UNIVERSITÉ DE GENÈVE

ANALYSE ET TRAITEMENT DE L'INFORMATION
14X026

TP5: Communications and probabilities

Author: Julien Python [120 Crédits]

E-mail: julien.python@etu.unige.ch

November 23, 2020

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'informatique

Exercice 1

1) $p(X,Y,Z) = p(X) \cdot p(Y|X) \cdot p(Z|X,Y)$

2) $p(X_1, X_2, X_3, X_4, X_5, X_6, X_7) =$
$p(X_1)\, p(X_2)\, p(X_3 | X_1, X_2)\, p(X_4)\, p(X_5) \cdot p(X_6 | X_3, X_4, X_5) \cdot$
$\cdot\, p(X_7 | X_6)$

Exercice 2

I) J'ai effectué le résultat pour 1000 experiences afin d'être plus précis, et on obtient que l'esperance du nombre de caractères corrects des ces 1000 experiences vaut 590.676, donc ~40% d'erreur.

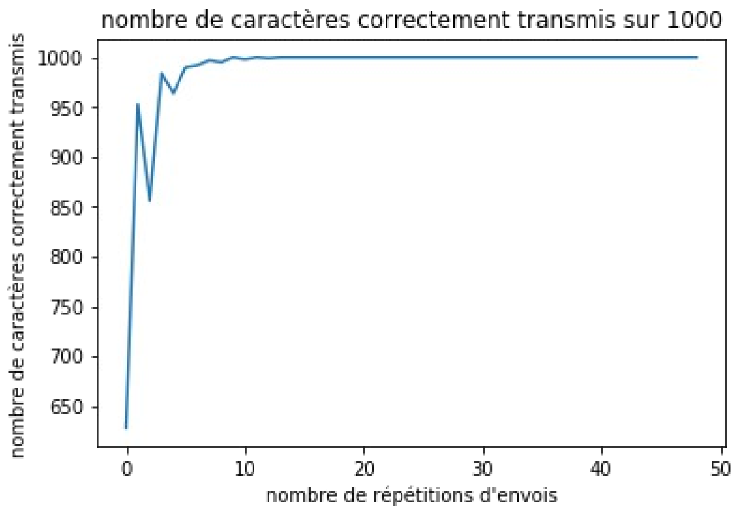
ii) J'ai répété la même experience pour n=3, où n est le nombre de répétitions, on obtient une esperance de 867 caractères corrects, ce qui représente un taux de ~13.3% d'erreur.

iii) J'ai répété la même experience pour n=5, où n est le nombre de répétitions, on obtient une espérance de 957.938 caractères correct, ce qui représente un taux de ~4.2% d'erreur

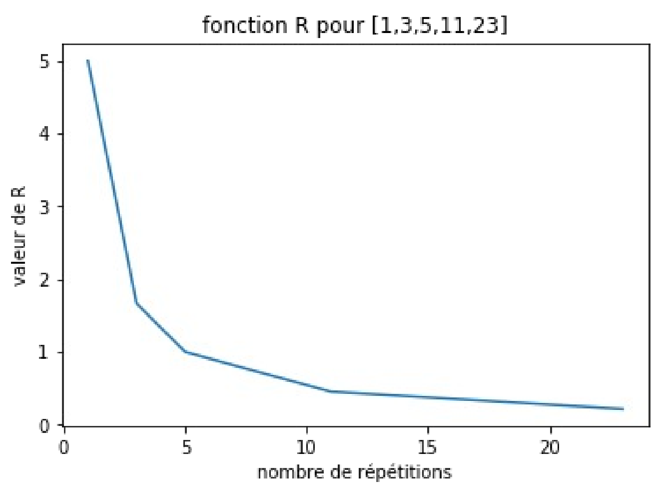
Pour n=11, on obtient une espérance de 998.549 caractères correct, ce qui représente un taux de ~0.145% d'erreur.

Pour n=23, on obtient une espérance de 1000 caractères correct, ce qui représente un taux de 0% d'erreur.

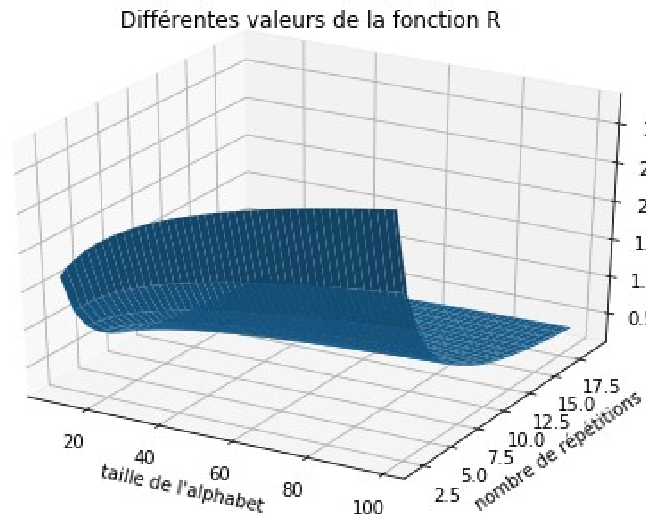
Comme on le remarque sur le graphe ci-dessous, peu après 10 l'erreur est très proche de 0.

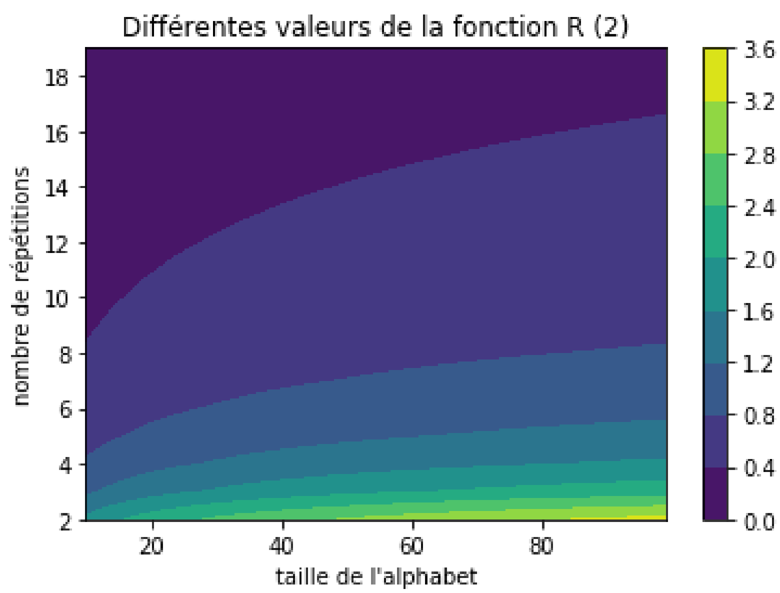


iv) Si on regarde seulement la valeur de R par rapport aux nombre de répétitions proposée dans le début de l'exercice, on obtient un graphe suivant. On remarque que plus le nombre de répétitions avance, plus la valeur de R converge vers zéro.



Mais afin d'y voir plus clair, j'ai décidé d'y afficher dans un plot en 3 dimensions qui a comme axe R, n et M. En procédant ainsi, on obtient les graphes suivants :



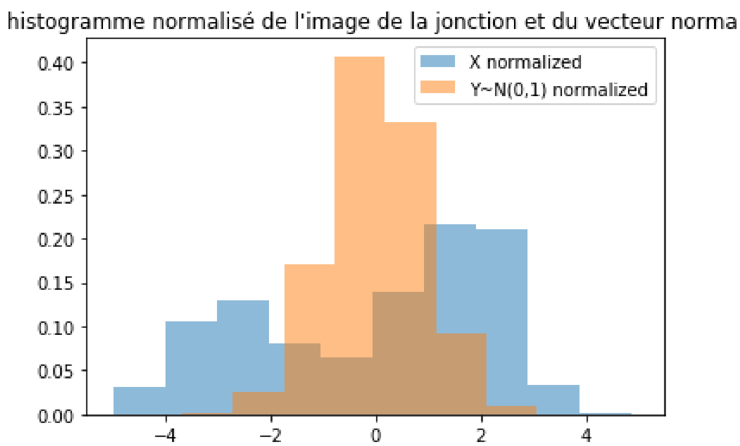


Sur le premier graphe, on voit bien la forme des donnée et l'influence du nombre de répétitions qui fait tendre R vers zéro, mais aussi en augmantant l'alphabet pour peu de répétition la valeur de R augmente (même si elle n'appartient pas à l'interval [0,1]). Sur le second graphe, on distinge clairement le comportement logarithmique de la valeur de R pour la relation entre la taille de l'alphaber et le nombre de répétitions. Ainsi, selon le taux de communication souhaité, on peut utilisé ce graphe pour choisir le nombre de répétition à effectuer pour un alphabet doonné (ou inversémment).

Exercice 3

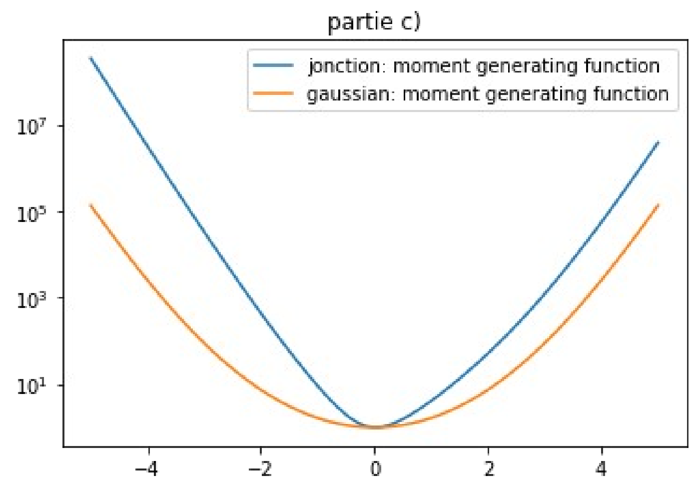
Dans cet exercice, nous allons travailler avec deux données, la première notée X est un vecteur qui est extrait d'une image de la jonction. La deuxième, Y, est un vecteur de même taille mais qui a été créé par une distribution gaussienne. De plus, ces deux vecteurs ont été normalisés selon les consignes de l'énoncé.

On obtient ainsi les histogrammes suivants :



On remarque que l'histogramme X présente distinguement 2 bosses, alors que l'histogramme Y une seule. Ainsi, on peut visualisé qu'on a deux clusters pour le vecteur X, et un seul pour le vecteur Y. On s'attend donc à retrouver des différences dans leurs attributs statistiques, et plus particulièrement dans leurs moments.

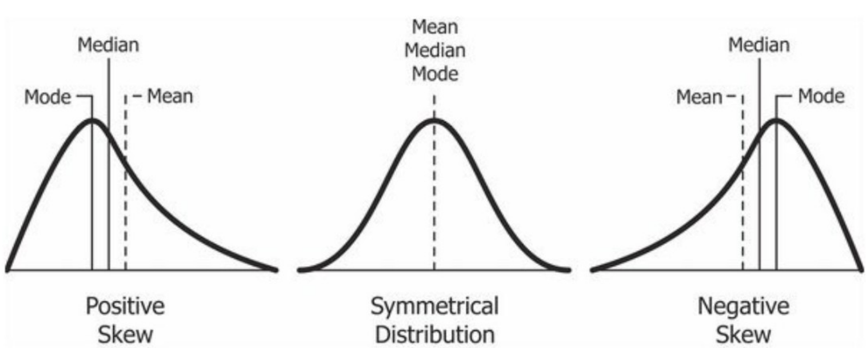
A l'aide des formules données dans l'énoncé du travail, on obtient le graphe suivant pour les approximations des fonctions génératrices de moments. On remarque ainsi une symmétrie axiale en zéro pour les données gaussiennes, tout comme son histogramme. Alors que pour les données de la photo de la jonction on remarque une légère asymétrie axiale en 0, qui peut s'expliquer par le fait que l'histogramme n'est pas symmétrique en zéro avec un cluster plus grand sur la partie droite.



Moments de X: [1. 0. 4.97 -4.62 45.06 -86.28 542.92]

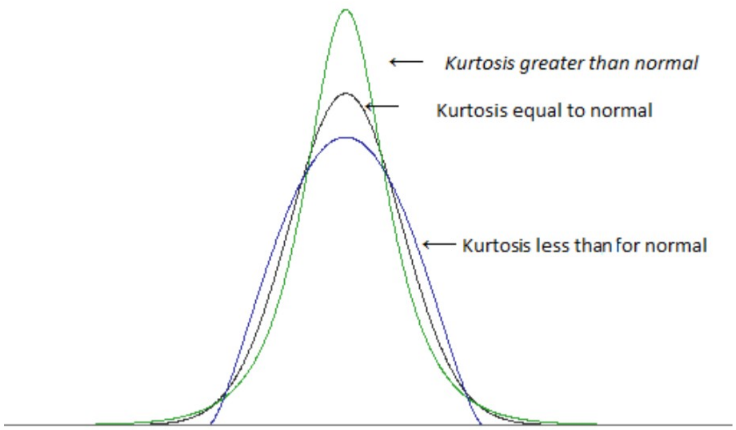
Moments de Y: [1. 0. 0.71 -0. 1.52 0. 5.41]

Le premier moment est l'expérance, et le second est la variance, on ne va pas appronfondir ces deuxx sujets ici car ils ont déjà été abordés dans les précédents tps. Le troisème moment nommé skewness, mesure l'asymmétrie de la distribution.

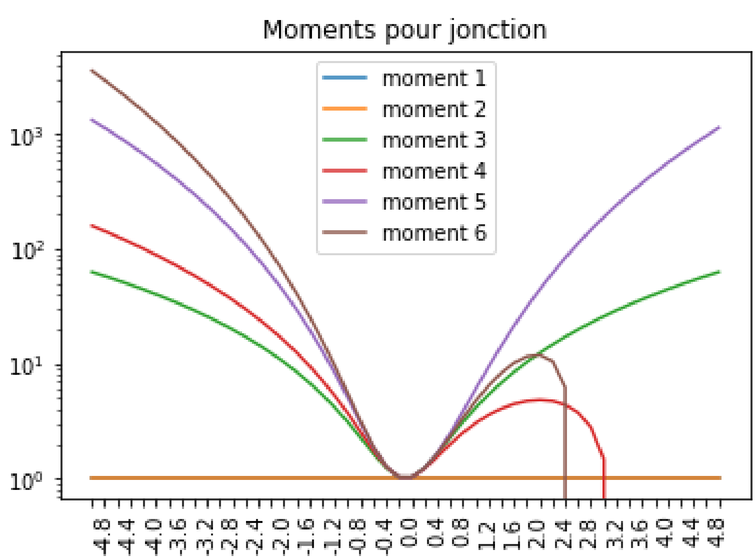


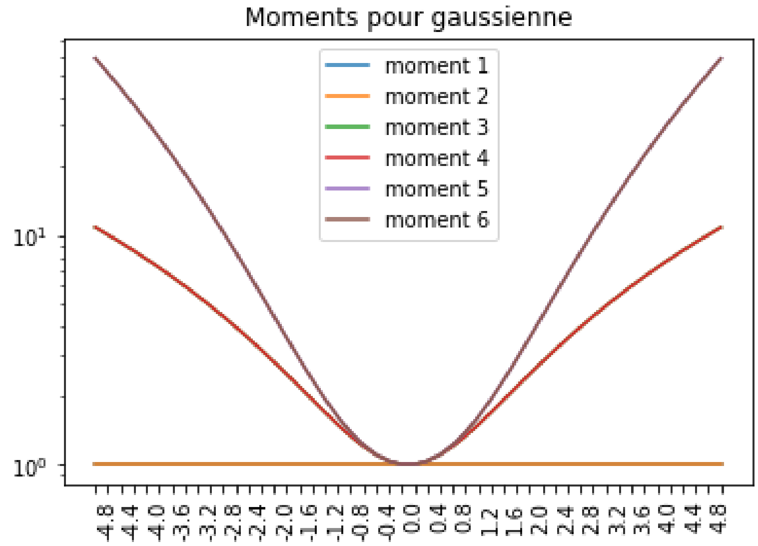
Etant donnée que nos deux valeurs sont positives, on est donc dans le premier cas de l'image ci-dessus. Mais on voit bien que le vecteur X a une asymétrie plus forte, alors que le vecteur Y est plus proche de 0.

Le 4ème moment, appelé kurtosis ou coefficient d'applatissement, il mesure la répartition des masses de probabilité autour de leur centre. Ainsi, il nous donne une idée à propos de comment est la fat tails de la distribution, c'est à dire à quelle frequence des valeurs apparaissent très éloignées de nos données (événements très peu probables). Ainsi, on voit que avec les données gaussiennes des données "extrêmes" peuvent toujours avoir lieu, tandis que c'est moins le cas avec le vecteur X. De plus, pour X, vu que la valeur est négative, ceci implique une queue plus fine et un haut de courbe plus plat, ou des déviations plus grandes de l'espérance.



Pour le 5ème moment, on peut le visualiser comme une représentation du skewness de la queue. Pour les deux derniers moments je n'ai pas trouvé plus d'informations mais je suppose que le 6ème moment est lié au 4ème moment car il a un comportement semblable, alors que le 7ème moment doit être lié au 5ème car il a aussi un comportement semble, c'est à dire en terme statistique que ces données semblent corrélées.





On sait que la fonction caractéristique d'une variable aléatoire définit complètement sa distribution probabiliste. Si une variable aléatoire admet une pdf, alors la fonction caractéristique est la transformée de Fourier de la pdf.

On sait aussi qu'il suffit des deux premier moments pour une variable aléatoire suivant une distribution gaussienne pour obtenir sa pdf. Mais, pour une variable aléatoire quelconque, il faudrait une infinité de moments pour obtenir sa pdf respective.

Ainsi, on obtient un résultat qui semble symmétrique pour la fonction génératrice des moments du vecteur Y, tandis qu'on obtient un résultat nettement asymmétrique pour le vecteur X. Ces résultats sont la conséquence de tout ce dont nous avons parlé dans le troisème exercice, et est un bon résumé de ce dernier.