Statistiques Appliquées

Luc Deneire

Université de Xidian
Programme International Polytech

Cahier d'Exercices

Nom (Pinyin):	Prénom :	No:

PREAMBULE IMPORTANT

Deux cahiers d'exercices sont proposés, un pour le premier semestre (Probabilités) et un pour le deuxième semestres (Statistiques).

Chaque chapitre est divisé en 10 Travaux Dirigés, un pour chaque séance, la dernière séance étant une séance de questions/réponses.

Pour chaque séance d'exercices, vous devez :

- préparer quatre exercices avant de venir à la séance. Ces exercices seront corrigés au tableau, collectivement ;
- pour une partie des exercices qui seront fait en cours, vous devez écrire proprement la correction dans le cahier ;
- vous pouvez utiliser la page de gauche comme cahier de brouillon.

DE LA PROPRETE DE CE CAHIER :

Ce cahier doit rester le plus propre possible :

- NE PAS ECRIRE sur le texte des énoncés, une marge à gauche est laissée pour les traductions (une traduction des énoncés est donnée en anglais, pour votre facilité).
- UTILISEZ la page de gauche pour vos brouillons.
- Ecrivez proprement sur la page de droite. Faites des phrases en FRANCAIS, il ne suffit pas d'écrire les calculs !

Chapter 1

Probabilités

1.1 TD 1 : Probabilités simples / Simple Probabilities

1.1.1 Cartes

Une expérience aléatoire a comme univers : { As, Roi, Dame, Valet, 10 }, avec les probabilités { 0.1 , 0.1, 0.2, 0.4, 0.2 }. On définit l'évènement $A = \{$ As, Roi, Dame $\}$ et l'évènement $B = \{$ Valet, 10, Dame $\}$.

Déterminez les probabilités :

A random experiment is made of the following universe : { Ace, King, Queen, Jack, 10 }, with the associated probabilities : { 0.1 , 0.1, 0.2, 0.4, 0.2 }. We define the event $A = \{$ Ace, King, Queen$\}$ and the event $B = \{$ Jack, 10, Queen $\}$.
Determine the following probabilities :

- $\mathbb{P}(A)$
- $\mathbb{P}(\bar{A})$
- $\mathbb{P}(A \cup B)$
- $\mathbb{P}(A \cap B)$
- $\mathbb{P}(A|B)$
- $\mathbb{P}(B|A)$

Solution 1.1.1

- $\mathbb{P}(A) = 0.1 + 0.1 + 0.2 = 0.4$
- $\mathbb{P}(\bar{A}) = 1 - 0.4 = 0.6$
- $\mathbb{P}(A \cup B) = 1$
- $\mathbb{P}(A \cap B) = \mathbb{P}(\text{Queen}) = 0.2$
- $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = 0.2/0.8 = 0.25$
- $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = 0.2/0.4 = 0.5$

1.1.2 Trouver l'univers

Find the Universe

Une expérience aléatoire consiste à jeter une balle dans un bac comprenant six trous, chacun de dimension identique et de taille plus grande que la taille de la balle. La balle tombe donc systématiquement à travers un trou, avec la même probabilité quelle que soit le trou. Quel est l'univers ? Quelle est la probabilité que la balle passe par les trous 1 ou 2 ? Quelle est la probabilité que la balle ne passe ni par le trou 4, ni par le trou 5 ?

A random experiment consists in throwing a ball in a tray with 6 holes. The ball always falls in a hole, with the same probability for all holes.
What is the universe ? What is the probablity that the ball falls either in hole 1 or in hole 2 ? What is the probability that the ball falls neither in hole 4, neither in hole 5 ?

Solution 1.1.2

> *The universe is :* $\{$"*The ball falls in hole 1*", "...2", "...3", "...4", "...5", "...6"$\}$ *and the associated probabilities are* $\{1/6, 1/6, ..., 1/6\}$.
> *The probability that the ball falls in hole 1 or 2 is 1/6 + 1/6 (because the two events are disjoint, hence* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$*).*
> *The probability that the ball falls neither in hole 4 nor 5 is 1-2/6 (*$\mathbb{P}(\Omega - A) = 1 - \mathbb{P}(A)$*).*

1.1.3 Papa à poux ...

Father with lice ...

Vos professeurs préférés se catégorisent en professeurs à poux, ou pas à poux, et professeurs papa et pas papa ... (et donc peut-être maman, mais ça, c'est une autre histoire). Sur 100 professeurs, on obtient les résultats suivants :

Prof.	à poux	pas à poux
papa	34	42
pas papa	16	8

On note l'événement A : professeur papa et l'événement B : professeur à pou. On demande d'abord de déterminer l'univers, et ensuite, de déterminer les probabilités suivantes :

Your beloved professors can be cast in different categories : professors with lice and professors without lice, professors that are fathers or not fathers ... (and hence perhaps mothers, but that's another story). On 100 professors, we have the following table :

Prof.	*with lice*	*without lice*
father	34	42
not father	16	8

We define the event A : professor and father and the event B : professor with lice. First determine the universe, then determine the following probabilities :

- $\mathbb{P}(A), \mathbb{P}(B)$
- $\mathbb{P}(\bar{A}), \mathbb{P}(\bar{B})$
- $\mathbb{P}(A \cup B), \mathbb{P}(A \cup \bar{B}), \mathbb{P}(\bar{A} \cup B)$
- $\mathbb{P}(A \cap B), \mathbb{P}(A \cap \bar{B}), \mathbb{P}(\bar{A} \cap B)$
- $\mathbb{P}(A|B), \mathbb{P}(A|\bar{B}), \mathbb{P}(\bar{A}|B)$

Solution 1.1.3

$\Omega = \{$*"Father with lice", "Father without lice", "not father with lice", not father without lice"*$\}$. *with associated probabilities* $\{0.34; 0.42; 0.16; 0.08\}$

- $\mathbb{P}(A) = 0.34 + 0.42 = 0.76$, $\mathbb{P}(B) = 0.34 + 0.16 = 0.5$
- $\mathbb{P}(\bar{A}) = 1 - 0.76 = 0.24$, $\mathbb{P}(\bar{B}) = 0.5$
- $\mathbb{P}(A \cup B) = 0.92$, $\mathbb{P}(A \cup \bar{B}) = 0.84$, $\mathbb{P}(\bar{A} \cup B) = 0.58$
- $\mathbb{P}(A \cap B) = 0.34$, $\mathbb{P}(A \cap \bar{B}) = 0.42$, $\mathbb{P}(\bar{A} \cap B) = 0.16$
- $\mathbb{P}(A|B) = \dfrac{0.34}{0.5} = 0.68$, $\mathbb{P}(A|\bar{B}) = \dfrac{0.42}{0.5} = 0.84$, $\mathbb{P}(\bar{A}|B) = \dfrac{0.16}{0.5} = 0.32$

1.1.4 Au tour des enfants ... problèmes classiques

Classical children problems

On suppose que lorsqu'un couple a un enfant, la probabilité d'avoir un garçon est égale à $p = 0.51$. On suppose par ailleurs que les naissances successives au sein d'un même couple sont mutuellement indépendantes. Dans un premier temps, exprimez clairement l'univers, ensuite, donnez les solutions aux questions.

Un couple a (exactement) deux enfants :

a. quelle est la probabilité que ces enfants soient deux garçons ?

b. quelle est la probabilité que ces enfants soient deux filles ?

c. quelle est la probabilité que le couple ait un garçon et une fille ?

d. on vous dit que l'aîné est une fille; quelle est la probabilité que le cadet soit une fille?

e. on vous dit qu'au moins l'un des enfants est une fille: quelle est la probabilité que les deux enfants soient des filles?

f. vous rencontrez cette famille avec un des enfants (l'autre enfant est absent): c'est une fille. Quelle est la probabilité que les deux enfants soient des filles ?

We assume that, when a couple has a child, the probability to have a boy is $p = 0.51$. We also assume that subsequent births for a same couple are independent. In a first phase, express (as clearly as possible) the universe. In a second phase, answer the following questions.
A couple has (exactly) two children :

a. what is the probability that the two children are both boys ?

b. what is the probability that the two children are both girls ?

c. what is the probability that the couple has one boy and one girl ?

d. you are told that the eldest child is a girl, what is the probability that the youngest child is a girl ?

e. you are told that at least one of the children is a girl : what is the probability that both children are girls ?

f. you meet the family with one of the children (the other child is absent) and the child you meet is a girl. What is the probability that both children are girls ?

Solution 1.1.4

Considering a couple with one child, the universe for the child is $\Omega = \{\text{"The child is a Boy"}, \text{"The child is a Girl"}\}$, which we can shortcut as $\Omega = \{B, G\}$
Considering a couple with two children, the universe can be written as :

$$\Omega = \left\{ \begin{array}{l} \text{"The Eldest child is a Girl AND The Youngest child is a Boy",} \\ \text{"The Eldest child is a Boy AND The Youngest child is a Boy",} \\ \text{"The Eldest child is a Girl AND The Youngest child is a Girl",} \\ \text{"The Eldest child is a Boy AND The Youngest child is a Girl"} \end{array} \right\},$$

which we can shortcut as $\Omega = \{GB, BB, GG, BG\}$
The associate probabilities are $\{p.(1-p), p^2, (1-p)^2, p.(1-p)\}$ (due to independence).
Then we have
a. $P = \mathbb{P}(BB) = p^2$
b. $P = \mathbb{P}(GG) = (1-p)^2$
c. $P = \mathbb{P}(GB + BG) = 2p(1-p)$
d. $P = \mathbb{P}(G) = 1-p$, we can also say that it is $\mathbb{P}(GG|(GG \cup GB)) = \frac{(1-p)^2}{(1-p)^2+p.(1-p)} = 1-p$
e. $P = \mathbb{P}(GG|(GG \cup BG \cup GB))$ also written as $\mathbb{P}(GG|GG, BG, GB) = (1-p)/(1+p)$ f. This question resembles strongly to the previous, but is VERY different.
A simple answer is : the probability that the family has two girls is the probability that "the other" child is a girl, that is $1-p$.
A more complicated answer is to come back to the notion of experiment and event and to rewrite the universe. Indeed, here there is an additional random experiment : the meeting of the child. Then we can express the meeting as "I meet the eldest child" or "I meet the youngest child". We can assume that we meet the eldest with probability q.
Then, an event can be written in the form : "I meet the eldest child, the eldest child is a Girl and the youngest child is a Boy", which we can shortcut as "EGB". The new universe is then :
$\Omega = \{EGB, YGB, EBB, YBB, EGG, YGG, EBG, YBG\}$ with associated probabilities $\{q.p.(1-p), (1-q).p.(1-p), q.p^2, (1-q).p^2, q.(1-p)^2, (1-q)(1-p)^2, q.p.(1-p), (1-q).p.(1-p)\}$.
In this frame, the probability asked for can be expressed as $\mathbb{P}(B|A)$ where $A = \{EGB, EGG, YGG, YBG\}$ and $B = \{EGG, YGG\}$, with $\mathbb{P}(A \cap B) = (1-p)^2$ and $\mathbb{P}(A) = (1-p)^2 + p.(1-p)$, hence $\mathbb{P}(B|A) =$

$$\frac{(1-p)^2}{(1-p)^2 + p(1-p)} = 1-p$$

1.2 TD 2

1.2.1 Probabilités conditionnelles simples

Simple conditional probabilities

Un fabricant de smartphones sélectionne 3 personnes dans une classe d'étudiants pour tester ses appareils. Dans cette classe, il y a 12 garçons et 20 filles. La sélection se fait selon des probabilités égales pour chaque individu. Déterminez :

- la probabilité que le deuxième testeur soit une fille si le premier est une fille ;

- la probabilité que le deuxième testeur soit un garçon si le premier est une fille ;

- la probabilité que le deuxième testeur est un garçon et que le premier est une fille, exprimez la en fonction de la probabilité du point précédent ;

- la probabilité que le troisième testeur soit une fille si le premier est un garçon ;

- la probabilité que le troisième testeur soit une fille si les deux premiers sont des garçons.

A smartphone selects 3 testers in a pool of 32 persons, of which 12 are boys and 20 girls. The selection follows a uniform distribution (equal probability for all individuals). Compute :

- *the probability that the second tester is a girl if the first one is a girl ;*
- *the probability that the second tester is a boy if the first is a girl ;*
- *the probability that the second tester is a boy and that the first is a girl, express it as a function of the probability of the previous item ;*
- *the probability that the third tester is a girl if the first is a boy ;*
- *the probability that the third tester is a girl if the two first are boys.*

Solution 1.2.1

- *19/31 ;*
- *12/31 ;*
- $\frac{20}{32} \cdot \frac{12}{31}$ *;* $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,.\mathbb{P}(B) = \frac{12}{31} \cdot \frac{20}{32}$ *(B = "the first tester is a girl", A = "the second tester is a boy").*
- *A first (easy) answer is to say that, the second being unknown that the solution is as on the first item : 20/31.*

  A second answer is to express A = "the third tester is a girl" and B = "the first tester is a boy", then $A \cap B$ *= "the third tester is a girl and the first tester is a boy',* $\mathbb{P}(A \cap B) = \frac{12}{32} \cdot \frac{20}{31}$*, and* $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \dfrac{\frac{12}{32} \cdot \frac{20}{31}}{\frac{12}{32}} = 20/31.$

  Another question can be "what about the second one". Lets denote B_1 *the event "*The first tester is a boy*".*

  Then, we look for $\mathbb{P}(F_3|B_1)$*. To do this we will go in the details of the universe* Ω *and express all joint probabilities.*

  We can express this as : $\mathbb{P}(F_3|B_1) = \mathbb{P}(F_3|B_1 \cap B_2) + \mathbb{P}(F_3|B_1 \cap F_2)$*, where obviously, the event* $(F_3|B_1 \cap B_2) = (F_3|B_1 \cap B_2|F_1)$ *(if you're not convinced, just draw the events as sets).*

  Then, applying $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$*, we have :*

  $\mathbb{P}(F_3|B_1) = \mathbb{P}(F_3|(B_1, B_2))\,\mathbb{P}(B_2|F_1) + \mathbb{P}(F_3|B_1, F_2)\,\mathbb{P}(F_2|B_1) = \frac{20}{30}\frac{12}{31} + \frac{19}{30}\frac{20}{31}$

  We can also detail of the universe Ω *and express all joint probabilities.*

  $\mathbb{P}((B_1, B_2, F_3)) = \frac{12.11.20}{32.31.30}$

  $\mathbb{P}((B_1, F_2, F_3)) = \frac{12.20.19}{32.31.30}$

  $\mathbb{P}((B_1, B_2, B_3)) = \frac{12.11.10}{32.31.30}$

  $\mathbb{P}((B_1, F_2, B_3)) = \frac{12.20.11}{32.31.30}$

  $\mathbb{P}((F_1, B_2, F_3)) = \frac{20.12.19}{32.31.30}$

  $\mathbb{P}((F_1, F_2, F_3)) = \frac{20.19.18}{32.31.30}$

  $A = \{(B_1, B_2, F_3); (B_1, F_2, F_3), (F_1, B_2, F_3), (F_1, F_2, F_3)\}$

  $B = \{(B_1, B_2, F_3); (B_1, F_2, F_3)(B_1, B_2, B_3); (B_1, F_2, B_3)\}$

  $A \cap B = \{(B_1, B_2, F_3); (B_1, F_2, F_3)\}$

  $\mathbb{P}(A \cap B) = \frac{12.11.20}{32.31.30} + \frac{12.20.19}{32.31.30} = \frac{12.20}{32.31}$

  $\mathbb{P}(B) = \frac{12.11.20}{32.31.30} + \frac{12.20.19}{32.31.30} + \frac{12.11.10}{32.31.30} + \frac{12.20.11}{32.31.30} = \frac{12}{32}$

  And finaly, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{20}{31}$
- *20/30 .*

1.2.2 Contrôle de qualité de circuits intégrés

Quality control of integrated circuits

Dans une usine de circuits intégrés, un circuit sujet à une contamination importante est défectueux avec une probabilité 0.20, et un circuit qui n'est pas sujet à une contamination est défectueux avec une probabilité 0.005. Quelle est la probabilité qu'un circuit issu de cette usine soit défectueux, sachant que le taux de contamination (nombre de circuits soumis à une contamination importante) est de 20 %.

> *In an integrated circuits factory, a circuit subject to important contamination turns out to be defect with a probability of 0.20 and a circuit which is not subject to contamination is defect with a probability of 0.005. What is the probability that a circuit is defect, if the contamination rate (number of circuits that are subject to an important contamination) is of 20 %?*

Solution 1.2.2

$\Omega = \{CF, \bar{C}F, C\bar{F}, \bar{C}\bar{F}\}$, where C means "contaminated" and F means "faulty". The event A is "ICs are contaminated" and B = "ICs are faulty". Then :

- $\mathbb{P}(B|A) = 0.20, \mathbb{P}\big(B|\bar{A}\big) = 0.005, \mathbb{P}(A) = 0.20$
- $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}\big(B \cap \bar{A}\big) = \mathbb{P}(B|A)\,.\mathbb{P}(A) + \mathbb{P}\big(B|\bar{A}\big)\,.\mathbb{P}\big(\bar{A}\big) = 0.2 \times 0.2 + 0.005 \times (1 - 0.2) = 0.044$

1.2.3 Polluants

Une nouvelle méthode analytique de détection des polluants dans l'eau est à l'essai. Cette nouvelle méthode d'analyse chimique est importante car, si elle était adoptée, elle pourrait être utilisée pour détecter trois contaminants différents - les polluants organiques, les solvants volatils et les composés chlorés - au lieu de devoir utiliser un seul test pour chaque polluant. Les fabricants du test affirment qu'il peut détecter des niveaux élevés de polluants organiques avec une précision de 99,7%, des solvants volatils avec une précision de 99,95% et des composés chlorés avec une précision de 89,7%. Si aucun polluant n'est présent, le test est négatif. Des échantillons sont préparés pour l'étalonnage de l'essai et 60% d'entre eux sont contaminés par des polluants organiques, 27% avec des solvants volatils et 13% avec des traces de composés chlorés. Un échantillon de test est sélectionné au hasard. begin itemize item Quelle est la probabilité que le test soit positif ? item Si le test est positif, quelle est la probabilité que des composés chlorés sont présents? end itemize

[a] *A new analytical method to detect pollutants in water is being tested. This new method of chemical analysis is important because, if adopted, it could be used to detect three different contaminants – organic pollutants, volatile solvents, and chlorinated compounds – instead of having to use a single test for each pollutant. The makers of the test claim that it can detect high levels of organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy, and chlorinated com- pounds with 89.7% accuracy. If a pollutant is not present, the test does not signal. Samples are prepared for the calibration of the test and 60% of them are contaminated with organic pollutants, 27% with volatile solvents, and 13% with traces of chlorinated compounds. A test sample is selected randomly.*

- *What is the probability that the test will signal?*
- *If the test signals, what is the probability that chlorinated compounds are present?*

[a]Taken from Montgomery's book

1.2.4 Conformité

Soient E_1, E_2, et E_3 des échantillons qui sont, pour un pourcentage donné, conforme à des spécifications (1) état solide (2) poids moléculaire et (3) de couleur. Un total de 240 échantillons sont classifiés dans les spécifications E_1, E_2, et E_3, où "yes" indique que l'échantillon est conforme à la spécification, en fonction du tableau suivant :

E_3	E_2	E_1	number of samples
yes	yes	yes	200
		no	5
	no	yes	1
		no	4
no	yes	yes	20
		no	6
	no	yes	4
		no	0

- Les événements E_1, E_2, et E_3 sont-ils mutuellement exclusifs ?
- Les événements $\overline{E_1}$, $\overline{E_2}$, et $\overline{E_3}$ sont-ils mutuellement exclusifs ?
- Que vaut $\mathbb{P}(\bar{E}_1 \text{ ou } \bar{E}_2 \text{ ou } \bar{E}_3)$?
- Quelle est la probabilité qu'un échantillon soit conforme à toutes les spécifications ?
- Quelle est la probabilité qu'un échantillon soit conforme à E_1 ou E_3 ?
- Quelle est la probabilité qu'un échantillon soit conforme à E_1 ou E_2 ou E_3 ?

[a] *Let E_1, E_2, and E_3 denote the samples that conform to a percentage of solids specification, a molecular weight specification, and a color specification, respectively. A total of 240 samples are classified by the E_1, E_2, and E_3 specifications, where yes indicates that the sample conforms.*

- *Are E_1, E_2, and E_3 mutually exclusive events ?*
- *Are $\overline{E_1}$, $\overline{E_2}$, and $\overline{E_3}$ mutually exclusive events ?*
- *What is $\mathbb{P}(\bar{E}_1$ or $\bar{E}_2$ or $\bar{E}_3)$?*
- *What is the probability that a sample conforms to all three specifications ?*
- *What is the probability tha a sampel conforms to E_1 or E_3 specifications ?*
- *What is the probability that a sample conforms to the E_1 or E_2 or E_3 specification?*

[a]Montgomery

Solution 1.2.3

- No, $\mathbb{P}(E_1 \cap E_2 \cap E_3) \neq 0$
- No, $\overline{E_1} \cap \overline{E_2} \neq \emptyset$
- $\mathbb{P}\big(\overline{E_1} \cup \overline{E_2} \cup \overline{E_3}\big) = \mathbb{P}\big(\overline{E_1}\big) + \mathbb{P}\big(\overline{E_2}\big) + \mathbb{P}\big(\overline{E_3}\big) - \mathbb{P}\big(\overline{E_1} \cap \overline{E_2}\big) - \mathbb{P}\big(\overline{E_1} \cap \overline{E_3}\big) - \mathbb{P}\big(\overline{E_2} \cap \overline{E_3}\big) + \mathbb{P}\big(\overline{E_1} \cap \overline{E_2} \cap \overline{E_3}\big) = 40/240$
- $\mathbb{P}(E_1 \cap E_2 \cap E_3) = 200/240$
- $\mathbb{P}(E_1 \cup E_3) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_3) = 234/240$
- $\mathbb{P}(E_1 \cup E_2 \cup E_3) = 1 - \mathbb{P}\big(\overline{E_1} \cap \overline{E_2} \cap \overline{E_3}\big) = 1 - 0 = 1$

1.2.5 Test positif

Positive test

Un test pour une maladie rare est supposé fiable dans 95% des cas: la probabilité que le test soit positif quand une personne est malade est égale à 0.95, et la probabilité qu'il soit négatif quand une personne n'est pas malade est aussi égale à 0.95. Cette fiabilité est notée p_f. La probabilité qu'une personne soit malade est égale à p_m= 0.001 (0.1% de la population).

a. Quelle est la probabilité qu'un test soit positif?

b. Une personne est testée positive. Quelle est la probabilité qu'elle soit atteinte de la maladie?

c. [Soit deux événements, A et B, et $\mathbb{P}(A) \neq 0, \mathbb{P}(B) \neq 0$. On dira que l'événement B suggère l'événement A si $\mathbb{P}(A|B) > \mathbb{P}(A)$.] Quelles sont les valeurs de fiabilité (prises entre 0 et 1) pour lesquelles le fait d'avoir un test positif suggère que la personne est malade? Interpréter le résultat.

Afin d'améliorer la fiabilité du dépistage, on répète le test dans le cas d'un résultat positif (évidemment, la fiabilité p_f de chaque test reste indépendante des résultats précédents).

d. Calculer la probabilité $P(P_n)$ que n tests successifs soient positifs.

e. Une personne est testée successivement n fois positive. Quelle est la probabilité qu'elle soit atteinte de la maladie?

f. On notera N la valeur minimale de n qui donne une réponse supérieure à 0.9 à la question précédente. Déterminer numériquement la valeur de N dans le cas: $p_f = 0.9$, $p_m = 0.1$.

g. Donner une explication de l'évolution de la probabilité de la question e. en fonction de n (en vous aidant d'un tableau ($1 \leq n \leq N$) de différentes probabilités que vous jugez utiles).

A test for a rare disease is assumed to be reliable in 95 % of the cases (i.e. the probability that the test is positive when a person is ill is equal to 0.95, and the probability that the test is negative when a person is not ill is also of 0.95). This reliability is denoted as p_f. The probability that a person is ill is p_m = 0.001 (0.1 % of the population).

a. What is the probability that a test is positive ?

b. A person is tested positive. What is the probability that this person is ill ?

c. [Let's have two events, A and B, with $\mathbb{P}(A) \neq 0$, $\mathbb{P}(B) \neq 0$. Event B suggests event A if $\mathbb{P}(A|B) > \mathbb{P}(A)$.] What are the reliability values (between 0 and 1) such that a positive test suggest the person is ill ? Interpret this result .

To improve the quality of the diagnostic, the test, if positive, is repeated (obviously, the reliability of each test p_f is independant of the previous test results).

d. Compute $\mathbb{P}(P_n)$, the probability that n successive tests are positive.

e. A person is tested n successive times positive. What is the probability that she is ill ?

f. Let N be the minimum value of n that gives an answer = 0.9 to the previous question. Détermine numerically the value of N for : $p_f = 0.9$, $p_m = 0.1$.

g. Explain the evultion of the probabilities as a function of n (with the help of a table ($1 \leq n \leq N$) of the different useful probabilities).

Solution 1.2.4

Let "P" denote positive and "I" denote ill, then $p_f = \mathbb{P}(P|I) = \mathbb{P}(\bar{P}|\bar{I}) = 0.95$. Furthermore, $p_m = \mathbb{P}(I) = 0.001$

a.
$$\begin{aligned}
\mathbb{P}(P) &= \mathbb{P}(P\cap I) + \mathbb{P}(P\cap \bar{I}) \\
&= \mathbb{P}(P|I)\,\mathbb{P}(I) + \mathbb{P}(P|\bar{I})\,\mathbb{P}(\bar{I}) \\
&= \mathbb{P}(P|I)\,\mathbb{P}(I) + (1-\mathbb{P}(\bar{P}|\bar{I}))\mathbb{P}(\bar{I}) \\
&= 0.95\times 0.001 + 0.05\times 0.999 = 0.0509
\end{aligned}$$

Note that $\mathbb{P}(P|\bar{I}) = 1-\mathbb{P}(\bar{P}|\bar{I})$ can be demonstrated as :

$$\mathbb{P}(P|\bar{I}) = \frac{\mathbb{P}(P\cap\bar{I})}{\mathbb{P}(\bar{I})} = \frac{\mathbb{P}(\bar{I})-\mathbb{P}(\bar{P}\cap\bar{I})}{\mathbb{P}(\bar{I})} = 1-\frac{\mathbb{P}(\bar{P}\cap\bar{I})}{\mathbb{P}(\bar{I})}$$

b. $\mathbb{P}(I|P) = \dfrac{\mathbb{P}(I\cap P)}{\mathbb{P}(P)} = \dfrac{\mathbb{P}(P|I)\,.\mathbb{P}(I)}{\mathbb{P}(P)} = \dfrac{0.95\times 0.001}{0.0509} = 0.0187$

c. The question is : find $\mathbb{P}(P|I)$ such that $\mathbb{P}(I|P) > \mathbb{P}(I)$. We will use $\mathbb{P}(I) = \mathbb{P}(I|P)\,\mathbb{P}(P) + \mathbb{P}(I|\bar{P})\,\mathbb{P}(\bar{P})$

$$\begin{aligned}
\mathbb{P}(I|P) &> \mathbb{P}(I) \\
&\Leftrightarrow \\
\tfrac{\mathbb{P}(P|I)\mathbb{P}(I)}{\mathbb{P}(P)} &> \mathbb{P}(I) \\
&\Leftrightarrow \\
\mathbb{P}(P|I) &> \mathbb{P}(P) \\
&\Leftrightarrow \\
\mathbb{P}(P|I) &> \mathbb{P}(P|I)\,\mathbb{P}(I) + (1-\mathbb{P}(\bar{P}|\bar{I}))(1-\mathbb{P}(I)) \\
&\Leftrightarrow \ (\text{with}\mathbb{P}(\bar{P}|\bar{I}) = \mathbb{P}(P|I) \\
\mathbb{P}(P|I) &> \mathbb{P}(P|I)\,(2.\mathbb{P}(I)-1) + 1 - \mathbb{P}(I) \\
&\Leftrightarrow \\
\mathbb{P}(P|I)\,.(2-2.\mathbb{P}(I)) &> 1-\mathbb{P}(I) \\
&\Leftrightarrow \\
\mathbb{P}(P|I) &> 0.5
\end{aligned}$$

Hence P suggests I if $\mathbb{P}(I|P) > 0.5$.

d. $\mathbb{P}(P_2) = \mathbb{P}($"first test Positive" $\cap$ "second test positive").
Independence of successive tests means $\mathbb{P}(\text{"}1^{st}\text{ test}|I\text{"} \cap \text{"}2^{nd}\text{ test}|I\text{"}) = \mathbb{P}(P|I)^2$.
Let's note the first test P^1 and the second test P^2.
then

$$\begin{aligned}
\mathbb{P}(P_2) &= \mathbb{P}(P^1\cap P^2) \\
&= \mathbb{P}(P_2|I)\,\mathbb{P}(I) + \mathbb{P}(P_2|\bar{I})\,\mathbb{P}(\bar{I}) \\
&= \mathbb{P}(P|I)^2\,\mathbb{P}(I) + \mathbb{P}(P|\bar{I})^2\,\mathbb{P}(\bar{I}) \\
&= \mathbb{P}(P|I)^2\,\mathbb{P}(I) + [1-\mathbb{P}(\bar{P}|\bar{I})]^2.[1-\mathbb{P}(I)] \\
&= 0.95^2\times 0.001 + 0.05^2\times 0.999 = 0.0034
\end{aligned}$$

and hence

$$\mathbb{P}(P_n) = 0.95^n\times 0.001 + 0.05^n\times 0.999$$

e.

$$\mathbb{P}(I|P_n) = \frac{\mathbb{P}(P_n|I)\,.\mathbb{P}(I)}{\mathbb{P}(P_n)} = \frac{\mathbb{P}(P|I)^n\,.\mathbb{P}(I)}{\mathbb{P}(P|I)^n\,.\mathbb{P}(I) + (1-\mathbb{P}(\bar{P}|\bar{I}))^n(1-\mathbb{P}(I))}$$

f. for $\mathbb{P}(I) = 0.01$ and $\mathbb{P}(P|I) = 0.9$

N	1	2	3	4	5
$\mathbb{P}(P_n\cap I) = \mathbb{P}(P|I)^n\,\mathbb{P}(I)$	0.009	0.0081	0.00729	0.006561	0.0059049
$\mathbb{P}(P_n\cap\bar{I})$	0.099	0.0099	0.00099	0.000099	0.0000099
$\mathbb{P}(P_n)$	0.108	0.018	0.00828	0.00666	0.0059148
$\mathbb{P}(I|P_n)$	0.0833333	0.45	0.8804348	0.9851351	0.9983262

Comment on the fact that $\mathbb{P}(P_n\cap\bar{I})$ is large when n is small, while with increasing n, the influence of "Ill people" increases on the probability that n

1.2.6 Indépendance conditionnelle : entre chien et loup

Conditional independance : between dog and wolf (means between dusk and dawn)

Soient A, B, C trois événements avec $\mathbb{P}(C) \neq 0$. On dit que A et B sont indépendants conditionnellement à C si A et B sont indépendant si C est réalisé :

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \,.\mathbb{P}(B|C)$$

Dans le Mercantour, alors que la nuit tombe, deux bergers aperçoivent un animal. Si l'animal est un loup, chacun des bergers, indépendemment, déclare que c'est un loup avec une probabilité de 0.7. Si l'animal est un chien, chacun des bergers, indépendemment, déclare que c'est un chien avec une probabilité de 0.8.

On suppose que la probabilité d'apercevoir un loup est de 0.6 et on note les événements A = “le premier berger déclare que c'est un loup” et B = “le deuxième berger déclare que c'est un loup”.

- Quelle est la probabilité que les deux bergers déclarent que c'est un loup ?
- Les événements A et B sont-ils indépendants ?

Let A, B, C be three events with $\mathbb{P}(C) \neq 0$. A and B are independent conditionnaly on C if A ant B are independent if C is true :

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \,.\mathbb{P}(B|C)$$

In the Mercantour mountain, when night falls, two sheperds see an animal. If the animal is a wolf, each sheperd, independently, says it is a wolf with probability 0.7. If the animal is a dog, each sheperd, independently, says it is a dog with probability 0.8. The animal is either a wolf or a dog.
We assume that the probability to see a wolf is 0.6 and we define the events A = “the first sheperd says the animal is a wolf” and B = “the second sheperd says the animal is a wolf”.

- *What is the probability that the two sheperds declare that the animal is a wolf.*
- *Are A and B independent events ?*

Solution 1.2.5

- $$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A \cap B \cap C) + \mathbb{P}\big(A \cap B \cap \bar{C}\big)\\ &= \mathbb{P}(A \cap B|C)\,\mathbb{P}(C) + \mathbb{P}\big(A \cap B|\bar{C}\big)\,\mathbb{P}\big(\bar{C}\big)\\ &= \mathbb{P}(A|C)\,\mathbb{P}(B|C)\,\mathbb{P}(C) + \mathbb{P}\big(A|\bar{C}\big)\,\mathbb{P}\big(B|\bar{C}\big)\,\mathbb{P}\big(\bar{C}\big)\\ & 0.7^2 \times 0.6 + 0.2^2 \times 0.4 = 0.34\end{aligned}$$

- $$\mathbb{P}(A \cap B) \stackrel{?}{=} \mathbb{P}(A)\,\mathbb{P}(B)$$
$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap C) + \mathbb{P}\big(A \cap \bar{C}\big)\\ &= \mathbb{P}(A|C)\,.\mathbb{P}(C) + \mathbb{P}\big(A|\bar{C}\big)\,\mathbb{P}\big(\bar{C}\big)\\ &= 0.7 \times 0.6 + 0.2 \times 0.4 = 0.5\end{aligned}$$

$\mathbb{P}(A)\,\mathbb{P}(B) = (0.5)^{=}0.25 \neq 0.34 \Rightarrow A$ and B are not strictly independant.

1.2.7 Pannes

Un dispositif se compose de 4 ensembles ; pendant le fonctionnement chaque ensemble tombe en panne avec une probabilité $p = 0.4$. Les ensembles subissent des défaillances indépendamment l'un de l'autre. Si plus de deux ensembles tombent en panne, le dispositif ne peut plus fonctionner ; si les défaillances affectent 1 ou 2 ensembles, il fonctionne, mais son efficacité baisse. Trouver la probabilité des événements :
$A = \{$aucun ensemble n'est tombé en panne$\}$;
$B = \{$le dispositif peut fonctionner$\}$;
$F = \{$ l'efficacité du dispositif en service est réduite$\}$.

Failures

A device is made up of 4 parts. Each parts fails with a probability $p = 0.4$, independently from the other parts. If (strictly) more than 2 parts fail, the device fails. If 1 or 2 parts fail, the device continues to work, but is not optimal. Find the probabilities of the following events :

- $A = \{$" *no part fails* "$\}$
- $B = \{$" *the device works* "$\}$
- $F = \{$" *the device works non optimally* "$\}$

Solution 1.2.6

For one part, let's denote $W = \{$"the part works$\}$", hence $\mathbb{P}(W) = 1 - p$ and $\mathbb{P}(\bar{W}) = p$. Then $\mathbb{P}(A) = \mathbb{P}(\text{"all parts work"}) = \mathbb{P}(W_1 \cap W_2 \cap W_3 \cap W_4) =$ due to independence $\mathbb{P}(W)^4 = (1-p)^4$
$\mathbb{P}(B) = \mathbb{P}(\text{"2, 3 or 4 parts work"}) = \mathbb{P}(\text{"2 parts work"}) + \mathbb{P}(\text{"3 parts work"}) + \mathbb{P}(\text{"4 parts work"}) = C_4^2 p^2 (1-p)^2 + C_4^1 p(1-p)^3 + (1-p)^4$
(here, possibility explain ("choose 2 among 4")).
$\mathbb{P}(F) = \mathbb{P}(\text{" 2 or 3 parts work"}) = C_4^2 p^2 (1-p)^2 + C_4^1 p(1-p)^3$

1.3 TD3 : probabilités conditionnelles

1.3.1 démonstration

Soient 3 événements quelconques A, B, C, de probabilité strictement positive, démontrez :

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B|A)\,\mathbb{P}(C|A \cap B)\,.$$

Prove that for any three events A, B, C, each having positive probability,

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B|A)\,\mathbb{P}(C|A \cap B)\,.$$

1.3.2 Poker

Au poker, John a une main très forte et parie 5 dollars. La probabilité que Marie ait une meilleure main est 0.04. Si Marie avait une meilleure main elle relancerait avec une probabilité de 0.9, mais avec une main plus pauvre elle relancerait seulement avec probabilité 0.1. Si Marie relance, quelle est la probabilité qu'elle ait un meilleure main que John?

> *In a poker hand, John has a very strong hand and bets 5 dollars. The probability that Mary has a better hand is .04. If Mary had a better hand she would raise with probability .9, but with a poorer hand she would only raise with probability .1. If Mary raises, what is the probability that she has a better hand than John does?*

Solution 1.3.1

- $$\mathbb{P}(M > J) = 0.04; \mathbb{P}(R|M > J) = \frac{\mathbb{P}(R \cup (M > J))}{\mathbb{P}(M > J)} = 0.9; \mathbb{P}(R|M <= J) = \frac{\mathbb{P}(R \cup (M > J))}{1 - \mathbb{P}(M > J)} = 0.1$$
- $\mathbb{P}(R \cup (M > J)) = 0.9 \times \mathbb{P}(M > J) = 0.036$
- $\mathbb{P}(R \cup (M <= J)) = 0.1 \times \mathbb{P}(M <= J) = 0.096$
- $\mathbb{P}(R) = \mathbb{P}(R \cup (M > J)) + \mathbb{P}(R \cup (M <= J)) = 0.0142$
- $\mathbb{P}((M > J)|R) = \dfrac{\mathbb{P}(R \cup (M > J))}{\mathbb{P}(R)} = 0.036/0.142 = 0.2536$

1.3.3 démonstration

Soient deux événements A et B tels que $\mathbb{P}(A|B) = \mathbb{P}(B|A), \mathbb{P}(A \cup B) = 1, \mathbb{P}(A \cap B)\, Q > 0$, démontrez que $\mathbb{P}(A) > 0.5$.

> *Let A et B be such that $\mathbb{P}(A|B) = \mathbb{P}(B|A)$, $\mathbb{P}(A \cup B) = 1, \mathbb{P}(A \cap B)\, Q > 0$, prove that $\mathbb{P}(A) > 0.5$.*

1.3.4 Phénotypes

Supposons que dans une population, on observe trois phénotypes différents notés A, B et C avec la répartition suivante parmi les mères de famille : 53% de phénotypes A, 23% de phénotypes B, 24% de phénotypes C. Le tableau suivant donne la proportion d'enfants de phénotype donné en fonction du phénotype de la mère.

	enfant A	enfant B	enfant C
mère A	35%	20%	45%
mère B	17%	46%	37%
mère C	18%	52%	30%

Par exemple, le tableau indique que 35% des enfants dont la mère est de phénotype A ont le phénotype A.

On choisit un enfant au hasard dans cette population.

- Quelle est la probabilité qu'il n'ait pas le phénotype A ?
- L'enfant choisi n'a pas le phénotype A. Quelle est la probabilité que sa mère ait le phénotype B ?

Let a population be such that 3 phenotypes, noted as A, B and C, are such that mothers are distributed according to : : 53% have phenotype A, 23% have phenotypes B, 24% have phénotype C. Le tableau suivant donne la proportion d'enfants de phénotype donné en fonction du phénotype de la mère.
The table hereabove gives the proportion of children having a given phenotype according to their mother's phenotype (here, the table indicates that 35 % of the children whose mother have a phenotype A also have a phenotype A).
A child is chosen at random in this population

- *What is the probability that he does not have phenotype A ?*
- *The child does not have phenotype A. What is the probability that his mother has phenotype B ?*

Solution 1.3.2

```
PMA=0.53
PMB=0.23
PMC=0.24
PEAifA=0.35
PEAifB=0.17
PEAifC=0.18
PEBifA=0.2
PEBifB=0.46
PEBifC=0.52
PECifA=0.45
PECifB=0.37
PECifC=0.30

PEA=PEAifA*PMA+PEAifB*PMB+PEAifC*PMC
PENA=1-PEA
print('PENA = ',PENA)
PENAifB=1-PEAifB
PMBifENA=PENAifB/PENA*PMB
print('PMBifENA =',PMBifENA)
----------------------------------------
PENA =  0.7322
PMBifENA = 0.26072111444960394
```

1.3.5 Phénotypes

Supposons que dans une population, on observe trois phénotypes différents notés A, B et C avec la répartition suivante parmi les mères de famille : 23% de phénotypes A, 23% de phénotypes B, 24% de phénotypes C et 30 % de phénotype D. Le tableau suivant donne la proportion d'enfants de phénotype donné en fonction du phénotype de la mère.

	enfant A	enfant B	enfant C	enfant D
mère A	35%	20%	25%	20 %
mère B	17%	46%	27%	10 %
mère C	18%	52%	25%	5 %
mère D	5%	48%	20%	27 %

Par exemple, le tableau indique que 35% des enfants dont la mère est de phénotype A ont le phénotype A.

On choisit un enfant au hasard dans cette population.

- Quelle est la probabilité qu'il ait le phénotype A si sa mère n'a pas le phénotype B?

- Quelle est la probabilité qu'il ait le phénotype A ?

- L'enfant choisi n'a pas le phénotype A. Quelle est la probabilité que sa mère ait le phénotype B ou le phénotype C?

- L'enfant choisi n'a pas le phénotype A. Quelle est la probabilité que sa mère n'ait pas le phénotype B ni le phénotype D?

Same problem with 23% de phénotypes A, 23% de phénotypes B, 24% de phénotypes C et 30 % de phénotype D.
and questions

- *Probability that the child has phenotype A if mother has phenotype B?*
- *Probability that the child has phenotype A ?*
- *Child does not have phenotype A, what is the probability that his mother has phenotype B or C ?*
- *Child does not have phenotype A, what is the probability that his mother has nor phenotype B nor D ?*

Solution 1.3.3

```
PMA=0.23 ;PMB=0.23 ;PMC=0.24 ;PMD=0.3
PEAifA=0.35;PEAifB=0.17;PEAifC=0.18;PEAifD=0.05;
PEBifA=0.2;PEBifB=0.46;PEBifC=0.52;PEBifD=0.48;
PECifA=0.25;PECifB=0.27;PECifC=0.25;PECifD=0.20;
PEDifA=0.2;PEDifB=0.1;PEDifC=0.05;PEDifD=0.27;

print('check mother probs', PMA+PMB+PMC+PMD)
print('check sum of conditional  probs',PEAifA+PEBifA+PECifA+PEDifA)
print('check sum of conditional  probs',PEAifB+PEBifB+PECifB+PEDifB)
print('check sum of conditional  probs',PEAifC+PEBifC+PECifC+PEDifC)
print('check sum of conditional  probs',PEAifD+PEBifD+PECifD+PEDifD)

PENAifA=1-PEAifA
PENAifB=1-PEAifB
PENAifC=1-PEAifC
PENAifD=1-PEAifD

print('PENAifB = ',PENAifB)
PEA=PEAifA*PMA+PEAifB*PMB+PEAifC*PMC+PEAifD*PMD
print('PEA = ',PEA)

PMAifENA=PENAifA/PENA*PMA
PMBifENA=PENAifB/PENA*PMB
PMCifENA=PENAifC/PENA*PMC
PMDifENA=PENAifD/PENA*PMD
PMBorCifENA=PMBifENA+PMCifENA
PMAorCifENA=PMAifENA+PMCifENA

print(PMBorCifENA)
PMBorDifENA=PMBifENA+PMDifENA

PMnoBnorDifENA=1-PMBorDifENA
print(PMnoBnorDifENA)
PMBnoBnorDifENA=PMAorCifENA
print(PMnoBnorDifENA)
-------------------------------------
check mother probs 1.0
check sum of conditional  probs 1.0
check sum of conditional  probs 1.0
check sum of conditional  probs 1.0
check sum of conditional  probs 1.0
PENAifB =  0.83
PEA =  0.1778
PMBorCifENA = 0.5295001365747064
PMnoBnorDifENA = 0.35004097241190935
PMnoBnorDifENA = 0.35004097241190935
```

1.3.6 Le jeu de Monty Hall

Le jeu oppose un présentateur à un candidat (le joueur). Ce joueur est placé devant trois portes fermées. Derrière l'une d'elles se trouve une voiture et derrière chacune des deux autres se trouve une chèvre. Il doit tout d'abord désigner une porte. Puis le présentateur doit ouvrir une porte qui n'est ni celle choisie par le candidat, ni celle cachant la voiture (le présentateur sait quelle est la bonne porte dès le début). Le candidat a alors le droit d'ouvrir la porte qu'il a choisie initialement, ou d'ouvrir la troisième porte.

Les questions qui se posent au candidat sont :

- Que doit-il faire ?
- Quelles sont ses chances de gagner la voiture en agissant au mieux ?

The game opposes a presenter to a candidate (the player). This player is placed in front of three closed doors. Behind one of them is a car and behind each of the other two is a goat. He must first designate a door. Then the presenter must open a door that is neither the one chosen by the candidate, nor that hiding the car (the presenter knows what is the right door from the beginning). The candidate then has the right to open the door he has chosen initially, or to open the third door.
The questions to the candidate are:

- *What should he do?*
- *What are her chances of winning the car by doing the best?*

1.3.7 Pluie

En Belgique, il pleut un jour sur deux. La météo prévoit correctement la pluie dans 70% des cas (la probabilité qu'il pleuve si la météo prévoit la pluie est de 70 %, et la probabilité qu'il ne pleuve pas si la météo prévoit qu'il ne pleut pas est de 70 %). Quand la météo prévoit de la pluie, Monsieur Bellemans prend sont parapluie. Si la météo ne prévoit pas de pluie, Monsieur Bellemans prend sont parapluie avec une probabilité de 1/3.

- Monsieur Bellemans n'a pas pris son parapluie, quelle est la probabilité qu'il pleuve ?
- Il pleut, quelle est la probabilité que Monsieur Bellemans n'ai pas pris son parapluie ?

In Belgium, it rains every other day. The rain is correctly forecast in 70% of the cases (the probability that it will rain if the weather forecasts rain is 70%, and the probability that it will not rain if the weather predicts that it will not rain is 70%). When the rain is forecasted , Mr. Bellemans takes his umbrella. If rain is not forecasted, Mr. Bellemans takes his umbrella with a probability of 1/3.

- *Mr. Bellemans did not take his umbrella, what is the probability that it rains?*
- *It's raining, what is the probability that Mr. Bellemans did not take his umbrella?*

1.4 TD4 : variable aléatoire, loi de probabilité

random variable, probability distribution

1.4.1 A votre santé !

Cheers

La meilleure bière du monde, et c'est un belge qui vous le dit, est la Westvleteren. Pendant une soirée, vous buvez 4 bières (vous êtes très sobre !), que vous prenez au hasard avec chaque fois une probabilité de 0.9 de choisir une Westvleteren, chaque "tirage" étant indépendant des autres. On appelle X le nofmbre de Westvleteren que vous avez bu sur la soirée. Donnez la loi de X (sa masse de probabilité et sa fonction de répartition).

The best beer in the world, and it's a belgian guy who tells you, is the Westvleteren. During a party, you drink 4 beers (you are sober !), that you take randomly. Each time, you have a probability of 0.9 to take a Westvleteren, independently from the other choices. Let X be the number of Westvleteren you have drunk during the party, give the distribution of X (it's cumulative density function and mass).

Solution 1.4.1

The domain of X is $R_X = \{0, 1, 2, 3, 4\}$, $p_X(x) = C_4^x p^x (1-p)^{4-x}$ and $F_X(x) = \sum_{i=0}^{x} C_4^i p(1-p)^{4-i}$. Note that $F_X(4) = 1$

1.4.2 Bonne voiture

Good car

Une voiture est formée de trois parties principales : l'habitacle, le moteur et l'électronique. Les fiabilités (probabilité de bon fonctionnement) de ces parties sont respectivement de 0.99, 0.98 et 0.94. Ces fiabilités sont supposées indépendantes. Déterminez la loi (masse de probabilité) que suit le nombre de parties qui fonctionnent correctement, ainsi que la fiabilité totale de la voiture.

Si l'amélioration des trois parties coute le même montant, quelle est la partie qu'il faut améliorer pour avoir le meilleur retour sur investissement (on considère par exemple que pour gagner un pour cent de fiabilité, on dépense 1000 Euros).

A car is made of three main parts : the interior, the motor and the computer (and associated electronics). The reliability (probability of working OK) of the different parts is respectively 0.99, 0.98 and 0.94. These reliabilities are considered independant. Determine the law of the number of parts that work correctly, as well as the global reliability of the car.
If improving the three parts has the same cost, which part should be improved to have the best return on investment ?

Solution 1.4.2

Let $p_1 = 0.99, p_2 = 0.98, p_3 = 0.94$, then

- $p_X(0) = \prod_{i=1}^{3}(1 - p_i)$
- $p_X(1) = \sum_{i=1}^{3} p_i \prod_{\substack{j\neq i \\ j=1}}^{3}(1 - p_j)$
- $p_X(2) = \sum_{i=1}^{3}(1 - p_i) \prod_{\substack{j\neq i \\ j=1}}^{3} p_j$
- $p_X(3) = \prod_{i=1}^{3} p_i$

The total reliability is $p_X(3) = 0.912$
Suppose we increment p_j to $p_j + \epsilon$, then, $p_X(3) = \prod_{\substack{j\neq i \\ j=1}}^{3} p_i(p_j + \epsilon) = \prod_{i=1}^{3} p_i p_j + \prod_{\substack{j\neq i \\ j=1}}^{3} p_i \epsilon$
Hence, the increase in reliability is $\prod_{\substack{j\neq i \\ j=1}}^{3} p_i \epsilon$
and we have to maximize $\prod_{\substack{j\neq i \\ j=1}}^{3} p_i$, and take j such that p_j is the smallest (here p_3)

1.4.3 Pièce de monnaie

Une pièce de monnaie non biaisée est jetée n fois. On note X la variable aléatoire qui est le nombre de fois que sort pile. Calculer la loi (masse de probabilité et fonction de répartition) de X.

Playing with coins

An unbiased coin is thrown n times. Let X be the random variable representing the number of times the outcome is heads, determine the probability mass of X.

Solution 1.4.3

$p_X(0) = p^n; p_X(1) = n \times p^n (p^n = p^{n-1}(1 - p)^1); ...; p_X(x) = C_n^x p^n$ (where $p = 0.5$, considering an unbiased coin).
Remark that $\sum_{i=0}^{n} C_n^i p^n = 1$ (binomial sum)

1.4.4 Normaliser et calculer

Une variable aléatoire X a comme fonction de probabilité:

$$p_X(x) = \begin{cases} x^2/a, & x \in \{-3, -2, -1, 0, 1, 2, 3\} \\ 0 & \text{ailleurs} \end{cases}.$$

Calculer:

1. la valeurs de a ,
2. la masse de probabilité $p_Z(z)$ de la v.a. $Z = X^2$,

Normalise and compute

A random variable X has the following density :

$$p_X(x) = \begin{cases} x^2/a, & x \in \{-3,-2,-1,0,1,2,3\} \\ 0 & \textit{otherwise} \end{cases}.$$

Compute :

1. *the value of a ,*
2. *the probability mass function $p_Z(z)$ of. $Z = X^2$,*

Solution 1.4.4

$$\sum_x p_X(x) = \frac{(-3)^2}{a} + \frac{(-2)^2}{a} + \frac{(-1)^2}{a} + \frac{(0)^2}{a} + \frac{(1)^2}{a} + \frac{(2)^2}{a} + \frac{(3)^2}{a} = \frac{28}{a} = 1 \Rightarrow a = 28$$

- $p_Z(0) = p_X(0) = 0$
- $p_Z(1) = p_X(-1) + p_X(1) = 2/a$
- $p_Z(z) = p_X(-z) + p_X(z) = \frac{2z}{a}$

1.4.5 Examen(s)

Un étudiant a le droit de se présenter jusqu'à m fois à un examen. La probabilité de réussir est chaque fois égale à p (indépendamment du nombre de fois qu'il s'est déjà présenté...). Calculer la masse de probabilité du nombre des essais (v.a. N), sachant que l'étudiant réussit à l'examen.

Exam(s)

A student has the right to present m times an exam. The probability to secceed is each time equal to p (and is independent from the other trials). What is the probability mass function of the number of trials (r.v; N) needed to succeed, knowing that the student eventually succeeds.

Solution 1.4.5

Here, we look for $p_N(n|n \leq m)$ where $n \leq m$ is the event :"the student eventually succeeds".

$$\begin{aligned} p_N(n|n \leq m) &= \frac{p_N(n \cap "n \leq m")}{\mathbb{P}(n \leq m)} \\ &= \frac{p_N(n)}{\mathbb{P}(n \leq m)} \qquad \text{for } 1 \leq n \leq m \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

1.4.6 Combien de questions ...

How many questions ...

Conseil : bien exprimer l'univers

Un prix est placé aléatoirement dans une boite parmi 10 (les boites sont numérotées de 1 à 10). On cherche à trouver le prix en posant des questions binaires (le prix est-il dans la boite numéro x ?). Donnez l'espérance mathématique du nombre de questions à poser sous l'hypothèse des deux stratégies suivantes :

1. La stratégie d'énumération : "*Le prix est-il dans la boite k ?*"

2. La stratégie dichotomique : vous éliminez à chaque fois la moitié des boites restantes avec une question du type : "*le prix est-il dans une boite de numéro inférieur ou égal à k ?*"

hint : express the universe correctly !
A price is place randomly in a box. There are 10 boxes numbered from 1 to 10. We try to find where the price is, by asking binary questions (is the price in box number x ?). Give the expected number of questions to ask for the two following strategies :

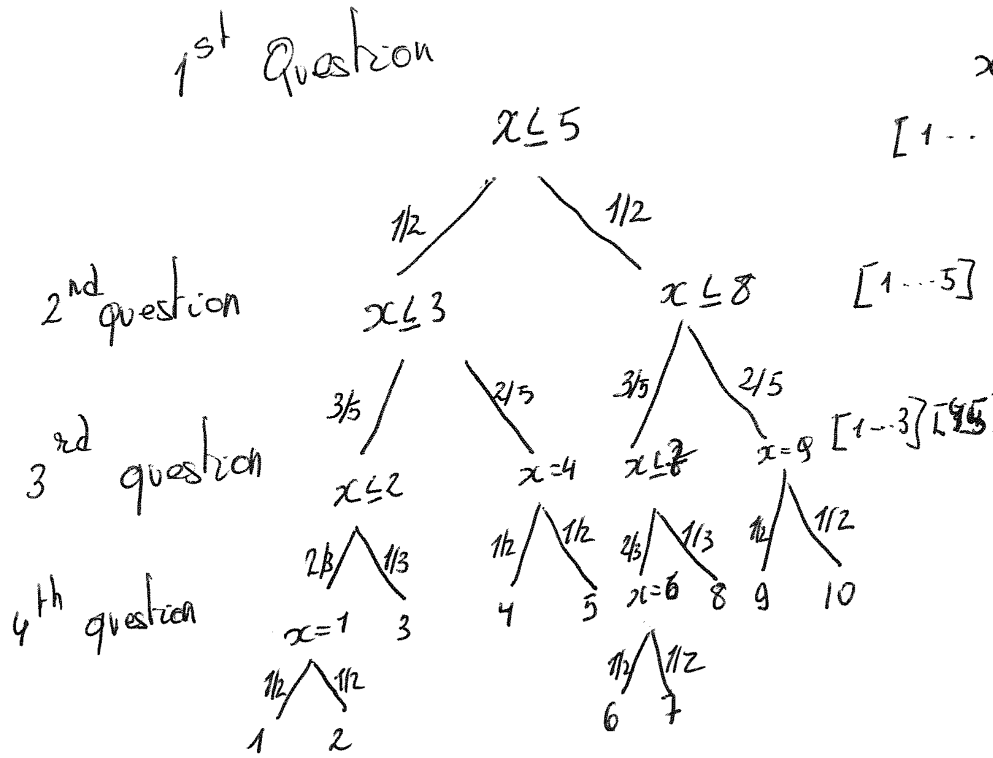
1. *Enumeration strategy :* "Is the price in box number k ? " *(for $k = 1...10$).*
2. *Dichotomic search strategy : you eliminate half of the possibilities at each question by asking* "Is the price in a box whose number is less or equal to k ?" *(where k is equal to 5 at the first question ...)*

Solution 1.4.6

The universe is $\Omega = \{1, 2, \cdots 10\}$, where "$x$" means "*The price is in box number* x.

1. The probability that $x = 1$ is 1/10, and it is the same for all values of x. Hence, $p_X(x) = \frac{1}{10}$ for $x = 1..8$ and $p_X(9) = 2/10 and EX = \frac{1}{10}\sum_{i=1}^{8} i + 9 * 2/10 = 5.4$

2. The tree with questions can be shown as follows :



Hence $\mathbb{P}(3 \text{ questions to find the price}) = \mathbb{P}(k \in [3, 4, 5, 8, 9, 10]) = 6/10$

and $\mathbb{P}(4 \text{ questions to find the price}) = \mathbb{P}(k \in [1, 2, 6, 7]) = 4/10$.

So $p_X(3) = \frac{6}{10}$ and $p_X(4) = \frac{4}{10}$ and $\mathrm{E}[X] = 3 \times \frac{6}{10} + 4 \times \frac{4}{10} = 3.4$, and the second strategy is better.

1.5 TD5 : variables aléatoires continues

1.5.1 Probabilités

Soit la v.a. X de fonction de répartition :

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.2x & 0 \le x \le 5 \\ 1 & x > 5 \end{cases}$$

Déterminez :

- $f_X(x)$

- $\mathbb{P}(X < 2.8)\,;\mathbb{P}(X > 1.5)\,;\mathbb{P}(X < -2)\,;\mathbb{P}(X > 6)\,;\mathbb{P}(2 < X < 3)$

let X be a r.v. with the cumulative distribution function given above, determine it's pdf and the probabilities here above

1.5.2 Ampoules ...

La durée de vie d'une ampoule électrique est donnée par la variable aléatoire X, exprimée en heures, de densité égale à:

$$f_X(x) = \begin{cases} a/x^3 & \text{si } 1500 \leq X \leq 2500 \\ 0 & \text{sinon} \end{cases}$$

Trouvez la constante a, la fonction de répartition de X, $\mathrm{E}[X]$ et $\mathrm{var}[X]$

Bulbs ...

The lifetime of an electric bulb is given by the random variable x, expressed in hours, whose pdf is :

$$f_X(x) = \begin{cases} a/x^3 & \textit{si } 1500 \leq X \leq 2500 \\ 0 & \textit{sinon} \end{cases}$$

Compute the constant a, the cumulative density function of X, $\mathrm{E}[X]$ and $\mathrm{var}[X]$

Solution 1.5.1

$$1 = \int f_X(x)\,dx = \int_{1500}^{2500} \frac{a}{x^3} dx = a\left[\frac{-1}{2x^2}\right]_{1500}^{2500} = 142.22210^{-9}$$

$a = 7,031,250$

$$\int_{1500}^{x} \frac{a}{y^3} dy = a\left[\frac{-1}{2x^2}\right]_{1500}^{x} = \frac{a}{2}\left[\frac{1}{1500^2} - \frac{1}{x^2}\right]$$

$$F_X\left(x\right) = \begin{cases} 0 & x \leq 1500 \\ a/2\left[\frac{1}{1500^2} - \frac{1}{x^2}\right] & 1500 \leq x \leq 2500 \\ 1 & x \geq 1500 \end{cases}$$

$$EX = \int_{1500}^{2500} \frac{a}{x^2} dx = 1875$$

$$EX^2 = \int_{1500}^{2500} \frac{a}{x} dx = a\left[\log(x)\right]_{1500}^{2500} = 3,591,742.7$$

$$\mathrm{var}[X] = \mathrm{E}\big[X^2\big] - EX^2 = 76,117.66$$

$$\sigma = 275.9$$

1.5.3 Fonction de répartition

Trouver la densité de probabilité de la v.a. **X** dont la fonction de répartition est :

$$F_X\left(x\right) = \begin{cases} 1 - \frac{a^3}{x^3} & \text{si } x \geq a, \\ 0 & \text{si } x < a \end{cases},$$

où a est une constante positive.

Cumulative density function

Find the density of the r.v. X whose cdf is defined by

$$F_X\left(x\right) = \begin{cases} 1 - \frac{a^3}{x^3} & si\ x \geq a, \\ 0 & si\ x < a \end{cases},$$

where a is a positive function.

Solution 1.5.2

$$P_X(x) = \frac{dF_X}{dx}(x) = \begin{cases} 3\frac{a^3}{x^4} & \text{si } x \geq a, \\ 0 & \text{si } x < a \end{cases},$$

1.5.4 De la normale vers l'uniforme

Soit $X \sim \mathcal{N}\mu, \sigma^2$. On veut approximer la loi normale par une loi uniforme sur $[a, b]$, tel que les caractéristiques de la loi normale (moynne et variance) sont conservées. Quelles valeur de a et b faut-il choisir.

From normal to uniform

Let $X \sim \mathcal{N}\mu, \sigma^2$. We want to approximate the Gaussian by a uniform on $[a, b]$, such that both have the same mean and variances. What values of a and b do we have to choose ?

Solution 1.5.3

$$\mu_{unif} = [a+b]/2, \sigma^2_{unif} = (b-a)/12 \Rightarrow a = \mu - \sigma\sqrt{3}, b = \mu + \sigma\sqrt{3}$$

1.5.5 Loi exponentielle

Soit X, une v.a. exponentielle de moyenne égale à 10, déterminez $\mathbb{P}(X > 10)$; $\mathbb{P}(X > 20)$; $\mathbb{P}(X > 30)$, trouvez x tel que $\mathbb{P}(X < x) = 0.95$

Let X, be an exponential r.v. of mean 10, determine $\mathbb{P}(X > 10)$; $\mathbb{P}(X > 20)$; $\mathbb{P}(X > 30)$, find x such that $\mathbb{P}(X < x) = 0.95$

1.5.6 Durée de vie

La durée de vie (en heures) des ventilateurs d'un PC est modélisée par une distribution exponentielle ($\lambda = 0.0003$).

- Quelle est la proportion de ventilateurs qui fonctionneront au moins 10.000 heures ?
- Quelle est la proportion de ventilateurs qui ne fonctionneront pas plus de 7.000 heures ?

The time to failure (in hours) of fans in a personal computer can be modeled by an exponential distribution with $\lambda = 0.0003$*. (a) What proportion of the fans will last at least 10,000 hours? (b) What proportion of the fans will last at most 7000 hours?*

1.6 TD6 : Espérance mathématique

Expected value

1.6.1 Variance et écart-type

Variance and standard deviation

Evaluer la variance et l'écart-type de la variable aléatoire dont la densité de probabilité est donnée par :

$$f_X(x) = \begin{cases} 4x(9-x^2)/81 & 0 \le x \le 3 \\ 0 & \text{ailleurs} \end{cases}$$

Evaluate the variance and standard deviation whose probability density function is given by :

$$f_X(x) = \begin{cases} 4x(9-x^2)/81 & 0 \le x \le 3 \\ 0 & \textit{elsewhere} \end{cases}$$

Solution 1.6.1

First check if $\int_x f_X(x)\,dx = 1$:

$$\int_0^3 4x(9-x^2)/81dx = \frac{4}{9}\frac{x^2}{2} - \frac{4}{4\times 81}x^4\Big]_0^3 = 1$$

$$\begin{aligned} \mathrm{E}[X] = & \int_0^3 x(4x(9-x^2))/81dx \\ = & \tfrac{4}{9}\tfrac{x^3}{3} - \tfrac{4}{81}\tfrac{x^5}{5}\Big]_0^3 \\ = & 5 - \tfrac{12}{5} = 1.6 \end{aligned}$$

$$\mathrm{var}[X] = \mathrm{E}\left[X^2\right] - E^2 X$$

$$\begin{aligned} EX^2 = & \int_0^3 x^2(4x(9-x^2))/81dx \\ = & \tfrac{4}{9}\tfrac{x^4}{4} - \tfrac{4}{81}\tfrac{x^6}{6}\Big]_0^3 = 9 - 6 = 3 \end{aligned}$$

$\mathrm{var}[X] = 3 - (1.6)^2 = 0.44; \sigma_X = \sqrt{44} = 0.663$

1.6.2 Espérance mathématique

Expected value

Soit X une variable aléatoire de densité égale à :

$$f_X(x) = \begin{cases} x/4 & \text{si } 1 < x \le 3 \\ 0 & \text{sinon} \end{cases}$$

soit également l'événement $A \stackrel{\triangle}{=} \{X \ge 2\}$.

1. Calculez $\mathrm{E}[X]$, $\mathbb{P}(A)$, $f_{X|A}(x)$ et $\mathrm{E}[X|A]$.
2. Soit $Y = X^2$, calculez $\mathrm{E}[Y]$ et $\mathrm{var}\{Y\}$.

Let X be a random variable whose probability density function is given by :

$$f_X(x) = \begin{cases} x/4 & \textit{si } 1 < x \leq 3 \\ 0 & \textit{sinon} \end{cases}$$

We define the following event : $A \triangleq \{X \geq 2\}$.

1. *Compute* $\mathrm{E}[X]$, $\mathbb{P}(A)$, $f_{X|A}(x|a)$ *and* $\mathrm{E}[X|A]$.
2. *Let* $Y = X^2$, *compute* $\mathrm{E}[Y]$ *and* $\mathit{var}\{Y\}$.

Solution 1.6.2

First check that $\int_x f_X(x)\,dx = 1$: $\int_1^3 \frac{x}{4}dx = \left.\frac{x^2}{8}\right]_1^3 = \frac{9-1}{8} = 1$

$$\mathrm{E}[X] = \int_1^3 \frac{x \times x}{4}dx = \left.\frac{x^3}{12}\right]_1^3 = \frac{27-1}{12} = \frac{13}{6}$$

$$\mathbb{P}(A) = \int_2^3 \frac{x}{4}dx = \left.\frac{x^2}{8}\right]_2^3 = \frac{9-4}{8} = \frac{5}{8}$$

$$\begin{aligned} f_{X|A}(x|A) = & \frac{f_X(x \cap A)}{\mathbb{P}(A)} = \frac{p_X(x)}{\mathbb{P}(A)} \qquad \text{for } 2 \leq x \leq 3 \\ = & \frac{8}{5}\frac{x}{4} = \frac{2x}{5} \end{aligned}$$

$$\mathrm{E}[X|A] = \int_2^3 x\frac{2x}{5}dx = \left.\frac{2}{5}\frac{x^3}{3}\right]_2^3 = \frac{2(27-8)}{15} = 2 + 8/15$$

$$Y = X \Rightarrow EY = EX^2 = \int_1^3 \frac{x^2 \times x}{4}dx = \left.\frac{x^4}{16}\right]_1^3 = \frac{81-1}{16} = 5.$$

$$\mathrm{E}\left[Y^2\right] = \mathrm{E}\left[X^4\right] = \int_1^3 \frac{x^4 \times x}{4}dx = \left.\frac{x^6}{24}\right]_1^3 = \frac{728}{24} = 30.33$$

$$\mathrm{var}[Y] = \mathrm{E}\left[Y^2\right] - \mathrm{E}[Y]^2 = 30.333 - 25 = 5.3333$$

1.6.3 Normaliser et calculer

Une variable aléatoire X a comme fonction de probabilité:

$$p_X(x) = \begin{cases} x^2/a, & x \in \{-3,-2,-1,0,1,2,3\} \\ 0 & \text{ailleurs} \end{cases}.$$

Calculer:

1. les valeurs de a et de $E[X]$,
2. la fonction de probabilité $p_Z(z)$ de la v.a. $Z = (X - E[X])^2$,

3. la variance σ_X^2 à partir de $p_Z(z)$,
4. la variance σ_X^2 à partir de $p_X(x)$.

Normalise and compute

A r.v. X has the follwing pdf :

$$p_X(x) = \begin{cases} x^2/a, & x \in \{-3,-2,-1,0,1,2,3\} \\ 0 & \textit{otherwise} \end{cases}.$$

compute :

1. *a and $E[X]$,*
2. *$p_Z(z)$ of. $Z = (X - E[X])^2$,*
3. *σ_X^2 from $p_Z(z)$,*
4. *σ_X^2 from $p_X(x)$.*

Solution 1.6.3

- $$\sum p_X(x) = \sum_{x=-3}^{x=3} \frac{x^2}{a} = 28/a \Rightarrow a = 28$$
  $$\mathrm{E}[X] = \sum_{x=-3}^{x=3} \frac{x^3}{a} = 0$$
- $Z = X^2 : Z \in \{0,1,4,9\}$ and $p_Z(z) = \frac{2z}{a}$
- $$\mathrm{E}[Z] = \sum_z \frac{2z^2}{a} = 7$$
  $$\mathrm{E}\left[Z^2\right] = \sum_z \frac{2z^3}{a} = 56.71$$
  $$\sigma_Z^2 = 56.71 - 7^2 = 7.71$$
- $$\mathrm{E}[Z] = \mathrm{E}\left[X^2\right] = \sum_{x=-3}^{x=3} \frac{2x^4}{a} = 7$$
  $$\mathrm{E}\left[Z^2\right] = \mathrm{E}\left[X^4\right] = \sum_{x=-3}^{x=3} \frac{2x^6}{a} = 56.71$$

1.6.4 Variance et écart-type

Soit une variable aléatoire X de densité de probabilité triangulaire, représentée dans la figure ci-dessous :

Cette densité peut être définie sur $x \in [-a, a]$ $a = 3$ dans l'exemple donné.

Donner la densité, la moyenne, la variance et l'écart-type de X. Représentez la fonction de répartition. Quelle est la probabilité que $X \in [-a/2, a]$.

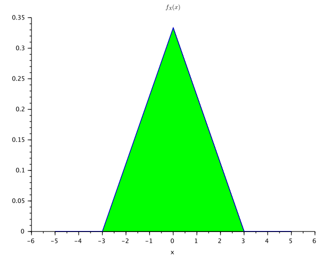


Figure 1.1: Densité de probabilité triangulaire

Variance and standard deviation

X is a r.v. of triangular pdf represented hereabove :
This density can be defined on r $x \in [-a, a]$ $a = 3$ as the example on the figure.
Determine the pdf, mean, variance and standard deviation of X, represent the cdf.
What is the probability of $X \in [-a/2, a]$.

Solution 1.6.4

1. $$f_X(x) = \frac{1}{a}\left(1 - \frac{|x|}{a}\right) \qquad \forall x \in [-a, a]$$
2. The cdf is made of two parabolic parts (ask them to draw it correctly).
3. $\mathrm{E}[X] = 0, \mathrm{var}[X] = a^2/6$
4. P=7/8

1.6.5 Espérance mathématique conditionnelle

Soit la densité exponentielle bilatérale suivante :

$$f_X(x) = \begin{cases} p\lambda e^{-\lambda x} & \text{si } x \geq 0 \\ (1-p)\lambda e^{\lambda x} & \text{si} x < 0 \end{cases},$$

où $\lambda > 0$ et $p \in [0, 1]$. Trouvez la moyenne et la variance de X de deux manières différentes :

- par calcul direct
- en utilisant une espérance conditionnelle et en vous basant sur la moyenne et la variance de la v.a. exponentielle unilatérale connue (la moyenne vaut $1/\lambda$ et la variance vaut $1/\lambda^2$).

Expected value

The pdf of a bilateral exponential r.v. is given by :

$$f_X(x) = \begin{cases} p\lambda e^{-\lambda x} & si\ x \geq 0 \\ (1-p)\lambda e^{\lambda x} & six < 0 \end{cases},$$

where $\lambda > 0$ *and* $p \in [0, 1]$. *Find the mean and variance of* X *by two different ways :*

- *by direct computation*
- *by using the concept of conditional expectation. Base yourself on the mean and variance of a unilateral exponential r.v.* Y, *which are known to be :* $\mathrm{E}[Y] = 1/\lambda$ *and* $\mathrm{var}[Y] = 1/\lambda^2$.

Solution 1.6.5

By direct computation

$$\mathrm{E}[X] = \int_{-\infty}^{0} (1-p)\lambda x e^{\lambda x} dx + \int_{0}^{\infty} p\lambda x e^{-\lambda x} dx$$

$\int_0^\infty \lambda x e^{-\lambda x} dx$ can be performed by parts with

$$\int u(x)v'(x)dx = u(x)v(x) - \int u'(x)v(x)dx$$

where

$$\begin{array}{ll} u(x) = \lambda x & v'(x) = e^{-\lambda x} \\ u'(x) = \lambda & v(x) = -\frac{1}{\lambda}e^{-\lambda x} \end{array}$$

Hence

$$\begin{aligned} \int \lambda x e^{-\lambda x} dx = & \quad -xe^{-\lambda x} - \int \lambda \left(-\frac{1}{\lambda}\right) e^{-\lambda x} dx = -xe^{-\lambda x} + \int -xe^{-\lambda x} dx \\ = & \quad -xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x} = e^{-\lambda x}\left(-x - \frac{1}{\lambda}\right) \end{aligned}$$

and

$$\int_0^\infty \lambda x e^{-\lambda x} dx = -xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}\Big]_0^\infty = 0 - \left(-\frac{1}{\lambda}\right) = \frac{1}{\lambda}$$

$$\Rightarrow \mathrm{E}[X] = (p-1)\frac{1}{\lambda} + p\frac{1}{\lambda} = \frac{2p-1}{\lambda}$$

$\int_0^\infty \lambda x^2 e^{-\lambda x} dx$ can be performed by parts with

$$\int u(x)v'(x)dx = u(x)v(x) - \int u'(x)v(x)dx$$

where

$$\begin{array}{ll} u(x) = \lambda x^2 & v'(x) = e^{-\lambda x} \\ u'(x) = 2\lambda x & v(x) = -\frac{1}{\lambda}e^{-\lambda x} \end{array}$$

Solution 1.6.6

Hence

$$\int \lambda x^2 e^{-\lambda x} dx = -x^2 e^{-\lambda x} - \int 2x\lambda \left(-\frac{1}{\lambda}\right) e^{-\lambda x} dx = -x^2 e^{-\lambda x} + \frac{2}{\lambda} \int \lambda x e^{-\lambda x} dx$$

and

$$\int_0^\infty \lambda x^2 e^{-\lambda x} dx = \left. -x^2 e^{-\lambda x} \right]_0^\infty - \frac{2}{\lambda} \int_0^\infty \lambda x e^{-\lambda x} dx = 0 + \frac{2}{\lambda}\left(\frac{1}{\lambda}\right) = \frac{2}{\lambda^2}$$

$$\Rightarrow \mathrm{E}\big[X^2\big] = (1-p)\frac{2}{\lambda^2} + p\frac{2}{\lambda^2} = \frac{1}{\lambda^2}$$

$$\mathrm{var}[X] = EX^2 - EX^2 = \frac{2}{\lambda^2} - \frac{(2p-1)^2}{\lambda^2} = \frac{3+4p-4p^2}{\lambda^2}$$

Using conditionals

$$\mathrm{E}[X|X \geq 0] = \frac{\mathrm{E}[X \cap X \geq 0]}{\mathbb{P}(X \geq 0)} = \frac{p\frac{1}{\lambda}}{p} = \frac{1}{\lambda}$$

likewise for $\mathrm{E}[X|X < 0]$.
Hence $\mathrm{E}[X] = \mathrm{E}[X \cap X < 0] + \mathrm{E}[X \cap X \geq 0]$

1.6.6 logs

On suppose que les ouvertures de session sur un réseau informatique suivent une loi de Poisson avec une moyenne de 3 ourvertures par minute.

- Quel est le temps moyen entre deux ouvertures?
- Quel est l'écart type du temps entre deux ouvertures ?
- Déterminez x de telle sorte que la probabilité qu'au moins une ouverture se produise avant que le temps x minutes ne soit égal à 0.95.

Suppose that the log-ons to a computer network follow a Poisson process with an average of 3 counts per minute. (a) What is the mean time between counts? (b) What is the standard deviation of the time between counts? (c) Determine x such that the probability that at least one count occurs before time x minutes is 0.95.

1.6.7 Visites sur site web

Les visites à un site Web à volume élevé sont supposées suivre une distribution de Poisson avec une moyenne de 10.000 visites par jour. Faites une approximation de chacun des éléments suivants:

- La probabilité de plus de 20 000 hits dans une journée
- La probabilité de moins de 9900 hits dans une journée
- La valeur telle que la probabilité que le nombre de hits dans une journée dépasser la valeur est 0.01

- Approximer le nombre de jours prévu dans une année (365 jours) qui dépassent 10 200 hits.
- Approximer la probabilité que sur un an (365 jours) plus de 15 jours chacun ont plus de 10 200 hits.

Hits to a high-volume Web site are assumed to follow a Poisson distribution with a mean of 10,000 per day. Approximate each of the following: (a) The probability of more than 20,000 hits in a day (b) The probability of less than 9900 hits in a day (c) The value such that the probability that the number of hits in a day exceed the value is 0.01
(d) Approximate the expected number of days in a year (365 days) that exceed 10,200 hits. (e) Approximate the probability that over a year (365 days) more than 15 days each have more than 10,200 hits.

Solution 1.6.7

> Let X denote the number of hits to a web site. Then, X is a Poisson random variable with a of mean 10,000 per day.$E(X) = \lambda = 10,000$ and V(X) = 10,000
> (d) P(X > 10200) = P(Z>2)=0.0228
> E(X)=365*0.0228 = 8.32 days per year
> (e) Y is a binomial with n=365 and p = 0.0228 E(Y) = 8.32 Var(Y) = 365*0.0228*0.9772 = 8.13
> P(Y>15) = P(Z > 2.34) = 0.0096

1.7 TD7 : Joint random variables

1.7.1 v.a. jointe simple (ref LLN - MAT1271)

La fonction $7 - x - y$ est positive sur le rectangle défini par $0 < x < 2, 2 < y < 5$.

1. Calculer k tel que $k(7 - x - y)$ soit une densité de probabilité pour (x, y)
2. Calculer $\mathbb{P}(X < 1, Y < 3)\,, \mathbb{P}(X + Y < 3)$ et $\mathbb{P}(X < 1|Y < 3)$.
3. Trouvez la densité de probabilité marginale de X.
4. Trouvez la densité conditionnelle $f_{X|Y}(x|y)$.

v.a. jointe simple (ref LLN - MAT1271)

The function $7-x-y$ is positive on the rectangle defined by $0 < x < 2, 2 < y < 5$.

1. *Compute k such that $k(7 - x - y)$ is a pdf for (x, y)*
2. *Compute $\mathbb{P}(X < 1, Y < 3)\,, \mathbb{P}(X + Y < 3)$ and $\mathbb{P}(X < 1|Y < 3)$.*
3. *Find the marginal pdf of X.*
4. *Find the conditional probability $f_{X|Y}(x|y)$.*

Solution 1.7.1

1.

$$\begin{aligned}
\int\int f_{X,Y}(()\,x,y)dxdy &= k\int_{x=0}^{2}\int_{y=2}^{5}(7-x-y)dydx \\
&= k\int_{x=0}^{2}\left[7y-xy-y^2/2\right]_{y=2}^{5}dx \\
&= k\int_{x=0}^{2}10.5-3xdx \\
&= k\left[10.5x-3x^2/2\right]_0^2 = k.15 \Rightarrow k=\frac{1}{15}
\end{aligned}$$

2.

$$\begin{aligned}
\mathbb{P}(X<1,Y<3) &= k\int_{x=0}^{1}\int_{y=2}^{3}(7-x-y)dydx \\
&= k\int_{x=0}^{1}\left[7y-xy-y^2/2\right]_{y=2}^{3}dx \\
&= k\int_{x=0}^{1}4.5-xdx \\
&= k\left[4.5x-x^2/2\right]_0^1 = 4k = \frac{4}{15}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(X+Y<3) &= k\int_{x=0}^{1}\int_{y=2}^{3-x}(7-x-y)dydx \\
&= k\int_{x=0}^{1}\left[7y-xy-y^2/2\right]_{y=2}^{3-x}dx \\
&= k\int_{x=0}^{1}\left\{7(3-x)-x(3-x)-(3-x)^2/2-(2\times 7-2x-2^2/2)\right\}dx \\
&= k\int_{x=0}^{1}(x^2/2-5x+9/2)dx = k.\frac{13}{6}
\end{aligned}$$

$$\mathbb{P}(Y<3) = 7k$$

$$\mathbb{P}(X<1|Y<3) = \frac{\mathbb{P}(X<1,Y<3)}{\mathbb{P}(Y<3)} = \frac{4}{7}$$

3. $f_X(x) = \displaystyle\int_y f_{X,Y}(x,y)\,dy = k(10.5-3x)$

$f_Y(y) = \displaystyle\int_x f_{X,Y}(x,y)\,dy = k(12-2y)$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{7-x-y}{12-2y}$$

1.7.2 Un point sur un demi-disque

Un point est choisi sur un demi-disque de rayon R. Le demi-disque est centré à l'origine et est situé dans le demi-plan supérieur. On demande :

1. La loi de probabilité conjointe de ses coordonnées **X** et **Y**
2. la loi de probabilité marginale **Y** et sa moyenne.
3. Vérifier (2) en calculant $\mathrm{E}(\mathbf{Y})$ sans utiliser la loi marginale de **Y**.

A point on a half disk

A point is chosen randomly (uniform distribution) on a half-disk of radius R. The half disk is centered at the origin and situated in the upper half plane. Determine :

1. *The pdf of it's coordinates* **X** *and* **Y**
2. *The law of the marginal* **Y** *and it's mean.*
3. *Verify (2) by computing* $\mathrm{E}[\mathbf{Y}]$ *without using the marginal o* **Y**.

Solution 1.7.2

1. Here, we assume a uniform distribution on the half disk, hence $f_{X,Y}(x,y) = \frac{2}{\pi R^2}$. **Note**

$$(\mathbf{X},\mathbf{Y}) = (\mathbf{R}\cos(\Theta), \mathbf{R}\sin(\Theta)),$$

and

$$(\Theta, \mathbf{R}) = (\arctan(\mathbf{Y}/\mathbf{X}), \sqrt{\mathbf{X}^2+\mathbf{Y}^2})$$

(with $\sqrt{\mathbf{X}^2+\mathbf{Y}^2} = R$)

2. To find the law of the marginal **Y** :

$$f_Y(y) = \int_{-A}^{A} \frac{2}{\pi R^2} dx = \begin{cases} \frac{4A}{\pi R^2} & \text{if } 0 \le y \le R, \\ 0 & \text{else.} \end{cases}$$

where $A = \sqrt{R^2-y^2}$. Then

$$\mathrm{E}[\mathbf{Y}] = \frac{4}{\pi R^2}\int_0^R y\sqrt{R^2-y^2}dy = \frac{4R}{3\pi}$$

by using a change of variable $z = R^2 - y^2$.

3. Noting D the half-disk :

$$\mathrm{E}[\mathbf{Y}] = \int_D y f_{X,Y}(x,y)\,dxdy = \int_0^\pi \int_0^R \frac{2}{\pi R^2} r\sin\theta r dr d\theta = \frac{4R}{3\pi}$$

1.7.3

Soient les v.a. X, Y, and Z dont la densité jointe est donnée par $f_{X,Y,Z}(x,y,z)$=8 x y z , for0 < x < 1, 0 < y <1, 0 < z < 1. Determinez :

- $\mathbb{P}(X < 0.5)$
- $\mathbb{P}(Z < 2)$
- $\mathbb{P}(X < 0.5, Y < 0.5)$

- $\mathbb{P}(X < 0.5, Z < 2)$
- $\mathrm{E}[X]$
- $\mathbb{P}(X < 0.5 | Y = 0.5)$
- $\mathbb{P}(X < 0.5, Y < 0.5 | Z = 0.8)$

Suppose the random variables X, Y, and Z have the joint probability density function $f_{X,Y,Z}(x, y, z)$=8 x y z , for0 < x < 1, 0 < y <1, 0 < z < 1. Determine the following (see above)

Solution 1.7.3

(1) 0.25 (2) 1 (3) 0.0625 (independence ...) (4) 0.24 (5) 2/3

1.7.4 v.a. jointes

Soient deux v.a. X et Y de densité de probabilité jointe :

$$f_{XY}(x, y) = \frac{a}{1 + x^2 + x^2.y^2 + y^2}$$

1. Calculez a
2. Calculez $f_X(x)$ et $f_Y(y)$. X et Y sont-ils indépendants ?
3. Calculer la probabilité que le point tombe dans un carré centré en $0, 0$ et de côté de longueur 2.

joint r.v.

Let X and Y be 2 r.v. such that :

$$f_{XY}(x, y) = \frac{a}{1 + x^2 + x^2.y^2 + y^2}$$

1. *Compute a*
2. *Compute $f_X(x)$ and $f_Y(y)$. Are X and Y indépendent ?*
3. *Comptue the probability that a point falls in a square centered in $0, 0$ and whose side is of length 2.*

Solution 1.7.4

1. $a = 1/\pi^2$
2. $f_X(x) = \frac{1}{\pi(1+x^2)}$,$f_Y(y) = \frac{1}{\pi(1+y^2)}$, hence, the two r.v. are independent.
3. $\mathbb{P}((X, Y) \in R) = 1/4$

1.7.5 Détection de signal

Un message binaire est transmis par les valeurs -1 ou +1. Le canal de communication corrompt le signal en ajoutant un bruit gaussien de moyenne μ et de variance σ^2. Le récepteur décide que le signal envoyé était -1 si le signal reçu est négatif, et +1 si le signal reçu est positif. Donnez la probabilité d'erreur (sous la forme d'une intégrale).

Signal detection

A binary message is transmitted with values -1 and +1. The communication channel corrupts the signal by adding a normally distributed noise of mean μ and variance σ^2. The receiver decides that the signal sent was -1 if the received signal is negative and +1 if the received signal is positive. Give the error probability in the form of an integral. forme d'une intégrale).

Solution 1.7.5

There is an error if

- the noise is larger than 1 if the transmitted signal is -1
- the noise is smaller than -1 if the transmitted signal is +1.

Let N be the Gaussian r.v. representing the noise. In the first case, the error probability is given by :

$$\mathbb{P}(N \geq 1) = 1 - \mathbb{P}(N < 1)$$

where $\mathbb{P}(N < 1)$ is the cdf of the normal r.v. Then,

$$\mathbb{P}(N < 1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{1} e^{-(v-\mu)^2/2\sigma^2} dv,$$

In the second case, by symetry, we have the same result.
For $\sigma = 1$, o$\mathbb{P}(N < 1)$ = 0.8413 and the error probability is 0.3174

1.8 TD 8 : Lois principales - principlement la normale !

TD 8 : Main laws ... mainly the normal law !

1.8.1 Intelligence

Intelligence

Les tests d'intelligence (QI) sont conçus de façon à ce que les résultats suivent une distribution Gaussienne $\mathcal{N}(100, 10^2)$.

Calculer en utilisant la table de la Fonction de répartition de la loi normale centrée réduite

1. Quelle proportion de la population obtient un résultat inférieur à 105.
2. Quelle proportion de la population obtient un résultat entre 95 et 105.

3. Quelles valeurs de QI, symétriques autour de la moyenne, couvrent 90% de la population.
4. Que représente un résultat supérieur à 120?

IQ tests are designed such that the results follow a Gaussian distribution $\mathcal{N}(100, 10^2)$.
Compute, by using the normalised cdf function :

1. *The proportion of the population that has an IQ less then 105.*
2. *The proportion of the population that has an IQ between 95 and 105.*
3. *Which values of IQ, symetric around the mean, cover* 90% *of the population.*
4. *What does it mean to have an IQ above 120 ?*

Solution 1.8.1

- $\mu = 100, \sigma = 10, \mathrm{IQ} < 105 \Leftrightarrow Z \leq \frac{105-\mu}{10} = 0.5$. $F_Z(0.5) = 0.6915 \Rightarrow$ 69 % of people have an IQ below 105 (note ... not really exact !).
- $95 \leq X \leq 105 \Leftrightarrow -0.5 \leq Z \leq 0.5$
  $\mathbb{P}(95 \leq X \leq 105) = \mathbb{P}(-0.5 \leq Z \leq 0.5) = F_Z(0.5) - F_Z(-0.5) = 0.38$ (draw the figures at the bboard).
- We look for $x_{0.95}$ and $x_{0.05}$. We search for $z_{0.95}$ in the table (such that $F_Z(z_{0.95}) = 0.95$). In the table, we find $z_{0.95} = 1.64$, which corresponds to $x_{0.95} = 100 + 10 * 1.64 = 116.4 \Rightarrow 83.6 \leq IQ \leq 116.4$.
- IQ = 120, $z_p = 2, F_Z(2) = 0.9772$, hence, having an IQ larger then 120 means you are roughly in the top 3 % most "intelligent" people.

1.8.2 Écart-type et quantiles de la normale

Soit une normale de moyenne égale à -50 et d'écart-type égal à 30. Calculez :
Les premier, deuxième et troisième quartiles.
Le cinquième et quatre-vingt quinzième percentile.

Écart-type et quantiles de la normale

Given a normal r.v. of mean -50 and standard deviation 30, compute the first, second and third quartiles as well as the fifth in ninety-fifth percentile.

Solution 1.8.2

- First quartile = $z_{0.25} \simeq -0.67 (= -z_{0.75}) \Rightarrow x_{0.25} = z_{0.25} \times 30 + (-50) = -70.2$
- Second quartile : $z_{0.5} = 0; x_{0.5} = -50$
- Third quartile : $z_{0.75} = 0.67; x_{0.75} = -29.8$
- 5^{th} percentile $z_{0.05} = -1.65; x_{0.05} = -99.2$
- 95^{th} percentile $z_{0.95} = 1.65; x_{0.95} = -0.8$

1.8.3 Intervalles

Soit X une variable aléatoire normale centrée d'écart-type σ, en utilisant les tables, calculez la probabilité des événements $\{X \geq k\sigma\}$ et $\{|X| \leq k\sigma\}$ pour $k = 1, 2, 3$.

Intervals

Let X be a normal r.v. of zero mean and standard deviation equal to σ, using the tables, compute the probabilities of following events : $\{X \geq k\sigma\}$ and $\{|X| \leq k\sigma\}$ for $k = 1, 2, 3$.

Solution 1.8.3

- $\mathbb{P}(X \geq k\sigma) = \mathbb{P}(Z \geq k) = 1 - F_Z(k) = [0.1587, 0.0228, 0.0023]$
- $\mathbb{P}(|X| \leq k\sigma) = \mathbb{P}(|Z| \geq k) = F_Z(k) - F_Z(-k) = [0.6826, 0.9544, 0.9974]$

1.8.4 Moyenne et intervalle de confiance

On s'intéresse à la mesure de la vitesse moyenne sur autoroute. La vitesse est une variable aléatoire de moyenne inconnue et d'écart-type de 15 km/h.

On effectue des mesures qui sont indépendantes les unes des autres. Donc, si la vitesse V est une variable aléatoire $\mathcal{N}(\mu, \sigma^2)$, les vitesses mesurées V_1 et V_2 sont également des v.a. $\mathcal{N}(\mu, \sigma^2)$, et V_1 est indépendant de V_2.

Quelle doit être la taille de l'échantillon (combien faut-il faire de mesures de vitesse) pour connaître la vitesse moyenne avec une précision de ± 1 km/heure pour un niveau de confiance de 95 % ?

Mean and confidence interval

We are interested in the mean speed measurement on a highway. This speed is a r.v. with unknown mean and standard deviation equal to 15 km/h.
What must be the size of the sample (how many measurements do we need) to know the mean of the speed with an accuracy of ± 1km/hour for a confidence level of 95 % ?

Solution 1.8.4

$$\bar{X} = \mathcal{N}(\mu, \frac{15^2}{n}); Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$z_{0.025} \leq Z \leq z_{0.975} \Rightarrow -1.96 \leq Z \leq 1.96$$

$\Rightarrow 1.96.\frac{15}{\sqrt{n}} = 1(km/h) \Rightarrow n = 865$

1.8.5 Moyenne

On mesure la température d'une pièce sortant d'un four, et on obtient l'échantillon suivant :

$$\begin{matrix} [345 & 358 & 344 & 350 & 334 & 344 \\ 346 & 343 & 357 & 343 & 343 & 353 & 342 \\ 355 & 361 & 347 & 348 & 360 & 348 & 348 & 364 \\ 370 & 353 & 357 & 351 & 328 & 348 & 357 & 354 & 349 &] \end{matrix}$$

calculez la moyenne et l'écart-type de l'échantillon.

Pouvez-vous donner la précision de l'estimation de la moyenne ?
Que pouvez-vous en déduire sur la densité de probabilité de la température ?
Mean

We measure the temparature of an oven and we get the following sample :
[345 358 344 350 334 344
346 343 357 343 343 353 342
355 361 347 348 360 348 348 364
370 353 357 351 328 348 357 354 349]
Compute the mean and standard deviation of the sample. Can you give the accuracy of the mean estimation ?
What can you say about the pdf of the temperature ?

Solution 1.8.5

$\bar{x} = 350; s = 8.66; n = 30 \Rightarrow$ accuracy of about $\frac{2s}{\sqrt{30}} = 3degrees$ with a confidence interval of 95 %.
Plot the histogram : looks roughly normal.

1.8.6 v.a. normale

Soit X et Y des variables aléatoires de moyennes 0 et 1 respectivement et de variances 1 et 4 respectivement. Trouvez :

1. $\mathbb{P}(X \leq 1.5)$ et $\mathbb{P}(X \leq -1)$
2. la densité de probabilité de $(Y-1)/2$
3. $\mathbb{P}(-1 \leq Y \leq 1)$
4. $\mathbb{P}(X \geq 2 | X \geq 0)$

normal r.v.

Let X and Y be two r.v. of, respectively means 0 and 1 and variances 1 and 4. Determine :

1. $\mathbb{P}(X \leq 1.5)$ *and* $\mathbb{P}(X \leq -1)$
2. *the pdf of* $(Y-1)/2$
3. $\mathbb{P}(-1 \leq Y \leq 1)$
4. $\mathbb{P}(X \geq 2 | X \geq 0)$

Solution 1.8.6

- $\mathbb{P}(X \leq 1.5) = F_Z(1.5) = 0.9332; \mathbb{P}(X \leq -1) = 1 - F_Z(1) = 0.1587$
- $Y \sim \mathcal{N}(1,4) \Rightarrow \frac{Y-1}{2} \sim \mathcal{N}(0,1)$
- $\mathbb{P}(-1 \leq Y \leq 1) = \mathbb{P}(-1 \leq Z \leq 0) = F_Z(0) - F_Z(-1) = 0.34$
- $\mathbb{P}(X \geq 2\ X \geq 0) = \frac{\mathbb{P}(X \geq 2, X \geq 0)}{\mathbb{P}(X \geq 0)} = \frac{1-0.9772}{5} = 0.0456$

1.8.7 Appels téléphoniques

Calls to a telephone system follow a Poisson distribution with a mean of five calls per minute. (a) What is the name applied to the distribution and parameter values of the time until the tenth call? (b) What is the mean time until the tenth call? (c) What is the mean time between the ninth and tenth calls? (d) What is the probability that exactly four calls occur within one minute? (e) If 10 separate one-minute intervals are chosen, what is the probability that all intervals contain more than two calls?

Solution 1.8.7

Y is the number of calls in one minute, is r.v. variable of $\lambda = 5$ calls per minute.
$P(Y = 4) = e^{(-5)}5^4/fact(4) = 0.1755$
$P(Y > 2) = 1 - P(Y <= 2) = 0.8754$
Let W denote the number of one minute intervals out of 10 that contain more than 2 calls. Because the calls are a Poisson process, W is a binomial random variable with n = 10 and p = 0.8754. Therefore,$P(W = 10) = C_{10}^{10} 0.8754^{10}(10.8754)^0 = 0.2643$

Fonction de répartition de la normale centrée réduite

Cumulative Density Function of a zero-mean unit-variance Gaussian r.v.

Fonction de répartition de $\mathcal{N}(0,1)$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

1.9 TD 9 : v.a. multiples / corrélation / covariance

TD 9 : multiple r.v. / correlation / covariance

1.9.1 Somme de v.a.

Les variables aléatoires X, Y et Z sont indépendantes et uniformément réparties sur $[0, 1]$. Trouvez le densité de probabilité de $X + Y + Z$.

Sum of r.v.

Three r.v. X, Y et Z are independent and uniformly distributed on $[0, 1]$. Find the pdf of $X + Y + Z$.

Solution 1.9.1

$f_X(x) = 1$ on $x \in [0,1]$, likewize for Y and Z.
Let $T = X + Y$, then $f_T(t) = f_X(x) * f_Y(y) = \int_0^1 f_T(t-x)\, f_Y(x)\, dx$.
for $0 \leq t \leq 1$:

$$\begin{aligned} f_T(t) &= \int_0^1 \Pi_{[0,t]}(x)\Pi_{[0,1]}(x)dx \\ &= \int_0^t dx = t \end{aligned}$$

for $1 \leq t \leq 2$:

$$\begin{aligned} f_T(t) &= \int_0^1 \Pi_{[t-1,1]}(x)\Pi_{[0,1]}(x)dx \\ &= \int_{t-1}^1 dx = 2 - t \end{aligned}$$

otherwise, $f_T(t) = 0$.

Hence $\begin{aligned} f_T(t) &= 0 & t < 0 \\ &= t & 0 \leq t \leq 1 \\ &= 2 - t & 1 \leq t \leq 2 \\ &= 0 & t > 2 \end{aligned}$

Let $U = T + Z$, $f_U(u) = f_Z(z) * f_T(t) = \int_0^2 f_Z(u-x)\, f_T(x)\, dx$.
Then :

- $0 \leq u \leq 1$:
  $f_U(u) = \int_0^u x dx = u^2/2$
- $1 \leq u \leq 2$:
  $f_U(u) = \int_{u-1}^1 x dx + \int_1^u (2-x)dx = 1/2 - (u-1)^2/2 + 2.(u-1) - (u^2/2 - 1/2) == 3u - u^2 - 3/2$
- $2 \leq u \leq 3$: $f_U(u) \int_{u-1}^2 (2-x)dx = 2(2-(u-1)) - 2^2 + (u-1)^2/2 = u^2/2 - 3u + 5/2$

1.9.2 Somme de v.a. REF Bertsekas

Une équipe de foot doit désigne trois tireurs de penalty, chaque tireur réussissant avec une probabilité p_i, indépendemment des autres tireurs. soit X le nombre de penalties marqués après que chaque tireur ait tiré une fois. Utilisez la convolution pour calculer la masse de probabilité de X.

1.9.3 Somme de v.a. REF Bogaerts

Soit $X \sim Y \sim Exp(\lambda)$, la durée de fonctionnement de deux machines identiques et indépendantes avant l'occurrence d'une première panne, alors $Z = X + Y$ est la durée totale de fonctionnement des deux machines si l'une se met en marche dès que l'autre est en panne. Trouvez la densité de probabilité de la durée totale de fonctionnement.

Sum of r.v.

Let $X \sim Y \sim Exp(\lambda)$ be the duration in which two identical and independent machines work correctly, before their first failure. Then, $Z = X + Y$ is the total operation duration if machine Y starts to operate when machine X fails. Find the pdf of this total duration.

Solution 1.9.2

$f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$

$$\begin{aligned} f_Z(z) &= f_X(x) * f_Y(y) \\ &= \int_{-\infty}^{\infty} f_X(z-x)\, f_Y(x)\, dx = \lambda^2 \int_0^z e^{-\lambda(z-x)} e^{-\lambda x} dx \\ &= \lambda^2 \int_0^z e^{-\lambda z} dx = \lambda^2 z e^{-\lambda z} \end{aligned}$$

1.9.4 Corrélation simple sur deux v.a. discrètes

Soient deux variables aléatoires X et Y, jointes, avec leur probabilités jointes données par :

- $p_{X,Y}(0,0) = 0.2$
- $p_{X,Y}(1,1) = 0.1$
- $p_{X,Y}(1,2) = 0.1$
- $p_{X,Y}(2,1) = 0.1$
- $p_{X,Y}(2,2) = 0.1$
- $p_{X,Y}(3,3) = 0.4$

Suggestion : représentez cette distribution jointe dans un graphique (x, y)
On vous demande l'écart type σ^2_{XY} et le coefficient de corrélation ρ_{XY}.

Simple correlation of 2 discrete r.v.

Let X and Y defined by their pmf :

- $p_{X,Y}(0,0) = 0.2$
- $p_{X,Y}(1,1) = 0.1$
- $p_{X,Y}(1,2) = 0.1$
- $p_{X,Y}(2,1) = 0.1$
- $p_{X,Y}(2,2) = 0.1$
- $p_{X,Y}(3,3) = 0.4$

hint : represent the pmf of (x, y) graphically
compute σ^2_{XY} and ρ_{XY}.

Solution 1.9.3

$\sigma^2_{xy} = 1.26, \rho_{xy} = 0.926$

1.9.5 Corrélation simple sur deux v.a. continues

Soient deux variables aléatoires X et Y, jointes, avec leur densité de probabilité donnée par :

$$f_{XY}(xy) = \frac{x.y}{16} \qquad ; 0 \leq X \leq 2 \qquad ; \qquad 0 \leq Y \leq 4$$

On vous demande l'écart type σ_{XY} et le coefficient de corrélation ρ_{XY}.

Que pouvez vous en déduire de X et Y. Est-ce que c'est conforme à l'intuition qu'on aurait pu avoir.

Corrélation simple sur deux v.a. continues

let X and Y, such that :

$$f_{XY}(xy) = \frac{x.y}{16} \qquad ; 0 \leq X \leq 2 \qquad ; \qquad 0 \leq Y \leq 4$$

Find σ_{XY} and ρ_{XY}.
What can you deduce about X and Y. Could you have found it intuitively?

Solution 1.9.4

$$\mathrm{E}[XY] = \int_0^2 \int_0^4 \frac{x^2y^2}{16} dydx = 32/9$$

$$\mathrm{E}[X] = \int_0^2 \int_0^4 \frac{x^2y}{16} dydx = 4/3$$

$$\mathrm{E}[Y] = \int_0^2 \int_0^4 \frac{xy^2}{16} dydx = 8/3$$

$$\Rightarrow \sigma^2_{XY} = \frac{32}{9} - \frac{4 \times 8}{3 \times 3} = 0 \Rightarrow \rho_{XY} = 0$$

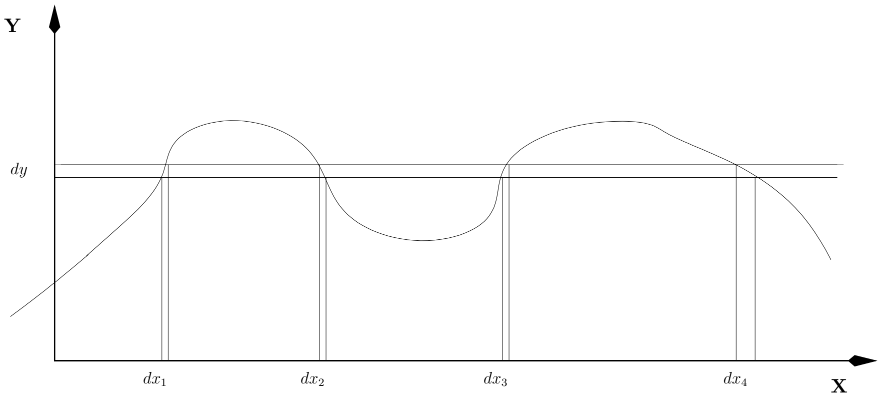
Hence, X and Y are decorrelated, and indeed $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$.

1.10 TD 10 :Changement de variables

Rappels sur les changements de variables

Cas uni-dimensionnel

On considère une variable aléatoire $\mathbf{Y}$, fonction de la variable aléatoire $\mathbf{X}$ telle que $\mathbf{Y} = g(\mathbf{X})$. Dans le cas général, la fonction n'est pas bijective, comme illustré dans la figure ci-dessous.



On sépare la fonction en tronçons i sur lesquels la fonction est bijective ($\mathbf{Y} = g_i(\mathbf{X}), i = 1, ..n$). Sur ces tronçons, on peut écrire $\mathbf{X} = g_i^{-1}(\mathbf{Y})$ où $g_i^{-1}(.)$ est la fonction inverse de $g(.)$.

Comme on l'a vu au cours, on a $P_Y(y)|dy| = \sum_{x_i|y=g(x_i)} P_X(x_i)|dx_i|$, par conséquent :

$$P_Y(y) = P_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + \cdots + P_X(g_n^{-1}(y)) \left| \frac{dg_n^{-1}(y)}{dy} \right|.$$

Exemple 1.10.1 *Changement de variable* $\mathbf{Y} = \mathbf{X}^2$

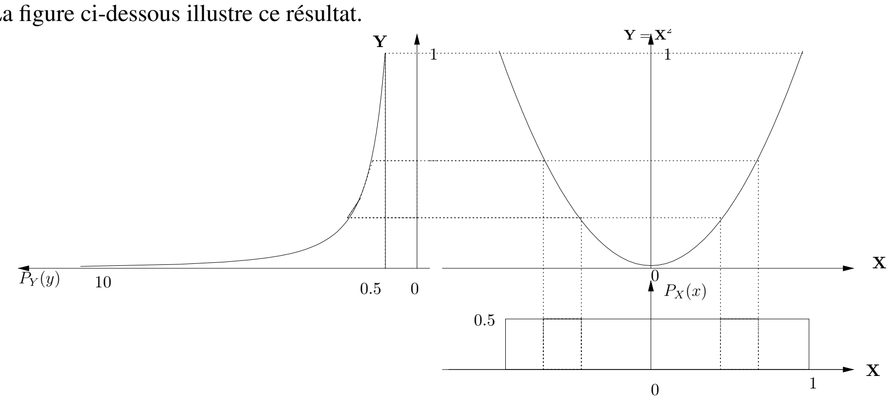
On considère la transformation $\mathbf{Y} = \mathbf{X}^2$ sur [-1,1]. Dans ce cas, on a deux tronçons :

1. sur $x \in [-1, 0]$: $\mathbf{X} = -\sqrt{\mathbf{Y}}$ $(g_1^{-1}(.) = -\sqrt{(.)})$
2. sur $x \in [0, 1]$: $\mathbf{X} = \sqrt{\mathbf{Y}}$ $(g_2^{-1}(.) = \sqrt{(.)})$

On obtient alors

$$\begin{aligned} P_Y(y) &= P_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + P_X(g_2^{-1}(y)) \left| \frac{dg_2^{-1}(y)}{dy} \right| \\ &= P_X(\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} + P_X(-\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}}. \end{aligned}$$

La figure ci-dessous illustre ce résultat.



◁

Cas multi-dimensionnel

On considère un changement de variables multi-dimensionnel que l'on peut écrire sous la forme :

$$\begin{aligned}\mathbf{Y}_1 &= g_1(\mathbf{X}_1, \ldots, \mathbf{X}_n) \\ &\vdots \\ \mathbf{Y}_n &= g_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)\end{aligned}$$

Pour des fonctions g_i continues et différentiables (et à condition que le jacobien défini ci-dessous soit non nul), on peut faire le même raisonnement que ci-dessous. Donc, dans un tronçon, on a :

$$P_Y(y_1, \ldots, y_n) = P_X(x_1, \ldots, x_n) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right|,$$

on notera que dans ce cas-ci, il est plus compliqué d'écrire l'expression en fonction de g_i^{-1}, mais on aurait plutôt des fonctions de type $\mathbf{X}_i = f_i(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$.

Exemple 1.10.2 *Précision de fabrication en microélectronique*

Dans le processus de fabrication de circuits intégrés, une des parties cruciales est la précision de la lithographie. On peut quantifier cette précision comme étant la déviation en coordonnées horizontales et verticales (x et y) par rapport à l'endroit à graver.
Dans le cas de technologies "70 nm", on peut considérer que les déviations en x et y sont des variables aléatoires indépendantes qui suivent des lois gaussiennes de moyenne nulle et de variance $\sigma^2 = 0.2\text{nm}^2$. La densité de probabilité conjointe des déviations $(\mathbf{X}, \mathbf{Y})$ est donnée par :

$$P_{XY}(x, y) = P_X(x)P_Y(y) = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}$$

La caractérisation de la précision en x et en y ne répond pas à la question suivante : "quelle est la loi de probabilité de la distance entre le point désiré et le point obtenu par la lithographie ?" Pour obtenir cette loi, on passe des coordonnées cartésiennes aux coordonnées polaires selon la transformation :

$$(\mathbf{X}, \mathbf{Y}) = (\mathbf{R}\cos(\Theta), \mathbf{R}\sin(\Theta)).$$

Le jacobien de la transformation vaut r :

$$\left|\frac{\partial(x, y)}{\partial(r, \theta)}\right| = \begin{vmatrix} \cos(\theta) & \sin(\theta) \\ -r\sin(\theta) & r\cos(\theta) \end{vmatrix} = r,$$

et donc

$$\begin{aligned} P_{\mathbf{R},\Theta}(r, \theta) &= \frac{r}{2\pi\sigma^2}e^{-[(r\cos(\theta))^2+(r\sin(\theta))^2]/2\sigma^2} \\ &= \frac{r}{2\pi\sigma^2}e^{-r^2/2\sigma^2} \end{aligned}$$

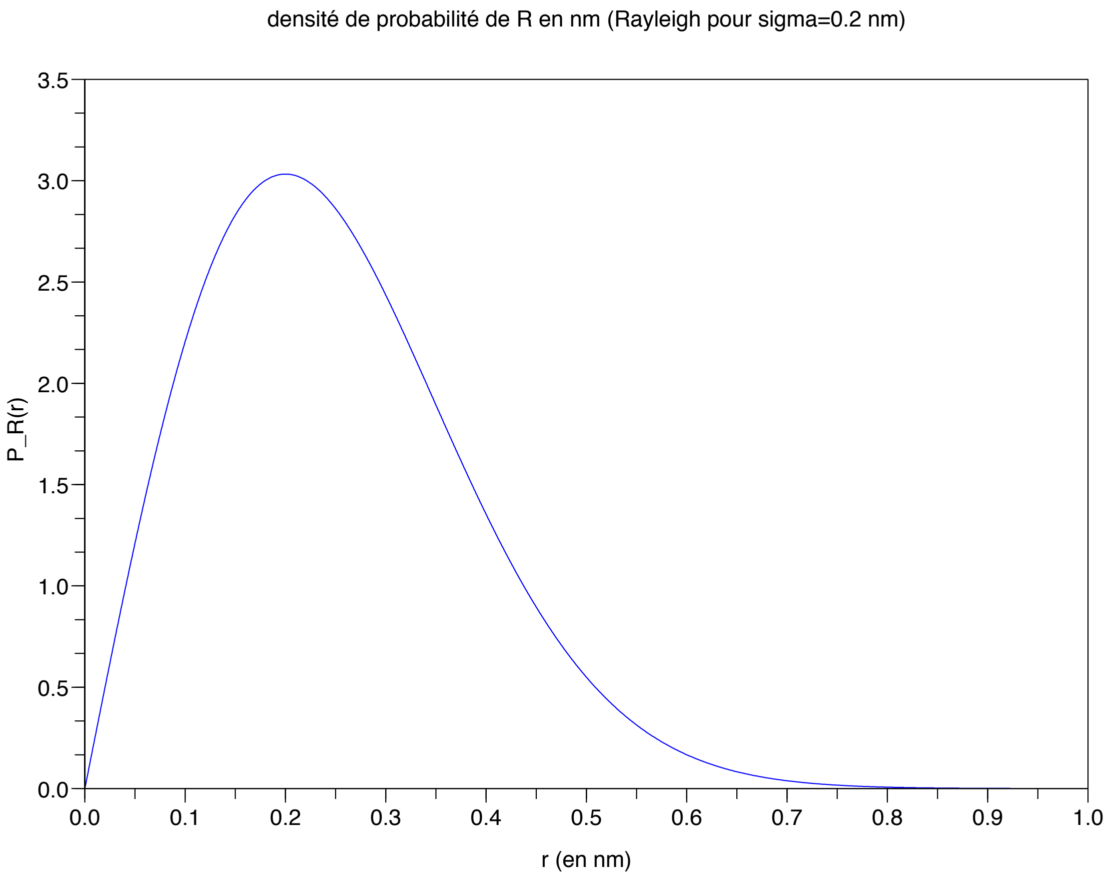
Pour trouver les marginales, il faut intégrer sur l'autre variable aléatoire, soit :

$$\begin{aligned} P_\Theta(\theta) &= \int_{r=0}^{\infty} P_{\mathbf{R},\Theta}dr \\ &= \int_{r=0}^{\infty} \frac{r}{2\pi\sigma^2}e^{-r^2/2\sigma^2}dr \\ &= \frac{1}{2\pi} \end{aligned}$$

L'angle est donc uniformément distribué sur $[0, 2\pi]$. D'autre part, on en déduit que

$$P_{\mathbf{R}}(r) = \frac{r}{\sigma^2}e^{-r^2/2\sigma^2}$$

On en déduit également que les variables aléatoires $\mathbf{R}$ et Θ sont indépendantes (la loi conjointe est donnée par le produit des lois marginales). La densité de probabilité suivie par la distance $\mathbf{R}$ est celle d'une variable dite de Rayleigh.
$\lhd$


densité de probabilité de R en nm (Rayleigh pour sigma=0.2 nm)
P_R(r)
r (en nm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

1.10.1 Transformations

Soit X une variable aléatoire dont les valeurs possibles et équiprobables sont les entiers entre 0 et 9. Calculer la fonction de probabilité suivante:

1. $p_Y(y)$ de $Y = X\text{mod}(3)$,

2. $p_Z(z)$ de $Z = 5 \mod (X+1)$.

Transformations

Let X be a uniform discrete r.v. on [0..9]. Compute the following probability functions :

1. *$p_Y(y)$ de $Y = X mod(3)$,*
2. *$p_Z(z)$ de $Z = 5 \mod (X+1)$.*

Solution 1.10.1

First $p_X(x) = 0.1$ for $x \in [0..9]$.
Hence, we can have the table

x	y	z
0	0	0
1	1	1
2	2	2
3	0	1
4	1	0
5	2	5
6	0	5
7	1	5
8	2	5
9	0	5

Hence $p_Y(0) = 0.4; p_Y(1) = 0.3; p_Y(2) = 0.3$
and $p_Z(0) = 0.2; p_Z(1) = 0.2; p_Z(2) = 0.1; p_Z(5) = 0.5$

1.10.2 changement de variable

Soit X une variable aléatoire uniformément distribuées sur $[-1, 1]$. Trouvez la densité de probabilité de $\sqrt{|X|}$ et de $-\ln|X|$.

change of variable

Let X be a uniform r.v. on $[-1, 1]$. Find the pdf of $\sqrt{|X|}$ and of $-ln|X|$.

Solution 1.10.2

$$Y = \sqrt{|X|} \quad \begin{array}{ll} \text{on } [-1,0] : X = -Y^2 & g_1^{-1}. = -(.)^2 \\ \text{on } [0,1] : X = Y^2 & g_2^{-1}. = (.)^2 \end{array}$$

$$f_Y(y) = f_X\big(-x^2\big)\,|-2y| + f_x\big(x^2\big)\,|2y| = \frac{1}{2} \times 2y + \frac{1}{2} \times 2y = 2y$$

$$Y = -\log(|X) \quad \begin{array}{ll} \text{on } [-1,0] : Y = -\log(-X), -X = e^{-Y} & g_1^{-1}. = -e^{-(.)} \\ \text{on } [0,1] : Y = -\log(X), X = e^{-Y} & g_2^{-1}. = e^{-(.)} \end{array}$$

$$\Rightarrow f_Y(y) = f_X\big(-e^{-y}\big)\,|-e^{-y}| + f_X\big(e^{-y}\big)\,|e^{-y}| = e^{-y}, y \in [0,\infty]$$

1.10.3 changement de variable

Soit une v.a. X de densité $f_X(x)$, et une autre variable aléatoire indépendante Y de densité $f_Y(y)$ trouvez la densité de

- $Z = e^X$, que devient cette densité si $X \sim \text{Un}([0,1])$
- $Z = |X|^{1/3}$
- $Z = |X|^{1/4}$
- Soit X et Y deux v.a. jointes et uniformément réparties sur $[0,1] \times [0,1]$, trouvez les fonction de répartition et densité de probabilité de $|X - Y|$.

change of variable

Let X be a r.v. of pdf $f_X(x)$, and Y of density $f_Y(y)$ find the density of

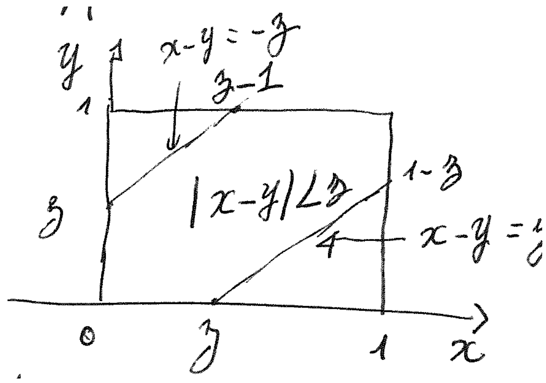
- $Z = e^X$, *what if $X \sim$ Un$([0,1])$?*
- $Z = |X|^{1/3}$
- $Z = |X|^{1/4}$
- *Let X and Y be jointly uniformly distributed on $[0,1] \times [0,1]$, find the cdf and pdf of $|X - Y|$.*

Solution 1.10.3

- e^x monotonous so $g^{-1}(.) = \log(.)$ and $z \geq 0$.
  $x = \log(z)$, $f_X(\log(z)) \left|\frac{1}{z}\right| = \frac{1}{z} f_X(\log(z))$
  if $X \sim \text{Un}([0,1])$, $f_Z(z) = \frac{1}{z}, z \in [1, e]$
-
$$Z = |X|^{1/3} \Rightarrow \begin{cases} x = -z^3 & \text{if } x \leq 0 \quad g_1^{-1}(z) = -z^3 \\ x = z^3 & \text{if } x \geq 0 \quad g_2^{-1}(z) = z^3 \end{cases}$$

$$\Rightarrow f_Z(z) = 3z^2\big(f_X(-z^3) + f_X(z^3)\big), z \geq 0$$

- same as previously
- $X \sim Y \sim \text{Un}[0,1] f_{X,Y}(x,y) = 1$ on $0 \leq x \leq 1$ and $0 \leq y \leq 1$
  Find the cdf of $|X - Y| = \mathbb{P}(|X - Y| < z)$



1.
$$F_Z(z) = \begin{cases} 0 & z < 0 \\ 1 - (z-1)^2 & 0 \leq y \leq 1 \\ 1 & z > 1 \end{cases}$$

Hence, in $0 \leq y \leq 1$, $f_Y(y) : \dfrac{dF_Z(z)}{dz} = 2(1 - z)$

2. *another way to do it*

$$\mathbb{P}(|X - Y| < z) = \mathbb{P}(-z < X - Y < z) = \mathbb{P}(X - Y < z) - \mathbb{P}(X - Y > -z)$$

1.10.4 Transformation cartésien vers polaire

Soit X et Y des variables aléatoires indépendantes de densités de probabilités gaussiennes centrées réduites. La paire (X, Y) peut être transformée en coordonnées polaires en termes de $R \geq 0$ et $\Theta \in [0, 2\pi]$ par les expressions :

$$X = R\cos\Theta, \qquad Y = R\sin\Theta$$

- Montrez que Θ est uniformément répartie sur $[0, 2\pi]$ et que R est distribué selon la loi :

$$f_R(r) = re^{-r^2/2}, \qquad r \geq 0$$

- Montrez que R^2 a une distribution exponentielle de paramètre $1/2$

Polar coordinates

Let X and Y be i.i.ds $\sim \mathcal{N}(0,1)$. (X,Y) is transformed to $R \geq 0$ and $\Theta \in [0,2\pi]$ by :

$$X = R\cos\Theta, \qquad Y = R\sin\Theta$$

- *Show that Θ is uniformely distributed on $[0,2\pi]$ and that R is distributed as :*

$$f_R(r) = re^{-r^2/2}, \qquad r \geq 0$$

- *Show that R^2 has an exponential distribution of parameter $1/2$*

Solution 1.10.4

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$$

Jacobian : $|J| = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$

Hence $f_{R,\Theta}(r,\theta) = r.f_{X,Y}(x = r\cos\theta, y = r\sin\theta) = r\frac{1}{2\pi}e^{-r^2/2}$, $r \geq 0, \theta \in [0,2\pi]$

$f_\Theta(\theta) = \displaystyle\int_0^\infty \frac{r}{2\pi}e^{-\frac{r^2}{2}}dr = \frac{1}{2\pi}\int_0^\infty e^{-u/2}du = \frac{1}{2\pi}$

$\Rightarrow f_R(r) : re^{-r^2/2}$,

and taking $z = r^2, r = \sqrt{z}, f_Z(z) = 1/2\frac{1}{\sqrt{z}}\sqrt{z}e^{-z/2} = \frac{1}{2}e^{-z/2}$

1.10.5 Un point sur un demi-disque

Un point est choisi sur un demi-disque de rayon R. Le demi-disque est centré à l'origine et est situé dans le demi-plan supérieur. On demande :

1. La loi de probabilité conjointe de ses coordonnées **X** et **Y**

2. la loi de probabilité marginale **Y** et sa moyenne.

3. Vérifier (2) en calculant E(**Y**) sans utiliser la loi marginale de **Y**.

Solution 1.10.5

1. La point étant choisi "au hasard" sur un demi-disque de rayon R, cela veut dire que l'on a une distribution uniforme sur le demi-disque et donc $P_{X,Y}(x,y) = \frac{2}{\pi R^2}$. **Note** : c'est la solution de Bertsekas, mais c'est peut-être un peu vite dit. J'aurais tendance à dire que $P_\Theta(\theta) = 1/\pi$ pour $\theta \in [-\pi/2, \pi/2]$ et $P_{\mathbf{R}}(r) = \partial_{rR}$. On a donc les relations

$$(\mathbf{X}, \mathbf{Y}) = (\mathbf{R}\cos(\Theta), \mathbf{R}\sin(\Theta)),$$

et, en tenant compte qu'on travaille uniquement sur le demi-disque supérieur :

$$(\Theta, \mathbf{R}) = (\arctan(\mathbf{Y}/\mathbf{X}), \sqrt{\mathbf{X}^2 + \mathbf{Y}^2})$$

(sous contrainte que $\sqrt{\mathbf{X}^2 + \mathbf{Y}^2} = R$)

2. Pour trouver la loi marginale de **Y**, on intègre la loi jointe sur **X** :

$$P_Y(y) = \int_{-A}^{A} \frac{2}{\pi R^2} dx = \begin{cases} \frac{4A}{\pi R^2} & \text{si } 0 \leq y \leq R, \\ 0 & \text{sinon.} \end{cases}$$

où $A = \sqrt{R^2 - y^2}$. On a alors

$$\mathrm{E}[\mathbf{Y}] = \frac{4}{\pi R^2} \int_0^R y\sqrt{R^2 - y^2} dy = \frac{4R}{3\pi}$$

en utilisant le changement de variable $z = R^2 - y^2$ pour l'intégration.

3. On peut trouver l'espérance en utilisant directement la loi conjointe. En notant D le demi-disque, on a :

$$\mathrm{E}[\mathbf{Y}] = \int_D y P_{X,Y}(x,y) dx dy = \int_0^\pi \int_0^R \frac{2}{\pi R^2} r \sin\theta r dr d\theta = \frac{4R}{3\pi}$$