Initiation aux statistiques et probabilités : recueil d'exercices

R. S. Stoica

Université Lille 1
Laboratoire Paul Painlevé
59655 Villeneuve d'Ascq Cedex, France

radu.stoica@math.univ-lille1.fr

Septembre, 2010

1 Introduction

Les exercices de ce recueil ont constitué les sujets d'examen du cours "Initiation aux statistiques et probabilités" que je donne aux étudiants en deuxième année de licence en "Sciences de la Vie et de la Terre" à l'Université Lille 1, depuis l'automne 2006.

Le cours comprend un volume horaire de 20 heures, et il est ce que l'on appelle un cours-td. C'est-à-dire que l'enseignement comprend à la fois des parties de cours magistral et des parties consacrées aux applications, aux exercices. La différence par rapport au cadre classique cours et td, c'est que dans ce cas c'est l'enseignant qui décide du poids qu'il doit donner à chacune de ces deux composantes.

Malgré la souplesse de cette nouvelle formule, je trouve le temps affecté à cet enseignement vraiment insuffisant. A vrai dire, à cause de cela j'ai eu l'intention d'appeler ce cours "'Éveil aux statistiques et probabilités". En effet, selon la formule classique 20 heures comprendraient 5 séances de cours de 2 heures, completées par 5 séances de travaux dirigés de 2 heures également. Suivant la même ancienne formule, considérons le semestre de 14 semaines d'enseignement avec un cours et un td par semaine. Il est facile de remarquer quelques faits notables, comme par exemple celui-ci : $14 \times 4 \neq 20$.

A quelques rares exceptions, le niveau général en mathématiques des étudiants rencontrés en cette filière est faible. Dans mon opinion, ceci est indépendant de leur intelligence, leur capacité d'apprendre et leur soif de vivre. Malgré toutes leurs qualités, ce cours pose beaucoup de difficultés aux étudiants car il s'appuie sur des pre-réquis mathématiques solides. En plus, c'est peut-être un des premiers enseignements où ils sont amenés à "réfléchir" et non plus "faire des exercices". Dans ce contexte, pendant ce cours j'ai décidé de limiter les dégâts. Cependant, souvent je me suis demandé si je ne suis pas en train d'enseigner le programme de seconde ...

L'Université a une double mission. La première est de préparer les étudiants à intégrer les structures socio-économiques d'aujourd'hui, qui sont toujours en perpétuelle transformation. La deuxième mission est de préparer les étudiants à trouver des réponses aux questions qu'aujourd'hui on trouve très difficiles et aussi à formuler les questions et les problèmes ouverts des années à venir.

Je ne peux pas m'empêcher de me demander comment nos étudiants vont trouver des solutions aux problèmes qui se poseront bientôt à eux ? Sans doute, en s'appuyant sur leur extraordinaires qualités. Peut-être aussi en utilisant un peu l'expérience de leurs enseignants. Ceci dit, je ne sais pas si avec le morcellement actuel de l'enseignement dans certaines filières universitaires auquel s'associe trop souvent la réduction du nombre d'heures, l'on donne aux étudiants les éléments les plus concrets pour avoir confiance dans la vie.

Pour la construction de cet enseignement j'ai beneficié de l'aide de mes collègues Jeanne Devolder, Nelly Hanoune, Philippe Heinrich et Bernard Lecocq qui ont généreusement mis à ma disposition leur ressources pédagogiques. J'ai également utilisé des exercices et des données provenant de pages Web d'autres collègues en France ou bien à l'étranger. Les documents mis en ligne par le collectif des statisticiens et probabilistes de l'Institut de Mathématiques de l'EPFL a constitué également une source importante d'exercices. J'ai aussi utilisé des exemples et des exercices des livres très connues comme [1] ou la première édition de [2]. Certaines des données utilisées proviennent de sites `http://www.insee.fr/` ou `http://www.statsci.org/`. A tous ces gens et à tous ceux que seulement par hasard j'ai dû oublier, j'adresse mes remerciements les plus sincères.

2 Exercices

Exercice 1.1. Les données suivantes représentent les valeurs d'un indicateur statistique calculé chaque année pour les pays de l'UE. Cet indicateur mesure l'émission totale de gaz à effet de serre. (Données INSEE).

Année	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Allemagne	88,9	89,6	86,6	84,6	81,9	81,8	83,2	81,7	81,7	80,9
Espagne	110,0	107,7	114,8	118,5	128,2	133,1	133,2	139,1	141,6	147,0
France	98,5	101,3	100,1	102,5	99,6	98,7	99,0	97,4	97,9	97,9

a) Calculez les quartiles empiriques $q_{25\%}$, $q_{50\%}$ et $q_{75\%}$ correspondantes.
b) Tracez et comparez les diagrammes en boîtes(-à moustaches) de données.
c) Est-ce que vous identifiez des valeurs aberrantes ?
d) Interprétez les résultats obtenus.

Exercice 1.2. Les données suivantes représentent les quantités de précipitations enregistrées pour des nuages naturels ainsi que pour des nuages traités chimiquement.

Nuages naturels :	7.09	6.72	5.91	5.84	5.77	5.49	5.09	4.99
	4.55	4.46	4.39	4.22	3.85	3.71	3.60	3.36
	3.35	3.26	3.26	3.19	3.07	2.85	2.44	1.58
	1.58	0.01						
Nuages traités :	7.91	7.43	7.41	6.88	6.55	6.19	6.06	5.81
	5.71	5.61	5.61	5.54	5.49	5.30	5.29	4.86
	4.77	4.77	4.74	4.52	3.70	3.48	3.44	2.86
	2.04	1.41						

a) Calculez les quartiles empiriques $q_{25\%}$, $q_{50\%}$ et $q_{75\%}$ correspondantes.
b) Tracez et comparez les diagrammes en boîtes(-à moustaches) de données.
c) Est-ce que vous identifiez des valeurs aberrantes ?
d) Interprétez les résultats obtenus.

Exercice 1.3. Les données suivantes représentent l'aire du corpus callosum (partie qui relies les deux hémisphères du cerveau) en $[cm^2]$ pour un échantillon de 10 femmes et 10 hommes.

Femmes	6.08	5.73	6.22	5.80	7.99	8.42	7.44	6.84	6.48	6.43
Hommes	7.99	8.76	6.32	6.32	7.60	7.62	6.03	6.59	7.52	7.67

a) Tracez les deux diagrammes tige-feuilles.
b) Calculez les quantiles empiriques $Q_{25\%}$, $Q_{50\%}$ et $Q_{75\%}$.
c) Tracez et comparez les diagrammes en boîtes(-à moustaches) de deux suites de données.

Exercice 1.4. Les pistons d'un moteur automobile comportent des anneaux forgés. Les données suivantes représentent le diamètre intérieur (en μm$\times$100) de ces pièces.

1030	1002	992	995	992	1014	1121	1002	1015	1009
1001	988	1024	1021	1005	994	997	1007	1015	985

a) Tracez l'histogramme.
b) Calculez les quantiles empiriques $Q_{25\%}$, $Q_{50\%}$ et $Q_{75\%}$, ainsi que la moyenne et l'écart-type de l'échantillon.
c) Après vérification, une valeur aberrante a été détéctée (1121). Après avoir éliminé cette valeur calculez la moyenne et la médiane. Comparez les résultats obtenus avec ceux du point b).

Exercice 1.5. Pour comparer l'effet de la vitamine C du jus d'orange et de l'acide ascorbique de synthèse, on a donné du jus d'orange à un group de 10 cobayes, et de la vitamine de synthèse à un groupe de 10 autres cobayes, pendant six semaines, et on a mesuré la longueur des odontoblastes des incisives. On a obtenu les résultats suivants :

Jus d'orange	8.2	9.4	9.6	9.7	10.0	14.5	15.2	16.1	17.6	21.5
Acide ascorbique	4.2	5.2	5.8	6.4	7.0	7.3	10.01	11.2	11.3	11.5

a) Tracez les deux diagrammes tige-feuilles.
b) Calculez les quantiles empiriques $Q_{25\%}$, $Q_{50\%}$ et $Q_{75\%}$.
c) Tracez et comparez les diagrammes en boîtes(-à moustaches) de deux suites de données.
d) Est-ce que vous identifiez des valeurs aberrantes ?
e) Interprétez les résultats obtenus.

Exercice 1.6. Une entreprise souhaite comparer les performances en terme de bruit des deux filtres pour automobiles. On a obtenu les mesures suivantes :

Filtre A	810	820	820	840	840	845	785	790	785	835
	835	835	845	855	850	760	760	770		
Filtre B	820	820	820	820	825	775	775	775	825	825
	815	825	825	770	760	765	820	825		

a) Tracez les deux diagrammes tige-feuilles.
b) Calculez les quantiles empiriques $Q_{25\%}$, $Q_{50\%}$ et $Q_{75\%}$ pour les deux filtres.
c) Tracez et comparez les diagrammes en boîtes(-à moustaches) de deux suites de données.
d) Est-ce que vous identifiez des valeurs aberrantes ?
e) Interprétez les résultats obtenus.

Exercice 1.7. Les données suivantes représentent le pourcentage de la force de travail employée en agriculture, industrie et services pour 10 pays européens pendant l'année 1960. Nous allons comparer la distribution de la force de travail dans ces trois secteurs.

Agriculture	14	4	18	15	20	6	20	36	27	44
Industrie	53	56	45	60	44	52	49	30	46	33
Services	33	40	37	25	36	42	32	34	28	23

a) Tracez les trois diagrammes tige-feuilles.
b) Calculez les quartiles empiriques $Q_{25\%}$, $Q_{50\%}$ et $Q_{75\%}$ correspondantes.
c) Tracez et comparez les diagrammes en boîtes(-à moustaches) de données.
d) Est-ce que vous identifiez des valeurs aberrantes ?
e) Interprétez les résultats obtenus.

Exercice 1.8. Les données suivantes représentent les pourcentages de femmes qui travaillent enregistrés dans 19 villes des Etats-Unis, pendant les années 1968 et 1972, respectivement.

1972	45	50	52	45	46	55	60	49	35	55	52	53	57	53	59	64	50
1968	45	50	52	45	43	55	45	34	45	54	42	51	49	54	50	58	49

a) Tracez les deux diagrammes tige-feuilles.
b) Calculez les quartiles empiriques $Q_{25\%}$, $Q_{50\%}$ et $Q_{75\%}$ correspondantes.
c) Tracez et comparez les diagrammes en boîtes(-à moustaches) de données.
d) Est-ce que vous identifiez des valeurs aberrantes ?
e) Interprétez les résultats obtenus.

Exercice 2.1. Les données ci-dessous représentent vingt mesures de l'épaisseur d'un fil en μm.

553	558	544	547	488	561	565	427	529	530
527	555	534	579	575	529	510	586	585	536

a) Représentez ces données à l'aide d'un histogramme.
b) Calculez la moyenne et la variance de l'échantillon.
Dans la suite de cet exercice, on suppose que les données proviennent d'une loi normale.
c) Superposez à l'histogramme la courbe de la densité normale avec cette moyenne et cette variance.
d) Calculez la probabilité que l'épaisseur soit inférieure à $510\mu m$.
e) Déterminez l'épaisseur du fil telle que 20% des valeurs lui soient supérieures.

Exercice 2.2. Les données ci-dessous représentent la taille (en millimètres) des oeufs de deux oiseaux, le *Pipit des arbres* et le *Troglodyte familier*.

Pipit des arbres	21.05	21.85	22.05	22.45	22.65	23.25	23.25	23.25
	23.45	23.45	23.65	23.85	24.05	24.05	24.05	
Troglodyte familier	19.85	20.05	20.25	20.85	20.85	20.85	21.05	21.05
	21.05	21.25	21.45	22.05	22.05	22.05		

a) Tracez les deux diagrammes tige-feuilles.
b) Calculez les quantiles empiriques $Q_{25\%}$, $Q_{50\%}$ et $Q_{75\%}$.
d) Tracez et comparez les diagrammes en boîtes(-à moustaches) de deux suites de données.
d) Parmi les données, y a-t-il des valeurs aberrantes ?

Exercice 2.3. Soit les observations suivantes :

X	9	3	2	5	10	1	9
Y	20	9	5	13	22	5	20

a) Dessinez le diagramme en nuage de points.
b) Calculez le coefficient de corrélation r.
c) Calculez les paramètres a et b de la droite de moindres carrés $Y = aX + b + \epsilon$.

Exercice 2.4. Les données suivantes sont issues d'un étude statistique du gouvernement Britannique. Dans 11 régions, on a enregistré les dépenses moyennes d'un foyer pour les boissons alcoolisées et le tabac :

Alcool	6.47	6.13	6.19	4.89	5.63	4.52	5.89	4.79	5.27	6.08	4.02
Tabac	4.03	3.76	3.77	3.34	3.47	2.92	3.20	2.71	3.53	4.51	4.56

a) Dessinez le diagramme en nuage de points.
b) Calculez le coefficient de corrélation r.
c) Calculez les paramètres a et b de la droite de moindres carrés $Y = aX + b + \epsilon$.
d) Est-ce que vous pouvez observer un couple de valeurs qui a une grande influence sur la forme de la droite de moindre carrés ?

Exercice 2.5. Les données suivantes représentent le nombre de buts marqués et encaissés par l'équipe Olympique Lyonnais lors de la première partie du saison jusqu'au 15 décembre 2008 de la Ligue 1 de Football 2008/2009. Le nombre de buts a été prélevé tous les 15 minutes.

Minutes	0 - 15	15 - 30	30 - 45	45-60	60 - 75	75 - 90
Buts marqués	2	2	8	4	4	2
Buts encaissés	2	3	1	2	2	2

a) Dessinez les histogrammes des deux jeux de données.
b) Comparez les deux distributions en utilisant la moyenne, la médiane et le mode.
c) Quelle est la loi de probabilité des buts marqués ? Et celle des buts encaissés ?
d) D'après ces données, quelle est la probabilité que Lyon marque dans le premier quart d'heure ou dans le dernier quart d'heure du match ? Et quelle est la probabilité que Lyon encaisse un but après un quart d'heure de jeu ?
e) Quelles sont les qualités de Lyon : défense ou attaque ? Répondez à cette question en vous laissant guider par la raison et surtout par l'analyse statistique des données de cet exercice.

Exercice 2.6. Les données suivantes représentent le nombre de buts marqués et encaissés par l'équipe de Bordeaux lors de la saison 2008/2009 de la Ligue 1 de Football. Le nombre de buts a été prélevé tous les 15 minutes.

Minutes	0 - 15	15 - 30	30 - 45	45-60	60 - 75	75 - 90
Buts marqués	7	6	9	12	12	16
Buts encaissés	4	5	7	8	3	6

a) Dessinez les histogrammes des deux jeux de données.
b) Comparez les deux distributions en utilisant la moyenne, la médiane et

le mode.
c) Quelle est la loi de probabilité des buts marqués ? Et celle des buts encaissés ?
d) D'après ces données, quelle est la probabilité que Bordeaux marque dans le premier quart d'heure ou dans le dernier quart d'heure du match ? Et quelle est la probabilité que Bordeaux encaisse un but après un quart d'heure de jeu ?
e) Quelles sont les qualités de Bordeaux : défense ou attaque ? Répondez à cette question en vous laissant guider par la raison et surtout par l'analyse statistique des données de cet exercice.

Exercice 2.7. Les données suivantes représentent les mesures effectuées par Henry Cavendish en 1798 pour évaluer la densité de la terre :

5.50	5.57	5.42	5.61	5.53	5.47	4.88	5.62	5.63	4.07
5.29	5.34	5.26	5.44	5.46	5.55	5.34	5.30	5.36	5.79
5.75	5.29	5.10	5.86	5.58	5.27	5.85	5.65	5.39	

a) Tracez le diagramme tige-feuilles.
b) Représentez les données à l'aide d'un histogramme.
c) Est-ce qu'il y a des données aberrantes ? Argumentez.
d) Les expériences montrent que la densité moyenne de la terre est 5.42. Dans ce cas, quel est le meilleur estimateur de cette quantité la moyenne empirique ou la médiane ? Argumentez.

Exercice 2.8. Les données suivantes représentent l'évolution du pourcentage d'emballages (verre) recyclés par rapport à la production d'emballages (Données INSEE - France métropolitaine).

Année	1985	1990	1995	2000	2005	2006	2007
Verre recyclé	16,7	26,8	39,5	49,7	59,7	59,5	61,6

a) Dessinez le diagramme en nuage de points.
b) Calculez le coefficient de corrélation r.
c) Calculez les paramètres a et b de la droite de moindres carrés $Y = aX + b + \epsilon$.
d) Faites une prédiction de ce pourcentage pour l'année 2010.

Exercice 3.1. a) Calculez la moyenne, la médiane et les quartiles ($Q_{25\%}$ et $Q_{75\%}$) d'une variable continue X qui suit une loi uniforme dans l'intervalle

$[10, 20]$.
b) Généralisez au cas d'une loi uniforme sur l'intervalle $[a, b]$.

Exercice 3.2. Les notes possibles données aux réponses d'un questionnaire en biologie vont de 0 à 10. La note moyenne est de 6.7 et l'écart-type de 1.2. En supposant que les notes soient distribuées suivant une loi normale, déterminez
a) la note que les 10% les plus mauvais de la classe n'ont pas dépassée.
b) la note que les 10% les meilleurs de la classe ont tous dépassée.

Exercice 3.3.
a) Déterminez les quantiles à 5%, 10%, 90% et 95% d'une densité normale avec paramètres $\mu = 0$ et $\sigma^2 = 1$.
b) Utilisez les résultats précédents pour déterminer ces mêmes quantiles pour une densité normale de paramètres $\mu = 5$ et $\sigma^2 = 4$.

Exercice 3.4. Soit X une variable aléatoire qui suit une loi normale de paramètres $\mu = 2$ et $\sigma^2 = 4$. Calculez :
a) $P(X < 2)$
b) $P(X + 2 > 2)$
c) $P(|2X + 2| \leq 2)$
d) $P(|X^2 + \cos(X) - 2| < 0)$

Exercice 3.5.
a) Déterminez les quantiles à 7%, 17%, 92% et 97% d'une densité normale avec paramètres $\mu = 0$ et $\sigma^2 = 1$.
b) Utilisez les résultats précédents pour déterminer ces mêmes quantiles pour une densité normale de paramètres $\mu = -3$ et $\sigma^2 = 4$.

Exercice 3.6.
a) Déterminez les quantiles à 8%, 19%, 91% et 99% d'une densité normale avec paramètres $\mu = 0$ et $\sigma^2 = 1$.
b) Utilisez les résultats précédents pour déterminer ces mêmes quantiles pour une densité normale de paramètres $\mu = -1.5$ et $\sigma^2 = 9$.

Exercice 3.7. Soit X une variable aléatoire réelle de loi $\mathcal{N}(\mu, \sigma^2)$, avec $P(X > 3) = 0.8413$ et $P(X > 9) = 0.0228$.
a) En utilisant la table de la loi normale centrée réduite, établir deux équations linéaires dont μ et σ sont solutions.

b) Déterminez μ et σ.

Exercice 3.8. Soit X une variable aléatoire réelle de loi $\mathcal{N}(\mu, \sigma^2)$, avec $\mu = 2$ et $\sigma^2 = 9$. Calculez.
a)$P(X < 0)$
b)$P(X + 2 \geq 2)$
c)$P(|X + 2| \leq 2)$
d)$P(|X^2 + \sqrt[3]{X} - 2| < 0)$

Exercice 4.1. Les densités normales sont souvent utilisés pour décrire les scores des tests de quotient intellectuel (QI). Les spécialistes admettent générallement comme paramètres $\mu = 100$ et $\sigma = 15$ pour répresenter la répartition de cette variable.
a) Calculez la probabilité qu'une personne ait un QI compris entre 100 et 130.
b) Déterminez la proportion de la population dont le QI est supérieur à 130.

Exercice 4.2. Soit X une variable aléatoire normale de paramètres $\mu = 3$ et $\sigma = 2$. Calculez
a) $P(2 \leq X < 5)$
b) $P(X \geq 0)$
c) $P(|X - 3| \geq 1)$
Soient maintenant μ et σ quelconques. Calculez
c) $P(\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma)$

Exercice 4.3. Soit une boîte qui contient cent allumettes dont dix sont défectueuses.
a) On choisit cinq allumettes au hasard. Quelle est la probabilité qu'aucune ne soit défectueuse ? Quelle est la probabilité qu'exactement une soit défectueuse ?
b) Si l'on choisit dix allumettes quelle est la probabilité qu'exactement deux soient défectueuses ?

Exercice 4.4. Soit P la probabilité définie sur l'espace de configurations Ω. Considérons deux événements quelconques, A et B des sous-ensembles de Ω. Prouvez les propriétes suivantes :
a) $P(A^c) = 1 - P(A)$
b) $A \subseteq B \Rightarrow P(A) \leq P(B)$
c) $0 \leq P(A) \leq 1$
d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Exercice 4.5. Dans une clinique, on effectue un test médical sur $n = 140$ personnes, pour savoir si elles sont séropositives ou pas. La probabilité qu'une personne soit détectée positive à ce test, alors qu'elle n'est pas porteuse du virus HIV, est $p = 0.005$. Soit X le nombre de personnes détectées positives au test, alors qu'elles sont en bonne santé.
a) Quelle est la loi de X ?
b) Calculez $P(X = 2)$ et $P(X \geq 1)$.

Exercice 4.6. Soit la densité de probabilité donnée par

$$f(x) = a\left[\frac{1+\cos(2x)}{2} + \frac{\tan^2(x)}{1+\tan^2(x)}\right]$$

et définie dans l'intervalle $[0, \pi/2)$.
a) Trouvez la constante a tel que $f(x)$ soit bien définie.
b) Calculez la $P(\pi/6 < X < \pi/3)$.
c) Calculez $\mathbb{E}[X]$ et $Var[X]$.

Exercice 5.1. On considère deux sous-ensembles disjoints A et B d'un ensemble Ω, pour lesquels $P(A) = a > 0$ et $P(B) = b > 0$. Ecrivez en fonction de a et b les probabilités suivantes.
a) $P(A \cup B)$
b) $P(A \cap B)$
c) $P(A^c \cup B)$
d) $P((A \cap B)^c)$
e) $P(A^c \cap B^c)$
f) $P(A \setminus B)$

Exercice 5.2. On considère deux sous-ensembles non-disjoints A et B d'un ensemble Ω. Nous savons également que $A \subset B$, $P(A) = a > 0$ et $P(B) = b > 0$. Ecrivez en fonction de a et b les probabilités suivantes.
a) $P(A \cup B)$
b) $P(A \cap B)$
c) $P(A^c \cap B)$
d) $P((A \cup B)^c)$
e) $P(A^c \cup B)$
f) $P(A^c \setminus B^c)$

Exercice 5.3. La durée de vie d'un composant d'une machine a une distribution continue sur l'intervalle $[0, 40]$ avec densité de probabilité $f(x) =$

$c(2x+5)$.
a) Trouvez la constante c tel que $f(x)$ soit bien une densité de probabilité.
b) Calculez la probabilité que la durée de vie du composant de la machine soit moins que 6.
c) Calculez $\mathbb{E}[X]$ et $Var[X]$.

Exercice 5.4. Soit la densité de probabilité donnée par $f(x) = c\sin(x)$ et définie dans l'intervalle $[0, \pi]$.
a) Trouvez la constante c tel que $f(x)$ soit bien définie.
b) Calculez la $P(\pi/4 < X < 3\pi/4)$.
c) Calculez $\mathbb{E}[X]$ et $Var[X]$.

Exercice 5.5. Soit la densité de probabilité donnée par $f(x) = a\cos(x)$ et définie dans l'intervalle $[-\pi/2, \pi/2]$.
a) Trouvez la constante a tel que $f(x)$ soit bien définie.
b) Calculez la $P(-\pi/4 < X < \pi/4)$.
c) Calculez $\mathbb{E}[X]$ et $Var[X]$.

Exercice 5.6. Soit la densité de probabilité donnée par $f(x) = a(\cos^2(x) + \sin^2(x))$ et définie dans l'intervalle $[0, \pi/2]$.
a) Trouvez la constante a tel que $f(x)$ soit bien définie.
b) Calculez la $P(\pi/6 < X < \pi/3)$.
c) Calculez $\mathbb{E}[X]$ et $Var[X]$.

Exercice 5.7. Soit la densité de probabilité donnée par $f(x) = a\exp[-2x]$ et définie dans l'intervalle $[0, \infty)$.
a) Trouvez la constante a tel que $f(x)$ soit bien définie.
b) Calculez les quartiles de cette distribution, c'est à dire $q_{25\%}, q_{50\%}$ et $q_{75\%}$.

References

[1] D. S. Moore and G. P. McCabe. *Introduction to the Practice of Statistics.* W. H. Freeman and Company, 1989.

[2] G. Saporta. *Probabilités, analyse des données et statistique, 2ème édition.* Technip, 2006.