Université Grenoble Alpes **Année 2020-2021**

UE MAP201 (L1)

Examen de première session

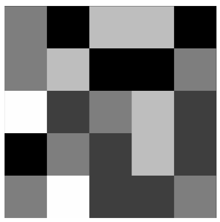
Une calculatrice et une feuille manuscrite A4 sont autorisées.
La note de l'examen sera la moyenne sur 20 des thèmes Image et Statistiques.

Durée : 2h

Thème image : 20 points

Exercice 1 (Images et histogrammes)

1) On considère une image de taille 5×5 avec 5 niveaux de gris, de 0 (noir) à 4 (blanc) :



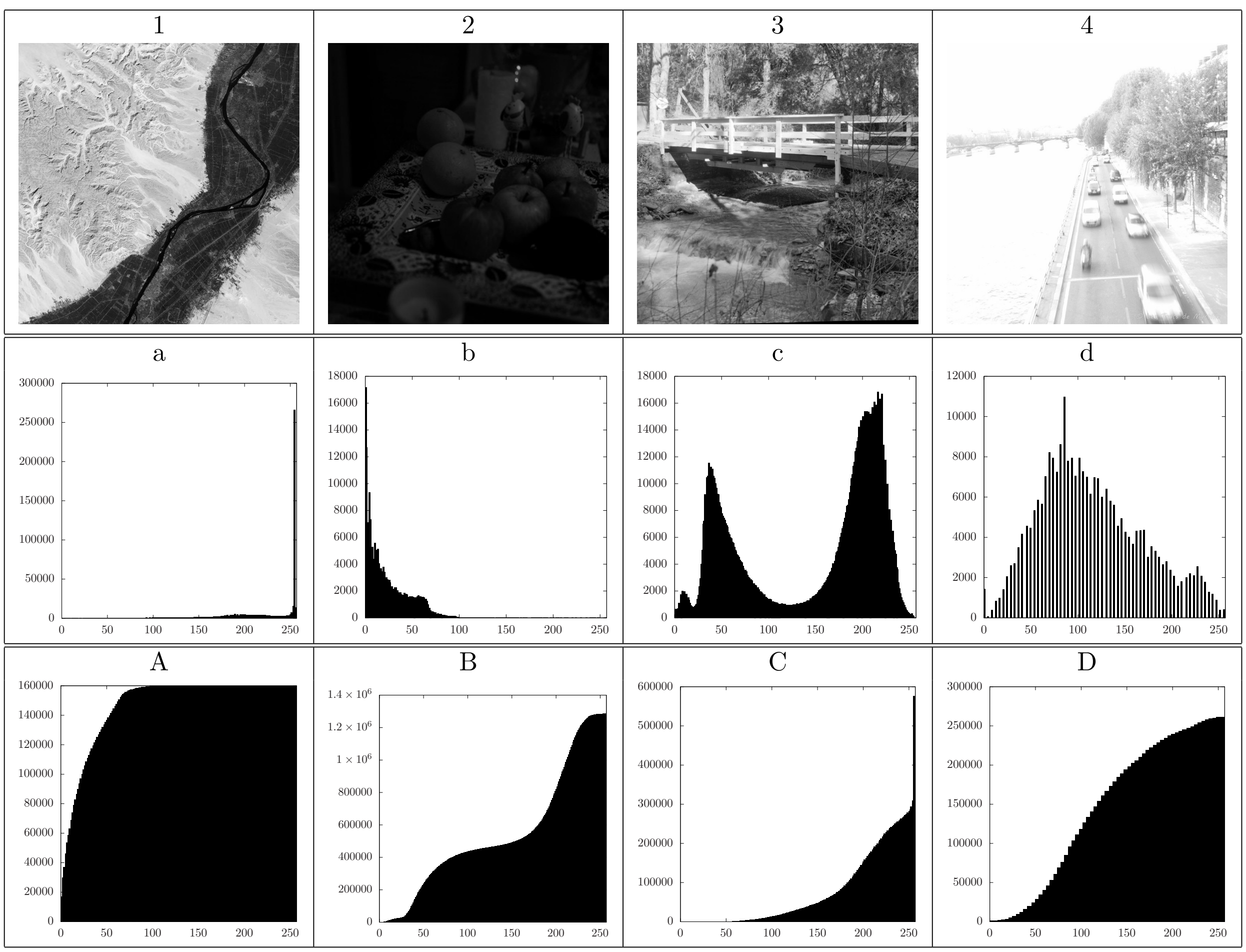
```
im  =
   2.   0.   3.   3.   0.
   2.   3.   0.   0.   2.
   4.   1.   2.   3.   1.
   0.   2.   1.   3.   1.
   2.   4.   1.   1.   2.
```

— Tracez l'histogramme de cette image. Notez bien que les valeurs des pixels vont de 0 à 4.

— Tracez son histogramme cumulé.

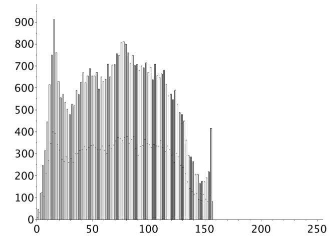
2) On considère maintenant 4 images à 256 niveaux de gris, ainsi que 4 histogrammes et 4 histogrammes cumulés. Déterminez l'histogramme et l'histogramme cumulé de chacune de ces images.

D'après les éléments dont vous disposez, quelle est l'image dont la taille (en nombre de pixels) est la plus grande ?



Exercice 2 (Histogramme et correction d'image)

On s'intéresse à une image I_1 en niveaux de gris codée sur 256 niveaux de gris, dont l'histogramme est donné ci-contre.



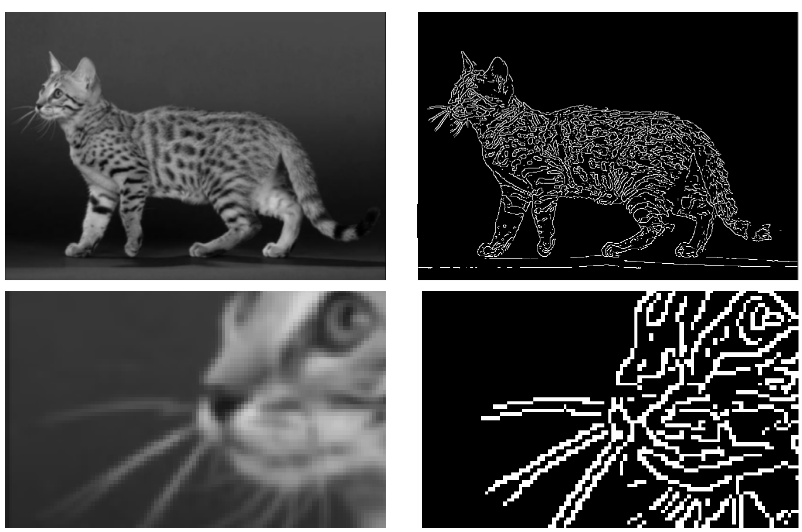
1. Comment qualifier la luminosité et le contraste de cette image ?
2. Afin d'améliorer la qualité de l'image on souhaite appliquer à ses pixels une transformation $T(p)$; l'image transformée sera alors $I_2(u,v) = T(I_1(u,v))$ pour tous les pixels (u,v) de l'image. On cherche un transformation linéaire, de la forme $T(p) = ap + b$, où a et b sont des constantes. A partir de l'histogramme, déterminez des valeurs de a et b qui améliorent le contraste.
3. Donner l'allure de l'histogramme de la nouvelle image.

Exercice 3 (Les moustaches du chat) On rappelle le principe de l'algorithme de détection de contours à partir du gradient d'une image $I(x,y)$, tel que vu en TP :

— Calcul du gradient de l'image $G(x,y) = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right)$ et de sa norme $\|G(x,y)\| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}$.

— Les contours de l'image sont définis comme les maxima locaux de la norme du gradient.

Voici un exemple du résultat obtenu avec cette méthode, avec en bas un zoom sur les moustaches du chat :



On peut remarquer que les moustaches sont dédoublées sur l'image du contour ; plus généralement, c'est le cas pour les traits (par opposition aux bords des objets pleins). L'objectif de cet exercice est de comprendre pourquoi. Pour simplifier l'analyse, on s'intéresse à une image I qui contient un trait noir vertical sur fond blanc, codée en niveaux de gris sur 8 bits :

```
I  =
    255   255   255   255   255   0   255   255   255   255   255
    255   255   255   255   255   0   255   255   255   255   255
    255   255   255   255   255   0   255   255   255   255   255
    255   255   255   255   255   0   255   255   255   255   255
    255   255   255   255   255   0   255   255   255   255   255
```

Les valeurs étant constantes sur chaque colonne, la dérivée en y est nulle et on ne s'intéressera qu'aux dérivées en x.

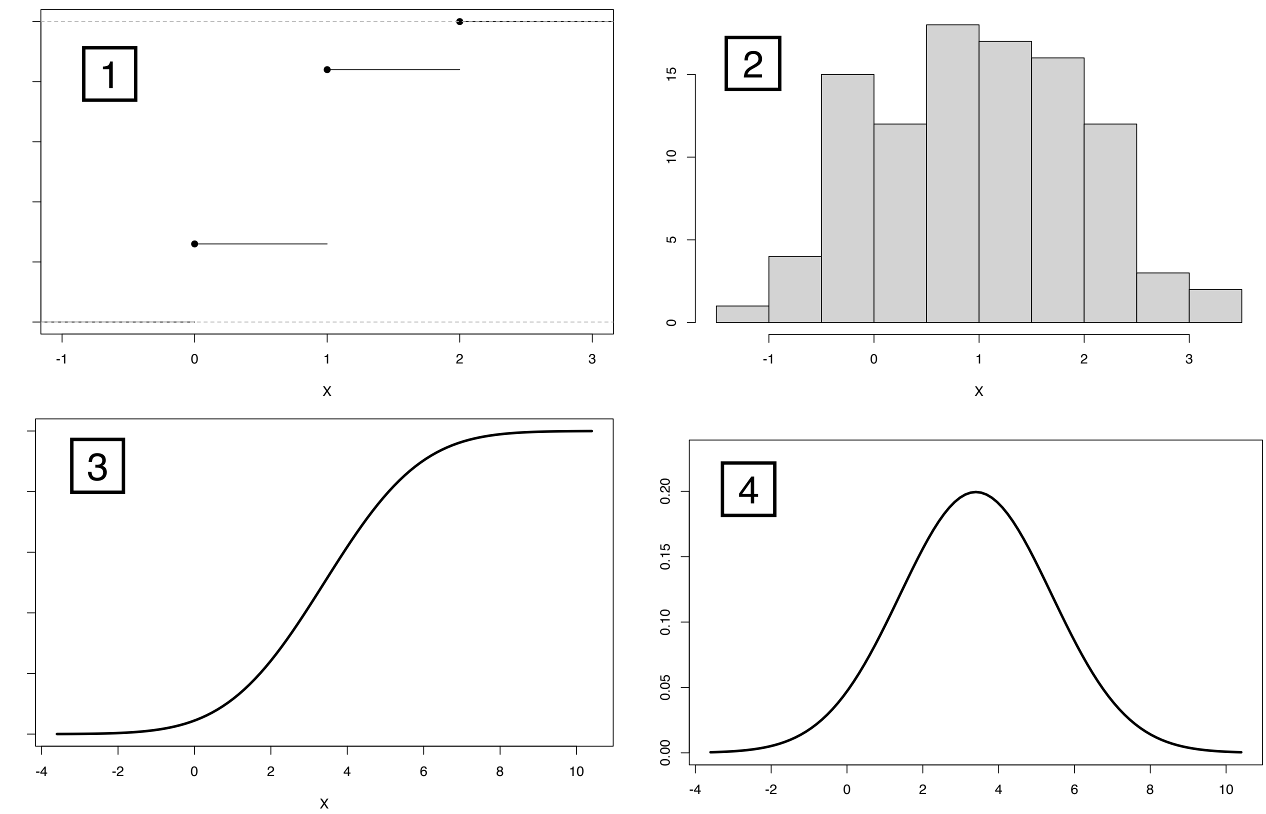
1. On donne la formule des différences finies *à droite* : soit f une fonction dérivable, alors $f'(x) \approx \frac{f(x+h)-f(x)}{h}$, pour $h \in \mathbb{R}$ suffisamment petit ; et, de manière similaire, la formule des différences finies *à gauche* : $f'(x) \approx \frac{f(x)-f(x-h)}{h}$. En déduire la formule des différences finies *centrée* :
$$f'(x) \approx \frac{f(x+h)-f(x-h)}{2h}.$$
2. On considère le filtre $D_x = 0.5[-1, 0, 1]$. Expliquer pourquoi l'application de ce filtre à une image correspond à une approximation de la dérivée en x selon la formule des différences finies centrée.

   Rappel : Etant donné un filtre horizontal $W = [w_1, w_2, w_3]$ et une image $I_1(u, v)$ où u est l'indice de ligne v l'indice de colonne, l'application du filtre W produit une nouvelle image I_2 telle que
$$I_2(u, v) = \sum_{j=-1}^{1} I_1(u, v + j) w_{j+2}$$
   pour les indices (u, v) correspondant à des pixels suffisamment loin du bord de l'image.
3. Appliquer le filtre D_x à l'image I pour calculer une approximation I_x de sa dérivée suivant x. Comme toutes les lignes sont identiques, vous n'indiquerez dans votre réponse qu'une ligne de I_x (et de même pour les images des questions suivantes). Pour le traitement des pixels du bord de l'image, on fait le choix de leur donner la valeur nulle : ainsi la première et la dernière colonne de I_x seront nulles et on n'appliquera la formule précédente que sur les autres colonnes.
4. Calculer l'image I_N correspondant à la norme du gradient de l'image I. L'image I_C contenant les contours est définie de la manière suivante :
$$I_C(u, v) = \begin{cases} 255 & \text{si } I_N(u, v) \text{ est un maximum local,} \\ 0 & \text{sinon.} \end{cases}$$
   Calculer I_C et conclure sur l'efficacité de la méthode pour détecter des traits.

Thème Statistique

Exercice 1 (Questions de cours (3 points))**.** Pour chacune des représentations graphiques ci-dessous, donner les informations suivantes : nom du graphe, dire si c'est un graphe empirique et/ou théorique, dire si cela correspond à une variable discète et/ou continue.

De plus, pour les graphiques 1 et 3, donner les valeurs minimales et maximales de l'axe des ordonnées.



Exercice 2 (Table de distribution (5 points))**.** On s'intéresse à la variable "Utilisation de l'ordinateur" (en nombre de jours de la semaine). Après sondage auprès d'un échantillon, nous avons recueilli les données brutes stockées dans l'objet `X` dans le logiciel `R`.
*Les questions **(A.N.)** se résolvent à l'aide des valeurs numériques données en fin d'exercice.*

1. Est-ce que la variable d'intérêt est discrète ou continue ?
2. Est-ce que la variable d'intérêt est quantitative ou qualitative ?
3. Donner le nom et la formule d'un estimateur sans biais de la moyenne. Donner une estimation sans biais de la moyenne **(A.N.)**.
4. Donner le nom et la formule d'un estimateur sans biais de la variance. Donner une estimation sans biais de la variance **(A.N.)**.

   Les données sont résumées dans le tableau suivant :

utilisation	0	1	2	3	4	5	6	7
effectif	2	2	4	6	8	8	12	8

5. Quelle est la taille de l'échantillon ?
6. Compléter le tableau de la feuille réponse.
7. Quels sont les trois quartiles de cet échantillon ?

Valeurs numériques.

```
> mean(X)
[1] 4.56
```

```
> median(X)
[1] 5
```

```
> var(X)
[1] 3.76
```

```
> (50/49)*var(X)
[1] 3.84
```

```
> (49/50)*var(X)
[1] 3.69
```

```
> sd(X)
[1] 1.94
```

Exercice 3 (Réduction de consommation ? (8 points))**.** On s'intéresse à la variable "Consommation d'essence" (en litre par 100 km). Nous avons recueilli des données auprès de $n = 25$ conducteurs et elles sont stockées dans l'objet `X1` dans le logiciel `R`. On suppose dans la suite que cette variable suit une loi gaussienne, de moyenne et variance inconnues.

Les questions ***(A.N.)*** *se résolvent à l'aide des valeurs numériques données en fin d'exercice.*

1. Donner la formule de l'intervalle de confiance que vous utiliseriez pour estimer la consommation moyenne d'essence au niveau de confiance $1 - \alpha$. Rappeler le nom ou la définition de toutes les quantités intervenant dans l'intervalle.

2. Donner l'intervalle de confiance à 95% calculé sur l'échantillon **(A.N.)**.

   L'entreprise *Echo-l'eau* a inventé un nouveau produit qui permettrait de réduire la consommation d'essence. Nous proposons à nos 25 conducteurs de refaire un tour après application du produit. Les données recueillies sur ce deuxième échantillon sont stockées dans l'objet `X2` dans le logiciel `R`. On suppose encore que cette variable est gaussienne.

3. Donner l'intervalle de confiance à 95% calculé sur l'échantillon **(A.N.)**.

4. Comparer les intervalles des questions 2 et 3.
   C'est une question ouverte qui admet plusieurs bonnes réponses. Sentez-vous libre de dire ce qui vous passe par la tête en lien avec le cours (sans toutefois en écrire un roman d'une page)

   Le statisticien de l'entreprise souhaite démontrer l'efficacité de ce nouveau produit. Il propose de faire un test statistique à partir des deux échantillons `X1` et `X2` décrits ci-dessus.

5. Quelles hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ le statisticien doit-il poser ?

6. Quel est le test que vous utiliseriez (en précisant la statistique et la zone de rejet) ?

7. Est-ce qu'il faut rajouter une hypothèse de modélisation pour être dans le cadre de validité du test que vous avez choisi ? Si oui, laquelle ?

8. Donner la p-valeur de votre test **(A.N.)**. Est-ce que le nouveau produit est efficace ?

Valeurs numériques.

```
> t.test(X1)
t = 67, df = 24, p-value <2e-16
95 percent confidence interval:
4.31 4.58
```

```
> t.test(X1, alternative = "less")
t = 67, df = 24, p-value = 1
95 percent confidence interval:
-Inf 4.55
```

```
> t.test(X1, alternative = "greater")
t = 67, df = 24, p-value <2e-16
95 percent confidence interval:
4.33 Inf
```

```
> t.test(X2)
t = 60, df = 24, p-value <2e-16
95 percent confidence interval:
4.17 4.47
```

```
> t.test(X2, alternative = "less")
t = 60, df = 24, p-value = 1
95 percent confidence interval:
-Inf 4.44
```

```
> t.test(X2, alternative = "greater")
t = 60, df = 24, p-value <2e-16
95 percent confidence interval:
4.2 Inf
```

```
> t.test(X1,X2)
t = 1, df = 48, p-value = 0.2
95 percent confidence interval:
-0.0754 0.3167
```

```
> t.test(X1,X2, alternative = "less")
t = 60, df = 24, p-value <2e-16
95 percent confidence interval:
-Inf 0.284
```

```
> t.test(X1,X2, alternative = "greater")
t = 1, df = 48, p-value = 0.1
95 percent confidence interval:
-0.0429 Inf
```

```
> t.test(X1-X2)
t = 3, df = 24, p-value = 0.007
95 percent confidence interval:
0.0365 0.2048
```

```
> t.test(X1-X2, alternative = "less")
t = 3, df = 24, p-value = 1
95 percent confidence interval:
-Inf 0.19
```

```
> t.test(X1-X2, alternative = "greater")
t = 3, df = 24, p-value = 0.003
95 percent confidence interval:
0.0509 Inf
```

```
> chisq.test(table(X1,X2))
X-squared = 600, df = 576, p-value = 0.2
```

Exercice 4 (Âge et jeux vidéos (4 points))**.** On s'intéresse aux variables “Notation du jeu vidéo *Mass Effect*” (note entre 0 et 5) et “Tranche d'âge” (plus ou moins de 25 ans). Nous avons recueilli des données auprès de $n = 400$ joueurs en leur demandant leur âge et de noter le jeu.

Les données sont présentées dans la table de contingence suivante :

Âge \ Note	0	1	2	3	4	5
moins de 25 ans	10	20	20	30	40	80
plus de 25 ans	10	10	10	20	50	100

1. Compléter la table de la feuille réponse.
2. Donner les deux distributions conditionnelles des notes selon l'âge (exprimer les fréquences en pourcentage).
3. Comparer les deux distributions de la question 2 entre elles. Pensez-vous qu'il y a un lien entre l'âge et la note attribuée au jeu ?
4. Quel est le nom du test que vous utiliseriez pour répondre plus précisément à la question précédente ? Quelles sont les hypothèse $\mathcal{H}_0$ et $\mathcal{H}_1$?
5. La p-valeur du test vaut `0.035`. Est-ce qu'il y a un lien significatif entre l'âge et la note attribuée au jeu ?

Feuille réponse

Exercice 2.

utilisation	0	1	2	3	4	5	6	7
effectif	2	2	4	6	8	8	12	8
effectif cumulé								
fréquence (%)								
fréquence cumulée (%)								

Exercice 4.

Âge \ Note	0	1	2	3	4	5	
moins de 25 ans	10	20	20	30	40	80	
plus de 25 ans	10	10	10	20	50	100	