Eléments de théorie des Probabilités Discrètes

Espaces de probabilité

Définition : une expérience aléatoire est une expérience dont on ne connaît pas le résultat à l'avance.

Définition : on appelle univers des possibles, ou espace de probabilité l'ensemble des résultats possibles d'une expérience aléatoire.

Remarque : On se limitera dans toute la suite au cas où l'univers des possibles est fini[1].

Événements

Soit E un espace de probabilités.

Définition : toute partie de E est appelée un événement

- Un événement qui contient un unique élément de E est un événement élémentaire.
- L'ensemble E tout entier est l'événement certain, l'ensemble vide est l'événement impossible.
- L'événement contraire $\overline{A}$ d'un événement A est l'événement constitué par le complémentaire de A dans E.
- Deux événements sont incompatibles s'ils ne peuvent être réalisés simultanément.

Distribution de probabilité

Définition : une distribution de probabilité sur $E = \{e_1, \ldots, e_n\}$ est une fonction $p : E \to I\!R$ telle que :

$$0 \leq p(e_k) \leq 1 \quad \text{et} \quad \sum_{i=1}^{n} p(e_i) = 1$$

[1]On peut passer au cas continu,mais alors il faut faire un usage massifde la théorie de la mesure et de l'intégration

Définition : étant donnée une distribution de probabilité p sur E, on appelle probabilité la fonction $P : \mathcal{P}(E) \to I\!R$ définie comme suit

$$P(A) = \sum_{e \in A} p(e) \tag{1}$$

Pour tout événement $A \subseteq E$, la probabilité de A est $P(A)$.

Opérations élémentaires

Soit E équipé d'une distribution p.

Théorème :

$$P(A_1 \cup A_2) \quad = \quad P(A_1) + P(A_2) - P(A_1 \cap A_2) \tag{2}$$

$$P(\overline{A}) \quad = \quad 1 - P(A) \tag{3}$$

Distribution uniforme

Définition : on appelle distribution uniforme la distribution qui assigne à tout événement élémentaire la même valeur

Il est immédiat de voir que si $|E| = n$, et p est uniforme, alors

– pour tout $e \in E$,

$$p(e) = \frac{1}{n}$$

– pour tout $A \subseteq E$,

$$P(A) = \frac{|A|}{|E|}$$

Exemples : le LOTO

Nous avons déjà calculé le nombre de combinaisons de 6 chiffres entre 1 et 49 :

$$\mathcal{C}_{49}^{6} = 13983816$$

Donc la probabilité de gagner le gros lot est

$$\frac{1}{13983816}$$

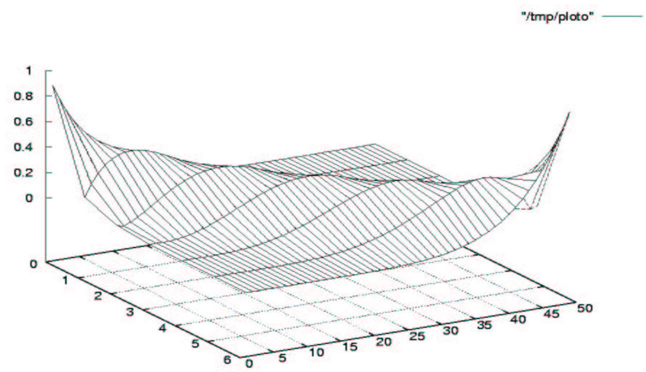
parce-que l'événement gagnant est élémentaire.

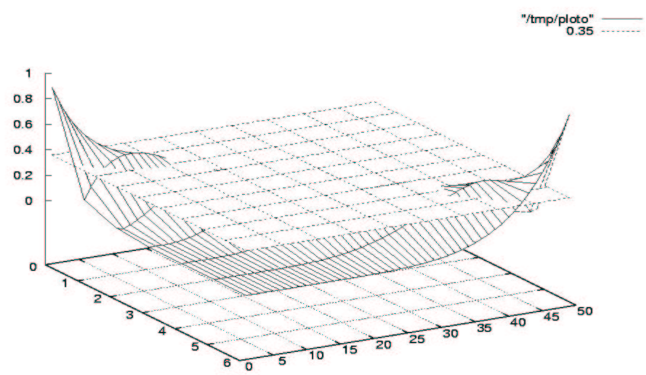
Exemples : le LOTO

Combien de chances a-t-on de trouver respectivement 6, 5, 4, 3, 2, 1 ou 0 bons numéros en jouant n chiffres distinctes ?

Notons $L(k, n)$ l'événement “on trouve k chiffres en jouant n chiffres”.

Combien y a-t-il d'éléments dans $L(k, n)$?





Le paradoxe du BigDil

Dans une variante du jeu du BigDil, le présentateur offre à la candidate de choisir un parmi 3 rideaux : l'un d'entre eux cache une voiture, les deux autres un verre en plastique.

Une fois que la candidate a choisi un rideau, le présentateur ouvre un des deux autres rideau et lui montre qu'il cache le verre en plastique.

Ensuite, il lui offre l'opportunité de changer son choix, si elle le souhaite.

Que feriez-vous à la place du candidat ?

Le paradoxe des anniversaires

Dans un Amphi il y a 50 étudiants. Quell'est la probabilité que au moins deux d'entre eux soient nés le même jour ?

FIG. 1 – Probabilité de 0 conflit en variant n

Nombre de personnes	P(anniversaires tous différents)
1	1
5	0.97
10	0.88
20	0.58
22	0.52
23	0.49
30	0.29
50	0.03

Donc, avec 23 étudiants, on peut déjà parier avantageusement sur la "coïncidence". **Vérification expérimentale**

Probabilité conditionnelle

Définition : la probabilité de l'événement A sachant B, ou *probabilité conditionnelle de A par rapport à B* est le rapport

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

Théorème : la fonction $p_A : E \rightarrow I\!R$ définie comme $p_A(e_i) = P(\{e_i\}|A)$ est une distribution de probabilité, quelque soit le choix de A événement non négligeable[2].

[2]i.e. $P(A) > 0$

Formule de Bayes

Théorème : soit $\{A_k\}_{k\in K}$ une partition[3] de E ; alors

$$P(B) = \sum_{k\in K} P(B|A_k)P(A_k)$$

Théorème :(Bayes) soit $\{A_k\}_{k\in K}$ une partition[4] de E ; alors pour tout événement non négligeable B on a

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k\in K} P(B|A_k)P(A_k)}$$

Preuve :

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} =^5 \frac{P(B|A_i)P(A_i)}{\sum_{k\in K} P(B|A_k)P(A_k)}$$

Événements indépendants

Définition : deux événements A et B sont indépendants ssi l'une des conditions équivalentes suivantes est vérifiée :
- $P(A|B) = P(A)$
- $P(B|A) = P(B)$
- $P(A \cap B) = P(A)P(B)$

Définition : les événements $e_1, \ldots, e_n$ sont conditionnellement indépendants par rapport à h si

$$P(e_1 \cap \ldots \cap e_n|h) = P(e_1|h) \ldots P(e_n|h)$$

Bayes pour événements indépendants

Théorème :(Bayes) soit $\{A_k\}_{k\in K}$ une partition[6] de E ; alors pour tout événements non négligeables et conditionnellement indépendants $e_1, \ldots, e_n$ on a

$$P(A_i|e_1 \cap \ldots \cap e_n) = \frac{P(e_1|A_i) \ldots P(e_n|A_i)P(A_i)}{\sum_{k\in K} P(e_1|A_k) \ldots P(e_n|A_k)P(A_k)}$$

[3] $A_i \cap A_j = \emptyset \cup_{k\in K}\{A_k\} = E$
[4] $A_i \cap A_j = \emptyset \cup_{k\in K}\{A_k\} = E$
[5] def de $P(B|A_i)$ et théorème précédent
[6] $A_i \cap A_j = \emptyset \cup_{k\in K}\{A_k\} = E$

Une vraie application : filtrage Bayesien du spam

On veut filtrer les message contenant du spam :

- on calcule la fréquence $M(auvais)$ des mots apparaissant dans une boîte-à-lettres contenant du spam
- on calcule la fréquence $B(on)$ des mots apparaissant dans une boîte-à-lettres contenant des vrais messages
- une fois reçu un nouveau message, on prend les 15 mots $m_1, \dots, m_{15}$ ayant la probabilité la plus éloignée de 0.5
- on calcule $P(M|m_1\ et \dots\ et\ m_{15})$ et on filtre le message si ell'est plus grande de 0.9

Problème : comment calculer $P(M|m_1\ et \dots\ et\ m_{15})$ en ne connaissant que les fréquences des m_i dans M et B ?

Paul Grapham a proposé en 2002, dans "A plan for spam" une stratégie basée sur ce raisonnement, avec des heuristiques qui ont été ensuite incorporés, par exemple, dans `SpamOracle`.

Variable aléatoire

Définition : une fonction $X : E \to \mathbb{R}$ est dite une variable aléatoire

Exemple : On tire une pièce 3 fois. Soit $X(t)$ le nombre de "faces" obtenues. Alors

$$X(FFF) = 3 \quad X(FFP) = X(FPF) = X(PFF) = 2$$

$$X(PPP) = 0 \quad X(FPP) = X(PPF) = X(PFP) = 1$$

Une variable aléatoire assigne une valeur à tout événement.

Définition : on écrira $p(X = v)$ pour

$$P(X^{-1}(v))$$

Espérance mathématique

Étant donnée une variable aléatoire $X : E \to \mathbb{R}$, on peut chercher à en trouver la "valeur moyenne" :

Définition : l'espérance mathématique d'une variable aléatoire $X : E \to \mathbb{R}$ définie sur l'espace de probabilité E, équipé de la distribution de probabilité P, est

$$E(X) = \sum_{e_i \in E} p(e_i) X(e_i)$$

On a bien que

$$E(X) = \sum_{e_i \in E} p(e_i) X(e_i) = \sum_{v \in X(E)} p(X = v) v$$

Distribution Binomiale

Définition : Une expérience de Bernoulli est une expérience pouvant avoir seulement deux valeurs[7], avec probabilité p de succès et $1-p$ d'échec.

Définition : Considérons une variable aléatoire X égale aux nombre de succès dans une suite de n expériences de Bernoulli indépendantes : la probabilité d'avoir exactement k succès suit la distribution binomiale

$$p(X = k) = \mathcal{C}_n^k q^k (1-q)^{n-k}$$

et on a

$$E(X) = nq$$

Propriétés de l'espérance

Théorème : Pour $X_1, \ldots, X_n$ variables aléatoires, on a

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) \quad (4)$$

$$E(aX_1 + b) = aE(X_1) + b \quad (5)$$

Indépendance

Définition : deux variables aléatoires X_1 et X_2 sont indépendantes si

$$p(X_1 = v \; et \; X_2 = w) = p(X_1 = v) p(X_2 = w)$$

Proposition : si deux variables aléatoires X_1 et X_2 sont indépendantes, alors

$$E(X_1 X_2) = E(X_1) E(X_2)$$

Variance et déviation standard

Définition : on définit la variance $V(X)$ d'une variable aléatoire X comme

$$V(X) = E((X - E(X))^2)$$

et la déviation standard σ de X comme

$$\sigma = \sqrt{V(X)}$$

Exercice : prouvez que $V(X) = E(X^2) - E(X)^2$

[7] typiquement, succès et échec

Inégalité de Chebischev

Théorème : pour X variable aléatoire, et pour tout $\alpha > 0$

$$p\left((X - E(X))^2 \geq \alpha\right) \leq V(X)/\alpha$$

D'où

$$p\left(|X - E(X)| \geq c\sigma\right) \leq 1/c^2$$

Donc, X tombe à plus de $c\sigma$ de $E(X)$ avec probabilité $\leq 1/c^2$. Cela dit, par exemple, que *au moins dans 75% de cas* X est à moins de 2σ de $E(X)$.