Thème : probabilités conditionnelles et indépendance Série 2

Exercice 1

À vue d'oeil, il fait beau sept fois sur dix à Yverdon–les–Bains le jour de la rentrée académique. Votre enseignant de probabilités et statistique dispose de deux sources de prévisions météorologiques indépendantes : le service météorologique suisse qui se trompe deux fois sur cent et une grenouille verte, qui se trompe une fois sur vingt[1].

Considérons les événements suivants :

- E : "il fait beau à Yverdon–les–Bains le jour de la rentrée académique";
- F : "le service météorologique annonce le beau temps";
- G : "la grenouille prévoit le beau temps".

a) À l'aide de l'énoncé, donner les probabilités

1. $P(F \mid E)$ 2. $P(F \mid \overline{E})$ 3. $P(G \mid E)$ 4. $P(G \mid \overline{E})$

b) Calculer les probabilités

1. $P(\overline{F} \cap G \mid E)$ 2. $P(\overline{F} \cap G \mid \overline{E})$

Exercice 2

Pour éviter un incendie, une alarme a été installée dans la cuisine d'un restaurant australien. Supposons que la probabilité que l'alarme retentisse par erreur un jour donné sans qu'il y ait incendie est 0.01 et celle qu'elle retentisse en cas d'incendie vaut 0.95. De plus, on suppose que la probabilité qu'un incendie se déclare dans la cuisine un jour donné est 0.005.

a) Calculer la probabilité que l'alarme retentisse un jour donné.
b) Déterminer la probabilité qu'un incendie se déclare dans la cuisine un jour donné en sachant que l'alarme l'a signalé.

Exercice 3

Un signal binaire valant 0 ou 1 est envoyé par câble électrique. La transmission est affectée par des perturbations dites "bruits" : le signal 0 est reçu en 1 avec probabilité 0.2 et le signal 1 est enregistré en 0 avec probabilité 0.1.

[1] Toutes ces probabilités ont été soigneusement estimées par votre enseignant. On les accepte telles quelles.

Définissons les événements suivants :

$$\begin{array}{lcl} E_0 & : & \text{“le signal émis vaut 0”;} \\ E_1 & : & \text{“le signal émis vaut 1”;} \\ R_0 & : & \text{“la valeur reçue est 0”;} \\ R_1 & : & \text{“la valeur reçue est 1”.} \end{array}$$

En supposant que le système de transmission émet le signal 0 avec probabilité 0.45, calculer les probabilités de transmission correctes, à savoir :

$$P(E_0 \,|\, R_0) \quad \text{et} \quad P(E_1 \,|\, R_1).$$

Exercice 4

On jette de suite deux dés équilibrés. Considérons les événements :

- A : “la somme des faces vaut 6”;
- B : “le premier dé donne 4”;
- C : “la somme des faces vaut 7”.

Les événements A et B sont-ils indépendants ? Qu'en est-il des événements B et C ? Donnez une interprétation intuitive des résultats.

Exercice 5

Un système électronique est formé des composants A, B, C et D placés selon le dispositif de la Figure 1. On suppose que les composants fonctionnent indépendamment les uns des autres et leurs probabilités de fonctionnement sont toutes égales à 0.8. Le système est opérationnel s'il existe un chemin entre les points I et O ne comprenant que des composants qui fonctionnent. Calculer la probabilité que le système soit défaillant.

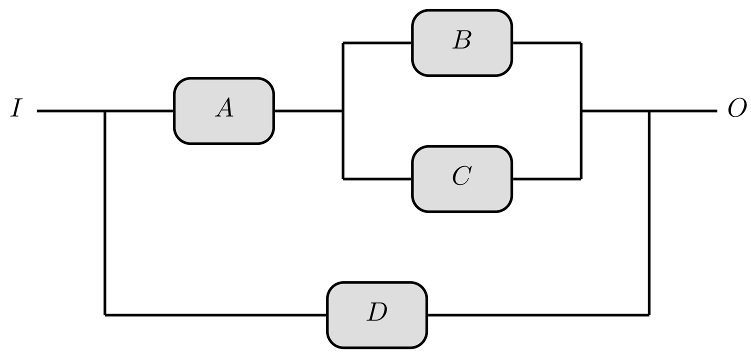

Figure 1: Système électronique formé des composants A, B, C et D.

Exercice 6

La société Gloup s'intéresse à la capacité de détection que possède le filtre bayesien anti-spam qu'elle vient d'installer. Dans la liste des mots et symboles permettant de détecter des messages publicitaires figurent les mots “viagra” et “miracle”. Par simplification, notons ces mots 1 et 2 et les probabilités

p_i	:	probabilité qu'un mot choisi au hasard dans un message électronique est le mot i en sachant que le message est un spam,
q_i	:	probabilité qu'un mot choisi au hasard dans un message électronique est le mot i en sachant que le message n'est pas un spam,

valent respectivement $p_1 = 0.05$, $p_2 = 0.1$ et $q_1 = 0.001$, $q_2 = 0.5$. Supposons que dans chaque type de messages électroniques, les mots de la liste se trouvent indépendamment les uns des autres dans les messages. La proportion de messages spam reçus par la société Gloup vaut 0.9. En sachant que les mots “viagra” et “miracle” figurent tous les deux une fois dans le même message électronique, déterminer la probabilité qu'il s'agisse d'un spam.

Exercice 7 (suite du problème de prévision météorologique)

En sachant que pour le jour de la rentrée académique, la météo avait annoncé de la pluie le matin alors que le comportement de la grenouille avait laissé prévoir du soleil, calculer la probabilité qu'il allait faire beau à Yverdon–les–Bains le matin de la rentrée.

Exercice 8

Lire puis commenter les deux articles écrits par F. Schütz.

Le monde est mathématiques

La formule miraculeuse du Révérend Bayes

Frédéric Schütz

Tout a commencé par un pasteur britannique du XVIIIe siècle féru de mathématiques, Thomas Bayes. Dans le domaine des probabilités, le révérend a découvert une formule qui porte son nom, le «théorème de Bayes», que tous les gymnasiens connaissent au moment de passer leur examen de maturité.
Ce résultat est lié à l'interprétation de la notion de probabilité. Si une pièce de monnaie a une chance sur deux de tomber sur le côté face, on peut le comprendre de la façon suivante: lançons la pièce de nombreuses fois, elle tombera sur face environ la moitié du temps. Mais comment parler, par exemple, de la probabilité que la théorie du réchauffement climatique soit correcte? C'est un événement unique, qu'il est impossible de répéter. Dans le domaine des statistiques

bayesiennes, une probabilité correspond au degré de confiance que l'on peut avoir dans une hypothèse. La formule de Bayes permet alors de prendre en compte des données nouvelles (par exemple des courbes de températures récentes) et de réévaluer cette confiance à la hausse ou à la baisse, à la lumière de ces informations – en d'autres termes, d'apprendre en fonction de ce qui est observé.
Les utilisations concrètes de ces méthodes couvrent quasiment tout l'éventail des sciences. Sur le site *www.mpt2013.fr*, Gilles Guillot, de l'Université technique du Danemark, décrit une application originale: les statistiques bayesiennes sont utilisées pour identifier l'origine des ivoires d'Afrique saisis par la douane aux aéroports. L'ADN prélevé sur les ivoires est comparé à celui d'éléphants dont l'origine géographique est bien identifiée; la formule de Bayes utilise ces informations pour calculer la probabilité que l'échantillon provienne d'une certaine latitude et longitude, et identifier ainsi son origine probable. A l'échelle du continent africain, la moitié des échantillons peuvent ainsi être localisés avec une erreur inférieure à 500 km.

A l'occasion de l'Année des mathématiques de la planète Terre 2013, *Le Temps* décrit l'environnement à travers les nombres et les formules.

“Le Temps”, 24 juillet 2013.

La rock star des statistiques; Nate Silver s'est assuré une renommée exceptionnelle en prédisant avec une rare exactitude les choix des 50 Etats américains au cours des présidentielles de 2008 et de 2012

AUTEUR: **Frédéric Schütz**

RUBRIQUE: OPINIONS/CHRONIQUES

LONGUEUR: 423 mots

Il est rare qu'un scientifique qui change d'employeur soit remarqué par la presse, qui s'intéresse généralement plus aux transferts entre clubs sportifs. L'attention médiatique portée au statisticien Nate Silver ces derniers mois aux Etats-Unis est d'autant plus inattendue.

Peu connu en Europe, Nate Silver est l'auteur du blog FiveThirthyEight, hébergé jusqu'à l'été par le New York Times, sur lequel il décortique les statistiques électorales américaines (le titre, 538, se réfère au nombre de grands électeurs dans le collège électoral des Etats-Unis). C'est le déménagement en juillet de ce site vers ESPN, chaîne de télévision sportive, qui a attiré l'attention des médias américains.

Plus fort que Paul le Poulpe, la pieuvre qui prédisait le résultat des matches de la Coupe du monde de football en 2010, Nate Silver a établi sa renommée en prédisant correctement le choix de 49 Etats sur 50 lors de l'élection de Barack Obama en 2008. Rebelote en 2012, où il prédit cette fois correctement le résultat de chacun des 50 Etats. Au vu de ces résultats et de sa notoriété, Silver a été invité récemment lors d'un important rassemblement de statisticiens, où il a été surnommé par certains la «rock star des statisticiens».

Quel est son secret? Alors que la majorité des commentateurs s'intéresse seulement au résultat des sondages, ces derniers ne sont qu'un des éléments des prédictions du statisticien. Les informations démographiques et historiques lui sont particulièrement précieuses: si un sondage donne 40% des voix à Obama dans une région donnée, cela paraît peu dans l'absolu. Mais si cette région n'a jamais voté à plus de 30% pour un démocrate, ça devient une excellente nouvelle pour le candidat. Une technique mathématique, les statistiques bayesiennes (LT du 24.07.2013), permet ensuite de combiner toutes ces informations pour obtenir les prédictions voulues. Ces dernières ne donnent pas simplement le résultat le plus probable, mais l'accompagnent d'une mesure de son incertitude. Celle-ci permet de prévoir si la lutte électorale sera serrée, et indique la probabilité que la prédiction soit fausse. Une information très utile, mais inaccoutumée dans un monde habitué à ne rechercher que des certitudes.

En Suisse, les sondages auxquels nous sommes confrontés dans les médias sont régulièrement controversés, à tort ou à raison. Beaucoup de ces statistiques bénéficieraient grandement d'un dépoussiérage et d'une analyse plus poussée. A quand un Nate Silver helvétique?

Statisticien au SIB Institut suisse de bioinformatique

“Le Temps”, 25 septembre 2013.