Feuille d'exercices

1 S'habituer au formalisme

1.1 Lois à densités – Manipulation

On a aborde quelques exercices simples de manipulation des lois à densités. Ces lois nécessitent une certaine aisance dans l'utilisation des intégrales généralisées.

Deux petits exemples.

Question 1-1. On considère une variable aléatoire X suivant la loi uniforme sur $[0, 1]$. Quelle est la loi de la variable aléaoitre $Y = X(X - 1)$?

Question 1-2. On considère une expérience aléatoire où l'on choisit au hasard un point du cercle de centre l'origine et de rayon l'unité. On note X et Y les variables aléatoires réelles qui correspondent au choix des coordonnées d'un tel point. Décrire les lois de X et Y.
Pourriez-vous généraliser au cas d'un point de la sphère ?

Loi exponentielle et temps d'arrêt. La date de connexion d'un client à votre après 00 : 00 est une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, de loi exponentielle de paramètre λ.

Question 1-3.

1. Rappeler et calculer les moments d'ordre 1 et 2 de d'une loi exponentielle de paramètre λ. En déduire la variance d'une telle loi.
2. Calculer la probabilité $\mathbb{P}(T > 1/\lambda)$.
3. On fixe $\epsilon > 0$. On considère pour $k \in \mathbb{N}$ les plages horaires $I_k = [k\epsilon, (k + 1)\epsilon[$. Calculer la probabilité $\mathbb{P}(T \in I_k)$.
4. Soit X la variable aléatoire à valeurs dans $\mathbb{N}$ définie pour tout $\omega in \Omega$ par
$$X(\omega) = k \Leftrightarrow T(\omega) \in I_k.$$
Quelle est la loi de X ?
5. Calculer pour tout $t > 0$ et tout $h > 0$, la probabilité $\mathbb{P}(t < T)$ ainsi que la probabilité conditionnelle $\mathbb{P}(T > t + h \mid T > t)$. Qu'est-ce que cela signifie ?

Somme de densités. On considère deux variables aléatoires X définies sur Ω, on construit à partir de celles-ci la variable aléatoire sur $\Omega \times \{0, 1\}$ définie par :

$$X(\omega, 0) = X_1(\omega) \quad \text{et} \quad X(\omega, 1) = X_2(\omega)$$

Question 1-4. Exprimer la loi de probabilités de X en fonction de X_1 et X_2. Si X_1 et X_2 sont des lois à densité montrer qu'il en va de même de X et calculer sa densité.

Transfert. On considère la variable aléatoire X à valeurs dans $]0,1[$ et dont la densité est donnée pour $x \in \mathbb{R}$ par

$$f_X(x) = \frac{1}{\ln(2)} \times \frac{1}{1+x} \mathbb{1}_{]0,1[}(x).$$

Question 1-5.

1. Déterminer la loi de $Y = 1/X$.
2. On note $\mathbb{E}(Y)$ la partie entière de Y, déterminer la loi de $Z = Y - \mathbb{E}(Y)$.

1.2 Lois à densités – Moments

Loi de Pascal. On considère la variable aléatoire X suivant la loi de densité $f_X(x) = \frac{1}{2}e^{-|x|}$.

Question 1-6. Calculer l'espérence de la variable aléatoire $Y = \mathbb{1}_{[1,+\infty[}X + \mathbb{1}_{[-\infty,1]}$.

Loi de Rayleigh On considère la variable aléatoire X suivant la loi de densité $f_X(x) = xe^{-x^2/2}\mathbb{1}_{[0,+\infty[}$.

Question 1-7.

1. Déterminer l'espérence $\mathbb{E}(X)$ de X. Calculer $\mathbb{P}\big(X > \mathbb{E}(X)\big)$.
2. Calculer $\mathbb{E}(X^2)$. En déduire $\mathbb{V}(X)$.
3. Calculer $\mathbb{R}\big(|X - \mathbb{E}(X)| > 1\big)$. Comparer le résultat avec l'inégalité de Tchebychev.

1.3 Lois à densités – Conditionnement

Manipulation des lois conjointes et marginales. On travaille trois exercices dont l'objectif est de vous habituer à la manipulation des lois conditionnelles et conjointes à denstiés.

Question 1-8. Soit (X,Y) un couple de variables aléatoires à dans $[0,+\infty[$, dont la loi conjointe a pour densité

$$f_{X,Y}(x,y) = \frac{1}{2\sqrt{x}}e^{-y}\mathbb{1}_D(x,y)$$

où D désigne le lieu de $\mathbb{R}^2$ $\{(x,y) \mid x > 0, y > \sqrt{x}\}$.

1. Déterminer les lois de X et Y.
2. Les variables X et Y sont elles indépendantes ?

Question 1-9. Soient X et Y deux variables aléatoires indépendantes de même densité $x \mapsto \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

1. Déterminer la loi de la variable aléatoire $R = \sqrt{X^2+Y^2}$.

2. Déterminer la valeur moyenne de R.
3. Les variables aléatoires X et R sont-elles indépendantes ?

Question 1-10. Soit (X,Y) une variable aléatoire à valeurs dans $]0,1[\times]0,1[$, de densité

$$f_{(X,Y)}(x,y) = 3\mathbb{1}_D$$

où D est le lieu de $\mathbb{R}^2$

$$D = \left\{(x,y) \mid 0 < x^2 < y < \sqrt{x} < 1\right\}.$$

Déterminer les lois conditionnelles et espérences condtionnelles de Y par rapport à X et de X par rapport à Y.

2 De la probabilité en ML

La ML construit des modèles dont l'objectif est un parmi :
- prédire une caractéristique d'intérêt d'un individu à partir de caractéristiques connues de celui-ci
- identifier des ***patterns*** dans un ensemble de données à l'étude, sans que cela soit conduit par un objectif particulier.

Les situations qu'on décrit par la suite apparaissent dans le premier contexte. Dans ce cadre on suppose qu'on a des données qui contiennent les caractéristiques d'un nombre d'individus ainsi que ***la*** caractéristique qu'on souhaite prédire.

2.1 *Score* d'un classificateur

On chercher dans cette section à quantifier la *qualité* de modèles de ML qu'on appelle les classificateur. Un classificateur est un modèle apparaît dans la situation où la caractéristique qu'on cherche à prédire est discrète ; par exemple des types de plantes, des couleurs de cheveux, des appréciations de goûts etc. C'est une fonction qui étant donné un certaint nombre de caractéristique en entrée renvoie une valeurs discrète, souvent codées entre 0 et le nombre de classes à prédire moins un.

Dans la formalisation qu'on propose ici on considère que les caractéristique en entrée constitue un espace d'états Ω. On considère de plus que celui-ci vient avec une fonction λ qu'on désigne par *label* et qui nous indique la caractéristique qu'on cherche à prédire. On a supposé que le dataset en entrée contient cette information. 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. 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.

Cas binaire. On se limite en un premier temps au cas d'un classificateur binaire, c'est-à-dire que $F = \{0, 1\}$.

Question 2-11. 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 concident.
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 2-12.

1. Exprmier 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 2-13.

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 resultats numériques obtenez-vous pour chacun des scores quand $p = q = 0.9$? Qu'en déduisez-vous ?

Cas général. On revient brièvement vers le cas général. On suppose désormais que F est l'ensemble discret $\{0, \dots, k-1\}$.

Question 2-14.

1. Exprimer la précision totale de X.
2. Étudiez la loi de $X - \lambda$. Est-elle aussi facilement interprétable que dans le cas binaire ?
3. Chercher une variable aléatoire plus adaptée à l'étude des problématiques de classifications qui ne sont pas binaires.

2.2 Classificateur Bayesien

2.3 *Trade-off* biais variance

Solutions des exercices

Vous trouverez dans la suite solutions et indications d'une partie des exercices de la feuille. Ceci étant majoritairement accessibles il vous est suffisant de comparer votre travail au résultats que vous retrouverez dans la suite.

Solution 1-1. L'expressions $x(x-1)$ décrit une parabole dont le sommet est atteint en $1/2$ et y prend la valeur $-1/4$. Sur $[0,1]$ cette parabole prend donc toutes les valeurs dans $[-1/4,0]$. Comme X prend ses valeurs dans $[0,1]$, la V.A. Y prend ses valeurs dans $[-1/4,0]$. On va décrire la fonction de répartition de Y. Soit $t \in [-1/4,0]$.

$$Y \leq t \iff X^2 - X - t \leq 0.$$

Pour décrire le lieu $Y \leq t$ on étudie donc le signe du trinôme du seconde degré $X^2 - X - t$ en X. Ce trinôme est négatif entre les racines. Une étude du discriminanet permet de voie que ces racines sont

$$\frac{1-\sqrt{1+4t}}{2} \quad \text{et} \quad \frac{1+\sqrt{1+4t}}{2}.$$

Ainsi

$$Y \leq t \iff \frac{1-\sqrt{1+4t}}{2} \leq X \leq \frac{1+\sqrt{1+4t}}{2}.$$

Donc

$$\mathbb{P}(Y \leq t) = \mathbb{P}\left(\frac{1-\sqrt{1+4t}}{2} \leq X \leq \frac{1+\sqrt{1+4t}}{2}\right) = \sqrt{1+4t}.$$

La densité f_Y de Y est ainsi donnée par

$$f_y(x) = \frac{2}{\sqrt{1+4x}}\mathbb{1}_{[-1/4,0]}(x).$$

Pour le voir il suffit de voir le fait que la densité est la dérivée de la fonction de répartition là où celle-ci est dérivable. Comme la valeur n'a pas d'importance dans le seul point où elle ne l'est pas (en 0) on obtient ce qu'on cherche.

Solution 1-2. On repère le point du cercle unité par l'angle que forme le rayon lui correspondant avec l'axe des abscisses. Autrement dit par ses coordonnées polaires. L'angle θ qui repère un point du cercle unité varie dans $[0,2\pi[$. On peut donc décrire une probabilité sur le cercle unité par l'espace d'états $\Omega = [0,2\pi[$, muni de la tribu des boréliens, et agrémenté de la probab $(1/2\pi)\ell$ où ℓ est déterminée sur les intervalles de Ω par la longueur de ceux-ci.
La variable aléatoire X est décrite par ce cadre par la fonction $\theta \mapsto \cos(\theta)$ et qui prend ses valeurs dans $[-1,1]$. La fonction de répartition de X est décrite pour tout $t \in [-1,1]$ par

$$\mathbb{P}(X \leq t) = \ell\left([\theta \mid \cos(\theta) \leq t]\right).$$

On pourrait s'arrêter à ce stade, car vous avez jamais vu la fonction arccos. On va, cela dit, aller un peu plus loin. La fonction arccos est la réciproque de la fonction cos quand cette dernière est restreinte à l'espace de départ $[0,\pi]$. Sur cet interval arccos est décroissante. On a donc, pour tout $t \in [-1,1]$,

$$\theta \in [0,\pi] \,\text{et}\, \cos(\theta) \leq t \iff \arccos(t) \leq \theta \leq \pi.$$

La symétrie axiale du graphe de cos par rapport à l'axe $x = \pi$ [a] permet de voir que

$$\theta \in [0, 2\pi] \text{ et } \cos(\theta) \leq t \Longleftrightarrow \arccos(t) \leq \theta \leq 2\pi - \arccos(t).$$

D'où, pour tout $t \in [-1, 1]$,

$$\begin{aligned}\mathbb{P}(X \leq t) &= \mathbb{P}\left(\arccos(t) \leq \theta \leq 2\pi - \arccos(t)\right) \\ &= 1 - \frac{1}{\pi} \arccos(t).\end{aligned}$$

Hors de $[-1, 1]$ la fonction de répartition de X est, ou bien nulle quand $t < -1$ ou égale à 1 sinon. Cela suffit pour déterminer la loi de X ainsi que celle de Y (par symétrie, c'est la même que la loi de X).

On peut ajouter à l'étude le calcul de la densité de la fonction de répartition F_X. C'est une fonction dérivable sauf éventuellement en -1 et 1. En dehors de ces points on obtient pour fonction de densité

$$f_X(x) = \frac{1}{\pi\sqrt{1-x^2}} \mathbb{1}_{[-1,1]}(x).$$

qu'on prolonge par 0 en -1 et 1. Cette relation vient du fait que

$$\arccos'(x) = \frac{1}{\sqrt{1-x^2}} \text{ [b]}.$$

a. Il faut un peu de travail et de dessin pour le voir. Je compte sur vous. (Oui, j'ai la flemme, mais venez me poser des questions).

b. Si vous voulez savoir pourquoi posez-moi la question.

Solution 1-6. D'après le théorème de transfert on a

$$\mathbb{E}(Y) = \int_{\mathbb{R}} \left(x\mathbb{1}_{[1,+\infty[}(x) + \mathbb{1}_{[-\infty,1]}(x)\right)\frac{1}{2}e^{-|x|}\,\mathrm{d}x.$$

D'où

$$\mathbb{E}(Y) = \frac{1}{2}\int_1^{+\infty} xe^{-x}\,\mathrm{d}x + \frac{1}{2}\int_{-\infty}^0 e^x\,\mathrm{d}x + \frac{1}{2}\int_0^1 e^{-x}\,\mathrm{d}x.$$

Une intégration par partie pour la première intégrale et un calcul direct pour la seconde donne

$$\mathbb{E}(Y) = 1 - \frac{1}{e}.$$

Solution 1-7. L'espérence de la V.A. X est donnée par l'expression

$$\mathbb{E}(X) = \int_0^{+\infty} x^2 e^{-x^2/2}\,\mathrm{d}x.$$

En décomposant l'intégrande en $u = x$ et $v' = xe^{-x^2/2}$ une intégration par parties donne

$$\mathbb{E}(X) = \sqrt{\frac{\pi}{2}}$$

par comparaison à l'espérence d'une variable aléatoire de loi gaussienne ayant une espérence nulle et une variance de 1.

En s'aidant d'une intégration par parties on a

$$\mathbb{P}\big(X \geq \mathbb{E}(X)\big) = \int_{\sqrt{\pi/2}}^{+\infty} xe^{-x^2/2}\,\mathrm{d}x = e^{-\pi/4}.$$

Une intégration par parties en deux temps qui prolonge le raisonnemenet à la première question permet d'avoir

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

donc

$$\mathbb{V}(X) = 2 - \frac{\pi}{2}.$$

Essayez de justifier sans calculette le fait que $e^{-\pi/4} > \mathbb{V}(X)$.
Un calcul similaire à ce qui a été effectué jusqu'à présent (toujours en complexifiant) donne

$$\mathbb{P}\big(|X - \mathbb{E}(X)|\big) = 1 + e^{-\pi/4}e^{-1/2}\left(e^{-\sqrt{\pi/2}} - e^{\sqrt{\pi/2}}\right).$$

La calculatrice vous permet de voir que c'est un bien meilleur encadrement que celui donné par l'inégalité de Tchebychev.

Solution 1-8. Les lois de X et Y ont des densités à supports dans $[0, +\infty]$ (elles sont nulles à l'extérieur de cet intervalle). Pour $x \in]0, +\infty]$, on a :

$$f_X(x) = \int_0^{+\infty} f_{(X,Y)}(x, y)\,\mathrm{d}y = \int_0^{+\infty} \frac{1}{2\sqrt{x}} e^{-y} \mathbb{1}_D(x, y)\,\mathrm{d}y = \frac{1}{2\sqrt{x}} \int_{\sqrt{x}}^{+\infty} e^{-y}\,\mathrm{d}y$$

donc

$$f_X(x) = \frac{1}{2\sqrt{x}} e^{-\sqrt{x}}.$$

Pour $y \in]0, +\infty]$, on a :

$$f_Y(y) = \int_0^{+\infty} \frac{1}{2\sqrt{x}} e^{-y} \mathbb{1}_D(x, y)\,\mathrm{d}x = \int_0^{y^2} \frac{1}{2\sqrt{x}} e^{-y}\,\mathrm{d}x$$

et ainsi,

$$f_Y(y) = ye^{-y}.$$

Dans la situation où les variables X et Y étaient indépendantes, le produit de leurs densités serait égal à la densité de la loi conjointe, ce qui n'est pas le cas.

Solution 1-9. On désigne respectivement par f_X, f_Y et $f_{(X,Y)}$ les densités de X, de Y et de la loi conjointe (X, Y). On cherche à calculer la fonction de répartition de R. Pour tout $t \in \mathbb{R}_+$, la condition

$$R \leq t \iff X^2 + Y^2 \leq t^2 \iff R \in D(\boldsymbol{o}, t)$$

où $D(\boldsymbol{o}, t)$ désigne le disque de centre l'origine et de rayon t. On a ainsi,

$$\mathbb{P}(R \leq t) = \int_{D(\boldsymbol{o},t)} f_{(X,Y)}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

Par indépendance la densité de la loi conjointe (X, Y) est le produits de celles de X et de Y. Donc

$$\mathbb{P}(R \leq t) = \int_{D(\boldsymbol{o},t)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,\mathrm{d}x\,\mathrm{d}y$$

Au vu du lieu d'intégration il est naturel de vouloir passer en coordonnées polaires. Je rappelle qu'on a, pour pouvoir effectuer ce changement de variables, on utilise dans la pratique les relations

$$x = r\cos(\theta), \quad y = r\sin(\theta) \quad \text{et} \quad \mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta.$$

On en déduit

$$\mathbb{P}(R \leq t) = \frac{1}{2\pi}\int_0^{2\pi}\int_0^t e^{-r^2/2} r\,\mathrm{d}r\,\mathrm{d}\theta = \int_0^t e^{-r^2/2} r\,\mathrm{d}r.$$

La loi de R est donc donné par la densité $f_R(r) = re^{-r^2/2}\mathbb{1}_{[0,+\infty]}(r)$.
L'espérence de R est par définition

$$\mathbb{E}(R) = \int_0^{+\infty} e^{-r^2/2}\,\mathrm{d}r$$

Par symétrie de la densité d'une loi gaussienne d'espérence 0 et de variance 1, l'espèrence de cette dernière est le double de celle qui nous intéresse. Ainsi

$$\mathbb{E}(R) = \sqrt{\frac{\pi}{2}}$$

On ne connaît ici la densité de la loi conjointe de X et R. Pour montrer que c'est deux lois ne sont pas indépendantes on peut procéder en exhibant une évènement de la loi conjointe dont la probabilités ne correspond pas au produit des évènements marginaux qui le composent. Ici il suffit de voir que

$$\mathbb{P}(X > 1, R < 1) = 0$$

mais $\mathbb{P}(X > 1) \neq 0$ et $\mathbb{P}(R < 1) \neq 0$.

Solution 1-10. La densité f_X de X est donnée par

$$f_X(x) = \int_D 3\mathbb{1}_D(x, y)\,\mathrm{d}y.$$

La variable x étant fixée, la fonction caractéristique $\mathbb{1}_D(x, y)$ vaut 1 ssi $y \in [x^2, \sqrt{x}]$ avec $x \in [0, 1]$. Donc

$$f_X(x) = \left(\int_{x^2}^{\sqrt{x}} 3\,\mathrm{d}y\right)\mathbb{1}_{]0,1]}(x)$$

On utilise ici la positivité de l'intégrale (Thm de Fubini) pour pouvoir intervertir les éléments d'intégration. On en déduit

$$f_X(x) = 3\left(\sqrt{x} - x^2\right)\mathbb{1}_{]0,1]}(x).$$

Pour le cas de Y il suffit de remarque qu'on peut également écrire D sous la forme

$$D = \left\{(x, y) \mid 0 < y^2 < x < \sqrt{y} < 1\right\}.$$

En manipulant l'encadrement dans la définitionde D. On en déduit donc que

$$f_Y(y) = 3\left(\sqrt{y} - y^2\right)\mathbb{1}_{]0,1]}(y).$$

On en déduit les densités conditionnelles

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \frac{1}{\sqrt{y} - y^2}\mathbb{1}_D(x, y)$$

car l'appartenance à D est plus contraignante que celle donnée par $\mathbb{1}_{[0,1]}$. La densité conditionnelle de Y sachant $X = x$ est donnée par la même expression au noms des variables près. À partir de là on trouve

$$\mathbb{E}(X \mid Y = y) = \int_{\mathbb{R}} x f_{X|Y=y}(x) \mathbb{1}_D(x, y) \, \mathrm{d}x = \int_{y^2}^{\sqrt{y}} \frac{x}{\sqrt{y} - y^2} \, \mathrm{d}x.$$

Un calcul direct donne

$$\mathbb{E}(X \mid Y = y) = \frac{\sqrt{y} + y^2}{2} \mathbb{1}_{]0,1]}(y) \, ^a.$$

On en déduit

$$\mathbb{E}(X \mid Y) = \frac{\sqrt{Y} + Y^2}{2} \mathbb{1}_{]0,1]}(Y).$$

Le calcul dans le cas de $\mathbb{E}(Y \mid X)$ est identique.

a. Rappel : $a^2 - b^2 = (a - b)(a + b)$.