appelée **entropie** de la variable aléatoire X (par convention, si l'un des p_i vaut 0, on déclare que $0 \log_2(0)$ vaut 0).

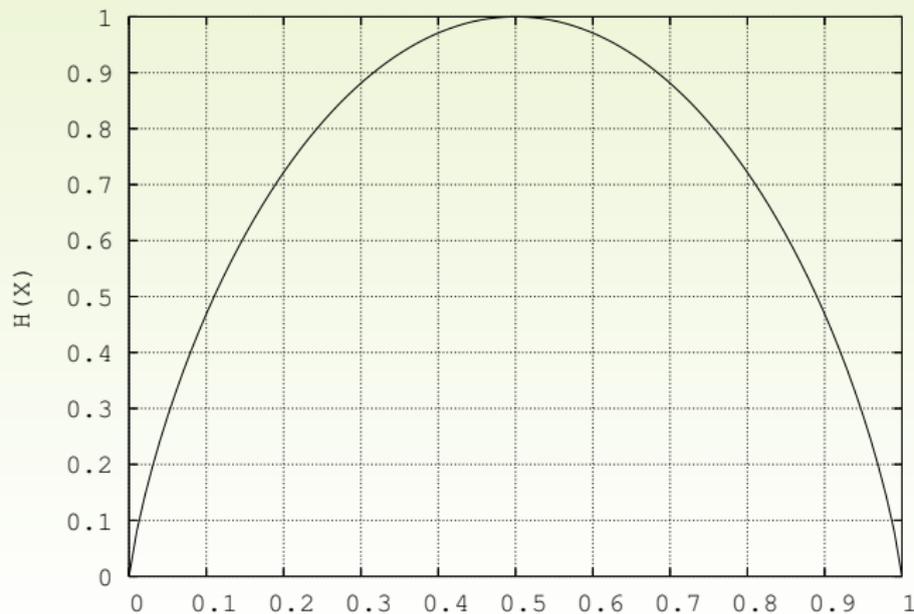
Entropie d'une variable aléatoire

- On peut montrer que $H(X)$ est maximale lorsque tous les p_i sont égaux.

Exemple de la pièce

$$P(X = \text{pile}) = p$$

$$P(X = \text{face}) = 1 - p$$



Entropie d'une variable aléatoire

- $H(X)$ représente la surprise moyenne associée à la découverte de la valeur de X .
- On peut aussi l'assimiler à la quantité d'incertitude relative à cette variable aléatoire.
- La théorie de l'information considère $H(X)$ comme l'**information** liée à l'observation de X .
- On peut donc dire que :
 - la surprise moyenne causée par X ,
 - l'incertitude liée à X
 - l'information moyenne recelée par X ,
- représentent en fait le même concept abordé par des points de vue un peu différents.