Variables Aléatoires Discrètes

Résumé

Cette période (relativement courte) est consacrée à l'étude du cas particulier des variables aléatoires discrètes. C'est un cadre propice à la mise en pratique des concepts abordés jusqu'à présent, sans la difficulté technique qui vient avec la manipulation des variables aléatoires à densité.

Table des matières

1 Germe d'une VA discrète

Dans l'ensemble de cette feuille $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé muni de sa tribu $\mathcal{A}$ et d'une probabilité $\mathbb{P}$.

Définition 1.1. Une ***variable aléatoire discrète*** sur $(\Omega, \mathcal{A}, \mathbb{P})$ est une variable aléatoire à valeurs dans un espace probabilisable $\big(\Pi, \mathcal{P}(\Pi)\big)$ où Π est un ensemble *dénombrable.*

Remarque 1. Dans la majorité des cas les VAs d'intérêt sont à valeurs dans un sous-ensemble de $\mathbb{R}$, on limitera les définitions dans la suite à ce cas. On suppose donc qu'une VA discrète est à valeurs dans $\mathbb{R}$. Les valeurs que prend la VA sont dans ce cas indexées par les entiers naturelles, elles forment un espace probabilisable muni de la tribu de l'ensemble de ses parties.

On appelle ***germe*** d'une VA aléatoire discrète $X : \Omega \to \mathbb{R}$ la fonction g donnée par la loi de X sur les issues. Plus formellement,

$$\forall x \in \mathbb{R}, \qquad g(x) = \mathbb{P}(X \in \{x\})$$

On parle dans ce contexte du ***germe de X en*** x.

Notation. On simplifie la notation précédente en notant $\mathbb{P}(X = x)$ la probabilité $\mathbb{P}(X \in \{x\})$.

Question 1-1. En utilisant les axiomes définissant une probabilité montrer que la loi d'une VA discrète $X : \Omega \to \mathbb{R}$ est caractérisée par ses germes en tout point.

2 Indépendance et conditionnement

Les notions d'indépendance et de conditionnement de VAs se décrivent relativement simplement dans le cas des VAs discrètes.

Question 2-2. Justifier le fait que deux VAs discrètes X_1 et X_2 sur $(\Omega, \mathcal{A}, \mathbb{P})$ sont indépendantes si et seulement si pour tout couple d'issues $(x_1, x_2) \in X_1(\Omega) \times X_2(\Omega)$ on a

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\, \mathbb{P}(X_2 = x_2).$$

On reprend les notations ci-dessus. On définit la ***loi de X_2 conditionnelle à l'évènement $(X_1 = x_1)$*** par le germe

$$\mathbb{P}_{X_2}^{(X_1=x_1)} := \mathbb{P}^{(X_1=x_1)}(X_2 = x_2) = \mathbb{P}(X_2 = x_2 | X_1 = x_1).$$

Question 2-3. Montrer que la loi du couple (X_1, X_2) est déterminée par $\mathbb{P}_{X_1}$ et $\mathbb{P}_{X_2}^{(X_1=x_1)}$.

3 Comparer deux VAs

Il arrive souvent qu'on ait à étudier la réalisation simultanée des issues de deux VAs $X_1 : \Omega_1 \to \mathbb{R}$ et $X_2 : \Omega_2 \to \mathbb{R}$ sur les espaces probabilisés respectivement donnés par $(\Omega_1, \mathcal{A}, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{A}, \mathbb{P}_2)$. Dans la démarche menée jusqu'à présent il nous fallait définir un espace probabilisable sur lequel on peut définir à la fois X_1 et X_2. En réalité, rendre explicite un tel espace n'est pas d'une grande importance dans la pratique ; il suffit de connaître les probabilités des issues simultanées et cela est uniquement déterminé par l'expérience aléatoire. La section qui suit a pour objectif de clarifier en quoi cette phrase fait sens.
Il est facile d'exhiber un espace d'états sur lesquelles les VAs X_1 et X_2 vivent naturellement : on considère tout simplement le produit cartésien $\Omega = \Omega_1 \times \Omega_2$. On est tenté de prendre pour tribu sur Ω l'ensemble des évènements $A \times B$ où $A \in \mathcal{A}$ et $B \in \mathcal{B}$.

Question 3-4. Montrer en quoi la définition ci-dessus ne définit pas une tribu sur Ω.

Pour résoudre ce problème on considère la plus petite tribu contenant les *carré* d'évènements précédents : la tribu engendrées par ces carrés [1]. Cette tribu peut-être construite en considérant les complémentaires et union dénombrables des *carrés* puis de même avec les objets obtenus à cette étape et ainsi de suite [2]. La magie de cette construction réside dans le fait qu'il suffit, pour définir une probabilité sur la tribu engendrée, de la définir sur les éléments qui l'engendre [3]. La vérification de ce fait est technique et ne fait pas partie du scope de ce cours, la conclusion seule nous importe ici. On peut désormais retrouver les deux VAs X_1 et X_2 sur Ω par composition à gauche avec les projections respectivement sur la première et seconde coordonnée.
Il s'agit maintenant de définir une probabilité sur Ω. On suppose pour cette période que les VAs X_1 et X_2 sont discrètes. D'après la discussion précédente il suffit de définir celle-ci sur les *carré*, les axiomes d'une probabilité permettent d'étendre cette définition à tout type d'évènements du produit. Dans

1. Penser aux sous-espaces vectoriels engendrés par une partie.
2. Pour une discussion éclairante sur ce point voir `Wikipedia - Tribu engendrée`.
3. Tout comme le fait que pour définir une application linéaire il suffit de connaître les valeurs qu'elle prend sur une base.

notre cas, on est uniquement concernés par la sous-tribu engendrée par les carrés du type $(X_1 = x_1) \times (X_2 \times x_2)$ pour des éléments $(x_1, x_2) \in \mathbb{R}^2$.

> **Question 3-5.** Supposons qu'on définisse la probabilité $\mathbb{P}$ sur Ω par
>
> $$\forall (x_1, x_2) \in \mathbb{R}^2, \qquad \mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}_1(X_1 = x_1)\,\mathbb{P}(X_2 = x_2).$$
>
> À quoi correspond un tel choix ?

La description précédente ne prend pas en compte le possible conditionnement de X_2 par rapport à X_1, on peut cependant modéliser une telle situation, et c'est rassurant. Supposons donné pour tout $x_1 \in \mathbb{R}$ un germe de probabilité $\mathbb{P}_2^{x_1}$ correspondant à la réalisation d'une issue de X_2 conditionnée à la réalisation de x_1 pour X_1.

Proposition 3.1. *L'expression définie pour tout* $(x_1, x_2) \in \mathbb{R}^2$, *par* $\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}_1(X_1 = x_1)\,\mathbb{P}_2^{x_1}(X_2 = x_2)$ *définit une probabilité sur* Ω *qui conditionne* X_2 *par* X_1.

Cette propriété nous apporte les garanties nécessaires pour comparer les réalisations simultanées de deux variables ; la généralisation au cas d'un nombre fini de VAs n'est qu'une difficulté technique. On renvoie à [Ouv98, pages 65-67 & 93].

4 VAs discrètes usuelles

On liste dans la suite les VAs usuelles, d'utilisations constantes dans les cas d'études pratiques et de modélisation.

4.1 Loi uniforme

Une variable aléatoire $X : \Omega \to \mathbb{R}$ ayant une image finie $\{x_1, \dots, x_n\}$ suit une ***loi uniforme*** si

$$\forall i \in \{1, \dots, n\}, \qquad \mathbb{P}(X = x_i) = \frac{1}{n}.$$

On écrit dans ce cas que $\mathbb{P}_X \rightsquigarrow \mathcal{U}(n)$.

> **Question 4-6.** Vérifier que cette description correspond à la notion de l'équiprobabilité de réalisation de chacune des issues de X.

4.2 Loi de Bernoulli

Une variable aléatoire $X : \Omega \to \mathbb{R}$ d'image $\{0, 1\}$ est dite ***loi de*** **Bernoulli** ***de paramètre p*** si

$$\mathbb{P}(X = 1) = p.$$

> **Question 4-7.** En quoi est-ce que la donnée de la valeur de $\mathbb{P}_X$ sur 1 est suffisante pour décrire $\mathbb{P}_X$?

4.2.1 *Score* d'un classificateur

On chercher dans cette section à quantifier la *qualité* d'un classificateur binaire. Les classifcateurs sont des modèles de ML utilisés lorsque l'on souhaite prédire une caractéristique discrète ; par exemple des types de plantes, des couleurs de cheveux, des appréciations de goûts etc. à partir d'observations en entrée. Un classificateur binaire n'a que deux valeurs de sorties qui correspond à l'existence ou non

d'une caractéristique que l'on cherche à observer ; avoir une maladie ou non, réussir un examen ou non, etc.
Dans la formalisation qu'on propose ici on considère que les observations en entrée constituent un espace d'états Ω. On considère de plus que celui-ci vient avec une fonction λ qu'on désigne par *label* qui code la caractéristique qu'on cherche à prédire. Autrement dit on suppose travailler avec un dataset en entrée qui contient cette information ; on est dans un cadre d'apprentissage supervisé. Dans ce contexte un classificateur est une variable aléatoire $X : \Omega \to F$ où F est l'espace des labels qu'on cherche à attribuer à nos entrées, c'est-à-dire $F = \{0, 1\}$. L'espace Ω est supposé fini probabilisé muni d'une probabilité $\mathbb{P}$. C'est cette probabilité qui fait office de *proportion*. On suppose que X est donnée et on souhaite étudier trois quantités qui sont liées à l'évaluation de la qualité d'un classificateur.

Question 4-8. On appelle ***précision totale*** d'un classificateur binaire X la proportion des individus bien classés, c'est-à-dire ceux où les réponses de X et λ coïncident.

1. Exprimer la précision totale de X en s'aidant de la variable aléatoire $X - \lambda$.
2. Déterminer la loi de $X - \lambda$ et interpréter ses deux autres valeurs.

Dans le cas d'un classificateur binaire deux autres quantités apparaissent relativement naturellement à l'évaluation :
— la ***précision*** : proportion des vrais positifs par rapport à l'ensemble des bonnes classifications.
— le ***rappel*** : proportion des vrais positifs par rapport à l'ensemble des positifs (donnés par λ).

Question 4-9.

1. Exprimer précision et rappel de X.
2. Étudier l'interconnexion entre précision et rappel ; qu'arrive-t-il à l'une si l'autre augmente ?

Dans les problématiques de classification on est toujours attentifs au fait d'être confrontés à des modèles dont la dépendance au label λ est très faible. C'est pour cette raison qu'on va souvent chercher à évaluer le ***score*** d'un classificateur X tels que X et λ définissent des variables aléatoires indépendantes.

Question 4-10.

1. Simplifier les expressions de la précision totale, précision et rappel dans le cas où X et λ sont indépendants.
2. On suppose que X et λ sont indépendants et suivent des loi de Bernoulli respectivement de paramètres p et q. Exprimer les différents scores de X dans ce cas.
3. Quels résultats numériques obtenez vous pour chacun des scores quand $p = q = 0.9$? Qu'en déduisez vous ?

4.3 Loi géométrique

On considère l'expérience aléatoire qui consiste à jeter une pièce truquée autant de fois qu'on veut. On suppose qu'on a $p \in]0, 1[$ chances d'avoir Pile. On note $X : \Omega \to \mathbb{R}$ la variable aléatoire qui consiste à obtenir le premier Face.

Question 4-11. Décrire la loi de la VA X. Quelle hypothèse implicite on utilise dans cette description ?

La VA X est dite suivre une ***loi géométrique de paramètre p***.

Question 4-12. Reprenez la question **??** et supposez que Brendan remets à chaque essai le briquet testé dans sa poche. Décrire la loi de la VA Brendan a réussi à allumer sa cigarette.

4.4 Loi binômiale

Une variable aléatoire $X : \Omega \to \mathbb{R}$ d'image dans $\{0, \dots, n\}$ suit une loi binômiale de paramètres n et p, qu'on écrit $\mathbb{P}_X \rightsquigarrow \mathcal{B}(n, p)$, si elle est de la forme

$$X = \sum_{i=1}^{n} B_i$$

pour des variables de Bernoulli B_i de paramètre p, indépendantes dans leur ensembles.

Question 4-13. Expliciter une formule pour la loi de X suivant une $\mathcal{B}(n, p)$.

Question 4-14. Décrire des expériences aléatoires modélisées par la loi précédente.

4.5 Loi hypergéométrique

On considère l'expérience aléatoire qui consiste à tirer au hasard k boules dans une urnes contenant n_b boules blanches et n_r boules rouges.

Question 4-15. Décrire la loi de la VA X qui compte le nombre de boules rouges.

Question 4-16. Revenez à la seconde partie de la question **??**.

4.6 Loi de Poisson

Vous pouvez retrouver la description suivante dans [Pol]. On considère une suite (p_n) avec $np_n \to \lambda \in \mathbb{R}_+^*$. On s'intéresse à la suite de VAs $X_n \rightsquigarrow \mathcal{B}(n, p_n)$. C'est une suite de VAs suivant des lois binômiales dont le paramètre de réalisation est de plus en petit. Ce processus est une tentative de modélisation d'évènements rares. Pour tout $k \in \mathbb{N}$

$$\mathbb{P}(X_n = k) = \begin{cases} \binom{k}{n} p_n^k (1 - p_n)^{n-k} & \text{si } k \leq n \\ 0 & \text{si } k \geq n + 1 \end{cases}$$

Question 4-17.

1. Montrer que $\lim_{n \to +\infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.
2. Vérifier que la donnée pour tout $k \in \mathbb{N}$, par $g(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ définit une loi de probabilité sur $\mathbb{N}$.

Une VA X qui suit la loi décrite précédemment est dite de Poisson de paramètre λ, on écrite $X \rightsquigarrow \mathcal{P}(\lambda)$.

Question 4-18. Vous retrouverez cet exemple dans [Ouv98, page 94]. On se place à un embranchement routier ; le nombre de véhicules arrivant pendant un intervalle d'une heure est une VA X, suivant une loi de Poisson de paramètre λ. Les véhicules ne peuvent prendre qu'une des

deux directions A ou B, et la VA Y représente le nombre de véhicules empruntant la direction A pendant cet intervalle de temps. Chaque véhicule prend la direction A avec probabilité p, et les choix sont faits de manière indépendante. C'est pour cette raison qu'on suppose que si n véhicules arrivent à l'embranchement pendant une heure donnée, la loi (conditionnelle) de $Y \rightsquigarrow \mathcal{B}(n, p)$..
La probabilité conditionnelle $\mathbb{P}^{(Y=k)}(X = n)$ est la probabilité que n véhicules soient arrivés à l'embranchement sachant que k véhicules ont emprunté la direction A. Calculer cette probabilité.

5 Moments d'une VA discrète

L'étude du comportement aléatoire d'une VA discrète s'étudie souvent par le biais de mesures de ***centrage*** et de ***dispersion***. Dans le jargon quotidien on parlera par exemple de *moyenne* ou *médiane* et de *variance* ou *écart-type*. Ces deux mesures, qu'on pense souvent en termes statistiques se calcule par le biais des ***moments*** d'ordre 1 et 2 des VAs qui modélisent les phénomènes à l'étude.

5.1 Espérance d'une VA discrète

L'équivalent de la notion de ***moyenne*** en théorie des probabilités est celui d'espérance.

Définition 5.1 (Espérance d'une VA discrète)**.** Soit $X : \Omega \to \mathbb{R}$ une VA discrète. Si $\sum_{x\in\mathbb{R}} |x|\,\mathbb{P}(X = x) < +\infty$ on appelle ***espérance*** de la VA X la quantité

$$\mathbb{E}(X) = \sum_{x\in X} x\,\mathbb{P}(X = x).$$

Question 5-19. Calculer l'espérance d'une VA discrète suivant une loi uniforme sur $\{0, \ldots, n\}$.

Question 5-20. Quelle est l'espérance d'une loi de Bernoulli de paramètre p ?

Question 5-21.

1. Soit X, Y deux VA discrètes sur $(\Omega, \mathcal{A}, \mathbb{P})$. Étant donnée $\lambda \in \mathbb{R}$, montrer que

$$\mathbb{E}(\lambda X + Y) = \lambda\,\mathbb{E}(X) + \mathbb{E}(Y). \qquad \text{(Linéarité de l'espérance)}$$

2. En déduire l'espérance d'une VA suivant une loi binômiale de paramètres n et p.

Question 5-22. Vérifier les propriétés suivantes de l'espérance d'une VA $X : \Omega \to \mathbb{R}$.

1. Si c est l'unique valeur que prend X alors $\mathbb{E}(X) = c$.
2. Si X est bornée X admet une espérance.
3. Si $X \geq 0$ alors $\mathbb{E}(X) \geq 0$.
4. $|\,\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Il arrive souvent qu'on s'intéresse à une fonction d'une VA $X : \Omega \to \Pi$, c'est-à-dire à la fonction composée $f(X) = f \circ X$ pour $f : \Pi \to \Delta$. Le théorème de transfert nous indique qu'on peut étudier la loi de $f(X)$ en travaillant toujours sur Π. On suppose que X est une VA discrète ; il en va de même de $f(X)$.

Théorème 5.1 (Théorème de transfert)**.** *Si la somme* $\sum_{x\in\Pi} |f(x)|\,\mathbb{P}(X=x)$ *est finie alors*

$$\mathbb{E}\left(f(X)\right) = \sum_{x\in\Pi} f(x)\,\mathbb{P}(X=x).$$

Question 5-23. Calculer les espérances des lois usuelles restantes dans la liste en amont.

5.2 Variance d'une VA discrète

La variance est une mesure de dispersion, autrement dit de variabilité autour de l'espérance d'une VA discrète.

Définition 5.2. Soit $X : \Omega \to \mathbb{R}$ une VA discrète. Si $\mathbb{E}(X^2) < +\infty$ on appelle ***variance*** de la VA X la quantité

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

On appelle ***écrat-type*** d'une VA X admettant une variance le scalaire

$$\sigma_X = \sqrt{\mathbb{V}(X)}$$

Question 5-24. Soit X une VA discrète qui admet une variance. Montrer que

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Question 5-25. Quel est l'équivalent de la relation (Linéarité de l'espérance) dans le cas de la variance d'une loi aléatoire discrète X sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Question 5-26. Calculer les variances des VAs discrètes suivant les lois usuelles listées en amont. Identifier parmi celles-ci celles qui vous posent problème.

Centrer et réduire ses données est un *pre-processing* standard pour différents types d'algorithmes de ML et pour différentes raisons. L'objectif peut par exemple être d'optimiser une descente de gradient ou d'étudier la corrélation de *features* différentes.

Définition 5.3. La VA ***centrée réduite*** associée à une VA $X : \Omega \to \mathbb{R}$ admettant une espérance et une variance est la VA

$$\overline{X} = \frac{X - \mathbb{E}(X)}{\sigma_X}.$$

Question 5-27. Vérifier que $\mathbb{E}(\overline{X}) = 0$ et $\mathbb{V}(\overline{X}) = 1$

5.3 Moments d'ordres supérieures d'une VA discrète

L'espérance et la variance sont des mesures correspondant au moment d'ordre 1 dans le premier cas et à partir de ceux d'ordres 1 et 2 dans le second.

Définition 5.4. Soient $X : \Omega \to \mathbb{R}$ une VA discrète et $p \geq 1$. Si $\mathbb{E}(|X|^p) < +\infty$ on appelle moment d'ordre p de X la quantité

$$\mu_p(X) = \mathbb{E}(X^p).$$

Avec les notations précédentes on peut écrire la variance de X sous la forme

$$\mathbb{V}(X) = \mu_2(X) - \mu_1(X)^2.$$

La manipulation des moments d'ordre supérieurs de variables discrètes n'est pas aisée avec les outils qu'on a en main. Il nous faudrait introduire dans ce but les séries génératrices de VAs discrètes. Notion qui, malgré son intérêt [4], nous pousserait loin du scope de ce cours. Il est cependant important d'en connaître l'existence et la définition ; ces moments sont utilisés dans les techniques de représentations de données.

6 Covariance et corrélation

La covariance de deux VAs est un concept qui cherche à quantifier les variabilités simultanées de ces deux VAs. Elle est liée à la notion d'indépendance de deux VAs par le fait que deux VAs indépendantes ont une covariance nulle, la réciproque est **fausse** ! Deux VAs ayant une covariance nulle (dites non corrélées) n'ont pas nécessairement une covariance nulle.

Définition 6.1. Soient X et Y deux VAs ayant des moments d'ordre 2. On appelle ***covariance*** de X et Y la quantité

$$\operatorname{cov}(X, Y) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)\big((Y - \mathbb{E}(Y)\big)\Big).$$

Question 6-28. En quoi est-ce que l'existence de moments d'ordre 2 de X et Y permet de justifier l'existence de l'espérance de $\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)$? [a]

a. À chercher du côté d'une certaine inégalité de Schwarz.

Question 6-29. Soient X et Y deux variables aléatoires ayant des moments d'ordre 2. Justifier les relations :

1. $\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)$;
2. $\mathbb{V}(X + Y) = \mathbb{V}(X) + 2\operatorname{cov}(X, Y) + \mathbb{V}(Y)$.

Déduire de la première question [a] que si X et Y sont indépendantes alors leur covariance est nulle.

a. Retour sur un travail déjà effectué en cours.

Remarque 2. Le fait que la covariance de deux VAs aient une covariance nulle n'implique pas l'indépendance de celle-ci. L'exercice (6-30) montre comment construire une famille de contre-exemples.

Question 6-30. On considère deux VAs indépendantes X et Y ayant des moments d'ordre 2. On s'intéresse aux VAs

$$\begin{cases} U &= aX + bY \\ V &= cX + dY \end{cases}$$

pour a, b, c et d des paramètres donnés. Trouver des contre-exemples à la proposition : *deux VAs aléatoires ayant une covariance nulle sont indépendantes*, sous la forme U, V précédente.

La covariance mesure en particulier le potentiel qu'on a à exprimer une VA en fonction d'une autre linéairement. C'est une manière de contextualiser la notion de covariance. Ce questionnement, exprimer linéairement une VA en fonction d'une autre, rentre dans le cadre générique de *régression linéaire.*
La ***régression linéaire*** *classique* est décrite par le problème d'optimisation

$$\min_{(a,b)\in\mathbb{R}^2} \mathbb{E}(Y - aX - b)^2.$$

4. Et son esthétisme !

La fonction objectif ci-dessus peut-être réécrite sous la forme

$$\mathbb{E}(Y - aX - b) = \mathbb{E}(\overline{Y} - a\overline{X})^2 + \big(\mathbb{E}(Y) - a\,\mathbb{E}(X) - b\big)^2. \tag{1}$$

Question 6-31. Quelle est la valeur minimale du membre de droite de l'égalité (1) à a fixé. En déduire un point optimal de notre problème de départ. Identifier le ***coefficient de corrélation*** [a] entre deux VAs ayant des moments d'ordre 2.

a. Commencer par en trouver une définition quelque part ..

Question 6-32. On note $\rho_{X,Y}$ le coefficient de corrélation entre les VAs X et Y. Montrer que

1. $|\rho_{X,Y}| \leq 1$ [a];
2. $\rho_{X,Y} = \mp 1$ si et seulement si il existe a, b, c des réels non nuls tels que

$$\mathbb{P}(aX + bY + c = 0) = 1.$$

a. Un peu d'inégalité de Schwarz cachée ici.

Références

[Ouv98] Jean-Yves Ouvrard. *Probabilités : Tome 1, Capes-Agrégation.* Enseignement des mathématiques. Cassini, 1998.

[Pol] Ecole Polytechnique. Aléatoire : introduction aux probabilités parties I. `https://www.coursera.org/learn/probabilites-1`.