UNIVERSITÉ DE SHERBROOKE

Faculté de génie

Département de génie électrique et génie informatique

Rapport APP5

Probabilités, Statistiques et Simulations Monte Carlo

Présenté à

Monsieur Eric Plourde

Présenté par

Pascal-Emmanuel Lachance – LACP3102

Anthony St-Laurent Cyr – STLA0801

Sherbrooke 4 avril 2023

Table des matières

[1. Analyse descriptive des données 1](#_Toc131534675)

[1.1. Tableau 1](#_Toc131534676)

[1.1.1. Tableau des caractéristiques 1](#_Toc131534677)

[1.1.2. Règle de Sturges 2](#_Toc131534678)

[1.1.3. Tableau des classes 2](#_Toc131534679)

[1.2. Graphique 3](#_Toc131534680)

[2. Inférence statistique sur la moyenne 4](#_Toc131534681)

[2.1. Estimateur non-biaisé et convergent de la moyenne 4](#_Toc131534682)

[2.2. Estimateur non-biaisé et convergent de la variance 5](#_Toc131534683)

[2.3. Vérification de l’homogénéité par coefficient de variation 6](#_Toc131534684)

[2.4. Estimateur par intervalle de confiance de l’écart-type 6](#_Toc131534685)

[2.5. Test de l’affirmation du collègue A 7](#_Toc131534686)

[2.5.1. Affirmation sur la moyenne 7](#_Toc131534687)

[2.5.2. Affirmation sur l’écart-type 8](#_Toc131534688)

[2.6. Test de Pearson (khi-deux) 9](#_Toc131534689)

[3. Calculs de probabilité 11](#_Toc131534690)

[3.1. Probabilité de non-respect de la garantie 11](#_Toc131534691)

[3.2. Loi binomiale 11](#_Toc131534692)

[4. Simulation de Monte-Carlo 13](#_Toc131534693)

[4.1. Générateur de nombre aléatoires 13](#_Toc131534694)

[4.2. Loi de Rayleigh 14](#_Toc131534695)

[4.2.1. Fonction de répartition du module de l’erreur 15](#_Toc131534696)

[4.2.2. Inversion de la fonction de répartition 15](#_Toc131534697)

[4.3. 10000 réalisations 16](#_Toc131534698)

[*4.4.* Nuage de point de *[r,* θ*]* 17](#_Toc131534699)

[4.5. Calcul de la distance radiale et de l’angle de visée 18](#_Toc131534700)

[4.6. Calcul des distances axiales à partir des coordonnées polaires 19](#_Toc131534701)

[4.7. Matrice de covariance des distances 21](#_Toc131534702)

[4.8. Analyse des fréquences relatives 21](#_Toc131534703)

Table des figures

[Figure 1 - Histogramme des fréquences des données 3](#_Toc131534704)

[Figure 2 : Histogramme du test de χ² 10](#_Toc131534705)

[Figure 3 - Graphique des différentes valeurs de θgénérées uniformément 13](https://usherbrooke-my.sharepoint.com/personal/stla0801_usherbrooke_ca/Documents/Fichiers%20de%20conversation%20Microsoft%20Teams/RapportApp5.docx#_Toc131534706)

[Figure 4 - Histogramme des distributions des nombres aléatoires 13](#_Toc131534707)

[Figure 5 - Loi de Rayleigh 14](https://usherbrooke-my.sharepoint.com/personal/stla0801_usherbrooke_ca/Documents/Fichiers%20de%20conversation%20Microsoft%20Teams/RapportApp5.docx#_Toc131534708)

[Figure 6 - 10000 valeurs de r avec un = 4 16](#_Toc131534709)

[Figure 7 - Comparaison des valeurs données par notre estimateur de la CDF inverse et la fonction Rayleigh de Python 16](#_Toc131534710)

[Figure 8 - r et θs'ils étaient correlés 17](https://usherbrooke-my.sharepoint.com/personal/stla0801_usherbrooke_ca/Documents/Fichiers%20de%20conversation%20Microsoft%20Teams/RapportApp5.docx#_Toc131534711)

[Figure 9 - Nuages de points de la CDF inverse par rapport à θ 17](#_Toc131534712)

[Figure 10 - Nuage de points de la distance radiale et l'angle de visée 18](#_Toc131534713)

[Figure 11 - Graphiques des coordonnées axiales 19](https://usherbrooke-my.sharepoint.com/personal/stla0801_usherbrooke_ca/Documents/Fichiers%20de%20conversation%20Microsoft%20Teams/RapportApp5.docx#_Toc131534714)

[Figure 12 - Nuage de points des coordonnées axiales 19](#_Toc131534715)

[Figure 13 - Histogrammes des distances Dx & Dy 20](#_Toc131534716)

# Analyse descriptive des données

## Tableau

### Tableau des caractéristiques

Tableau - Tableau des caractéristiques

|  |  |
| --- | --- |
| *Statistique Descriptive* | *Valeur* |
| Moyenne | 9054 |
| Coefficient de variation | 2,34% |
| Médiane | 9050 |
| Mode | 9000 |
| Écart-type | 212,09 |
| Variance | 44983,67 |
| Étendu | 1300 |
| Minimum | 8500 |
| Maximum | 9800 |
| Nombre d'échantillons | 50 |
| Nombre de classes | 7 |
| Étendue des classes | 190 |

### Règle de Sturges

La règle de Sturges arrondi à la hausse nous donne le nombre de classes et l’arrondi à la hausse de l’étendue divisé par le nombre de classes nous donne l’étendue des classes.

### Tableau des classes

Tableau - Tableau des classes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Borne inférieure*** | ***Milieu de classe*** | ***Borne supérieure*** | ***Classes*** | ***Fréquences Absolues*** | ***Fréquences relatives*** | ***Fréquences cumulées croissantes*** |
| 8500 | 8595 | 8690 | 8500 <= x < 8690 | 2 | 4% | 4% |
| 8690 | 8785 | 8880 | 8690 <= x < 8880 | 3 | 6% | 10% |
| 8880 | 8975 | 9070 | 8880 <= x < 9070 | 20 | 40% | 50% |
| 9070 | 9165 | 9260 | 9070 <= x < 9260 | 21 | 42% | 92% |
| 9260 | 9355 | 9450 | 9260 <= x < 9450 | 2 | 4% | 96% |
| 9450 | 9545 | 9640 | 9450 <= x < 9640 | 1 | 2% | 98% |
| 9640 | 5E+98 | 1,00E+99 | x > 9640 | 1 | 2% | 100% |
|  |  |  | Somme | 50 | 100% |  |

## Graphique

Figure - Histogramme des fréquences des données

On peut remarquer que la distribution semble être normale, avec une moyenne proche de la médiane et un faible coefficient de variation. La moyenne est également inférieure à celle que le collègue A propose.

# Inférence statistique sur la moyenne

## Estimateur non-biaisé et convergent de la moyenne

L’estimateur sélectionné V suivant nous permet d’estimer efficacement µ, car il est non-biaisé et convergent tels que démontré ci-dessous.

CQFD !

CQFD!

## Estimateur non-biaisé et convergent de la variance

L’estimateur sélectionné est le U suivant nous permettant d’obtenir une variance très précise. De plus, elle est convergente, car on obtient 0 si on fait tendre N vers l’infini dans U on obtient une variance de 0. De plus, il est non-biaisé, car on obtient effectivement σ², lors de la preuve.

CQFD !

CQFD!

## Vérification de l’homogénéité par coefficient de variation

L’homogénéité est valide si on a un coefficient de variation inférieur à 15% ce qui est le cas, puisqu’il est de 2.34%.

## Estimateur par intervalle de confiance de la moyenne

L’estimateur par intervalle de confiance est obtenu par l’équation suivante. Pour obtenir une moyenne μ entre 8995.211 et 9112.789 avec une erreur de 58.789.

## Estimateur par intervalle de confiance de l’écart-type

L’estimateur par intervalle de confiance est obtenu par l’équation suivante. Pour obtenir une moyenne σ entre 177.169 et 264.297 avec une erreur de 58.789.

## Test de l’affirmation du collègue A

### Affirmation sur la moyenne

Le collègue A a posé comme hypothèse que la moyenne est de 9150, de cette information nous posons H0 est la moyenne vaut 9150 et H1 la moyenne ne vaut pas 9150. Nous connaissons le seuil de signification α de 5% ainsi nous pouvons trouvés les z limites pour rejeter H0 qui sont de 1.96. Dans notre cas, le z trouvé vaut 3,201 ce qui nous fait rejeter H0 et confirmer que le collègue A avait tords. Pour ce qui est de l’erreur nous la connaissons puisqu’il s’agit de la même valeur que le seuil de signification α. Pour ce qui est de la seconde espèce il s’agit de la probabilité de rejeté H0 alors que H0 est vrai. Pour ce faire on calcul les bornes avec notre α centré à la valeur moyenne estimé et on les centre-réduit par rapport à la moyenne échantillons pour obtenir leur z on trouve la probabilité entre-elles. Ce qui nous donne un β de 10.75 %.

, rejet de H0

### Affirmation sur l’écart-type

Le collègue A posé comme hypothèse que l’écart-type est inférieur ou égal à 1000, de cette information nous posons H0 étant l’écart-type plus petit ou égal à 1000 et H1 l’écart type est supérieur à 1000. Nous connaissons le seuil de signification α de 5%, ainsi nous pouvons trouver le χ² maximum pour rejeter H0, qui est de 66.3387, avec un degré de liberté ν de 49. Dans notre cas, le χ² trouvé vaut 2.2042 ce qui nous permet d’accepter H0 et confirmer que le collègue A avait raison sur son affirmation concernant l’écart-type.

## Test de Pearson (khi-deux)

Pour le test de χ², nous devons tout d’abord poser une hypothèse nulle H0, qui est que la distribution est normale, et H1 que la distribution ne soit pas une distribution normale. Nous devons par la suite créer une classification respectant le fait d’avoir minimalement 5 éléments par classes, pour pouvoir obtenir notre k. Ensuite avec le k et le r (notre nombre de variable mis à 2 (écart-type et moyenne)), nous pouvons trouver notre degré de liberté ν. Nous connaissons notre niveau de confiance de 95% avec lequel on peut trouver notre χ²α;ν qui ici vaut 5.9915. Avec le Tableau 3 on peut obtenir notre χ² qui est la somme des pour obtenir 5.6613, qui est inférieur au Khi-deux théorique, ce qui nous permet de garder l’hypothèse H0 : il s’agit bien d’une distribution normale.

Tableau - Test de Pearson (khi-deux)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classes | Limites | Borne gauche | Borne droite | Borne gauche | Borne droite | Pi(zg<z<zd) | foi | Relative | fti | (Fgi-fti)²/fti |
| 1 | x < 8900 | 0 | 8900 | -42,69 | -0,73 | 0,2327 | 5 | 10% | 11,635 | 3,7836893 |
| 2 | 8900 <= x < 9000 | 8900 | 9000 | -0,73 | -0,255 | 0,1686 | 8 | 16% | 8,43 | 0,02193357 |
| 3 | 9000 <= x < 9100 | 9000 | 9100 | -0,26 | 0,22 | 0,1819 | 12 | 24% | 9,095 | 0,92787521 |
| 4 | 9100 <= x < 9200 | 9100 | 9200 | 0,22 | 0,69 | 0,1717 | 11 | 22% | 8,585 | 0,67935061 |
| 5 | x >= 9200 | 9200 | 1,00E+99 | 0,69 | 4,7149E+96 | 0,2451 | 14 | 28% | 12,255 | 0,24847205 |
|  |  |  |  |  | Somme | 1 | 50 | 100% | 50 | 5,66132074 |

Figure 2 : Histogramme du test de χ²

# Calculs de probabilité

## Probabilité de non-respect de la garantie

On sait que la moyenne estimée est de 1770 heures/par ans et qu’on offre une garantie de 5 ans, on obtient donc une garantie couvrant 8850 heures. Pour obtenir la probabilité d’erreur nous considérons une loi normale, puisque nous avons plus de 30 échantillons. Avec la moyenne et l’écart-type trouvé en 1 nous pouvons centrer-réduire la valeur limite de la garantie pour obtenir un z (8850) de -0.962 ce qui nous donne une probabilité de 16.85% qu’un radar ne respecte pas la garantie.

## Loi binomiale

Avec la probabilité d’échec de 16.85% nous pouvons utiliser cette probabilité pour produire une fonction binomiale avec le nombre d’itération de 50 et un nombre d’erreur selon le cas. Nous obtenons ainsi nos quatre valeurs d’erreurs. Pour ce qui est de la moyenne et de la variance de la loi binomiale nous savons que la moyenne est N × P, donc 8,425 et la variance est N × P ×(1-P), ce qui nous donne 7,005.

# Simulation de Monte-Carlo

## Générateur de nombre aléatoires

Le générateur de nombres aléatoires a été fait avec la fonction numpy.random.rand de Python. Pour obtenir , chaque résultat du générateur U[0, 1] a été multiplié par .

La distribution graphiquement semble être très uniforme, autant en regardant la Figure 3 que la Figure 4.

Figure - Graphique des différentes valeurs de θgénérées uniformément

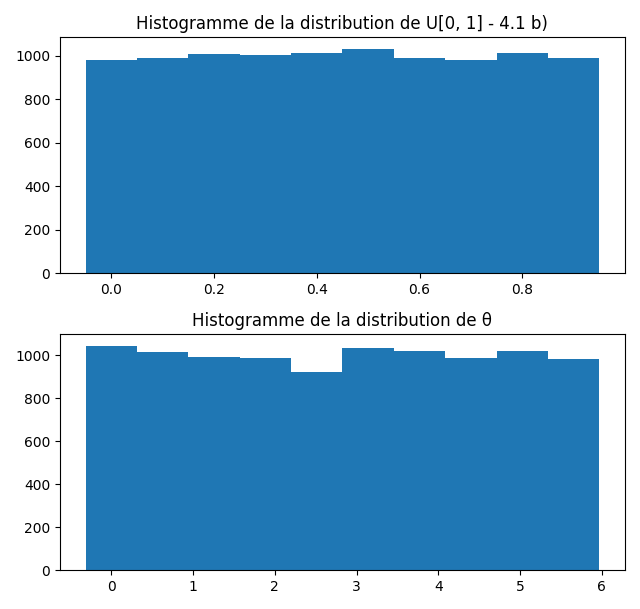


Figure - Histogramme des distributions des nombres aléatoires

## Loi de Rayleigh

La Figure 5 présente en premier la loi de Rayleigh théorique pour différentes valeurs de . Ces valeurs proviennent de l’équation , traduites en l’expression Python :  
r / sigma2 \* np.exp(-r\*\*2 / (2 \* sigma2)).

Par la suite, la CDF (calculée en 4.2.1) est présentée pour les mêmes valeurs de , , donnant l’expression :  
1 - np.exp(-r\*\*2 / (2 \* sigma2)).

Finalement, notre fonction inverse de la CDF (calculée en 4.2.2) est présentée,

np.sqrt(-2 \* sigma2 \* np.log(1 - Fr)).

Figure - Loi de Rayleigh

### Fonction de répartition du module de l’erreur

Notre CDF de la fonction de distribution de Rayleigh a été calculé en faisant l’intégrale de la PDF, en substituant :

### Inversion de la fonction de répartition

L’inverse de la CDF est calculée en isolant r de :

## 10000 réalisations

Pour tester l’inverse de la fonction de répartition, 10000 valeurs uniformément distribuées entre 0 et 1 ont été rentrées dans notre CDF inversée, pour nous donner un graphique présenté à la Figure 6.

Pour la Figure 7, 10000 valeurs ont été mises dans notre fonction de CDF inverse, ainsi que dans la fonction Python numpy.random.rayleigh, pour & . Il n’y a pas de différence visuelles majeures entre ce qui est généré par la fonction de numpy et celle que nous avons créé.

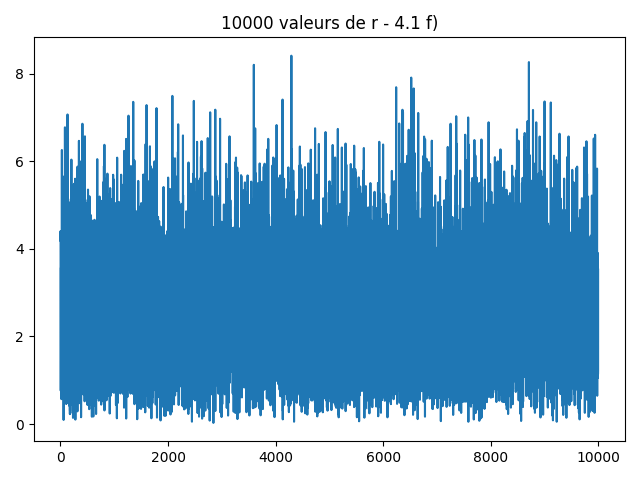


Figure - 10000 valeurs de r avec un = 4

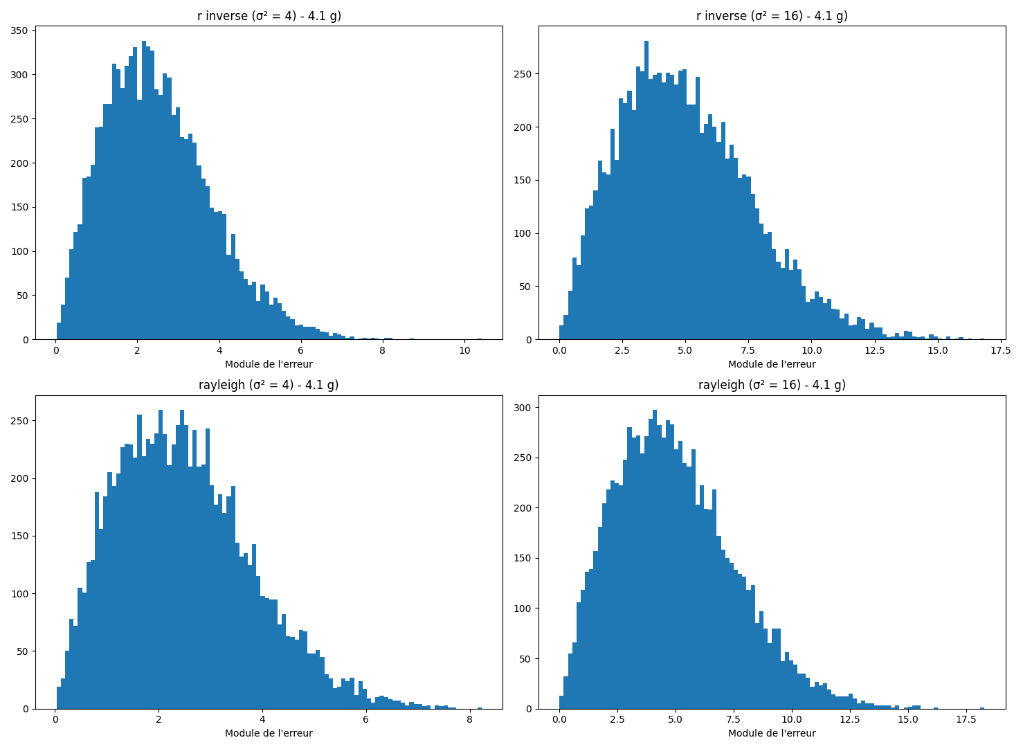


Figure - Comparaison des valeurs données par notre estimateur de la CDF inverse et la fonction Rayleigh de Python

## Une image contenant graphique Description générée automatiquementNuage de point de *[r,* θ*]*

La Figure 9 présente un nuage de points avec sur l’axe des x et sur l’axe des y, avec 10000 points. Il n’y a aucun lien graphiquement visible entre et r sur les deux graphiques, qui ressemble fort à celui de la Figure 6. Le nuage de point est très dispersé, avec un nombre de points égal pour chaque valeur de et de moins en

Figure - r et θs'ils étaient correlés

Si et r avaient été générés par le même générateur d’origine, on verrait une corrélation claire comme à la Figure 8.

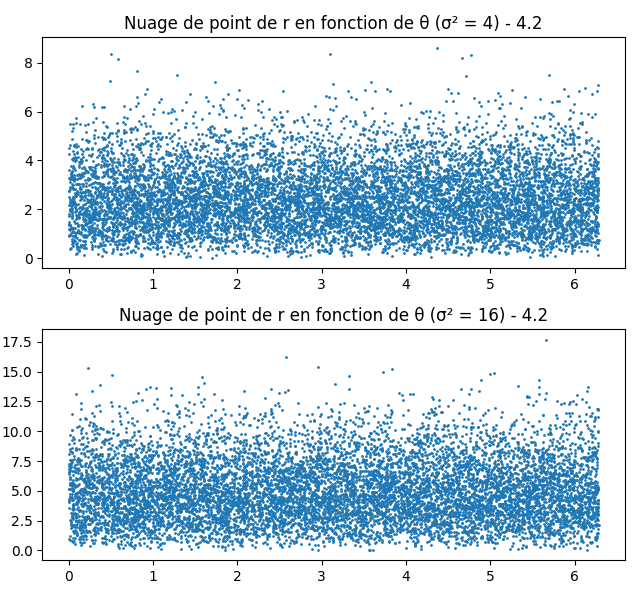


Figure - Nuages de points de la CDF inverse par rapport à θ

## Calcul de la distance radiale et de l’angle de visée

La distance D et l’angle ont été calculés en utilisant Pythagore et de la trigonométrique de base plutôt qu’en utilisant les formules proposées dans le guide étudiant.

Le nuage de points est plutôt circulaire et n’a pas d’orientation significative, ce qui démontre que les données de distance et d’angle ne sont pas corrélées.

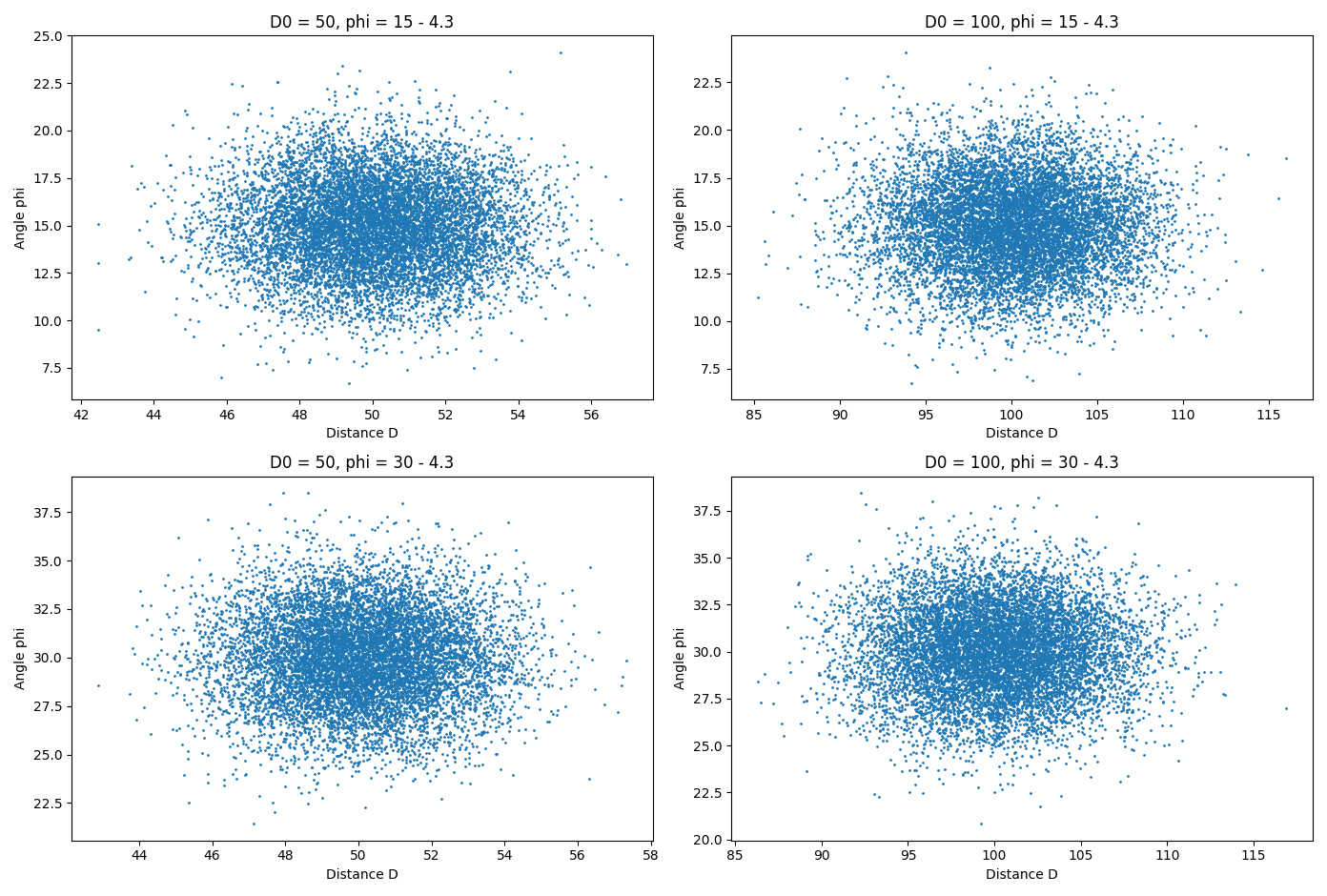


Figure - Nuage de points de la distance radiale et l'angle de visée

## Calcul des distances axiales à partir des coordonnées polaires

Nos graphiques 1D des coordonnées axiales & (Figure 11) ne donne pas d’information particulièrement pertinentes, à part une tendance visible centrées autour de leurs moyennes.

Figure - Graphiques des coordonnées axiales

Les nuages de points (Figure 12) sont beaucoup plus intéressants à analyser. Avec le plus petit, on voit que les valeurs de gauche pour les distances sont plus serrés, et que la forme presque circulaire du nuage de point n’est pas oblique, ce qui semble confirmer qu’il n’y a pas de corrélation entres les données.

Les distances axiales sont aussi simplement une décomposition de la distance radiale D et de l’angle de visée (4.5), selon les équations :

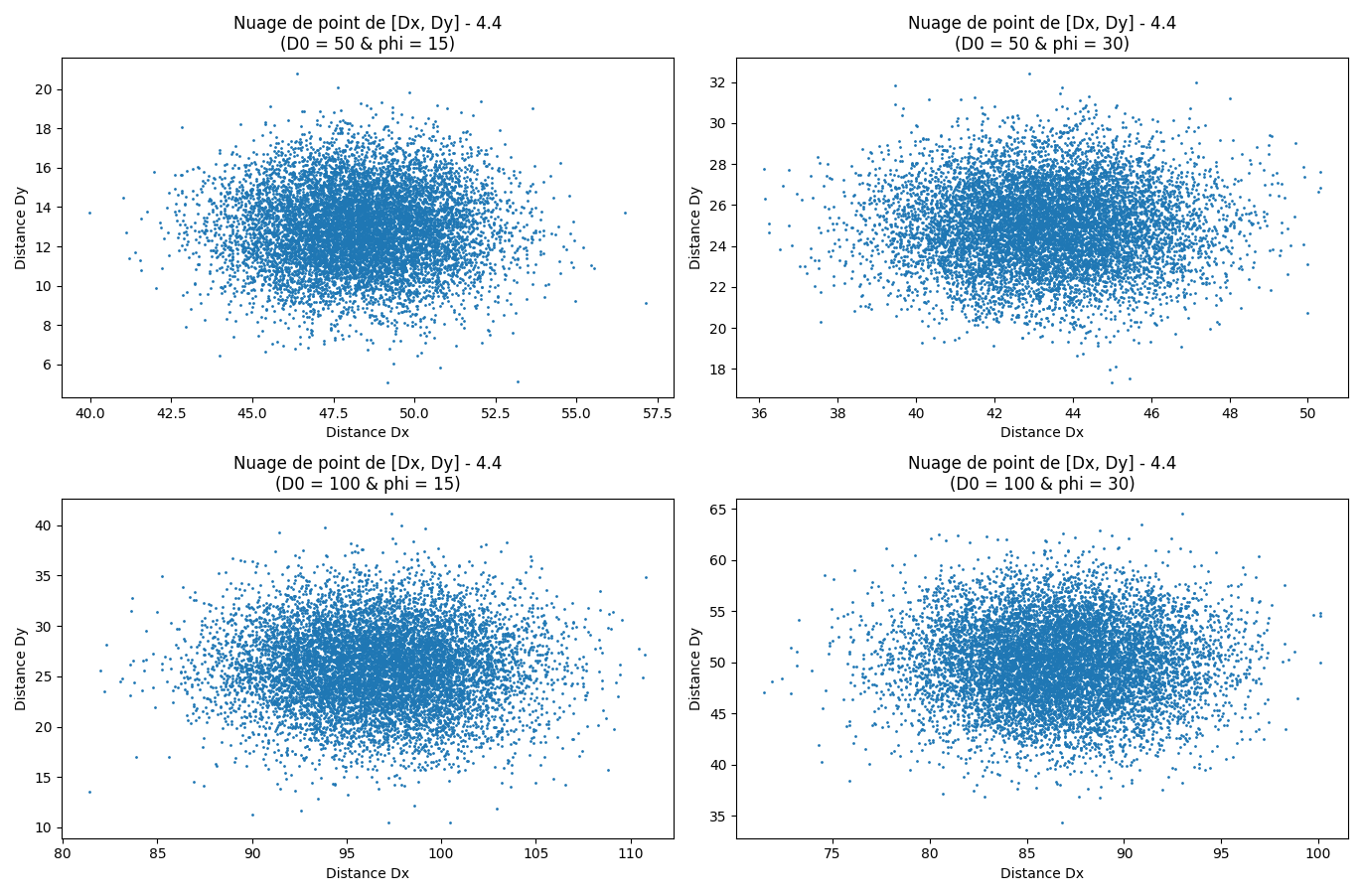


Figure - Nuage de points des coordonnées axiales

Les histogrammes des distances & (Figure 13) ont une allure suivant une distribution normale. Un Pythagore des moyennes X̄ redonne une valeur proche du , et l’écart-type semble cohérent avec les valeurs de & .Les courbes de fréquence relatives ressemblent à une gaussienne. Un test de Pearson (khi-deux) pourrait permettre de s’assurer que la distribution est bien gaussienne. Pour par exemple, la PDF pour serait décrite par l’équation d’en-dessous (avec un test pour similaire avec un .

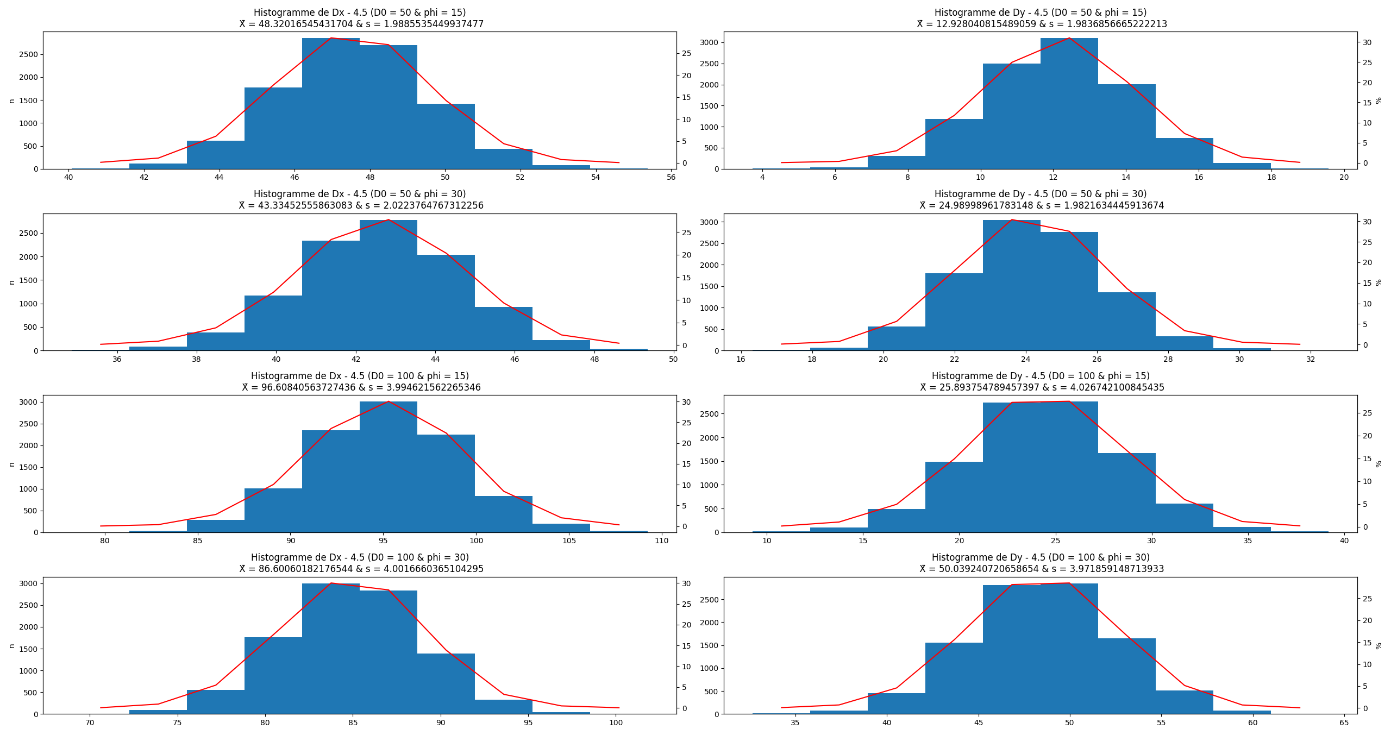


Figure - Histogrammes des distances Dx & Dy

## Matrice de covariance des distances

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

## Analyse des fréquences relatives

Pour obtenir les longueurs des axes majeurs et mineurs de l’ellipse considérant que et ne sont pas corrélé, nous savons que la longueur de l’axe majeur est et la longueur de l'axe mineur , notre S pour un niveau de confiance de 95% est de 5.991. De plus, l’angle de l’ellipse est nul.