Statistique et Informatique (3I005)

2016-2017

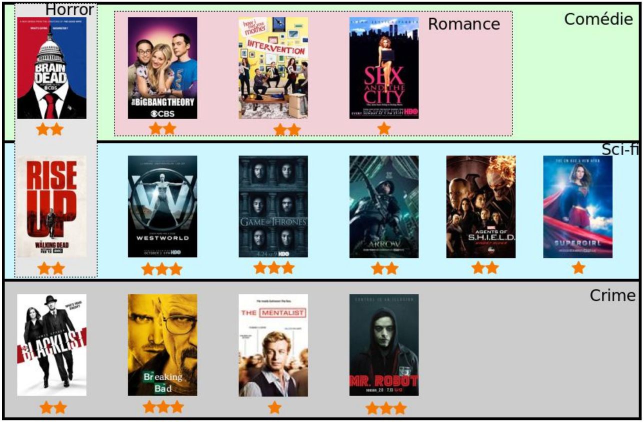
Nicolas Baskiotis

Université Pierre et Marie Curie (UPMC)
Laboratoire d'Informatique de Paris 6 (LIP6)
http://3I005.lip6.fr

Cours 3 :

Variables aléatoires
Fonctions de répartition

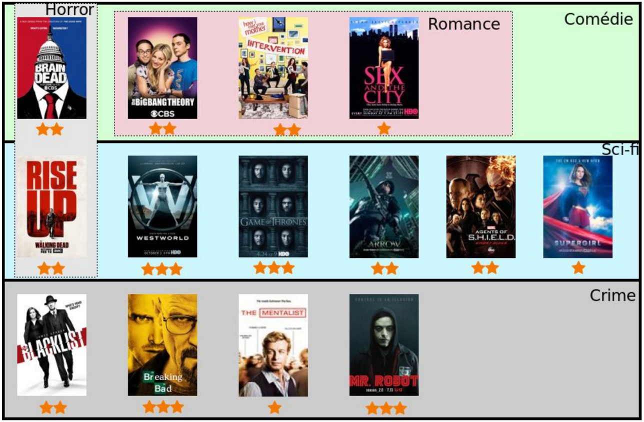
Rappel des cours précédents



Notions fondamentales

- Univers, Événement, Mesure de probabilité, Espace probabilisé
- Incompatibilité, Indépendance, Conditionnement

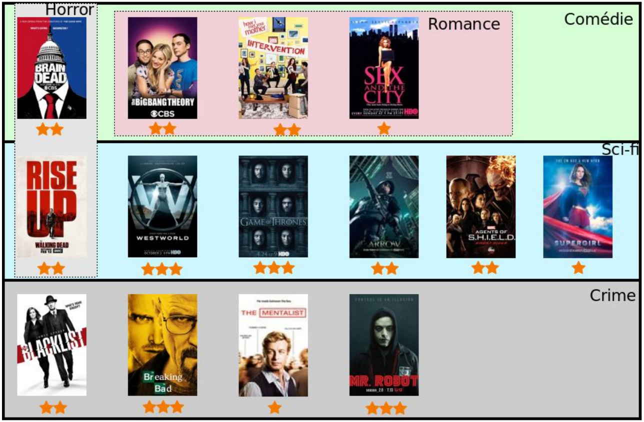
Rappel des cours précédents



Notions fondamentales

- On tire une série au hasard :
  - événement élémentaire ?
  - événements incompatibles ?
  - événements indépendants ?

Rappel des cours précédents



Notions fondamentales

- On tire une série au hasard :
  - événement élémentaire ? une série
  - événements incompatibles ? Comédie et Crime
  - événements indépendants ? Score$=2$ (S_2) et Sci-Fi

Plan

1. Théorème de Bayès
2. Variable aléatoire
3. Fonction de répartition

Probabilités Conditionnelles

Considérons deux événements E et F, Supposons qu'on ne s'intéresse à la réalisation de E, étant donnée la réalisation de F. Cela revient à estimer la réalisation de $E \cap F$ par rapport à F

Définition

Soit Ω un ensemble dénombrable et P une mesure de probabilisé sur Ω. Soit F un événement *de probabilité non nulle*. On appelle probabilité conditionnelle sachant F l'application:

$$P(. \mid F) : \mathcal{P}(\Omega) \to [0, 1]$$

définie par

$$\forall E \in \mathcal{P}(\Omega), P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Cette application est une mesure de probabilité sur Ω.
Note : $P(E \mid F)$ se lit "probabilité de E sachant F".

Formule de Bayes, théorème des probabilités totales

Formule de Bayes

Soient E et F deux événements de probabilité non nulle. Alors :

$P(E \cap F) = P(F \mid E) \times P(E) = P(E \mid F) \times P(F)$, soit

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

Théorème des probabilités totales

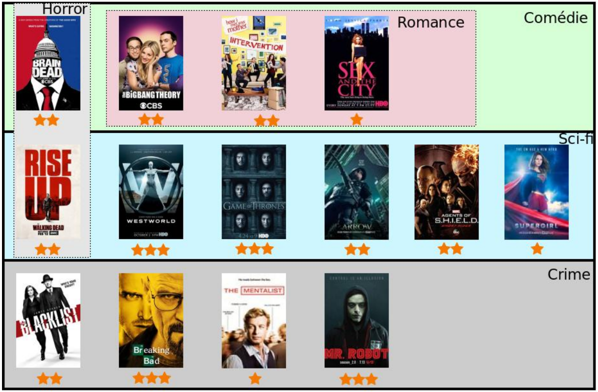
Soit $(F_i)_i$ une partition de Ω (aussi appelé ensemble complet d'événements) :

- si $i \neq j$ alors $F_i \cap F_j = \emptyset$ (F_i et F_j sont incompatibles),
- $\bigcup_i F_i = \Omega$.

Alors $\forall E \subset \Omega, P(E) = \sum_i P(E \cap F_i) = \sum_i P(E|F_i)P(F_i)$.

De plus, pour tout i, $P(F_i|E) = \dfrac{P(E \mid F_i) \times P(F_i)}{\sum_{j=1}^{N} P(E \mid F_j) \times P(F_j)}$.

Formule de Bayès : exemple



Conditionnement, indépendance

- $P(S_2|\text{Comédie}) = P(S_2 \cap \text{Comédie})/P(\text{Comédie}) = \frac{3}{4}$
- $P(S_2 \cap \text{Comédie}) = P(S_2|\text{Comédie}) \times P(\text{Comédie}) = P(\text{Comedie}|S_2) \times P(S_2)$

$\Rightarrow$ $P(S_2|\text{Comédie}) = P(\text{Comédie}|S_2) \times P(S_2)/P(\text{Comédie})$ (formule de Bayes)

Probabilités totales : exemple

Décomposition de la probabilité de score $= 2$

$$P(S_2) = P(S_2 \cap \text{Comédie}) + P(S_2 \cap \text{SciFi}) + P(S_2 \cap \text{Crime})$$
$$= P(S_2|\text{Comédie})P(\text{Comédie}) + P(S_2|\text{SciFi})P(\text{SciFi}) + P(S_2|\text{Crime})P(\text{Crime})$$

Probabilités Conditionnelles

Application en chaîne de la formule des probabilités conditionnelles

- Par définition, si $P(F) \neq 0$, on a $P(E \cap F) = P(E|F)P(F)$
- Plus généralement, si $E_1, ..., E_n$ sont n événements, on a :

$$P(E_1 \cap E_2 \cap ... \cap E_n) = P(E_1) \prod_{i=2}^{n} P(E_i \mid E_1 \cap ... \cap E_{i-1})$$

Exemple

Quelle est la probabilité de tirer trois boules de la même couleur dans une urne contenant 7 boules rouges et 5 boules bleues, en tirant les trois boules l'une après l'autre et sans remise?

Probabilités Conditionnelles

Application en chaîne de la formule des probabilités conditionnelles

- Par définition, si $P(F) \neq 0$, on a $P(E \cap F) = P(E|F)P(F)$
- Plus généralement, si $E_1, ..., E_n$ sont n événements, on a :

$$P(E_1 \cap E_2 \cap ... \cap E_n) = P(E_1) \prod_{i=2}^{n} P(E_i \mid E_1 \cap ... \cap E_{i-1})$$

Exemple

Quelle est la probabilité de tirer trois boules de la même couleur dans une urne contenant 7 boules rouges et 5 boules bleues, en tirant les trois boules l'une après l'autre et sans remise?

Posons

- R_i=`La` i^{eme} `boule tirée est rouge`, $i \in \{1,2,3\}$
- B_i=`La` i^{eme} `boule tirée est bleue`, $i \in \{1,2,3\}$

On a alors $P(R_1 \cap R_2 \cap R_3) = P(R_1)P(R_2|R_1)P(R_3|R_2 \cap R_1) = \frac{7}{12} \times \frac{6}{11} \times \frac{5}{10}$.

De même, $P(B_1 \cap B_2 \cap B_3) = \frac{5}{12} \times \frac{4}{11} \times \frac{3}{10}$

Formule de Bayes : exemple

Exemple

On tire successivement et sans remise 4 lettres du mot “ATTACHANT” Quelle est la probabilité d’obtenir “CHAT” ?

Exemple

On enlève aléatoirement une carte d’un jeu de 52 cartes, et on ignore laquelle. On tire ensuite au hasard une carte dans ce jeu incomplet et c’est un coeur. Quelle est la probabilité pour que la carte manquante soit un coeur?

Formule de Bayes : exemple

Exemple

On enlève aléatoirement une carte d'un jeu de 52 cartes, et on ignore laquelle. On tire ensuite au hasard une carte dans ce jeu incomplet et c'est un coeur. Quelle est la probabilité pour que la carte manquante soit un coeur?

On considère les événements suivants:

- CP: La carte perdue est un coeur
- TC: Tirer un coeur du jeu incomplet

On cherche $P(CP|TC)$.

Formule de Bayes : exemple

Exemple

On enlève aléatoirement une carte d'un jeu de 52 cartes, et on ignore laquelle. On tire ensuite au hasard une carte dans ce jeu incomplet et c'est un coeur. Quelle est la probabilité pour que la carte manquante soit un coeur?

On considère les événements suivants:

- CP: La carte perdue est un coeur
- TC: Tirer un coeur du jeu incomplet

On cherche $P(CP|TC)$.
Nous avons alors $P(CP) = \frac{1}{4}$ et $P(TC \mid CP) = \frac{12}{51}$
TC peut s'écrire comme: $TC = (TC \cap CP) \cup (TC \cap \bar{C}P)$ et

$$P(CP \mid TC) = \frac{P(TC \mid CP) \times P(CP)}{P(TC \mid CP) \times P(CP) + P(TC \mid \bar{C}P) \times P(\bar{C}P)} = \frac{\frac{12}{51} \times \frac{1}{4}}{\frac{12}{51} \times \frac{1}{4} + \frac{13}{51} \times \frac{3}{4}} = \frac{12}{51}$$

Exercice

Problème d'urnes

On dispose de deux urnes :

- l'urne 1 contient 3 boules blanches et 1 noire,
- l'urne 2 une boule blanche et 2 noires.

On lance un dé non truqué, si le dé donne 1 ou 2 on tire une boule dans l'urne 1; sinon dans l'urne 2.

- Calculer la probabilité de tirer une boule blanche.
- On a tiré une boule blanche; probabilité qu'on ait tiré dans l'urne 1 ?

Exercice

Rat de laboratoire

Une expérience est conduite pour étudier la mémoire des rats. Un rat est mis devant trois couloirs. Au bout de l'un d'eux se trouve de la nourriture qu'il aime, au bout des deux autres, il reçoit une décharge électrique. Cette expérience élémentaire est répétée jusqu'à ce que le rat trouve le bon couloir. Sous chacune des hypothèses suivantes, avec quelle probabilité la première tentative réussie est-elle la k-ème ?

- le rat n'a aucun souvenir des expériences précédentes,
- le rat se souvient uniquement de l'expérience précédente,
- le rat se souvient des deux expériences précédentes.

Plan

Variable aléatoire discrète

Intuition

Lorsqu'on est face à une expérience aléatoire, on s'intéresse plus souvent à une *valeur* attribuée au résultat qu'au résultat lui-même.

- problème du dénombrement : trop long à énumerer, peu informatif.
- solution : "traduire" l'univers en évènements "compréhensibles" et ordonnés (avec une valeur pouvant faire sens).

$\Rightarrow$ variable aléatoire discrète : application de l'univers vers un espace discret .

- intérêt : enfin pouvoir "calculer" autre chose que des probabilités.

Exemples

- Lors d'un jeu, on s'intéresse plus au gain que l'on peut obtenir qu'au résultat exact du jeu.
- Lorsqu'on joue au blackjack, on s'intéresse plus à la probabilité de faire un 21 que des configurations élémentaires donnant 21.

Exemple

Exemple du lancer de dé

On lance un dé après avoir misé 1€. Si le résultat est un 5 ou un 6 on double la mise, sinon perd la mise. Dans ce cas:

- $\Omega = \{D1, D2, D3, D4, D5, D6\}$
- Card $\Omega = 6$, Card $\mathcal{P}(\Omega) = 2^6$, et $\forall e \in \Omega, P(e) = \frac{1}{6}$
- Comment calculer la probabilité de gagner 1€?

Exemple

Exemple du lancer de dé

On lance un dé après avoir misé 1€. Si le résultat est un 5 ou un 6 on double la mise, sinon perd la mise. Dans ce cas:

- $\Omega = \{D1, D2, D3, D4, D5, D6\}$
- Card $\Omega = 6$, Card $\mathcal{P}(\Omega) = 2^6$, et $\forall e \in \Omega, P(e) = \frac{1}{6}$
- Comment calculer la probabilité de gagner 1€?
- Soit X la $v.a.$ qui associe à tout résultat du dé un *gain*:
  $X(D1) = X(D2) = X(D3) = X(D4) = -1$
  $X(D5) = X(D6) = (2-1) = 1$
  X est à valeur dans l'ensemble noté $\mathcal{X} = \{-1, 1\} \subset \mathbb{R}$
  $X : \Omega \rightarrow \mathcal{X}$
- $X^{-1}(1)$: l'ensemble des évènements élémentaires correspondant au gain d'1€
- $P(X^{-1}(\{1\})) = P($*Le résultat du dé est* 5 *ou* $6)$.

$\Rightarrow$ Définir une probabilité sur $\mathcal{X}$, notée $\mathbb{P}$, en retournant dans l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$
$\mathbb{P}(\{1\}) = P(X^{-1}(\{1\})) = P($*Le résultat du dé est* 5 *ou* $6)$.

Variable aléatoire à valeurs discrètes

Définition

Soit Ω un ensemble dénombrable, et P une mesure de probabilité sur Ω.
Soit Ω', un ensemble discret.
Une variable aléatoire est une fonction X de Ω muni de la mesure P vers Ω'.

Exemples

- Lancer d'un dé :
  Soit $\Omega = \{1, ..., 6\}$ muni de la probabilité uniforme P.
  $$X : i \mapsto \begin{cases} 1 \text{ si } i \text{ est pair} \\ 0 \text{ sinon} \end{cases}$$
  est une variable aléatoire de (Ω, P) vers $\Omega' = \{0, 1\}$.
- Lancer de deux dés :
  Soit $\Omega = \{1, ..., 6\}^2$ muni de la probabilité uniforme P.
  $X : (i,j) \mapsto i + j$
  est une variable aléatoire de (Ω, P) vers $\Omega' = \{2, ..., 12\}$

Loi de probabilité

Définitions

Soit (Ω, P) un espace probabilisé où Ω est dénombrable.
Soit Ω' un ensemble discret, et X une v.a. de (Ω, P) vers Ω'.

- X définit une mesure de probabilité sur Ω', notée P_X, par:
  pour tout sous-ensemble E' de Ω' :

  $$P_X(E') = P\big(X^{-1}(E')\big)$$

  avec $X^{-1}(E') = \{\omega \in \Omega | X(\omega) \in E'\}$
- L'ensemble des valeurs $P_X\big(\{\omega'\}\big)$ pour $\omega' \in \Omega'$ s'appelle la *loi de probabilité* de X.

Notations

- L'événement $X \in]-\infty, a]$ sera noté par $X \leq a$
- L'événement $X \in]a, b]$ sera noté par $a < X \leq b$
- L'événement $X \in \{a\}$ sera noté par $X = a$
- On a donc $P_X(B) = P(X^{-1}(B)) = P(X \in B)$

Loi d'une variable aléatoire

Propriété

Une variable aléatoire est totalement définie par sa loi de probabilité, caractérisé par :

- son domaine de définition : l'ensemble des valeur qu'elle peut prendre,
- les probabilités attribuées à chacune de ses valeurs $P(X = x)$.

Questions :

soit Ω un ensemble de cardinal n,

- quel est le plus grand cardinal de l'ensemble des valeurs d'une application de Ω ?
- combien d'applications de Ω vers $\{1, ..., n\}$ différentes existe-t-il ?

Exemple

Jeux de hasard

- On lance un dé après avoir misé 3 euros. Si le résultat est $1, 2, 3$ ou 4, on perd la mise. Sinon, on triple la mise.
  Quelle est la probabilité de gagner 6 euros ? de perdre sa mise ?
- Le joueur décide de jouer deux fois de suite. Quelle est la loi de probabilité du gain sur l'ensemble des deux lancers ?

Exemple

Jeux de hasard

- On lance un dé après avoir misé 3 euros. Si le résultat est $1, 2, 3$ ou 4, on perd la mise. Sinon, on triple la mise.
  Quelle est la probabilité de gagner 6 euros ? de perdre sa mise ?

  $\Omega = \{1, ..., 6\}$ muni de la probabilité uniforme. On note X la v.a. qui représente le gain:
  $$P(X = 6) = P\big(\{5, 6\}\big) = \frac{1}{3}$$
  $$P(X = -3) = \frac{2}{3}$$
- Le joueur décide de jouer deux fois de suite. Quelle est la loi de probabilité du gain sur l'ensemble des deux lancers ?

Exemple

Jeux de hasard

- On lance un dé après avoir misé 3 euros. Si le résultat est $1, 2, 3$ ou 4, on perd la mise. Sinon, on triple la mise.
  Quelle est la probabilité de gagner 6 euros ? de perdre sa mise ?

  $\Omega = \{1, ..., 6\}$ muni de la probabilité uniforme. On note X la v.a. qui représente le gain:
  $$P(X = 6) = P\big(\{5, 6\}\big) = \frac{1}{3}$$
  $$P(X = -3) = \frac{2}{3}$$

- Le joueur décide de jouer deux fois de suite. Quelle est la loi de probabilité du gain sur l'ensemble des deux lancers ?

  L'univers est $\{1, ..., 6\}^2$, muni de la loi de probabilité uniforme.
  Le gain total G suit la loi :
  $$P(G = -6) = P\big(\{1, ..., 4\}^2\big) = \frac{16}{36}, \quad P(G = 3) = ?,\ P(G = 12) = ?$$

Plan

Fonction de répartition

Définition

Soit X une $v.a.$
on appelle **fonction de répartition** de X, notée F la fonction:

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0,1] \\ x &\mapsto F(x) = P(X \leq x) \end{aligned}$$

Propriétés

- $P(a < X \leq b) = F(b) - F(a)$
- F est croissante bornée:
  - $F(-\infty) = \lim\limits_{x \rightarrow -\infty} F(x) = 0$
  - $F(+\infty) = \lim\limits_{x \rightarrow +\infty} F(x) = 1$

Fonction de répartition (suite)

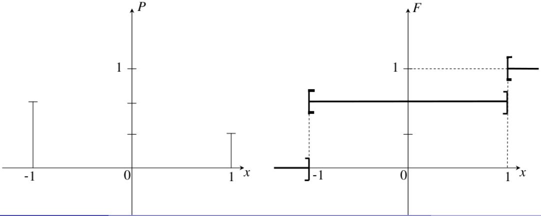
Exemple: Lancer de dé

Avec l'exemple du lancer de dé précédent nous avons $\Omega = \{1, 2, 3, 4, 5, 6\}$, et $X : \Omega \to \mathbb{R}$ définie par:

$$X(e) = -1 \text{ si } e \in \{1, 2, 3, 4\}, \text{ et } X(e) = +1 \text{ si } e \in \{5, 6\}$$

L'ensemble des valeurs possibles est $\mathcal{X} = \{-1, 1\}$. On peut alors caractériser la loi de X par sa fonction de répartition $F : F(x) = P(X \leq x)$

- si $x < -1, \mathcal{X} \cap (]-\infty, x]) = \emptyset \Rightarrow F(x) = 0$
- si $x \in [-1, 1[, \mathcal{X} \cap (]-\infty, x]) = \{-1\} \Rightarrow F(x) = \frac{2}{3}$
- si $x \in [1, \infty[, \mathcal{X} \cap (]-\infty, x]) = \{-1, 1\} \Rightarrow F(x) = \frac{2}{3} + \frac{1}{3} = 1$



Variables aléatoires indépendantes

Définitions

Soit (Ω, P) un espace probabilisé.

- Soient X et X' deux v.a. de Ω vers Ω'.
  Les variables X et X' sont indépendantes si :

$$\forall A \subset \Omega', \forall B \subset \Omega', P(X \in A \cap X' \in B) = P(X \in A)P(X' \in B)$$

- Soient X_1, ..., X_n, n v.a. de Ω vers Ω'.
  $X_1, ..., X_n$ sont mutuellement indépendantes si, pour tous sous-ensembles $E_1, ..., E_n$ de Ω', on a:

$$P\left(\bigcap_{i\in\{1,\cdots,n\}} X_i \in E_i\right) = \prod_{i\in I} P(X_i \in E_i)$$

Retour à l'exemple précédent

Si on lance deux fois un dé, avec à chaque fois un gain si le résultat est 5 ou 6, les gains obtenus à chacun des lancers sont indépendants.

Exemple

Résultats de n répétitions indépendantes d'une expérience aléatoire

Soit (Ω, P) un espace probabilisé, X une v.a. sur Ω vers Ω'.
On note $\Omega_n = \Omega^n$, et P_n la mesure produit sur Ω_n :

$$\forall \omega = (\omega_1, ..., \omega_n) \in \Omega_n, P_n(\{\omega\}) = \prod_{i=1}^{n} P(\omega_i)$$

On note $X_i : \omega \in \Omega_n \mapsto X(\omega_i)$.
Les v.a. X_i sont mutuellement indépendantes et suivent la même loi que X :

$$\forall i, \forall E' \subset \Omega', P_n(X_i \in E') = P(X \in E')$$

Retour à l'exemple précédent (2)

Si on lance n fois un dé, avec à chaque fois un gain si le résultat est 5 ou 6.
Ω_n est l'ensemble des réalisations possibles des n lancers.
Pour un événement élémentaire $\omega = (\omega_1, ..., \omega_n)$:

- ω_i est le résultat du i-ième lancer,
- X_i représente le gain obtenu au i-ième lancer.

Deux lois de probabilités discrètes importantes

Loi de Bernoulli

La loi de Bernoulli est la loi d'une $v.a.$ X à valeur dans $\{0, 1\}$.
$X = 1$ représente le "succès" de l'expérience, et $X = 0$ l'"échec".

$$\forall x \in \{0, 1\}, P(X = x) = p^x(1 - p)^{1-x}$$

La probabilité de succès $p = P(X = 1)$ est le paramètre de la loi.

Loi binomiale

Soit X, le nombre de succès d'une épreuve de Bernoulli de paramètre p, répétée n fois indépendamment. La loi de X est appelée la *loi binomiale* de paramètres n et p :

$$\forall k \in \{0, ..., n\}, P(X = k) = C_n^k p^k(1 - p)^{n-k}$$

Loi conjointe

Definition

Soit (Ω, P) un espace de probabilité, et soient X et Y deux $v.a.$ sur cet espace, à valeur resp. dans F et G. (X, Y) est une $v.a.$, appelée loi conjointe de X et Y; les valeurs de (X, Y) sont dans $F \times G$.

Propriétés

- la connaissance uniquement de X et de Y ne suffit pas à connaitre la loi jointe, sauf si X est indépendant de Y.
- $\forall x \in F, P(X = x) = \sum_{y \in G} P(X = x, Y = y)$

$\Rightarrow$ la connaisance de la loi jointe permet de déduire la loi de X, appelée dans ce cas *loi marginale*.

Exemple

Soit (X, Y) un couple de $v.a.$ de loi telle que $P((X, Y) = (i,j)) = 1/9$ ssi $0 \leq i \leq 2$ et $-i \leq j \leq i$.

- Quelle est la représentation graphique de la loi ?
- Quelles sont les lois marginales de X et de Y ?

Application

Probabilités infinis

On lance un dé équiprobable à 6 faces, de manière indépendante, et on note

- E_n l'évènement "le premier 5 est au rang n"
- A l'évènement "le 5 apparaît avant le 2"

Calculer A en calculant $P(A \cap E_n)$.

Test de primalité de Fermat

1. répéter k fois :
2. soit a un nombre aléatoire entre 1 et $n-1$
3. si a^{n-1} mod n $!=\ 1$ alors
4. retourner "n n'est pas premier"
5. finsi
6. fin de la boucle
7. retourner "n est probablement premier"

Principe

- Algorithme probabiliste :
  fait appel à un générateur de nombre aléatoire durant son exécution.
- Si n est premier, $\forall a \in \{1, ..., n-1\}$, a^{n-1} mod $n == 1$,
  si n n'est pas premier et n'est pas un nombre de Carmichael (très rares), plus de la moitié des entiers a entre 1 et $n-1$ vérifient : a^{n-1} mod n $!=\ 1$.

Test de primalité de Fermat (2)

1. répéter k fois :
2.   soit a un nombre aléatoire entre 1 et $n-1$
3.   si a^{n-1} mod n $\; != \; 1$ alors
4.   retourner "n n'est pas premier"
5.   finsi
6. fin de la boucle
7. retourner "n est probablement premier"

Question

Supposons que n n'est pas premier, ni un nombre de Carmichael.
Pour quelles valeurs de k l'algorithme retourne-t'il

"n n'est pas premier"

avec une probabilité plus grande que $1 - 10^{-6}$?