

Zététique

Fiche de TD

L3 Informatique 2021–2022

Probabilités et hasard

Exercice 1. (Probabilités élémentaires)

Un *univers* est un ensemble (fini ici) $\Omega = \{e_1, \dots, e_n\}$ où les e_i sont des *événements*. Chaque événement $e \in \Omega$ vient avec une *probabilité* $\mathbb{P}(e)$ qui mesure le risque pour e de se produire. On a les axiomes suivants

$$0 \leq \mathbb{P}(e) \leq 1, \quad \text{pour tout } e \in \Omega,$$
$$\sum_{e \in \Omega} \mathbb{P}(e) = 1.$$

De plus, la notation $\mathbb{P}$ est étendue aux sous-ensembles d' Ω : pour tout $E \subseteq \Omega$, on définit

$$\mathbb{P}(E) := \sum_{e \in E} \mathbb{P}(e).$$

Cette valeur quantifie donc le risque pour qu'un événement de E se produise.

1. Étant donnés deux ensembles d'événements E_1 et E_2 tels que $E_1 \cap E_2 = \emptyset$, exprimer la probabilité $\mathbb{P}(E_1 \cup E_2)$ de $E_1 \cup E_2$. Il s'agit, par définition, de la probabilité qu'un événement de E_1 ou de E_2 se produise.
2. Étant donnés deux ensembles d'événements E_1 et E_2 , exprimer la probabilité $\mathbb{P}(E_1 \cap E_2)$ de $E_1 \cap E_2$. Il s'agit, par définition, de la probabilité qu'un événement qui appartienne à la fois à E_1 et à E_2 se produise. Exprimer également cette probabilité dans les cas particuliers où E_1 et E_2 sont disjoints et où E_1 est un sous-ensemble de E_2 .
3. Étant donné un ensemble d'événements E , exprimer la probabilité $\mathbb{P}(\bar{E})$ de $\bar{E} := \Omega \setminus E$. Il s'agit, par définition, de la probabilité qu'un événement qui n'appartienne pas à E se produise.
4. Étant donnés deux ensembles d'événements E_1 et E_2 , exprimer la probabilité $\mathbb{P}(E_2 \setminus E_1)$. Il s'agit, par définition, de la probabilité qu'un événement qui appartienne à E_2 mais pas à E_1 se produise. Exprimer également cette probabilité dans le cas particulier où E_1 est un sous-ensemble de E_2 .
5. Les événements d' Ω sont *équiprobables* s'ils ont tous la même probabilité. Exprimer la probabilité $\mathbb{P}(e)$ de tout événement e d' Ω .

Exercice 2. (Indépendance)

Soit $\Omega := \{e_1, \dots, e_n\}$ un univers. Deux ensembles d'évènements E_1 et E_2 sont *indépendants* si $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2)$.

1. On décide de lancer un dé à 6 faces non pipé. Soit E_1 l'ensemble d'évènements « le nombre obtenu est pair » et E_2 l'ensemble d'évènements « le nombre obtenu est un multiple de 3 ». Formaliser le problème en décrivant explicitement Ω , E_1 et E_2 . Ensuite, déterminer si E_1 et E_2 sont indépendants.
2. On décide de lancer deux dés à 6 faces non pipés D et D' . Soit E_1 l'ensemble d'évènements « la somme des nombres obtenus sur D et D' est 6 » et E_2 l'ensemble d'évènements « le nombre obtenu sur D est pair ». Formaliser le problème en décrivant explicitement Ω , E_1 et E_2 . Ensuite, déterminer si E_1 et E_2 sont indépendants.
3. Une urne contient 12 boules numérotées de 1 à 12. Soit E_1 l'ensemble d'évènements « la boule tirée porte un numéro pair » et E_2 l'ensemble d'évènements « la boule tirée porte un numéro multiple de 3 ». Formaliser le problème en décrivant explicitement Ω , E_1 et E_2 . Ensuite, déterminer si E_1 et E_2 sont indépendants.
4. Répondre à la question précédente mais où il y a maintenant 13 boules numérotées de 1 à 13.

Exercice 3. (Espérance)

Soit Ω un univers dans lequel chaque évènement e est accompagné d'un gain $g(e) \in \mathbb{R}$. L'espérance d' Ω est la valeur

$$\mathbb{E}(\Omega) := \sum_{e \in \Omega} \mathbb{P}(e) g(e).$$

Intuitivement, $\mathbb{E}(\Omega)$ est la valeur gagnée en moyenne en tirant au hasard (en suivant bien entendu les probabilités) un évènement d' Ω .

1. Un ami vous propose de lancer un dé à 6 faces non pipé. Vous gagnez $k \in$ si le dé affiche le côté k pour $1 \leq k \leq 5$ et perdez 6 € si le dé affiche le côté 6. Devez-vous jouer ?
2. Un ami vous propose de tirer deux cartes d'un jeu de 52 cartes. Si les deux cartes tirées sont de la même couleur (parmi les quatre couleurs habituelles : pique, trèfle, cœur et carreau), vous gagnez 3 € et perdez 1 € dans le cas contraire. Devez-vous jouer ?
3. Même question que la précédente, à la différence que c'est un jeu de 32 cartes qui est proposé.
4. Un ami vous propose de lancer un dé à 6 faces non pipé. Vous gagnez 100000 € si le dé affiche le côté 6 et perdez 10000 € dans les autres cas. Devez-vous jouer ?

Exercice 4. (Expériences répétées)

Soit $\Omega := \{e_1, \dots, e_n\}$ un univers. On s'intéresse ici au nouvel univers Ω^k dont les événements sont les suites de longueur $k \geq 0$ fixée d'événements d' Ω . Comme par définition les événements d' Ω sont indépendants, la probabilité de l'événement $(f_1, \dots, f_k)$ d' Ω^k vérifie

$$\mathbb{P}((f_1, \dots, f_k)) = \mathbb{P}(f_1) \dots \mathbb{P}(f_k).$$

1. Soit $\Omega := \{p, f\}$ l'univers dans lequel p (resp. f) est l'événement dans lequel une pièce équilibrée tombe sur pile (resp. face). Calculer la probabilité de l'événement (p, p, p, p) et l'événement (p, f, p, f) d' Ω^4 .
2. Soit $\Omega := \{r_1, r_2, r_3, r_4, r_5, r_6\}$ l'univers dans lequel r_i , $1 \leq i \leq 6$, est l'événement dans lequel un dé pipé à 6 faces montre la face i . On sait ici que l'on a

$$\mathbb{P}(r_1) = \frac{1}{4}, \quad \mathbb{P}(r_2) = \frac{1}{4}, \quad \mathbb{P}(r_3) = \frac{1}{8}, \quad \mathbb{P}(r_4) = \frac{1}{16}, \quad \mathbb{P}(r_5) = \frac{1}{8}, \quad \mathbb{P}(r_6) = \frac{3}{16}.$$

Calculer la probabilité de l'événement (r_3, r_1, r_6) d' Ω^3 et de l'événement $(r_1, r_2, r_1, r_6, r_3)$ d' Ω^5 .

Exercice 5. (Hasard et suites de bits)

Une suite de bits (c'est-à-dire, un mot fini sur l'alphabet $\{0, 1\}$) u est dite *aléatoire* (ou survenue au *hasard*) si u ne renferme aucune régularité. Autrement dit, il n'est possible de décrire u qu'en l'explicitant littéralement. L'objectif ici est d'étudier une heuristique qui permet d'affirmer avec un indice de confiance élevé qu'une suite de bits n'est pas aléatoire.

1. Soient les suites de bits

$$u^{(1)} := 10101011010101010100100110010100101010101001011010010101,$$

$$u^{(2)} := 01001101100110101101010011100101101100111010001011010101010101,$$

$$u^{(3)} := 100001101110110000100011010111011000110111101111110110011000101,$$

$$u^{(4)} := 0110111001011101111000100110101011110011011110111110000100011001010011.$$

Essayer, intuitivement, de discerner parmi ces suites lesquelles sont aléatoires et lesquelles ne le sont pas.

Un *run* dans une suite de bits u est un facteur $u_i u_{i+1} \dots u_j$ tel que $u_i = u_{i+1} = \dots = u_j$. Par exemple, la suite de bits $u := 110001011$ contient neuf runs de longueur 1, quatre runs de longueur 2, un run de longueur 3, aucun run de longueur 4 et donc aucun run de longueur supérieure à 4.

2. Démontrer qu'en moyenne, la probabilité qu'une suite de bits u commence par un run de longueur $1 \leq k \leq |u|$ est

$$\frac{1}{2^{k-1}}.$$

3. Si u est une suite de bits de longueur n et $1 \leq k \leq n$ est un entier, donner le nombre de positions de u à partir desquelles un run de longueur k peut potentiellement commencer.
4. Dédurre des deux questions précédentes qu'étant donné une suite de bits u de longueur n , le nombre moyen de runs de longueur $1 \leq k \leq n$ est

$$\frac{n - k + 1}{2^{k-1}}.$$

5. En comparant les nombres moyens attendus de runs de toute longueurs raisonnables possibles et ceux effectifs dans les suites de bits $u^{(1)}$, $u^{(2)}$, $u^{(3)}$ et $u^{(4)}$ précédentes, distinguer celles qui sont aléatoires de celles qui ne le sont pas. Comparer ces résultats avec ceux de la première question.
6. Que peut-on dire en particulier de la suite $u^{(4)}$ et donc de la méthode proposée dans le cas général?

Exercice 6. (Hasard et complexité de Kolmogorov)

La *complexité de Kolmogorov* d'un objet x est la taille d'un plus petit programme permettant de construire x . Cette complexité dépend du langage utilisé pour exprimer l'algorithme, de la façon de mesurer la taille d'un programme et de la façon de représenter x . Cette mesure de complexité sert donc principalement à comparer les complexités d'objets plus qu'à les déterminer dans l'absolu.

1. Comparer les complexités de Kolmogorov des listes d'entiers

$$\ell^{(1)} := [1, 1, 2, 3, 5, 8, 13, 21, 34],$$

$$\ell^{(2)} := [1, 1, 2, 3, 5, 8, 10, 21, 34],$$

$$\ell^{(3)} := [6, 1, 9, -1, 26, 24, 25, 101, 11, 9, 4],$$

$$\ell^{(4)} := [1, 1, 2, 5, 14, 42, 132, 429, 1430].$$

La complexité de Kolmogorov offre une définition de la notion d'objet aléatoire : un objet x est *aléatoire* si sa complexité de Kolmogorov est maximale (autrement dit, il n'est pas possible de compresser x par un programme strictement plus petit en taille qui le génère).

2. En se basant sur cette notion d'objets aléatoires, déterminer quelles suites de bits de l'exercice 5 sont aléatoires. Même question pour les suites d'entiers de la question précédente.

Exercice 7. (Théorème de Bayes)

Soient Ω un univers et E_1 et E_2 deux ensembles d'évènements dans Ω . La probabilité qu'un évènement de E_2 se produise sachant qu'un évènement de E_1 s'est produit est notée $\mathbb{P}(E_2|E_1)$. Il est possible de voir $\mathbb{P}(E_2|E_1)$ comme la probabilité de E_2 **parmi** E_1 . Le *théorème de Bayes* indique la relation

$$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1|E_2)\mathbb{P}(E_2)}{\mathbb{P}(E_1)}$$

entre probabilités.

Rappelons un fait utile pour comprendre et démontrer cette formule. Lorsque Ω est un univers fini et E un ensemble d'évènements dans Ω , nous avons

$$\mathbb{P}(E) = \frac{\#E}{\#\Omega}$$

où $\#X$ désigne le cardinal de tout ensemble fini X .

1. En supposant qu' Ω est fini,

- (a) exprimer la probabilité de l'intersection $E_1 \cap E_2$, qui est donc la probabilité qu'un évènement commun à E_1 et à E_2 se produise ;
- (b) exprimer la probabilité de $E_2|E_1$, qui est donc la probabilité qu'un évènement de E_2 se produise dans l'univers local E_1 ;
- (c) exprimer la probabilité de $E_1|E_2$, qui est donc la probabilité qu'un évènement de E_1 se produise dans l'univers local E_2 ;
- (d) calculer $\mathbb{P}(B)\mathbb{P}(A|B)$ et donner une interprétation de cette quantité ;
- (e) calculer $\mathbb{P}(B)\mathbb{P}(A|B)$ et donner une interprétation de cette quantité ;
- (f) en déduire le théorème de Bayes.

2. On dispose de la collection de dés non pipés suivante : un dé à 6 faces, un dé à 12 faces, un dé à 16 faces, un dé à 32 faces et un dé à 64 faces. On sait que le numéro 5 vient d'être tiré par l'un des dés de la collection et on cherche à connaître pour chaque dé la probabilité qu'il ait été lancé. Utiliser le théorème de Bayes pour formaliser ce problème et y répondre. Même chose pour le numéro 19 et le numéro 33.

Exercice 8. (Faux positifs)

Une maladie M mortelle se diagnostique par le biais d'un test T fiable à 95%. Ceci signifie que 95% des malades sont déclarés par le test comme tels et que 95% des sujets sains sont déclarés par le test comme tels. Il s'avère que T vous déclare positif.

- 1. Quelle est la probabilité que vous soyez condamné ? Est-il possible de répondre à cette question ?
- 2. Répondre à la question précédente dans laquelle la prévalence de M est d'une personne sur 10000.
- 3. Répondre à la première question dans laquelle la prévalence de M est d'une personne sur 7.

Exercice 9. (Sondage inavouable)

On souhaite réaliser un sondage pour connaître la proportion, parmi les utilisateurs de France d'Internet, de ceux qui ont déjà visionné illégalement une série. Pour cela, on tire au sort un échantillon de 10000 individus parmi l'ensemble des utilisateurs du service et on leur envoie un lien vers un site sur lequel on leur demande de cocher la réponse à la question « Avez-vous déjà visionné une série illégalement sur Internet ? ».

1. Identifier le biais le plus évident dans le protocole donné.

Pour éviter que certains individus donnent l'aveu de leur activité illégale (ce sondage n'est pas répressif, il est juste à but informatif), on décide de mettre en place le mécanisme suivant. On demande à chaque utilisateur sondé de lancer un dé non pipé à 6 faces. Si celui tombe sur 6, l'utilisateur doit répondre « oui » à la question précédente. Sinon, il doit répondre honnêtement à la question.

2. Expliquer en quoi ce mécanisme pare le biais précédent.
3. Expliquer dans quelle mesure il est important que le dé soit lancé par l'utilisateur et non pas simulé par le site Internet du sondage.
4. Avec le protocole ainsi établi, on recueille 8883 réponses « oui » et 1117 réponses « non ». Comment pouvons-nous retrouver la véritable réponse à la question ?

Nous cherchons à améliorer encore le protocole en stipulant — toutes choses restant égales par ailleurs — que si le dé tombe sur 1, l'utilisateur doit répondre « non » à la question.

5. Donner le principal intérêt de cette variante par rapport au protocole précédent.
6. Expliquer dans ce nouveau cas comment retrouver les véritables proportions obtenues à l'issue du sondage.