



## RFIDEC

# Reconnaissance des **F**ormes et Introduction à la **D**écision

Fascicule de TD/TME

Année 2012-2013

# 1 Statistique descriptive

## Exercice 1 – Salaires

Deux PDG se rencontrent et discutent de leurs entreprises respectives. Le premier, de la société A, annonce fièrement au second, de la société B : « Je paye mieux mes salariés que vous. Mes 150 ouvriers ont un salaire moyen de €1500 et mes 50 cadres gagnent €3000 ». Le second lui rétorque : « Vous vous trompez mon ami, le salaire moyen de mon entreprise est plus élevé que le votre. Mes 100 ouvriers gagnent en moyenne €1000 et mes cadres ont €2500 pour salaire moyen ».

A quelle condition le PDG de la société B a-t-il raison ?

## Exercice 2 – Housing

En 1993, l'association des chercheurs de la chambre de commerce américaine a recensé le prix moyen des maisons d'environ 150m<sup>2</sup> dans diverses villes. Voici les résultats obtenus :

Huntsville	105111\$	Anchorage	155913\$	Phoenix	100748\$
Little Rock	92750\$	San Diego	223600\$	Denver	125200\$
Orlando	104112\$	Bloomington	105613\$	Ames	110325\$
La Nouvelle Orléans	88000\$	Minneapolis	128112\$	Lincoln	92925\$
Manchester	125750\$	Albuquerque	130550\$	Réno-Sparks	142300\$
Albany	116363\$	Charlotte	115800\$	Cincinnati	121329\$
Salem	108656\$	Sioux Falls	99890\$	Memphis	91297\$
Houston	96900\$	Salt Lake City	92281\$	Charleston	126000\$
Green Bay	114500\$				

**Q 2.1** Regrouper les modalités de la variable « prix » en 5 classes de tailles identiques et déterminer l'effectif et les fréquences de chaque classe. Déterminer l'histogramme correspondant.

NB : chercher des *chiffres ronds* pour les bornes d'intervalles.

**Q 2.2** Regrouper les 3 dernières classes en une seule. Déterminer l'histogramme correspondant.

## Exercice 3 – Roulette

L'observation des résultats de  $n = 100$  tirages successifs à la roulette, où la roue comporte 37 cases numérotées de 0 à 36, a donné les résultats suivants :

$x_l$	0	1	2	3	4	5	6	7	8	9	10	11	12
$n_l$	2	2	3	2	2	2	2	4	3	2	6	3	8
$x_l$		13	14	15	16	17	18	19	20	21	22	23	24
$n_l$		4	2	2	3	5	5	1	2	2	2	3	1
$x_l$		25	26	27	28	29	30	31	32	33	34	35	36
$n_l$		5	2	0	0	0	3	1	5	4	2	4	1

**Q 3.1** Si la roulette n'est pas truquée, la loi de probabilité de la variable aléatoire « numéro tiré »,  $X$ , est la loi uniforme sur l'ensemble  $\{0, 1, 2, \dots, 35, 36\}$ . Quelles sont alors l'espérance mathématique et la médiane de  $X$  ? Comparez-les à la moyenne  $\bar{x}$  et la médiane  $M$  de la distribution des observations.

**Q 3.2** Regrouper les observations d'abord en une classe de 4 numéros suivie de 11 classes de 3 numéros, puis en une classe de 7 numéros suivie de 5 classes de 6 numéros, et tracer à chaque fois l'histogramme correspondant. Pensez-vous que la roulette était truquée ?

---

**Exercice 4 – Automobiles**


---

D'après le Business Week daté du 3 août 1992, le gouvernement américain a accordé en 1991 aux 12 firmes automobiles listées ci-dessous le nombre de brevets suivant :

General Motors	506	Ford	234	Chrysler	97
Nissan	385	Porsche	50	Honda	249
Mazda	210	Volvo	23	Daimler-Benz	275
Toyota	257	Mitsubishi	36	BMW	13

**Q 4.1** Déterminer la médiane de la variable « nombre de brevets », ainsi que les premiers et derniers quartiles.

---

**Exercice 5 – Assurance**


---

Une société d'assurances veut proposer à ses clients une nouvelle police d'assurance. Pour déterminer son tarif, elle a sélectionné les dossiers de 40 de ses clients, et a estimé ce qu'elle aurait dû rembourser aux dits clients l'année dernière. Voici les remboursements (en euros) qu'elle a calculés :

500	150	0	0	2500	250	0	100000	1000	500
0	0	5500	560	0	230	150	1000	0	1250
60	0	1500	0	750	1500	2000	0	750	0
700	1800	200	0	600	0	20000	300	0	0

**Q 5.1** Calculer la moyenne et l'écart-type de la variable « remboursement ».

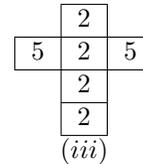
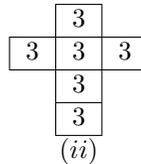
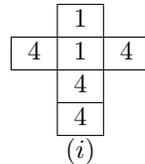
**Q 5.2** Calculer le premier et le dernier décile.

**Q 5.3** Quelle est la mesure de dispersion la plus appropriée ?

## 2 Notions élémentaires de probabilités

### Exercice 6 – Dés de Gardner

Dans un numéro de la revue *Scientific American* de 1974, M. Gardner proposait un jeu consistant à choisir un dé parmi les trois dés à 6 faces non pipés ci-dessous, de manière à essayer d'obtenir le nombre le plus élevé en lançant le dé une seule fois.



**Q 6.1** On vous propose de jouer au jeu à 2 joueurs suivant : chaque joueur mise  $M$  euros. Puis on vous demande de choisir un des dés ci-dessus, votre adversaire en choisit ensuite un autre et enfin chacun lance son dé. Celui qui obtient le nombre le plus élevé remporte la mise.

**Q 6.1.1** Calculez, pour chaque couple  $(x, y)$  de dés la probabilité qu'en jouant avec le dé  $x$  on obtienne un résultat plus élevé qu'avec  $y$ .

**Q 6.1.2** Sachant que la mise est de 30 euros, devez-vous accepter de jouer et, le cas échéant, quel dé devez-vous choisir ?

### Exercice 7 – Dutch book

Un bookmaker (B.) doit afficher les cotes des paris avant une course à laquelle participeront 3 chevaux. On désigne par  $e_i$  l'événement « le cheval  $i$  gagne » ( $i = 1, 2, 3$ ). Un et un seul des  $e_i$  se réalisera.

Les parieurs peuvent parier des sommes quelconques sur un ou plusieurs des événements  $e_1, e_2, e_3$ ,  $e_{1,2} = e_1 \cup e_2$ ,  $e_{1,3} = e_1 \cup e_3$  et  $e_{2,3} = e_2 \cup e_3$ ; ils peuvent aussi parier contre ces événements.

Parier *sur* l'événement  $E$  à la cote «  $k$  contre 1 » signifie : miser, puis recevoir du B.  $\text{€}(k + 1)$  pour  $\text{€}1$  misé si  $E$  se produit et rien sinon.

Parier *contre* l'événement  $E$  à la cote «  $k$  contre 1 » signifie : recevoir du B. une mise, puis lui payer  $\text{€}(k + 1)$  pour  $\text{€}1$  misé si  $E$  se produit et rien sinon.

**Q 7.1** Le B. attribue des cotes quelconques.

**Q 7.1.1** Montrer que si le B. offre des cotes  $k_1 = 2$  pour  $e_1$  et également  $k_{2,3} = 2$  pour  $e_2 \cup e_3$ , un parieur peut gagner de l'argent à coup sûr.

**Q 7.1.2** Montrer qu'il en est de même si le B. offre des cotes  $k_1 = k_2 = 3$  pour  $e_1$  comme pour  $e_2$  et  $k_{1,2} = 0.5$  pour  $e_1 \cup e_2$ .

**Q 7.2** Le B. attribue aux événements  $e_i$  des probabilités  $p_i$  ( $i = 1, 2, 3$ ) :

$$p_i \geq 0 \quad (i = 1, 2, 3); \quad p_1 + p_2 + p_3 = 1$$

et fixe les cotes  $k_i$  et  $k_{i,j}$  de sorte que tous les paris soient *équitable*s, c'est-à-dire qu'ils aient une espérance mathématique nulle.

**Q 7.2.1** Quelle relation existe-t-il entre les  $k_i, k_{i,j}$ , d'une part, et les  $p_i$  d'autre part ?

**Q 7.2.2** Montrer que, quelles que soient les quantités totales *nettes* mises  $x_i$  ( $x_i =$  quantité totale mise sur  $e_i$  moins quantité totale mise contre  $e_i$ ) et  $x_{i,j}$ , il est impossible que le B. perde de l'argent à coup sûr.

**Exercice 8 – La roulette**

Dans les casinos, la roulette contient 37 numéros : 18 rouges, 18 noirs et un vert. Quand la roulette tourne, la bille a autant de chances de tomber sur chacun des 37 numéros. Si l'on mise €1 sur le rouge et que ce dernier sort, on gagne €1, sinon on perd la mise de €1.

**Q 8.1** La roulette vous sera-t-elle profitable ?

**Q 8.1.1** Soit  $X$  la variable aléatoire représentant le résultat d'une mise de €1. Quelle est la distribution de probabilité de  $X$  ? Quelle est l'espérance de  $X$  ?

**Q 8.1.2** En moyenne combien gagnerez-vous ou perdrez-vous par mise ?

**Q 8.1.3** Combien gagnerez-vous ou perdrez-vous si vous jouez 100 fois en misant €1 à chaque fois ? 1000 fois ? Peut-on en déduire que la roulette n'est pas un jeu profitable ? Justifiez votre réponse.

**Exercice 9 – Fiabilité**

Un chef d'entreprise a rempli une base de données contenant la liste des pannes des machines de l'entreprise. En consultant cette base, il a pu déterminer que la variable aléatoire  $X = \ll$  nombre de pannes pour un jour donné  $\gg$  suivait la distribution de probabilité :

$x$	0	1	2
$P(X = x)$	0,80	0,15	0,05

**Q 9.1** À quelle famille de lois appartient la loi suivie par la variable  $X$  ? Déterminer la moyenne et l'écart-type de  $X$ .

**Q 9.2** En moyenne, combien de pannes peut-on espérer (craindre plutôt ?) sur une période d'un an, en supposant qu'il y a 250 jours de travail par an ?

**Exercice 10 – Paradoxe de Simpson**

Le recensement des jugements prononcés dans l'état de Floride entre 1973 et 1978 a permis d'établir le tableau suivant, qui présente les sentences en fonction de la couleur de peau de l'accusé :

meurtrier	peine de mort	autre sentence
noir	59	2547
blanc	72	2185

**Q 10.1** Calculez la probabilité d'obtenir la peine de mort sachant que l'on est noir, puis sachant que l'on est blanc. Qu'en concluez-vous ?

**Q 10.2** En fait le tableau ci-dessus est une synthèse du tableau ci-dessous :

victime	meurtrier	peine de mort	autre sentence
blanche	noir	48	238
	blanc	72	2074
noire	noir	11	2309
	blanc	0	111

Calculez la probabilité d'obtenir la peine de mort conditionnellement à la couleur de peau de l'accusé et de la victime. La justice est-elle clémente envers les noirs dans l'état de Floride ? Justifiez votre réponse.

**Exercice 11 – Sport et age**

Dans un échantillon aléatoire de 240 personnes, on a recueilli l'information suivante sur l'âge et sur le type de sport le plus fréquemment pratiqué à Jussieu :

âge \ activité sportive	moins de 20 ans	[20; 25[ ans	[25; 30[ ans	plus de 30 ans
jogging	15	20	15	30
natation	15	10	20	25
ping pong	20	10	30	30

**Q 11.1** Quelles sont les deux variables aléatoires étudiées ?

**Q 11.2** Calculer la loi jointe de ces deux variables.

**Q 11.3** Calculer la probabilité qu'un individu de faire de la natation (dans cet échantillon). Quelle est la probabilité qu'un individu de cet échantillon tiré au hasard ait entre 20 et 25 ans ?

**Q 11.4** Calculer la probabilité qu'un individu qui fait du jogging d'avoir plus de 30 ans (dans cet échantillon).

**Q 11.5** Ces deux variables aléatoires semblent-elles indépendantes ?

**Définition : Coefficient de corrélation linéaire**

Soit  $X, Y$  deux variables. Le coefficient de corrélation linéaire entre  $X$  et  $Y$  est :

$$r = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y},$$

où  $\sigma_X$  et  $\sigma_Y$  sont les écart-types des variables  $X$  et  $Y$ .

**Exercice 12 – Automobiles**

L'expert d'un journal a noté sur une échelle de 0 à 10 la tenue de route (variable  $X$ ) et le confort (variable  $Y$ ) des modèles automobiles proposés par 4 constructeurs dans la même catégorie (« familiales ») et à des prix voisins :

Marque	$X$	$Y$
Hyundai	7	4
FIAT	5	7
Peugeot	8	9
Renault	9	8

Calculer le coefficient de corrélation linéaire  $r$  entre  $X$  et  $Y$ .

**Exercice 13 – Poids vs Taille**

Des chercheurs américains ont calculé le coefficient de corrélation linéaire  $r$  entre la taille  $X$  et le poids  $Y$  d'un échantillon de  $n = 1000$  hommes américains et trouvé  $r = 0.87$ . Ils ont mesuré les tailles en pouces et les poids en livres. Qu'auraient-ils trouvé s'ils avaient utilisé le système métrique (cm et kg) ?

**Exercice 14 – Notes d'examen**

Les étudiants d'un même groupe de TD ont obtenu chacun à une UE une note de partiel (sur 50) et une note globale (sur 100, qui intègre note de CC, de partiel et d'examen) ; les données sont les suivantes :

N°Etudiant	<i>partiel</i>	<i>global</i>
1	45	92
2	23	86
3	50	97
4	46	95
5	33	87
6	21	76
7	13	72
8	30	84
9	34	85
10	50	98

**Q 14.1** Calculer le coefficient de corrélation linéaire  $r$  entre  $X$  (note de partiel) et  $Y$  (note globale).

**Q 14.2** Représenter graphiquement le nuage de points correspondant aux données. Est-ce que la relation entre les variables semble être linéaire? Y a-t-il des points aberrants?

**Q 14.3** On a oublié les notes d'un 11<sup>ème</sup> étudiant qui sont : 40 au partiel et 70 de note globale. Si on les incorpore, quel sera l'effet sur  $r$ ?

### Exercice 15 – Indépendances conditionnelles

La loi de probabilité jointe de 3 variables aléatoires  $X$ ,  $Y$  et  $Z$ , est donnée par le tableau suivant dans lequel, par exemple, la case  $1/12$  représente la probabilité  $P(X = x_2, Y = y_1, Z = z_1)$  :

		$Y = y_1$	$Y = y_2$	$Y = y_3$
$Z = z_1$	$X = x_1$	$1/24$	$1/15$	$1/8$
	$X = x_2$	$1/12$	$7/120$	$1/8$
		$Y = y_1$	$Y = y_2$	$Y = y_3$
$Z = z_2$	$X = x_1$	$3/40$	$1/20$	$13/120$
	$X = x_2$	$1/20$	$3/40$	$17/120$

On note respectivement «  $X \perp\!\!\!\perp Y$  » et «  $X \perp\!\!\!\perp Y | Z$  » l'indépendance probabiliste entre  $X$  et  $Y$ , et l'indépendance probabiliste entre  $X$  et  $Y$  conditionnellement à  $Z$ .

**Q 15.1** D'un point de vue probabiliste, a-t-on  $X \perp\!\!\!\perp Y$ ,  $X \perp\!\!\!\perp Z$ ,  $Z \perp\!\!\!\perp Y$ ? Rappel : si  $A$  et  $B$  sont indépendants,  $P(A, B) = P(A) \times P(B)$ .

**Q 15.2** A-t-on  $X \perp\!\!\!\perp Y | Z$ ,  $X \perp\!\!\!\perp Z | Y$ ,  $Z \perp\!\!\!\perp Y | X$ ?

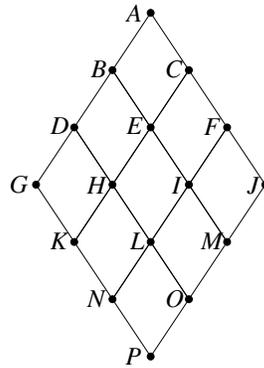
### Exercice 16 – Robotique

Un robot doit se rendre du point  $A$  au point  $P$  en passant par les arêtes du graphe ci-dessous. Le robot est limité dans ses mouvements, aussi ne peut-il que descendre (par exemple, lorsqu'il est en  $E$ , il ne peut aller qu'en  $H$  ou en  $I$ , mais pas en  $B$ ). Lorsqu'il est sur un nœud du graphe, il peut descendre soit sur l'arête de gauche, soit sur celle de droite. Son programme lui fait choisir 7 fois sur 10 l'arête de gauche et 3 fois sur 10 celle de droite.

**Q 16.1** Calculez la probabilité que le robot passe en  $B$  pour aller vers  $P$ . Faites de même avec  $C$ . Soit  $X_1$  la variable aléatoire de modalités  $\{B, C\}$  représentant le « point de passage du robot sur le niveau en dessous de  $A$  ». Quel est le type de distribution de probabilité suivie par  $X_1$  (binomiale, Poisson, normale, etc)? Quels sont les paramètres de cette loi?

**Q 16.2** Notez, pour chaque chemin menant à  $D$ , le nombre de fois où le robot a été à gauche ou à droite. Faites de même avec  $E$  et  $F$ . Déduisez-en la probabilité que le robot passe en  $D$ ,  $E$  et  $F$  pour aller vers  $P$ .

**Q 16.3** Calculez la probabilité que le robot passe en  $G$  pour aller vers  $P$ . Faites de même avec  $H$ ,  $I$ , et enfin



*J.*

**Q 16.4** Soit  $X$  une variable aléatoire valant 0 si le robot est passé en  $G$ , 1 s'il est passé en  $H$ , 2 en  $I$  et 3 en  $J$ . Quelle est la loi de probabilité suivie par  $X$ ? Justifiez votre réponse.

**Exercice 17 – Indépendances**

Soit trois variables aléatoires  $X, Y, Z$ , ayant respectivement pour domaines  $\{x_1, x_2\}$ ,  $\{y_1, y_2\}$  et  $\{z_1, z_2\}$ . On a pu déterminer la probabilité jointe  $P(X, Y, Z)$  de ces 3 variables :

	$y_1$		$y_2$	
	$x_1$	$x_2$	$x_1$	$x_2$
$z_1$	0,060	0,140	0,060	0,140
$z_2$	0,192	0,048	0,288	0,072

Montrez que  $X$  est indépendante de  $Y$  conditionnellement à  $Z$ .

## Introduction à GNU Octave 3

### POUR TOUS LES TPs. Structure des programmes octave en rfidec :

- Vous créez un répertoire par TME contenant toutes les ressources du TME. Ce répertoire sera compressé puis
  - Vos programmes comporteront toujours un fichier `main.m` regroupant vos calculs pour un tp donné. La console ne doit être utilisée que pour vérifier les valeurs des variables courantes. Pour appeler le programme sur lequel vous travaillez, appeler `main` (sans le `.m`) depuis la console octave dans le répertoire courant.
  - Pour exécuter Octave, vous taperez la commande `/opt/local/bin/octave` dans une console.
  - Afin de palier les problèmes d'installation et les problèmes de persistance, vous démarrerez tous vos programmes par les commandes suivantes :
 

```
clear all; % effacer toutes les variables
close all; % fermer toutes les fenêtres
putenv('GNUTERM','x11'); % bon terminal pour converser avec octave
```
  - Lorsque vous cherchez le fonctionnement d'une fonction, utilisez la commande `help + nomFonction`
  - `%` ou `#` correspondent à un commentaire sur la ligne
- Cet énoncé comporte beaucoup de documentation et peu de réelles questions. Votre compréhension du sujet sera évaluée sur les questions [18.2](#), [19.2](#), [19.3](#), [19.4.2](#), [20.2](#), [21.2](#), [22.1](#), [23.1](#).

### Exercice 18 – Gestion des entrées/sorties, déclaration des variables

Les variables sont créées automatiquement lors de leur déclaration. Toutes les variables sont de type double et toutes les variables sont des matrices (un scalaire est une matrice 1x1).

**Q 18.1** Création de variables, les approches classiques Taper les lignes de code suivantes et visualiser les résultats. Pour afficher une variable, il suffit de taper son nom sans point virgule.

#### Q 18.1.1 Création directe

```
a = 2; % matrice 1x1 contenant la valeur 2.0
m0 = [1 2; 3 4]; % espace (ou virgule) -> valeur suivante
                    % point-virgule -> ligne suivante
v1 = 1:10; % [1 2 3 4 5 6 7 8 9 10]
v2 = 1:2:10; % de 1 à 10 par pas de 2 : [1 3 5 7 9]
v3 = (1:10)'; % ' donne la transposée et nous créons donc un vecteur colonne
```

Les `[]` permettent de définir une matrice, l'opérateur `:` est une boucle implicite.

#### Q 18.1.2 Création de matrices spéciales à l'aide de fonctions dédiées Tester les fonctions ci-dessous.

```
m1 = ones(10,2); % matrice de 1, args = dimensions
m2 = zeros(5,4); % matrice de 0, args = dimensions
m3 = eye(4); % matrice identité carrée, arg = dimension
m4 = rand(5,6); % matrice de nombres aléatoires indépendants, args = dimensions
m5 = randn(5,6); % tirages selon une gaussienne(mu=0,var=1), args = dimensions
```

**Q 18.1.3** Chargement/sauvegarde de matrices de valeurs

Les fonctions `save` et `load` d'octave permettent de sauver et charger des matrices de données.

Saisir dans un fichier texte `college.dat` les deux colonnes de nombres suivantes (vous utiliserez `TextEdit`, `nedit`, `gedit`, `emacs` ...). Ces nombres correspondent à des notes (2 épreuves) d'élèves (15) sur lesquelles nous travaillerons par la suite.

Utiliser la commande `notes = load("college.dat");` puis vérifier le contenu de la variable `notes`.

Tenter ensuite de sauvegarder la matrice `notes` dans le fichier `college_save.dat` en utilisant la commande `save` (aide de octave : `help save`)

```
14.5  8.5
15.5  8.5
9      14.5
9.5    15.5
11     9.5
3.5    6
11.5   11
8.5    5.5
3       2
17     12
6       13
10     12.5
10      4
11.5   5.5
13.5   8
```

**Q 18.1.4** Etat de la mémoire

```
long = length(v1);      % acces à la longueur de v1
dims = size(m0);       % renvoie un vecteur n lignes, n colonnes : [2, 2]
nlines = size(m0,1);
ncolones = size(m0,2);
```

La commande `who -variables` permet de connaître les variables existantes dans la mémoire courante. Taper cette commande directement dans la console et vérifier que toutes les variables précédentes sont bien présentes. NB : `whos` permet en plus de connaître la taille des matrices créées.

**Q 18.2** Récapitulatif

Nous souhaitons créer une matrice 10x3 dont la première colonne contient les indices 1 à 10 dans l'ordre. La seconde colonne contiendra des nombres aléatoires entre 0 et 1. La troisième colonne ne contiendra que des 0.

Vous ajouterez ensuite une ligne en haut de la matrice contenant les indices de colonne 1 à 3.

NB : vous pouvez créer des matrices dans des matrices, c'est-à-dire faire appel à des fonctions dans les `[]`.

```
1.00000  2.00000  3.00000
1.00000  0.03479  0.00000
2.00000  0.66074  0.00000
3.00000  0.15187  0.00000
4.00000  0.03640  0.00000
5.00000  0.62497  0.00000
6.00000  0.54774  0.00000
7.00000  0.68919  0.00000
8.00000  0.86146  0.00000
9.00000  0.72030  0.00000
10.00000 0.84590  0.00000
```

**Exercice 19 – Fonctions de base et boucles****Q 19.1** Syntaxe**Q 19.1.1** Opérations sur les matrices

ATTENTION : TOUS LES INDICES DE MATRICES DEMARRENT A 1

Sur une matrice, vous disposez des opérateurs suivants :

```
% soient les variables définies dans la question précédente
v1(1) = 3;      % affectation
a = m0(2,3);   % accès en lecture
v1 = v1+ 3;    % ou v1 += 3 % matrice + scalaire
              % changement sur les toutes les valeurs de v1
              % NB: le - fonctionne pareil
m1tr = m1';    % transposée
m1 = ones(10,1)*[1 2 3]; % Attention, produit matriciel
```

```

m2 = ones(10,3)*2;      % multiplication par un scalaire
m3 = m1 .* m2;         % multiplication terme à terme
                        % NB: il existe également ./ (division) et .^ (puissance)

```

### Q 19.1.2 Modification de valeur et sous-matrices

```

v1(1:5) = v1(1:5)+3;   % changement sur les 5 premières valeurs de v1
v2(end-1:end) += 20;   % changement des 2 dernières valeurs de v2
m1(1:3,1:2) = 1000;    % affectation sur un bloc
m3(2,:)               % retourne la 2ème ligne, toutes les colonnes
m3(:,3)               % retourne la 3ème colonne (toutes les lignes)
m4([1 3 6], 1)        % retourne les 3 cases (1,1), (3,1) et (6,1)

```

### Q 19.1.3 Opérations de base proposées par Octave (sur scalaires et matrices)

```

min(a,b) % renvoie la valeur min (terme à terme pour des matrices a et b)
max(a,b)
ceil(a)  % arrondi au dessus
floor(a) % arrondi au dessous
sort(a)  % renvoie un tableau trié (chaque colonne est triée)
sort(a,2) % renvoie un tableau trié (chaque ligne est triée)
mean(a)  % moyenne de chaque colonne de a
mean(a,2) % moyenne de chaque ligne de a
sum(a)   % somme de chaque colonne de a
sum(a,2) % somme de chaque ligne de a
std(a)   % écart-type de chaque colonne de a
mod(a,b) % renvoie a modulo b (reste de la division entière)

```

## Q 19.2 Application 1

Calculer la moyenne et l'écart-type des deux épreuves de la question [18.1.3](#).

## Q 19.3 Application 2

Nous souhaitons générer aléatoirement les notes de la question [18.1.3](#) sachant que :

- Le nombre d'élèves est  $n = 15$
- Les notes sont tirées selon une loi gaussienne d'écart-type 4. Utiliser la commande `randn` pour générer les tirages puis multiplier par l'écart-type.
- La première épreuve à une moyenne de 10, la seconde, une moyenne de 8 (on veut donc décaler la première colonne de 10 et la seconde de 8)
- On veut être sûr que les notes sont supérieures ou égales à 0 (utiliser `max`)
- On veut être sûr que les notes sont inférieures ou égales à 20 (utiliser `min`)
- On veut des notes entières (utiliser `round`)

## Q 19.4 Boucles et conditions

### Q 19.4.1 La syntaxe est la suivante :

```

for i=1:10      % pour i allant de 1 à 10
  for j=1:10    % pour j allant de 1 à 10
    maMatrice(i,j) = i+j;
  endfor
endfor

```

NB : le pas de la boucle peut être réglé comme suit : `for i=1:pas:100`. Il existe des boucles `while` et `do-until` si besoin.

La syntaxe du IF est la suivante :

```

if i<10
    i = i + 10;
else
    ...
endif

```

**Q 19.4.2** Application des boucles et conditionnelles

Générer une matrice M 10x10 avec les caractéristiques suivantes (i désigne les lignes et j les colonnes) :

- si (i+j) pair alors M(i,j) pair (tirage aléatoire)
- si (i+j) impair alors M(i,j) = i+j

**Exercice 20 – Fonctions d’affichage**

**Q 20.1** Premier contact : tracé d’un vecteur y par rapport à un vecteur x. Les deux vecteurs doivent faire la même taille.

```

x = 1:0.1:10; y = x.^2; % chaque terme de x au carré
figure                % création d’une nouvelle fenêtre
plot(x,y);           % affichage des x en abscisse, y en ordo.
plot(x,y,"r+");     % modification de style (cf help plot)
clf;                 % clear
figureplot(x,y);
hold on;             % pour tracer plusieurs courbes sur la
                    % même figure

y2 = 3*x+1;
plot(x,y2,"r");

```

La fonction bar permet de tracer des histogrammes

```

bar([2 4 6 5]);      % affichage histogramme
bar([1 2 4 5], [2 4 6 5]); % les barres sont centrées sur les abscisses précisées

```

**Q 20.2** Application

Tracer la fonction  $f(x) = x^2 + x - 1$  pour x entre -5 et 10. La fonction sera tracée en vert et le style sera une succession d’étoiles.

**Exercice 21 – Fonctions de recherche avancées**

**Q 21.1** Syntaxe La fonction find permet de trouver tous les éléments d’une matrice qui satisfont une certaine clause.

```

v = 1:2:10;          % 1  3  5  7  9
index = find(v>4);  % 3  4  5
v(index)            % affichage des valeurs  5  7  9
length(v)           % affiche 3 : il y a 3 valeurs supérieures à 4 dans v

```

**Q 21.2** Application

Soient les notes de la question 18.1.3, à l’aide d’une boucle for et d’un find, trouver le nombre d’élèves qui ont des notes dans :  $[0, 2[$ , ...  $[18, 20]$ .

Tracer l’histogramme associé.

Nous proposons ensuite de regrouper les notes sur des ensembles de tailles différentes :  $[0, 5[$ ,  $[5, 8[$ ,  $[8, 10[$ ,  $[10, 12[$ ,  $[12, 15[$  et  $[15, 20]$ . Tracer l’histogramme en respectant bien sûr la normalisation des hauteurs.

---

**Exercice 22 – Aménagement du code en fonctions**


---

Afin de ne pas trop encombrer le main, vous développerez des fonctions à raison d'une fonction par fichier (fonction et fichier portant le même nom). Exemple : fichier `maFonction.m` :

```
% documentation de la fonction
%
function [retour1, retour2] = maFonction(arg1, arg2, arg3)
    retour1 = mean(arg1);
    retour2 = mean(arg2)+mean(arg3);
endfunction
```

Usage de `maFonction` dans le fichier `main.m` :

```
...
[moy1, moy2] = maFonction(x, y, z);
...
```

Tous les fichiers de codes auront un nom en `.m`, utiliser une extension `.dat` pour les fichiers de données.

**Q 22.1 Application**

Ecrire une fonction qui prend en argument le tableau de notes et retourne les deux moyennes et les deux écarts-types correspondant aux deux épreuves. Cette fonction sera appelée depuis votre `main.m` et les résultats seront affichés à l'aide de la commande `fprintf` (cf help dans octave) qui permet de mettre en forme du texte dans la console.

---

**Exercice 23 – Debuggage**


---

**Q 23.1** Dimensions, opérations matricielles et bugs usuels Corriger les expressions suivantes (elles sont toutes défaillantes) :

```
a = rand(10,4); b = [1 2 3 4]; c = a*b;
```

Nous souhaitons construire une loi discrète dont les valeurs  $p(X = n)$  sont stockées dans le vecteur  $p$ . Nous proposons la stratégie suivante : tirer  $n$  valeurs aléatoires puis normaliser  $p$  de sorte que  $\sum_{i=1}^n p_i = 1$ .

```
n=10;      p = rand(n,1);      p = p/mean(p);
```

Soit une loi discrète jointe dont les valeurs  $p(X = i, Y = j)$  sont stockées dans la matrice  $p$ . Nous souhaitons calculer les  $p(X = i) = \sum_j p(X = i, Y = j)$ .

```
pX = sum(p)
```

Soit une matrice  $A$  de dimensions  $(d1, d2)$  et un vecteur ligne  $B$  de dimension  $d2$ . Nous voulons multiplier chaque colonne  $j$  de  $A$  par la valeur  $B_j$  sans utiliser de boucles.

```
C = A * B
```

Soit une matrice  $A$  quelconque, nous voulons la transformer de sorte que chaque ligne somme à 1 :

```
A = A ./ (ones(1, size(A,1))*sum(A,2))
```

**Q 23.2** Outils de débbugage de GNU Octave

La console est très utile pour le débbugage : elle donne accès à toutes les variables courantes. Cependant, dans les fonctions que vous développez, l'environnement est local et les bugs sont plus durs à identifier. La fonction `keyboard` vous permet de stopper l'exécution du programme et d'avoir accès à l'environnement local de la fonction.

Insérer la commande `keyboard` dans la fonction de la question 22.1 et vérifier que vous avez accès à l'environnement local.

Une alternative intéressante consiste à taper la commande

```
debug_on_error(true)
```

qui active automatiquement le débbugeur en cas d'erreur.

Les commandes utiles lors du débbugage sont : `dbcont` (continuer l'exécution du programme) et `quit`. Il existe des `dbstep` mais il est en général plus intéressant de copier du code dans la console.

**Exercice 24 – Tracés en 3D, isocontour et fonctionnalités avancées**

On souhaite dans cet exercice tracer des courbes en 3D  $f(x, y)$ . Vous allez donc créer une telle fonction dans la question 24.1, puis la dessiner dans la question 24.2.

**Q 24.1** On souhaite créer la fonction  $f(x, y) = x^2 + y^2$ , pour  $x \in [1, 3]$  et  $y \in [-1, 1]$ . Afin d'exploiter les vecteurs et matrices d'Octave, on discrétise cet espace en n'en sélectionnant que 5 valeurs par axe :

```
x = [ 1.0000  1.5000  2.0000  2.5000  3.0000 ]; % définir l'axe des x
y = [ -1.00000 -0.50000  0.00000  0.50000  1.00000]; % définir l'axe des y
```

Pour tous  $x, y$ , on peut aisément définir les fonctions  $g(x, y) = x$  et  $h(x, y) = y$  à l'aide d'Octave :

```
[xgrid, ygrid] = meshgrid(x, y); % création des combinaisons
```

```
xgrid =          ygrid =
 1.0000  1.5000  2.0000  2.5000  3.0000  -1.00000 -1.00000 -1.00000 -1.00000 -1.00000
 1.0000  1.5000  2.0000  2.5000  3.0000  -0.50000 -0.50000 -0.50000 -0.50000 -0.50000
 1.0000  1.5000  2.0000  2.5000  3.0000   0.00000  0.00000  0.00000  0.00000  0.00000
 1.0000  1.5000  2.0000  2.5000  3.0000   0.50000  0.50000  0.50000  0.50000  0.50000
 1.0000  1.5000  2.0000  2.5000  3.0000   1.00000  1.00000  1.00000  1.00000  1.00000
```

Pour construire la matrice correspondant à la fonction  $k(x, y) = x^2$ , il suffit alors de taper `xgrid .* xgrid`. Par conséquent, pour construire  $f$ , il suffit de taper :

```
f = xgrid .* xgrid + ygrid .* ygrid;
```

**Q 24.2**

Pour pouvoir tracer  $f$  en 3D, la variable `f` doit avoir les mêmes dimensions que `xgrid` et `ygrid`.

```
surf(xgrid, ygrid, f); % affiche la surface f
contour(xgrid, ygrid, f); % affiche les courbes de niveau de f
```

**Exercice 25 – Soumission de TME**

Créez une archive contenant les fichiers que vous avez manipulés tout au long du TME. Pour soumettre cette archive à votre chargé de TD/TME, il suffit de remplir le formulaire de l'URL :

<http://www-desir.lip6.fr/~gonzales/teaching/rfidec/soumission/soumission.php>

### 3 Échantillonnage et intervalles de confiance

#### Exercice 26 – Estimation de la variance d’une loi de probabilité

Soit  $X = (X_1, X_2, \dots, X_k, \dots, X_n)$  l’échantillon i.i.d. empirique tiré de  $X_0$ , variable dont l’espérance  $E(X_0) = m$  et la variance  $V(X_0) = \sigma^2$  sont deux paramètres inconnus ; on note  $\theta = (m, \sigma^2)$  le paramètre bi-dimensionnel.

**Q 26.1** Quelle est l’espérance  $E_\theta(\bar{X})$  de la moyenne empirique  $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$  ?

**Q 26.2** Quelle est sa variance  $V_\theta(\bar{X})$  ?

**Q 26.3** En déduire que  $\bar{X}$  est un estimateur sans biais et convergent de  $m$ , c’est-à-dire que :

$$E_\theta(\bar{X}) = m \text{ et } \lim_{n \rightarrow \infty} E_\theta[(\bar{X}_n - m)^2] = 0.$$

[*rappel* : lorsque  $n$  varie, on écrit  $\bar{X}_n$  au lieu de  $\bar{X}$ ].

**Q 26.4** Montrer que  $\frac{1}{n} E_\theta[\sum_{k=1}^n (X_k - m)^2] = \sigma^2$ .

**Q 26.5** Pourquoi ne peut-on pas prendre  $\frac{1}{n} \sum_{k=1}^n (X_k - m)^2$  comme estimateur de  $\sigma^2$  ?

**Q 26.6** On considère la statistique  $Y = \sum_{k=1}^n (X_k - \bar{X})^2$ . En utilisant la décomposition  $X_k - \bar{X} = (X_k - m) - (\bar{X} - m)$ , montrer que  $Y = \sum_{k=1}^n (X_k - m)^2 - n(\bar{X} - m)^2$  puis que  $E_\theta(Y) = (n-1)\sigma^2$ .

**Q 26.7** En déduire que la *variance empirique corrigée*  $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$  est un estimateur sans biais de  $\sigma^2$ .

#### Exercice 27 – Confiance en la bière

On vérifie si le volume moyen de bière versée dans les bouteilles par une machine respecte bien les standards de la compagnie. L’écart-type du volume de liquide versé par la machine est de 3 ml. On prélève un échantillon au hasard de 100 bouteilles provenant de la production de cette machine et on obtient une moyenne de 337 ml. Construisez l’intervalle de confiance de niveau de confiance 95% pour le volume moyen de bière versée dans les bouteilles.

#### Exercice 28 – Estimation par intervalles de confiance

Chaque année, plusieurs milliers de candidats se présentent au concours d’entrée aux ENSI. Par expérience, on sait que chaque année la variance sur les notes obtenues par les candidats est (à peu près) de 16. Un correcteur a corrigé 100 copies dont la note moyenne (sur 20) est  $\bar{x} = 8.75$  et s’intéresse à la moyenne  $m$  de l’ensemble des copies du concours. Il suppose que les notes des copies suivent (approximativement) une loi normale  $\mathcal{N}(m; 16)$ .

[ **Notation** : pour  $U$  var. de loi  $\mathcal{N}(0; 1)$ ,  $u_\alpha$  est le  $\alpha$ -quantile, caractérisé par  $P(U > u_\alpha) = \alpha$  ]

**Q 28.1** Montrer que pour un échantillon empirique  $X$  de taille  $n = 100$  tiré de  $X_0$  de loi  $\mathcal{N}(m; 16)$ ,

$$P\left(\bar{X} - u_{\frac{\alpha}{2}} \cdot \frac{4}{10} \leq m \leq \bar{X} + u_{\frac{\alpha}{2}} \cdot \frac{4}{10}\right) = 1 - \alpha,$$

$\bar{X}$  étant la moyenne empirique.

**Q 28.2** En déduire l’expression d’un intervalle de confiance à  $100(1 - \alpha)\%$  pour  $m$ . Calculer les valeurs numériques des bornes d’intervalles à 95% et 99%.

#### Exercice 29 – Estimation par intervalles de confiance

[Les notations sont celles de l'exercice précédent]

On suppose l'échantillon  $X$  tiré de  $X_0$ , variable suivant une loi normale  $\mathcal{N}(m; \sigma^2)$  dont les deux paramètres sont inconnus.

On sait que, dans ce cas, la variable  $T = \frac{\bar{X} - m}{S/\sqrt{n}}$  suit une loi de STUDENT à  $n - 1$  degré de liberté.

[Notation : on note  $t_{n-1, \alpha}$  le  $\alpha$ -quantile, caractérisé par  $P(T > t_{n-1, \alpha}) = \alpha$ ]

**Q 29.1** Donner l'expression d'un intervalle de confiance à  $100(1 - \alpha)\%$  pour  $m$ .

**Q 29.2 – Application :** L'UPMC possède un parc important de stations de travail Sun. Un des ingénieurs système du LIP6 a sous sa responsabilité 49 stations. La fréquence des pannes de ses Sun est en moyenne de 54 mois, avec un écart-type de 8 mois. Le responsable matériel informatique de l'UPMC désire acheter du matériel informatique et lui demande quelle est la fréquence des pannes des Suns.

L'ingénieur système considère donc ses machines comme un échantillon de 49 individus et suppose que  $X$ , la fréquence des pannes de la population totale (l'ensemble des Sun dans le monde entier), suit une loi normale.

Quel intervalle de confiance à 95% peut-il annoncer au responsable ?

---

### Exercice 30 – Pointures

---

Dans cet exercice, on cherche à savoir quelle est la pointure de chaussure moyenne d'une population d'étudiants. Pour cela, une étude préalable a montré que l'écart-type sur les pointures de la population était de 3. On a par ailleurs extrait l'échantillon de pointures suivant :

41	39	43	42	42	45	43	41	40
----	----	----	----	----	----	----	----	----

Donnez une estimation de la moyenne  $\mu$  par intervalle de confiance de niveau de confiance 95%.

## TME : Corrélation, indépendances et $\chi^2$

### Exercice 31 – Corrélation et $\chi^2$

La taille d'un athlète peut jouer un rôle important dans ses résultats en saut en hauteur. Les données utilisées ici présentent la taille et la performance de 20 champions du monde.

NB : les données sont disponibles à l'adresse suivante : <http://www-connex.lip6.fr/~guigue/wikihomepage/uploads/Course/dataSautHaut.dat>.

Observation	Nom	Taille	Performance
1	Jacobs (EU)	1.73	2.32
2	Noji (EU)	1.73	2.31
3	Conway (EU)	1.83	2.40
4	Matei (Roum)	1.84	2.40
5	Austin (EU)	1.84	2.40
6	Ottey (Jam)	1.78	2.33
7	Smith (GB)	1.84	2.37
8	Carter (EU)	1.85	2.37
9	McCants (EU)	1.85	2.37
10	Sereda (URSS)	1.86	2.37
11	Grant (GB)	1.85	2.36
12	Paklin (URSS)	1.91	2.41
13	Annys (Bel)	1.87	2.36
14	Sotomayor (Cuba)	1.96	2.45
15	Sassimovitch (URSS)	1.88	2.36
16	Zhu Jianhua (Chine)	1.94	2.39
17	Brumel (URSS)	1.85	2.28
18	Sjoeberg (Suède)	2.00	2.42
19	Yatchenko (URSS)	1.94	2.35
20	Povarnitsine (URSS)	2.01	2.40

#### Q 31.1 – Calcul de corrélation :

**Q 31.1.1** Calculer la moyenne des deux variables Taille et Performance.

**Q 31.1.2** Calculer l'écart-type des deux variables Taille et Performance.

**Q 31.1.3** Écrire une fonction de calcul de taux de corrélation entre deux vecteurs. Pour cela, la fonction prendra en argument une matrice à 2 colonnes, celles-ci représentant les deux vecteurs. Appliquer cette fonction aux deux séries présentées ci-dessus.

NB : le but de cet exercice est aussi de vérifier que vous savez créer une nouvelle fonction (cf tutoriel).

**Q 31.1.4** Tracer les points dans un graphique Taille (abscisse) / Performance (ordonnée). Analyser grossièrement (à l'oeil) la linéarité des données.

#### Q 31.2 – Calcul de $\chi^2$ :

**Q 31.2.1** Écrire une fonction prenant en paramètre une matrice à deux colonnes représentant des valeurs de taille et de performance des athlètes ainsi que deux nombres  $N$  et  $M$ . Cette fonction retournera le tableau de contingence correspondant à cette matrice de données en discrétisant les tailles en  $N$  classes et les performances

en  $M$  classes (le tableau de contingence est donc de taille  $N \times M$ ). Appliquer cette fonction sur les données des athlètes en utilisant  $N = M = 3$ .

**Q 31.2.2** Écrire une fonction de calcul du  $\chi^2$  sur un tableau de contingence. Appliquer la au tableau précédent.

**Q 31.2.3** Refaire les calculs pour 4 classes pour la variable Taille et 2 classes pour la variable Performance.

### Exercice 32 – Probabilités jointes, marginales, conditionnelles et indépendance

**Rappel :** Soit 3 variables aléatoires  $X, Y$  et  $Z$ . On peut parler de :

- probabilités marginales  $P(X), P(Y), P(Z)$ ,
- probabilités jointes  $P(X, Y), P(Z, Y), P(X, Y, Z)$ ...
- probabilités conditionnelles  $P(X|Z), P(Y|Z, X)$ ...

**Données :** Soit un modèle ( $M$ ) constitué de 3 variables aléatoires  $X, Y$  et  $Z$  dont on connaît :

- la marginale de  $X$  :

$X$	$X = 0$	$X = 1$
$P(X)$	0.3	0.7

- la conditionnelle de  $Y$  sachant  $X$  :

$P(Y X)$	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	0.2	0.2	0.6
$X = 1$	0.4	0.5	0.1

- la conditionnelle de  $Z$  sachant  $X$  :

$P(Z X)$	$Z = 0$	$Z = 1$
$X = 0$	0.7	0.3
$X = 1$	0.4	0.6

**Q 32.1** Écrire une fonction prenant en paramètres un vecteur correspondant à une probabilité  $P(A)$  ainsi qu'une matrice correspondant à une probabilité  $P(B|A)$ , et retournant un vecteur correspondant à la probabilité  $P(B)$ . Utiliser cette fonction pour calculer les marginales  $P(Y)$  et  $P(Z)$  du modèle ( $M$ ).

NB : le code de la fonction est trivial. Si vous maîtrisez la syntaxe de création des fonctions, écrire directement tout le code dans le fichier main.m

**Q 32.2** Écrire une fonction prenant en paramètres un vecteur correspondant à une probabilité  $P(A)$  ainsi qu'une matrice correspondant à une probabilité  $P(B|A)$ , et retournant la probabilité  $P(A, B)$ . Calculer les probabilités  $P(X, Y)$  et  $P(X, Z)$  du modèle ( $M$ ).

**Q 32.3** Calculer  $P(Y, Z) = \sum_X P(X, Y, Z) = \sum_X P(X)P(Y|X)P(Z|X)$ .

NB : on fait l'hypothèse que  $Y$  et  $Z$  sont indépendantes conditionnellement à  $X$ .

**Q 32.4** Écrire une fonction qui, à partir de  $P(A), P(B)$  et  $P(A, B)$  teste l'indépendance exacte des variables aléatoires  $A$  et  $B$ .

NB : on vérifiera que la somme des différences absolues entre les deux calculs ne dépasse pas  $10^{-6}$ .

Tester l'indépendance des trois couples de variables  $(X, Y), (X, Z)$  et  $(Y, Z)$ .

### Exercice 33 – Intervalles de confiances en octave

Reprendre l'exercice précédent sur les pointures.

**Q 33.1** Tracer l'histogramme des pointures en utilisant la fonction `bar`.

**Q 33.2** Calculer l'intervalle de confiance à 95% autour de la moyenne  $\mu$  des pointures.

NB : la question est plutôt simple, elle vise à vous faire utiliser les fonctions existantes sur les lois usuelles : `normrnd`, `norminv`...

## 4 Tests d'hypothèses, d'ajustement et d'indépendance

### Exercice 34 – Test entre hypothèses simples

Parmi les personnes atteintes d'une certaine maladie, que l'on ne sait pas traiter, 36% guérissent spontanément, les 64% restant devenant des malades chroniques.

Un laboratoire pharmaceutique propose un remède très coûteux avec lequel, affirme-t-il, le pourcentage de guérison passe à 50%.

Un service hospitalier doute de l'efficacité de ce remède; pour le tester, il l'administre à un échantillon de 100 patients atteints de la maladie; les patients sont numérotés de  $k = 1$  à  $k = 100$ .

#### Q 34.1 – Mise en place du test d'hypothèse :

**Q 34.1.1** Quelles sont les hypothèses simples en présence ? (on appellera  $\theta$  le paramètre). Laquelle doit-on prendre comme hypothèse  $H_0$  ?

**Q 34.1.2** Au patient  $k$  est associée la variable  $X_k$  qui prend la valeur 1 si ce patient guérit et la valeur 0 sinon. Quel est le type de loi suivie par  $X_k$  dans les deux hypothèses  $H_0$  et  $H_1$  ?

Vérifier que les probabilités élémentaires des deux lois peuvent se mettre sous la forme :

$$P_\theta(X_k = x_k) = \theta^{x_k}(1 - \theta)^{1-x_k}, \quad x_k \in \{0, 1\}$$

avec  $\theta = \theta_0$  pour l'une,  $\theta = \theta_1$  pour l'autre.

**Q 34.1.3** En déduire l'expression de la vraisemblance [= la probabilité élémentaire de la loi de l'échantillon  $(X_1, \dots, X_k, \dots, X_n)$ ],

$$L(x, \theta) = L(x_1, \dots, x_k, \dots, x_n, \theta) = \prod_{k=1}^n P_\theta(X_k = x_k).$$

Montrer qu'elle s'exprime comme une fonction de la moyenne empirique  $\bar{x}$  et de  $\theta$ .

**Q 34.1.4** En déduire que, pour tout test du rapport de vraisemblance  $L(x, \theta_0)/L(x, \theta_1)$ , il existera un nombre positif  $\lambda$  tel que :

$$\begin{aligned} \bar{x} < \lambda &\Rightarrow \text{accepter } H_0 \\ \bar{x} > \lambda &\Rightarrow \text{rejeter } H_0. \end{aligned}$$

#### Q 34.2 – Réalisation effective du test :

**Q 34.2.1** Que représente la variable  $Y = n\bar{X}$  ? Quelle est la loi de  $Y$  dans l'hypothèse  $H_0$  ?

**Q 34.2.2** La table ci-dessous donne les probabilités exactes d'observer  $k$  guérisons au moins parmi les 100 malades sous l'hypothèse  $H_0$  (de  $k = 42$  à  $k = 50$ ); les valeurs obtenues par l'approximation normale sont données au-dessous.

$k$	42	43	44	45	46	47	48	49	50
$P_{\theta_0}(Y \geq k)$	0.126	0.089	0.060	0.040	0.025	0.015	0.009	0.0052	0.0029
<i>val. appr.</i>	0.125	0.089	0.059	0.038	0.023	0.014	0.008	0.0047	0.0025

Au niveau de signification  $\alpha = 0.05$ , quelles sont les valeurs de  $k$  pour lesquelles on doit : accepter l'hypothèse  $H_0$  ? rejeter  $H_0$  ? rejeter  $H_0$  avec une certaine probabilité ? (que l'on précisera).

**Q 34.2.3** Sous l'hypothèse  $H_1$ ,

$$\begin{aligned} P_{\theta_1}(Y \leq 43) &= 0.0967 \text{ et } P_{\theta_1}(Y = 44) = 0.039 \\ P_{\theta_1}(Y \leq 46) &= 0.242 \text{ et } P_{\theta_1}(Y = 47) = 0.066 \end{aligned}$$

En déduire la puissance du test de niveau de signification  $\alpha = 0.05$ .

**Q 34.2.4** Mêmes questions que précédemment mais cette fois au niveau de signification  $\alpha = 0.01$ .

**Q 34.2.5** L'approximation normale est-elle bonne ici ? Quel est le théorème de convergence qui laisse prévoir ce fait ?

**Q 34.3** On suppose que le chef du service hospitalier est capable :

– d'attribuer des probabilités *a priori* aux deux hypothèses

$$\pi_0 = P(H_0); \pi_1 = P(H_1) = 1 - \pi_0;$$

– et d'estimer le coût d'une erreur de 1<sup>ère</sup> espèce,  $C_0$  et de 2<sup>ème</sup> espèce  $C_1$ , ce qui permet de se placer dans le cadre de la statistique bayésienne.

Soit un test  $T_W$ , caractérisable par sa région critique  $W$ , c'est-à-dire tel que :

$$\begin{aligned} H_0 \text{ rejeté} &\Leftrightarrow x \in W \\ H_0 \text{ accepté} &\Leftrightarrow x \notin W \end{aligned}$$

**Q 34.3.1** Montrer que l'espérance mathématique du coût de ce test est :

$$\pi_0 C_0 \sum_{x \in W} L(x, \theta_0) + \pi_1 C_1 \sum_{x \notin W} L(x, \theta_1).$$

**Q 34.3.2** En déduire que, s'il existe  $x \in W$  tel que :

$$L(x, \theta_0) > \frac{\pi_1 C_1}{\pi_0 C_0} L(x, \theta_1),$$

alors il existe un autre test d'espérance de coût strictement inférieure à celle du test  $T_W$ .

**Q 34.3.3** Montrer de même que s'il existe  $x \notin W$  tel que

$$L(x, \theta_0) < \frac{\pi_1 C_1}{\pi_0 C_0} L(x, \theta_1),$$

alors il existe un autre test d'espérance de coût strictement inférieure à celle du test  $T_W$ .

**Q 34.3.4** En déduire qu'un test optimal bayésien est nécessairement un test du rapport de vraisemblance. Comment  $\lambda$  varie-t-il avec  $\pi_0$  et  $\pi_1$  d'une part et  $C_0$  et  $C_1$  d'autre part ?

Donner un test optimal bayésien lorsque  $\pi_0 = 0.95$ ,  $C_0 = 1$  et  $C_1 = 10$  (unités monétaires).

### Exercice 35 – Test d'indépendance du $\chi^2$

Un échantillon de 200 contribuables est prélevé afin de vérifier si le revenu brut annuel d'un individu est un caractère dépendant du niveau de scolarité de l'individu. Les observations recueillies sont données dans le tableau suivant :

scolarité (années) $\rightarrow$ revenu (kF) $\downarrow$	[0; 7[	[7; 12[	[12; 14[	[14; $\rightarrow$ [	total
[0; 75[	17	14	9	5	45
[75; 120[	12	37	11	5	65
[120; 200[	7	20	20	8	55
[200; $\rightarrow$ [	4	9	10	12	35
total	40	80	50	30	200

**Q 35.1** On admet que les fréquences relatives déduites des marges du tableau donnent les vraies lois de probabilité,  $p_r$  et  $p_s$  des variables  $R(\text{evenue})$  et  $S(\text{colorité})$ . Donner le tableau des fréquences théoriques,  $200 \times p_{rs}$ , correspondantes en cas d'indépendance des deux variables.

**Q 35.2** Calculer le  $\chi^2$ . Expliquez pourquoi il y a 9 degrés de liberté. Doit-on rejeter l'hypothèse d'indépendance au risque  $\alpha = 5\%$  ?

### Exercice 36 – Test d'ajustement du $\chi^2$

Dans un supermarché, on maintient 8 caisses de plus de 10 articles en opération durant les nocturnes du jeudi. Normalement, la clientèle devrait se répartir uniformément entre les caisses. Afin de vérifier cela, on a recensé le nombre de clients passés à chacune des caisses un jeudi soir. Les résultats observés ont été les suivants :

Numéro de la caisse	Nombre de clients
1	72
2	70
3	71
4	52
5	45
6	59
7	67
8	48
Total	484

Hypothèse  $H_0$  à tester : « la clientèle se répartit uniformément entre les 8 caisses ».

**Q 36.1** Sous l'hypothèse  $H_0$ , quels sont les effectifs théoriques  $\nu_i$  dans chaque classe ?

**Q 36.2** Calculer la statistique d'ajustement :

$$A = \sum_{i=1}^I \frac{(n_i - \nu_i)^2}{\nu_i}.$$

**Q 36.3** Quel est le nombre de degrés de liberté ? Au niveau de signification  $\alpha = 0.05$ , doit-on accepter  $H_0$  ?

### Exercice 37 – Boules de couleur

Soit une urne contenant des boules de 5 couleurs différentes : (R)ouges, (B)leues, (V)ertes, (J)aunes, (N)oirs. On suspecte que la distribution de probabilité sur les couleurs des boules de l'urne est la suivante :

$$P(R) = 0,2 \quad P(B) = 0,4 \quad P(V) = 0,1 \quad P(J) = 0,2 \quad P(N) = 0,1.$$

Par ailleurs, on a tiré un échantillon i.i.d. de 20 boules et on a noté le nombre de boules de chaque couleur :

Couleur	R	B	V	J	N
Nb boules	2	9	4	5	0

Faites un test d'ajustement avec un niveau de confiance  $1 - \alpha = 90\%$  pour déterminer si, oui ou non, la distribution de probabilité sur les couleurs des boules est celle indiquée ci-dessus.

### Exercice 38 – Notation en RFIDEC

On sait, par expérience, que les notes de partiel de RFIDEC suivent une loi normale  $\mathcal{N}(\mu; 6^2)$ . On considère

l'échantillon de notes i.i.d. suivant : 

10	8	13	20	12	14	9	7	15
----	---	----	----	----	----	---	---	----

 .

**Q 38.1** Estimez la moyenne de la population des notes de partiel de RFIDEC grâce à un intervalle de confiance de niveau de confiance  $1 - \alpha = 95\%$ .

**Q 38.2** Par expérience, les années précédentes, la moyenne au partiel de RFIDEC était égale à 14. Dressez un test d'hypothèse de niveau de confiance  $1 - \alpha = 95\%$  pour confronter les hypothèses  $H_0 = \ll$  la moyenne est égale à 14  $\gg$  et  $H_1 = \ll$  la moyenne a baissé, i.e., elle est inférieure à 14  $\gg$ .

**Q 38.3** Calculez la puissance du test pour une moyenne de 12 ( $H_1$  : la moyenne est égale à 12).

---

### Exercice 39 – Il faut assurer

---

La loi oblige tout automobiliste à contracter une assurance. La prime exigée annuellement d'un assuré dépend de plusieurs facteurs : la zone habitée, le type de véhicule, l'utilisation à des fins commerciales ou non, la distance estimée que parcourra l'assuré. . . Il est presque impossible d'estimer la distance parcourue par un automobiliste pour une année donnée. Voilà pourquoi tous les assurés d'un véhicule non utilisé à des fins commerciales se voient imposer le même montant sur ce point. Celui-ci est fonction de la distance moyenne parcourue annuellement par les automobilistes de cette catégorie. Des études ont montré par le passé que celle-ci était de 18000 km avec un écart-type de 5000 km. Le montant que l'on prévoit d'exiger est de 2 centimes du km, autrement dit  $18000 \times 0,02 = 360\text{€}$ . Le montant de cette prime a continuellement augmenté ces dernières années, de telle sorte que l'opinion publique commence à être très mécontente et à exercer de fortes pressions sur les compagnies d'assurances pour qu'elles baissent leurs tarifs.

C'est ainsi que la MAIF est priée de réévaluer tous les facteurs considérés dans le calcul de la prime. Le plus vulnérable de ces facteurs est précisément la distance parcourue annuellement. Un statisticien est donc chargé de réexaminer le bien-fondé de l'estimation à 18000 km de la moyenne contestée. La démarche qu'il compte suivre est de prélever rapidement un échantillon de 400 individus afin de tester si la moyenne a effectivement diminué. Si tel est le cas, une étude plus exhaustive, menée sur un grand nombre d'assurés, sera entreprise afin d'estimer très précisément la valeur de la moyenne. Sinon, ce facteur ne sera pas révisé.

La variable qu'étudie le statisticien est  $X$  : la distance parcourue en 2011 (dernière année complète sur laquelle on peut fonder l'étude), par un véhicule utilisé à des fins non commerciales. Il décide pour l'instant de ne pas remettre en cause l'estimation de l'écart-type  $\sigma$  de  $X$  (5000 km). En revanche, il veut réestimer la moyenne  $\mu$  et vous demande donc de l'aider :

**Q 39.1** Dans un test d'hypothèses, quelles hypothèses  $H_0, H_1$  formulerez-vous pour tester s'il faut revoir les tarifs de l'assurance à la baisse? Quelle serait la forme de la région critique?

**Q 39.2** Le statisticien s'interroge sur les conséquences qu'aurait le fait de rejeter  $H_0$  alors que celle-ci est vraie. Cela entraînerait la réalisation de l'étude exhaustive pour rien, donc une dépense inutile, et porterait atteinte à la réputation du statisticien. Il décide donc de ne pas prendre de risque et de fixer la probabilité de commettre une telle erreur à  $\alpha = 0,01$ . Exprimez  $\alpha$  en fonction de la région critique.

**Q 39.3** Quelle est la loi suivie par  $\bar{X}$  sous  $H_0$ ?

**Q 39.4** À partir de quelle valeur le test nous indique-t-il de rejeter  $H_0$ ?

**Q 39.5** Notre statisticien veut maintenant examiner la puissance de son test afin de voir si sa règle de décision est « solide ». Il réfléchit alors sur les conséquences de rejeter  $H_1$  alors que  $H_1$  est vraie. Si une telle erreur se produisait, les automobilistes n'obtiendraient pas une réduction du prix de la prime alors qu'il y aurait droit. Si la diminution à laquelle ils avaient droit se chiffrait à 20 euros ou moins, on ne pourrait pas parler de conséquences sérieuses. à quel nombre moyen  $k$  de kilomètres parcourus correspond une baisse de 20 euros de l'assurance?

**Q 39.6** Calculez la puissance du test pour  $k$ . Cette puissance nous indique la probabilité que la règle de décision soit « fiable » pour une moyenne de  $k$  kilomètres, c'est-à-dire lorsque les assurés devraient commencer à percevoir une différence au niveau du prix de leur assurance. Est-ce que le statisticien peut procéder au recueil des données auprès des 400 personnes ou bien son test d'hypothèses n'est-il pas sûr?

## TME : Génération de données

---

### Exercice 40 – Génération de données

---

**Q 40.1** Écrire une procédure générant un échantillon de taille  $N$  pour une variable  $A$  dont on donne la loi sous la forme d'un vecteur (comme dans l'exercice 32 par exemple).

**Q 40.2** Écrire une procédure pour calculer la probabilité estimée d'une variable  $A$  dont on connaît un échantillon  $N$ .

**Q 40.3** Afin de tester les méthodes précédentes, reprenons les variables  $X$  et  $Y$  de l'exercice 32. Générer 1000 échantillons de chacune des variables. Calculer les probabilités estimées de  $X$  et de  $Y$  sur cet échantillon. Comparer avec les lois marginales théoriques.

**Q 40.4** On utilise la procédure de la question 40.1 pour générer un échantillon `DATA` des variables  $(X, Y, Z)$  : un échantillon est la combinaison du tirage de  $X$ , de  $Y$  et de  $Z$  (variables de l'exercice 32) suivant la procédure précédente.

**Q 40.4.1** Peut-on retrouver les marginales (modulo les erreurs dues à l'échantillonnage) ?

**Q 40.4.2** Est-ce que `DATA` représente la loi jointe de  $X$ ,  $Y$  et  $Z$  ? Comment le tester ?

**Q 40.5** En remarquant que  $P(A|B)$  peut se lire «  $A$  suit la loi  $P(A|B = 1)$  quand  $B = 1$  et  $A$  suit la loi  $P(A|B = 0)$  quand  $B = 0$  », écrire une procédure calculant `DATA2`, un échantillon de taille 1000 représentant la loi jointe de  $X$ ,  $Y$  et  $Z$ .

**Q 40.6** Écrire une fonction qui, à partir d'un échantillon de la loi jointe  $(A, B)$  teste l'indépendance de  $A$  et de  $B$  ; ou plus précisément, fournit la valeur du  $\chi^2$  ainsi que le degré de liberté du système.

NB : vous ré-utiliserez la fonction que vous avez développée précédemment. Vous pourrez aussi tester le fonctionnement des méthodes incluses dans octave : `chi2***`

**Q 40.7** Tester l'indépendance des trois couples de variables  $(X, Y)$ ,  $(X, Z)$  et  $(Y, Z)$ .

## 5 Maximum de vraisemblance et maximum *a posteriori*

### Exercice 41 – MAP

Soit  $X$  une variable aléatoire définie sur l'ensemble des nombres entiers positifs.  $X$  suit la loi géométrique de paramètre  $p \in [0, 1]$  si  $P(X = n) = (1 - p)^{n-1}p$ . On a observé 5 réalisations (obtenues indépendamment les unes des autres) d'une variable  $X$  suivant la loi géométrique : 

4	2	6	5	8
---	---	---	---	---

.

**Q 41.1** Estimez par maximum de vraisemblance la valeur du paramètre  $\theta = p$  de la loi.

**Q 41.2** Avant le tirage de l'échantillon, nous avons une connaissance a priori sur le paramètre  $\theta$  : ce dernier suivait a priori une loi Beta de paramètres 4 et 5, autrement dit  $\pi(\theta) \propto \theta^3(1 - \theta)^4$ . Estimez la valeur du paramètre  $\theta = p$  par maximum a posteriori.

### Exercice 42 – MAP 2

Soit  $X$  une variable aléatoire suivant la loi binomiale  $\mathcal{B}(K, \theta)$ , où  $K$  est une constante supposée connue. On observe un échantillon  $\{x_1, \dots, x_n\}$  de taille  $n$  d'instanciations de cette variable aléatoire.

**Q 42.1** Calculez la valeur de  $\theta$  par maximum de vraisemblance. Bien entendu, vous démontrerez mathématiquement votre résultat.

**Q 42.2** Des études statistiques nous indiquent que  $\theta$  suit une loi Beta *a priori*  $\pi(\theta) = \text{Beta}(\theta, a, b)$ . Quelle est la valeur a posteriori de  $\theta$ ? Justifiez mathématiquement votre réponse.

### Exercice 43 – Maximum a posteriori, maximum de vraisemblance

Une pièce de monnaie peut être plus ou moins biaisée en faveur de *Pile* ou de *Face*. On prend pour paramètre  $\theta$  la probabilité de *Pile* :

$$P_\theta(\text{Pile}) = \theta.$$

L'ensemble des valeurs possibles pour  $\theta$  est  $\Theta = \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}\}$ ; les probabilités a priori  $\pi(\theta)$  de la v.a.  $\tilde{\theta}$  sont :

$\theta$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$
$\pi(\theta)$	0.1	0.2	0.4	0.2	0.1

On effectue 5 lancers indépendants de la pièce et observe le nombre  $x$  de résultats *Pile* obtenu; la v.a.  $X$  a donc pour valeurs possibles  $\mathcal{X} = \{0, 1, 2, 3, 4, 5\}$ .

**Q 43.1** Quelle est la loi suivie par  $X$  conditionnellement à l'hypothèse  $\tilde{\theta} = \theta$ ? Calculer tous les éléments du tableau des probabilités conditionnelles  $p(x/\theta)$ ,  $(x, \theta) \in \mathcal{X} \times \Theta$ . (on pourra se servir d'une table et mettre à profit les symétries des données)

**Q 43.2** Dédurre de la question précédente les valeurs des éléments du tableau des probabilités jointes  $\pi(x, \theta)$ ,  $(x, \theta) \in \mathcal{X} \times \Theta$ . À partir de ce tableau, comment peut-on retrouver la loi a priori  $\{\pi(\theta)\}$  de la v.a.  $\tilde{\theta}$ ? comment trouve-t-on la loi a priori de  $X$ ? Calculez-la.

**Q 43.3** Dédurre de ce qui précède les valeurs des éléments du tableau des probabilités a posteriori  $\pi(\theta/x)$ ,  $(x, \theta) \in \mathcal{X} \times \Theta$ .

**Q 43.4** Donner les valeurs d'acceptation des diverses hypothèses sur la valeur du paramètre :

**Q 43.4.1** quand la règle de décision est celle de la *probabilité d'erreur minimum*;

**Q 43.4.2** quand la règle de décision est celle du *maximum de vraisemblance*.

**Q 43.4.3** Quand ces deux règles donnent-elles le même résultat?

## TME : roulette

---

### Exercice 44 – Roulette

---

Est il possible de gagner à la roulette à coup sûr :

- en théorie ?
- en pratique ?

Nous souhaitons construire un système pour gagner à coup sur, en pratique ! Pour cela, nous introduisons une hypothèse : le tirage obtenu dépend de trois facteurs.

- la vitesse de rotation de la table
- la vitesse de lancé de la bille par le croupier
- la position de la table au moment du lancé de la bille

Deux photos successives très rapides sont prises discrètement et permettent d'estimer ces trois caractéristiques. Plusieurs complices ont pratiqué des relevés et les résultats sont stockés dans le fichier `tirageRoulette2010.dat`<sup>1</sup>. Pour simplifier les traitements ultérieurs, les vitesses ont été discrétisées en 3 et 5 classes (la vitesse normale est au milieu et nous retrouvons moins fréquemment des vitesses plus lentes ou plus rapides). De même, le problème de prévision d'un nombre étant trop complexe, nous jouerons à Passe et Manque (la mise porte sur  $> 18$  ou  $\leq 18$ ), le 0 correspond toujours à une perte.

Le fichier est constitué de 5000 tirages présentés par ligne  $X = [VRR, VL, CL, F]$  avec :

- VRR : vitesse de rotation de la roulette  $\sim \mathcal{B}(2, 0.5)$
- VL : vitesse de lancé de la bille  $\sim \mathcal{B}(4, 0.5)$
- CL : chiffre en face de la bille au moment du lancé  $\mathcal{U}(36)$
- F : tirage : 0=0, 1=manque, 2=passe

**Q 44.1** Estimer l'espérance de gain correspondant à une mise aléatoire.

NB : vous simulerez une mise aléatoire de pour chacun des 5000 exemples et calculerez le gain moyen.

**Q 44.2** En se basant sur les 1000 premières observations, construire la table  $p(VRR, VL, CL|F)$ .

NB : Vous utiliserez soit une matrice à 4 dimensions pour stocker les résultats soit plusieurs matrices 3D en différenciant les cas  $F=0, 1, 2$ . Parcourir ensuite la liste des réalisations pour remplir la ou les matrices.

**Q 44.3** Peut-on se baser sur cette première table pour construire un modèle de prédiction ?

- Construire un tel modèle
- Appliquer ce modèle sur les données proposées
- Calculer le gain obtenu sur les 1000 échantillons d'apprentissage
- Calculer le gain obtenu sur les données non-utilisées en apprentissage

**Q 44.4** La table construite précédemment est très grande... Que pensez vous du nombre de la taille de cette table en fonction du nombre de données dont vous disposez ?

- Pour réduire la taille de cette table, nous proposons de réduire la taille de CL en 6 classes.
- Calculer le gain obtenu sur les 1000 échantillons d'apprentissage
- Calculer le gain obtenu sur les données non-utilisées en apprentissage

**Q 44.5** Afin d'améliorer les performances, nous voulons introduire les connaissances a priori dont nous disposons sur VRR, VL, CL et F.

- Introduire le calcul du maximum a posteriori dans la boucle for précédente en utilisant le théorème de Bayes.  
NB : la méthode `pVRR = binopdf([0 1 2], 2, 0.5)` permet de calculer le tableau de la loi de probabilité de la binomiale  $\mathcal{B}(2, 0.5)$
- Calculer le gain obtenu sur les 1000 échantillons d'apprentissage
- Calculer le gain obtenu sur les données non-utilisées en apprentissage

**Q 44.6** Construction d'un modèle de roulette

---

1. téléchargeable sur le site de l'UE

Essayons maintenant de générer des données en construisant un modèle de roulette dans la méthode `tirage.m`.

**Q 44.6.1** Pourquoi avoir choisi la loi binomiale pour modéliser les vitesses de rotation et de lancer de la bille ?

**Q 44.6.2** Génération des données selon plusieurs lois

- générer  $N$  données  $VRR \sim \mathcal{B}(2, 0.5)$  en utilisant `binornd`
- générer  $N$  données  $VL \sim \mathcal{B}(4, 0.5)$  en utilisant `binornd`
- générer  $N$  données  $CL$

Utiliser le code suivant pour calculer la sortie :

```
vitesseRotation = [10:0.2:10+(nVRR)*1]'; % vitesse VRR
nbTour = (2:0.2:2+(nVL))'; % nb tour de la bille suite
                    % a la vitesse de lancé

precisionMesureVRR = 0.01;
precisionVitesseVL = 0.005;
precisionCL= 1;

nbTours = (vitesseRotation(VRR)+randn(n,1)*precisionMesureVRR*3) .* ...
           (nbTour(VL)+randn(n,1)*precisionVitesseVL*3);
chiffre = (nbTours-floor(nbTours))*36;
F = round(CL+randn(n,1)*precisionCL*3 + chiffre);

F = mod(F, 37);
```

**Q 44.7** Vérifier que les tirages sont bien uniformes à l'aide d'un test du  $\chi^2$  (avec une signification de 5%)

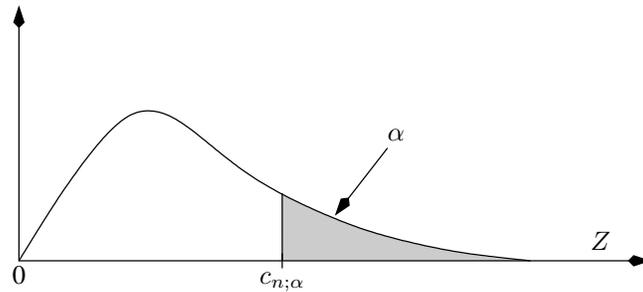
NB1 : la fonction `[yy,xx]=hist(t, categories)` permet pour une variable discrète unidimensionnelle  $t$  de compter les occurrences de toutes les categories. Le nombre d'occurrences est stocké dans `yy`

NB2 : la fonction `chi2inv(population,DL)` remplace votre table du  $\chi^2$ . Pour connaître la valeur limite à un niveau de signification 0.05 dans un problème à DL degrés de liberté, il faut utiliser : `chi2inv(0.95,DL)`



Table de la loi du  $\chi^2$ 

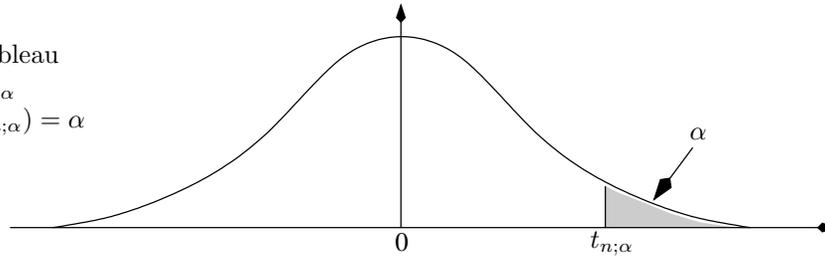
valeurs dans le tableau  
 ci-dessous : les  $c_{n;\alpha}$   
 tels que  $P(Z > c_{n;\alpha}) = \alpha$



$n \setminus \alpha$	0,995	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	0,0000393	0,000157	0,000982	0,00393	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	40,3	43,8	47,0	50,9	53,7

## Table de la loi de Student

valeurs dans le tableau  
 ci-dessous : les  $t_{n;\alpha}$   
 tels que  $P(Z > t_{n;\alpha}) = \alpha$



$n \setminus \alpha$	0,10	0,05	0,025	0,01	0,005	0,001	$n \setminus \alpha$	0,10	0,05	0,025	0,01	0,005	0,001
1	3,078	6,314	12,706	31,821	63,657	318,309	41	1,303	1,683	2,020	2,421	2,701	3,301
2	1,886	2,920	4,303	6,965	9,925	22,327	42	1,302	1,682	2,018	2,418	2,698	3,296
3	1,638	2,353	3,182	4,541	5,841	10,215	43	1,302	1,681	2,017	2,416	2,695	3,291
4	1,533	2,132	2,776	3,747	4,604	7,173	44	1,301	1,680	2,015	2,414	2,692	3,286
5	1,476	2,015	2,571	3,365	4,032	5,893	45	1,301	1,679	2,014	2,412	2,690	3,281
6	1,440	1,943	2,447	3,143	3,707	5,208	46	1,300	1,679	2,013	2,410	2,687	3,277
7	1,415	1,895	2,365	2,998	3,499	4,785	47	1,300	1,678	2,012	2,408	2,685	3,273
8	1,397	1,860	2,306	2,896	3,355	4,501	48	1,299	1,677	2,011	2,407	2,682	3,269
9	1,383	1,833	2,262	2,821	3,250	4,297	49	1,299	1,677	2,010	2,405	2,680	3,265
10	1,372	1,812	2,228	2,764	3,169	4,144	50	1,299	1,676	2,009	2,403	2,678	3,261
11	1,363	1,796	2,201	2,718	3,106	4,025	51	1,298	1,675	2,008	2,402	2,676	3,258
12	1,356	1,782	2,179	2,681	3,055	3,930	52	1,298	1,675	2,007	2,400	2,674	3,255
13	1,350	1,771	2,160	2,650	3,012	3,852	53	1,298	1,674	2,006	2,399	2,672	3,251
14	1,345	1,761	2,145	2,624	2,977	3,787	54	1,297	1,674	2,005	2,397	2,670	3,248
15	1,341	1,753	2,131	2,602	2,947	3,733	55	1,297	1,673	2,004	2,396	2,668	3,245
16	1,337	1,746	2,120	2,583	2,921	3,686	56	1,297	1,673	2,003	2,395	2,667	3,242
17	1,333	1,740	2,110	2,567	2,898	3,646	57	1,297	1,672	2,002	2,394	2,665	3,239
18	1,330	1,734	2,101	2,552	2,878	3,610	58	1,296	1,672	2,002	2,392	2,663	3,237
19	1,328	1,729	2,093	2,539	2,861	3,579	59	1,296	1,671	2,001	2,391	2,662	3,234
20	1,325	1,725	2,086	2,528	2,845	3,552	60	1,296	1,671	2,000	2,390	2,660	3,232
21	1,323	1,721	2,080	2,518	2,831	3,527	61	1,296	1,670	2,000	2,389	2,659	3,229
22	1,321	1,717	2,074	2,508	2,819	3,505	62	1,295	1,670	1,999	2,388	2,657	3,227
23	1,319	1,714	2,069	2,500	2,807	3,485	63	1,295	1,669	1,998	2,387	2,656	3,225
24	1,318	1,711	2,064	2,492	2,797	3,467	64	1,295	1,669	1,998	2,386	2,655	3,223
25	1,316	1,708	2,060	2,485	2,787	3,450	65	1,295	1,669	1,997	2,385	2,654	3,220
26	1,315	1,706	2,056	2,479	2,779	3,435	66	1,295	1,668	1,997	2,384	2,652	3,218
27	1,314	1,703	2,052	2,473	2,771	3,421	67	1,294	1,668	1,996	2,383	2,651	3,216
28	1,313	1,701	2,048	2,467	2,763	3,408	68	1,294	1,668	1,995	2,382	2,650	3,214
29	1,311	1,699	2,045	2,462	2,756	3,396	69	1,294	1,667	1,995	2,382	2,649	3,213
30	1,310	1,697	2,042	2,457	2,750	3,385	70	1,294	1,667	1,994	2,381	2,648	3,211
31	1,309	1,696	2,040	2,453	2,744	3,375	71	1,294	1,667	1,994	2,380	2,647	3,209
32	1,309	1,694	2,037	2,449	2,738	3,365	72	1,293	1,666	1,993	2,379	2,646	3,207
33	1,308	1,692	2,035	2,445	2,733	3,356	73	1,293	1,666	1,993	2,379	2,645	3,206
34	1,307	1,691	2,032	2,441	2,728	3,348	74	1,293	1,666	1,993	2,378	2,644	3,204
35	1,306	1,690	2,030	2,438	2,724	3,340	75	1,293	1,665	1,992	2,377	2,643	3,202
36	1,306	1,688	2,028	2,434	2,719	3,333	$\infty$	1,282	1,645	1,960	2,326	2,576	3,090
37	1,305	1,687	2,026	2,431	2,715	3,326							
38	1,304	1,686	2,024	2,429	2,712	3,319							
39	1,304	1,685	2,023	2,426	2,708	3,313							
40	1,303	1,684	2,021	2,423	2,704	3,307							